Estimating The Returns to Charitable Fundraising*

Danyal Khan, Word count: 9950
London School of Economics

July 31, 2021

Abstract

In this paper, I introduce a novel system of dynamic equations that capture the revenue-generation and decision-making process of charitable fundraising. Exploiting plausibly exogenous variation in the national minimum wage, I estimate the returns to fundraising for large "green" charities based in the UK. I find that for the charity in the sample with the median fundraising ratio (defined as the ratio of fundraising revenue to expenditure) every additional £1 spent on fundraising raises £1.45 in return. I also find tentative evidence that the collective returns to fundraising, which account for the interplay between rival charities, are greater than individual returns to fundraising, that a charity could earn in isolation. I also use the estimated equations to characterise and estimate the bias in the fundraising ratio (ROI) as a measure of the true fundraising efficiency of a charity, i.e. the inherent ability of the fundraising managers to raise money irrespective of factors such as the size of their budget, the popularity of their cause or the influence of their rival charities. I find that whilst the fundraising ratio (ROI) is heavily biased by factors such as these, it remains a strong predictor of fundraising efficiency, and therefore remains a useful tool for charity managers and regulators. However, due to weak identification, heterogeneous causal effects and persistent sample selection bias, the existence and magnitude of the true parameters, in addition to the direction of average bias in the fundraising ratio, is very uncertain and explains a lack of statistical significance of the estimated parameters.

*I would like to thank Dr Levy, Lucy, Akmal, Ibby and my EME peers for their invaluable feedback and guidance.

Contents

1 Introduction

The Oxfam Scandal in 2012 brought renewed scepticism of large charities, in particular regarding the efficient and effective use of donor funds. In addition, austerity measures, the COVID19 pandemic and rapidly changing donor preferences have forced charities in every subsector to explore new fundraising strategies that innovatively and effectively capture donors.

Green Charities play an increasingly important part in raising awareness for man-made climate change, tackling environmental exploitation and initiating conservation projects - strongly enabled by large-scale fundraising efforts. In this paper, I introduce a system of equations that capture the revenue-generation and decision-making process of a charity engaging in fundraising. Exploiting plausibly exogenous variation in the national minimum wage, I use an instrumental variables strategy to identify these equations and estimate the monetary returns to charitable fundraising for large Green Charities [1] based in the UK, using extensive data from the Charity Commission.

The identification of the proposed system of equations allows me to separately estimate two types of *returns to fundraising* (defined as the causal effect of fundraising spending on fundraising revenue): (1) the individual returns to fundraising, which are earned when a charity fundraises in isolation, and (2) the collective returns, which additionally account for the interplay between rival charities. The estimated equations also allow me to characterise the bias in the fundraising ratio, or rate of return on investment (ROI) as a measure of fundraising efficiency.

Charitable fundraising is an inherently economic phenomenon, comparable to investment or advertising, in that it describes the allocation of current resources towards obtaining future resources. To illustrate, a charity - Charity A - may choose to spend £500 to organise a fundraising gala that generates £2000 in donations, whilst its rival[2] - Charity B - spends £5000 on a marketing campaign that generates £10,000 in the same year.

Using the ROI as an indicator for the efficiency of fundraising, [3] Charity A is said to be more fundraising *efficient* than Charity B, since for every £1 spent on fundraising, they receive £4 ($\frac{2000}{500} = 4$) back in donations, compared to the £2 ($\frac{10000}{5000} = 2$) received by Charity B. The National Council for Voluntary Organisations (NCVO) and Charity Navigator- a charity assessment organisation- both consider the ROI to be a useful indicator of financial performance. This raises the question: *if the ROI measures the efficiency of fundraising, should charities seek to maximise their fundraising ROI*?

Employing the fundraising ROI, henceforth the fundraising ratio, as a measure of how *inherently* efficient a charity is at generating income ignores much of the economics underpin-

[1] I return to the definition of "Green" Charities in the Data Section

[2] The preferred term by charity managers is "peer" but I use "rival" throughout this paper to capture the competitive connotations.

[3] $fundraising\ ratio\ (ROI) = \frac{fundraising\ revenue}{fundraising\ spending}$

ning the fundraising process. For example, Charity B may employ more innovative fundraising techniques than Charity A, making it intuitively more efficient, but has a smaller ratio simply because their cause is significantly less fashionable or well-known amongst the donor population[4], perhaps due to the fundraising inactivity of charities with the same cause. This means Charity B requires more effort than Charity A to raise the same amount of money, resulting in a lower fundraising ratio. Whilst *true* fundraising efficiency is something charity fundraisers should strive to improve, attempting to maximise the fundraising ratio, which is not entirely not under the control of the charity, can lead to unintended consequences.

Secondly, the fundraising ratio ignores the net return on fundraising, which indicates how much better off a charity is from fundraising. In the example above, Charity B is economically better off despite a lower fundraising ratio, since their net revenue (£10,000-£5000= £5000) is higher than Charity A's net revenue (£2000-£500=£1500). A higher net revenue is analogous to higher profit and by implication, Charity B has more to spend on charitable activities that benefit their beneficiaries (Steinberg (1994)). Hence, as an indicator of performance, a stronger case could be made to strive for *optimality*[5] of fundraising efforts rather than efficiency.

[4] I treat a charity's beneficiaries as exogenous to charity managers, hence a less popular cause as a driver of a lower fundraising ratio should not indicate true inefficiency.

[5] The optimal fundraising effort or spending is the level that maximises the net revenue. We return to the concept of optimality in section 2 and Equation 2

2 Background

Figure 1 presents a graphical illustration of the stylised interplay between the observable variables F_{it} - representing the fundraising revenue (donations) of a charity i in time t, B_{it} - representing the fundraising expenditure (budget) of charity i in time t and A_{it} - representing the average fundraising expenditure of charity i's rivals in time t. I explain the interpretation of the graphical parameters in Figure 1 in this section.

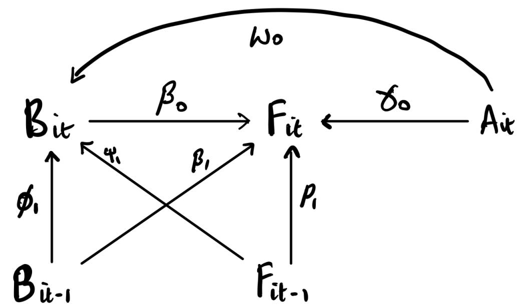

Figure 1: Interplay between own fundraising expenditure, rival fundraising expenditure and fundraising revenue across time

The most fundamental causal effects of interest are the contemporaneous and lagged effects of own charity fundraising spending on donations, represented by the β_0 and β_1 coefficients respectively in Figure 1. The existence of β_1 is highly plausible since it is likely that the individual returns to fundraising are not limited to the same year; for example, a marketing campaign or investment in a new fundraising channel may expect to yield returns over a longer time period.

The other causal effects of interest relate to the interplay between rival charities, characterised by two main interactions: *spillover effects* (e.g. the effect of Charity A's spending on Charity B's donations) and *strategic effects* (e.g. the effect of Charity A's spending on Charity B's spending). Spillover effects are represented by γ_0 whilst strategic effects are represented by ω_0 in Figure 1.

In addition, ϕ_1 and ρ_1 represent *inertia* in the fundraising process, whilst ψ_1 represents the *enabling* effect of higher income in time $t-1$ on fundraising expenditure in time t. I return to interpretation and identification of these parameters in section 3.

Jacobs and Marudas (2006) estimate the individual returns to fundraising for a subset of UK and US based charities and use estimates of the fundraising elasticity to categorise the fundraising efforts of the subset into "excessive", "optimal" and "insufficient", based on their fundraising behaviour. Jacobs and Marudas (2006) builds on the work of Okten and Weisbrod (2000), Posnett and Sandler (1989), Steinberg (1994) and Khanna et al. (1995) which, taken together, find evidence of net-revenue-maximising behaviour across large US and UK charities to be mixed and vary widely across charity subsectors. Whilst the results are convincing, these papers fail to account for the impact of rival charities on fundraising efforts, (via strategic effects) and fundraising outcomes (via spillover effects), hence potentially suffer from significant omitted variable bias.

Literature on the effect of advertising on sales surveyed in Bagwell (2007) present important hypotheses regarding the existence and direction of fundraising spillover effects. Bagwell (2007) reports that the existence of positive strategic effects (advertising inducing rivals to advertise), negative spillover effects (advertising negatively impacting the sales of rivals) and contemporaneous and lagged own brand advertising effects on sales were first evidenced by Jean-Jacques Lambin in 1976. Vardanyan and Tremblay (2006) builds on the work of Lambin, finding evidence that positive spillover effects of advertising are more likely to occur in emerging markets, as advertising increases total demand (increasing the total pie) whereas negative spillovers are more likely in mature markets, as advertising simply results in business stealing (a redistribution of the existing pie). Hence, it is likely that the direction of fundraising spillovers will vary across different charity subsectors: in emerging charity subsectors such as the environmental sector, fundraising may be *inclusive*, generating positive spillover effects on other charities as public sensitivity towards the environment grows, whereas in more mature charity subsectors such as housing, the donor pool is relatively stable hence fundraising may be *exclusive*, generating negative spillover effects as charities steal donors from each other.

Arulampalam et al. (2015) investigate the returns to fundraising for large UK-based overseas development charities, and identify strong positive fundraising spillovers (positive externalities) indicating that the efforts of one development charity may increase contributions to other development charities. However, in only investigating fundraising revenues, Arulampalam et al. (2015) are not able to fully characterise the interactions between rival charities, which are not limited to spillover effects. This is depicted in Figure 1, where the total effect of rivals fundraising spending A_{it} on own fundraising revenue F_{it} is given by the direct spillover effect γ_0 plus the indirect effect through B_{it}, via the strategic effects ω_0. For example, an increase in fundraising efforts by Charity A may directly affect Charity B's donations, but also induce Charity B to respond by intensifying their own fundraising efforts [6]. Rose-Ackerman (1982) presents a theoretical framework for the existence of positive strategic effects (strategic complementarity) in the fundraising process between rival charities; in the extreme, they predict a race-to-the-bottom as competition for donors incentivises a vicious cycle of ever-increasing fundraising spending, depleting the net revenues of charities and thus their capacity for charitable impact in the process.

In this paper, I present a novel system of equations that impose a structure on the stylised graphical model in Figure 1. A model that captures the decision making process and the revenue generating process, including own fundraising effects, strategic and spillover effects has not been explored before empirically in the context of large charities. The empirical strategy involves exploiting plausibly exogenous variation in lagged fundraising revenues and expenditures, as well as in the national minimum wage. The causal identification of the parameters illustrated in Figure 1 will allow me to characterise the bias in the fundraising

[6] These strategic effects ω_0 are analogous to general equilibrium effects.

ratio as a measure of fundraising efficiency.

3 Theoretical Framework

I first specify the following model that treats fundraising revenue F_{it} as the outcome variable of a dynamic production function:

$$F_{it} = E_{it} F_{it-1}^{\rho_1} B_{it}^{\beta_0} B_{it-1}^{\beta_1} A_{it}^{\gamma_0} \tag{1}$$

Where B_{it} and B_{it-1} represent current and lagged fundraising expenditures (inputs), A_{it} represents the current average fundraising expenditure of charity i's rivals and E_{it} represent unobservable (Hicks-neutral) shocks to fundraising revenue. All parameters are interpreted as partial elasticities. Applying the log-transformation yields:

$$f_{it} = \rho_0 + \rho_1 f_{it-1} + \beta_0 b_{it} + \beta_1 b_{it-1} + \gamma_0 a_{it} + \varepsilon_{it} \qquad \text{(Production Function)}$$

Where ε_{it} is a mean-zero error term assumed to be uncorrelated across charities and ρ_0 is a constant that captures the average autonomous fundraising revenue across time and across charities. The own charity fundraising elasticities β_0 and β_1 capture the impact of fundraising efforts on fundraising outcomes across time; these parameters nest different causal mechanisms linking fundraising efforts to outcomes, including through increased brand awareness, donor retention and reliability signals. Persistence in fundraising revenue over time is captured by ρ_1, henceforth the brand effect, which may exist due to donor inertia or the signalling effect of high income in time $t-1$, encouraging high donations in time t. γ_0 captures contemporaneous spillover effects outlined in section 2.

Next, I assume that charities choose their fundraising expenditure (B_{it}) to maximise the expected net fundraising revenue ($\pi_{it} = F_{it} - B_{it}$) in the same period, conditional on their past and current information at time t. If it exists, this implies that the optimal fundraising expenditure B_{it}^{opt} satisfies:

$$B_{it}^{opt} = arg \max_{B_{it}} \left[\mathbb{E}(\pi_{it} | F_{it-1}, B_{it-1}, A_{it}) \right]$$

Solving the first order condition and setting the marginal product of B_{it} on F_{it} equal to 1 yields the unconstrained optimum level of fundraising expenditure:

$$B_{it}^{opt} = \left[\beta_0 \mathbb{E}(E_{it}|.) B_{it-1}^{\beta_1} F_{it-1}^{\rho_1} A_{it}^{\gamma_0}\right]^{\frac{1}{1-\beta_0}} \tag{2}$$

which can be represented in the reduced form:

$$B_{it} = G_{it} B_{it-1}^{\phi_1} F_{it-1}^{\psi_1} A_{it}^{\omega_0}$$

taking logs yields:

$$b_{it} = \phi_0 + \phi_1 b_{it-1} + \psi_1 f_{it-1} + \omega_0 a_{it} + g_{it} \qquad \text{(Decision Function)}$$

where g_{it} is a mean zero error term assumed to be uncorrelated across charities and ϕ_0 is a constant. These reduced-form parameters ϕ_1, ψ_1 and ω_0 also nest the causal interpretations introduced in section 2, as persistence, enabling and strategic effects respectively. Therefore, I choose not to impose the restrictions relating the parameters in the Fundraising Production Function to the parameters in the Fundraising Decision Function[7]. Figure 2 depicts the stylised system with the parameters of both models alongside each other, ignoring the unobservables[8].

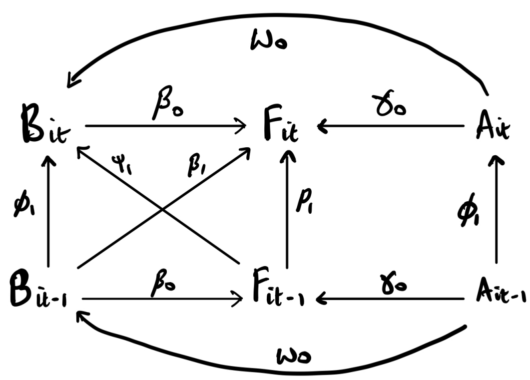

Figure 2: Complete model

I assume that that the *individual* returns to fundraising for charity i in any given time t are defined as the relative changes in fundraising revenue as a result of a relative change

[7]Imposing the parameter restrictions would assume that all charities achieve static net-revenue maximisation and are unconstrained in fundraising expenditure - this is not likely to hold for all charities. See section 10 for a deeper discussion about heterogeneity amongst charities

[8]In the complete system, I make several simplifying assumptions, including: (1) contemporaneous effects are constant over time, (2) own fundraising effects, enabling effects persistence effects and brand effects are at most one period lagged, (3) spillover and strategic effects are at most contemporaneous (lagged spillover effects γ_1 and lagged strategic effects ω_1 do not exist in reality) and (4) strategic effects are uni-directional - a single charity cannot influence the average expenditure of its rivals. The assumption that persistent effects ϕ_1 are incident on average rival expenditure is provided for completeness and is irrelevant for the analysis.

Parameter	Label
β_0	Contemporaneous own fundraising effects
β_1	Lagged own fundraising effects
ω_0	Strategic effects
γ_0	Spillover effects
ψ_1	Enabling effects
ϕ_1	Persistence effects
ρ_1	Brand effects

in fundraising expenditure by charity i in time t, whilst the *collective* returns additionally account for the relative changes in fundraising revenue that accrue due to a relative change in the average fundraising expenditure of *rivals* to charity i in time t. In terms of elasticities, the relationship between the individual returns and collective returns can be represented as follows:

$$\overbrace{\underbrace{\frac{dF_{it}}{dB_{it}}\frac{B_{it}}{F_{it}}}_{\text{Individual returns}} + \frac{dF_{it}}{dA_{it}}\frac{A_{it}}{F_{it}}}^{\text{Collective returns}} \tag{3}$$

Given the interpretation of the parameters of interest as partial elasticities: $\frac{\partial F_{it}}{\partial B_{it}}\frac{B_{it}}{F_{it}} = \beta_0$, $\frac{\partial B_{it}}{\partial A_{it}}\frac{A_{it}}{B_{it}} = \omega_0$ and the total derivative of F_{it} with respect to A_{it}: $\frac{dF_{it}}{dA_{it}} = \gamma_0\frac{F_{it}}{A_{it}} + \beta_0\frac{\partial B_{it}}{\partial A_{it}}\frac{F_{it}}{B_{it}}$, substituting in the derivatives into the above condition yields:

$$\overbrace{\underbrace{\beta_0}_{\text{Individual returns}} + \beta_0\omega_0 + \gamma_0}^{\text{Collective returns}} \tag{4}$$

Which implies that the size of the collective returns to fundraising relative to the individual returns depends crucially on the direction of the direct spillover effects γ_0 and indirect effect $\beta_0\omega_0$ of the strategic and own fundraising effect combined [9].

Strictly speaking, F_{it} and B_{it} are both conditional expectations, conditional on the observables in each equation, such that the conditional unobservables do not vary with the observable variables[10]. Hence it is useful at this point to examine the unobservable mean-zero error term ε_{it} and its descendant g_{it}[11] which are assumed to be uncorrelated across charities. ε_{it} captures the idiosyncratic determinants of fundraising revenue ignored by the

[9] The direct and indirect effects that constitute the total elasticity of average rival fundraising expenditure with respect to fundraising revenue are represented by the two paths from A_{it} to F_{it} in Figure 2.

[10] This requires $\mathbb{E}[E_{it}|F_{it-1}, B_{it}, B_{it-1}, A_{it}] = e^{\rho_0}$ and $\mathbb{E}[G_{it}|B_{it-1}, F_{it-1}, A_{it}] = e^{\phi_0}$, which is similar to strict exogeneity.

[11] g_{it} is a descendant of ε_{it} since G_{it} could be interpreted as the *expected* component of E_{it} conditional on a charity's past observables - as can be seen from Equation 2

model, which can be broadly categorised into charity-driven and donor-driven unobservable factors. Charity-driven unobservables consist of factors which are under the control of the charity, including inputs such as managerial quality, fundraising techniques and transparency, whereas donor-driven unobservables are not under the control of the charity and taken to be exogenous, including donor preferences and external events affecting the real or perceived vulnerability of the charity's beneficiaries. For example, an increase in extreme weather events leading to natural disasters can raise public awareness of man-made climate change, leading to an exogenous increase in donations to Green Charities.

I posit that it is the charity-driven unobservable factors that best characterise an inherent fundraising efficiency that the fundraising ratio R_{it} is designed to measure, analogous to Total Factor Productivity (TFP) in the context of production functions. Substituting the fundraising production function for F_{it} into the definition of the fundraising ratio R_{it} and F_{it-1} for $R_{it-1}B_{it-1}$ yields:

$$R_{it} = \frac{F_{it}}{B_{it}} = E_{it}(R_{it-1}B_{it-1})^{\rho_1} B_{it}^{\beta_0 - 1} B_{it-1}^{\beta_1} A_{it}^{\gamma_0}$$

taking logs and assuming that ε_{it} can be decomposed linearly into a charity-driven component ε_{it}^c and a donor-driven component ε_{it}^d yields:

$$r_{it} - \varepsilon_{it}^c = \varepsilon_{it}^d + (\beta_0 - 1)b_{it} + (\beta_1 + \rho_1)b_{it-1} + \gamma_0 a_{it} + \rho_1 r_{it-1} \quad (5)$$

Where the left-hand-side gives the bias in the fundraising ratio in logarithms as a measure of the true fundraising efficiency ε_{it}^c and the right-hand-side shows the factors that determine the size and direction of the bias. Equation 5 shows that unobservable donor shocks ε_{it}^d can directly affect the bias, whilst the bias induced by rival spending A_{it} depends on the sign of the spillover effect γ_0. In addition, if we assume that fundraising expenditures are stable over time ($b_{it} \approx b_{it-1}$), there exist diminishing returns to scale ($\beta_0 + \beta_1 < 1$) and fundraising revenues are very volatile ($\rho_1 << 1$) then Equation 5 predicts that larger charities (with higher fundraising budgets) have artificially low fundraising ratios, ceteris parabis.

Finally, given an estimate of the parameters in the Production Function and a method for purging ε_{it}^d from ε_{it}, it is possible to estimate the true fundraising efficiency ε_{it}^c for each charity in every time period.

4 Data

I have extracted data from the Charity Commission (CC) Register, which holds yearly information on all registered charities in England and Wales including their stated beneficiaries or causes, and detailed financial information for all charities reporting over £500,000 in total income. The data is made publicly available as a data dump and is updated every month starting in 2007. This information is taken from the Statement of Financial Activities (SoFA) submitted to the CC by individual charities every financial year. I have used these data extracts to construct a panel dataset, in which the charity identifier i and year t define an observation.

This paper will focus on so called "Green Charities", which are defined as those that list *Animals* (code 111) or *Environment/Conservation/Heritage* (code 112) as one of their main beneficiaries or causes. In practice, this definition characterises a large subsector of charities that describe themselves as supporting, perhaps among others, an environment-related charitable cause; this may encompass climate change advocacy groups, wildlife protection groups and renewable energy research charities in addition to animal rights charities to list but a few.

In general, charities have two main income categories[12]:

1. Generated Income- including donations, in-kind gifts, investments, non-charitable sales and legacies

2. Charitable Income- which includes sales resulting directly from providing charitable services (e.g. care home fees)

Charities have three main sources of expenditure:

1. Costs associated with generating funds - including fundraising costs, investment management costs and trading costs associated with non-charitable sales

2. Charitable expenditure - which includes any spending deemed to impact the beneficiaries directly

3. Governance costs

Since many charity projects and personnel are multi-functional, charities are given reasonable discretion as to the allocation of costs between different activities; for example, a charity manager may split their time between fundraising activities, staff management and charitable activities, hence their total wage costs are evenly split between the expenditure categories.

[12] This information is taken from Charities SORP

The Fundraising Revenue of charity i in time t, F_{it}, is defined as the total voluntary income (donations, legacies and endowments) and activity generated funds. [13] The Fundraising Expenditure of charity i in time t, B_{it} is defined as the cost of raising voluntary income and the trading costs, which include the operating costs associated with commercial (non-charitable) activities like charity shops.

These definitions enable the fundraising activity (income and expenditure) to be isolated from any other activities of the charity, strictly pertaining to activities that involve raising voluntary income from the public. This includes commercial activities, but excludes any activities associated with the charity's beneficiaries, investment or governance activities.

The definition of A_{it}, the average fundraising expenditure of charity i's rivals in time t, requires strong assumptions as to which other charities constitute rivals to charity i in time t and thus which other charities *exclusively* have the ability to influence the fundraising decision-making (via strategic effects) and fundraising outcomes (via spillover effects) of charity i in time t.

I assume three main criteria to identify which charities should constitute rival charities to a given charity i in time period t:

1. Same area of operation[14] as charity i
2. Same beneficiaries[15] as charity i
3. Similar size to charity i in time t

The first two criteria are relatively non-contentious. The final criterion turns the definition of A_{it} into a weighted-average of other charity's fundraising expenditure, in which the weights assigned to other charities critically depend on definition of size, the strength of similarity between rivals and stability over time. I define rival charities to be those in the same area of operation with the same beneficiaries, and at most ± 30 places away from charity i in the ranking of charitable expenditure in a given year t. The justification for this is as follows: due to the high discretion given to charities as to the allocation of their costs, there exist large incentives for charities to strategically choose their level of charitable expenditure (Dang and Owens (2020)) in keeping with the charitable expenditure of other charities they consider to be peers. This enables them to maintain a similar programme ratio [16] to their peers thereby avoiding public scrutiny. In addition, a report[17] from the Charity Commission suggests that around 23% of charities collaborate with charities of a

[13] I am not able to distinguish between restricted income (which can only be spent on charitable activities) and unrestricted income using this dataset, although typically the cost of raising unrestricted income is significantly higher than raising restricted income.

[14] the areas of operation are either international or national

[15] the three groups of beneficiaries are only animals, only environment or both

[16] the ratio of charitable expenditure to total expenditure

[17] Strength in Numbers

similar size. It can be calculated[18] that any given charity considers around 60 other charities to be potential collaborators of a *similar size* on average. This justifies using similarity in charitable expenditure as a measure of the *distance* between charities and ± 30 places to specify the maximum distance between rivals.

This definition implies the following about rival charities: (1) any charity has at most 60 and at least 30 rivals in any time period, (2) rivals may differ every year and (3) if charity i and charity j are rivals at time t, they may have shared rivals and non-shared rivals.

A_{it} has also been constructed to only average the *cost of raising voluntary income* over rivals, which excludes trading costs. This is because it is unlikely that trading activity (e.g. operating charity shops) induces spillover and strategic effects across rival charities, hence including these expenditures introduces unnecessary noise.

I use minimum total labour cost C_{it} instrument for fundraising expenditure to estimate the Production Function[19]. C_{it} is calculated by multiplying the total number of employees in charity i in time t by the national minimum/living[20] wage in nominal terms, thereby indicating the minimum possible wage bill for charity i in time t.

From the population of all registered Green Charities with over £500k in income, (approximately 2600 in total across the years 2007-2020) I have selected a sample of approximately 1700 charities across the years 2007-2015 in order to improve balance in the panel and account for misreporting. This involves restricting the time period to 2007 - 2015, as fundraising expenditure is not reported independently of other expenditure after 2015, and selecting only charities above 14 years of age, so as to avoid new charities entering into the dataset creating unnecessary panel imbalance. [21]. Charities reporting negative fundraising expenditures and revenues have been excluded, in addition to charities reporting financial year lengths outside of the range (359, 371).

Table 1 summarises the four variables of interest in addition to the fundraising ratio $R_{it} = \frac{F_{it}}{B_{it}}$ in the sample, whilst Table 2 provides a breakdown of the variation within and between charities in the sample. With an average fundraising expenditure of around £670k a year and revenue of around £2.5m a year, the charities in the sample raise £52 for every

[18]The report suggests that in 2010, 45% of small charities have collaborated at some point over the last 2 years, 51% of whom are collaborating with charities of a *similar* size. I assume that: (1) charities only collaborate with *potential collaborators* with positive probability, where potential collaborators are defined as having the same beneficiaries and area of operation, and (2) that the probability of collaborating with another charity, given that they are a potential collaborator of a similar size, is equal to 1. Using the law of conditional probability, this probability $Pr(collaborate|potential\ collaborator\ of\ similar\ size) = 1$ can be rewritten as $\frac{Pr(potential\ collaborator\ of\ similar\ size|collaborate).Pr(collaborate)}{Pr(potential\ collaborator\ of\ similar\ size)} = 1$, where the first term in the numerator is approximated as 0.51 and the second term in the numerator is approximated as 0.45 using the data from the report. Assuming there are x potential collaborators of *similar* size out of a sample of approximately 250 (the average number of other *potential collaborators* of a charity in a given year in my sample), the probability in the denominator can be approximated as $\frac{x}{250}$. Rearranging yields $x = 57$ which can be approximated as 60. However, this is analysis is approximate and therefore is only used as a benchmark: I also consider variations of this definition in section 7.

[19]I discuss the relevance and validity of the Production Function in section 5

[20]The minimum wage was renamed the living wage in 2016

[21]the implications for sample selection bias will be discussed in section 5

£1 spent on fundraising. Rival Expenditure (A_{it}) is on average £ 470k lower than own charity fundraising expenditure (B_{it}) as B_{it} also includes trading costs, which are excluded from A_{it}. Whilst the amount of between variation and within variation is similar for F_{it} (Revenue), there exists substantially less variation in B_{it} (Fundraising Expenditure) and A_{it} (Rival Expenditure) and C_{it} (Minimum total labour cost) within charities (across time) than between different charities, suggesting that fundraising expenditure and staff costs may be relatively stable across time.

Table 1: Sample Summary Statistics

	Mean	SD	Min	Max	Number of observations
B_{it} (Expenditure)	669,999	3,147,112	0	57,938,000	9,794
F_{it} (Revenue)	2,492,707	10,677,858	0	494,871,392	9,801
A_{it} (Rival Exp.)	197,102	349,561	0	2,157,057	9,796
R_{it} (Ratio)	52	804	0	55,539	6,356
C_{it} (Min. labour cost)	391	1,455	0	39,523	9,800

Table 2: Standard Deviation between and within sampled charities

	Standard Deviation		
Variable	Within	Between	Overall
B_{it} (Expenditure)	627,532	2,521,730	3,147,112
F_{it} (Revenue)	5,776,477	7,291,418	10,700,000
A_{it} (Rival Expenditure)	83,952	314,388	349,561
R_{it} (Ratio)	666	615	804
C_{it} (Min. labour cost)	397	1,142	1,455

Table 3 reports summary statistics for the population of registered Green Charities with over £500k in total income. Relative to the population, the sampled charities appear to spend more and gain less from fundraising on average, reporting a substantially smaller fundraising ratio. However, the reduction in sample size (from around 2500 to 1700 charities) allows for more balance, with the average number of time periods reported per charity increasing from 5.20 to 5.85.

Table 3: Summary Statistics for all Green Charities

	Mean	SD	Min	Max	Number of observations
B_{it} (Expenditure)	638,159	3,080,555	-7,565	57,938,000	11,075
F_{it} (Revenue)	2,592,065	11,484,093	-30,743	637,600,000	16,990
A_{it} (Rival Exp.)	192,191	341,721	0	2,157,057	11,074
R_{it} (Ratio)	85	3,097	-147,258	182,134	7,074
C_{it} (Min. labour cost)	426	1,666	0	61,271	16,988

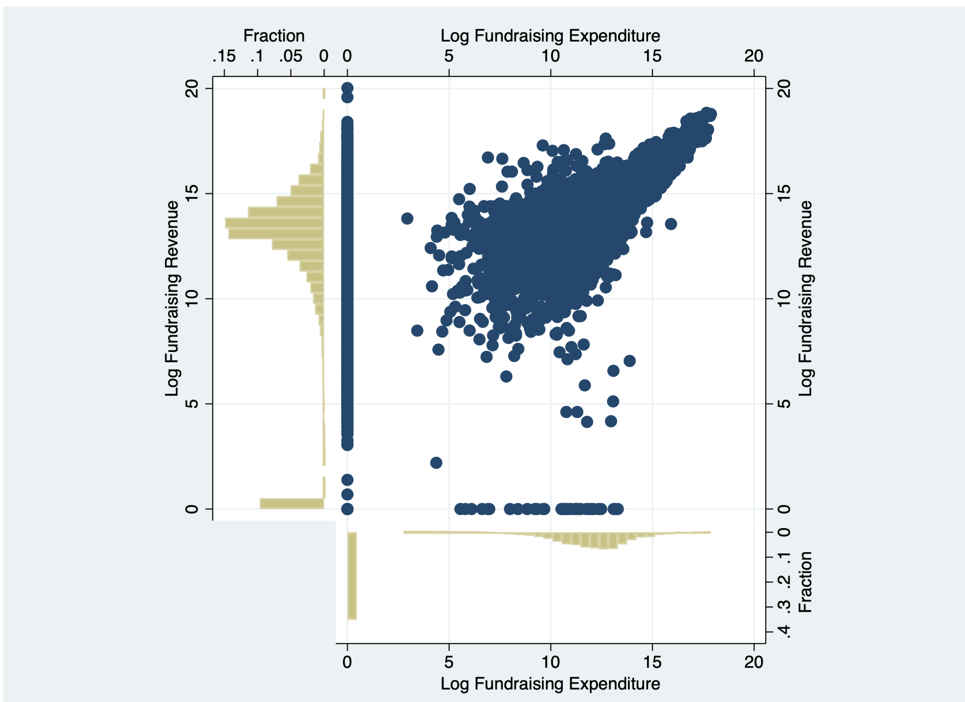

Figure 3: Correlation between f_{it} and b_{it}

Figure 4: Correlation between b_{it} and a_{it}

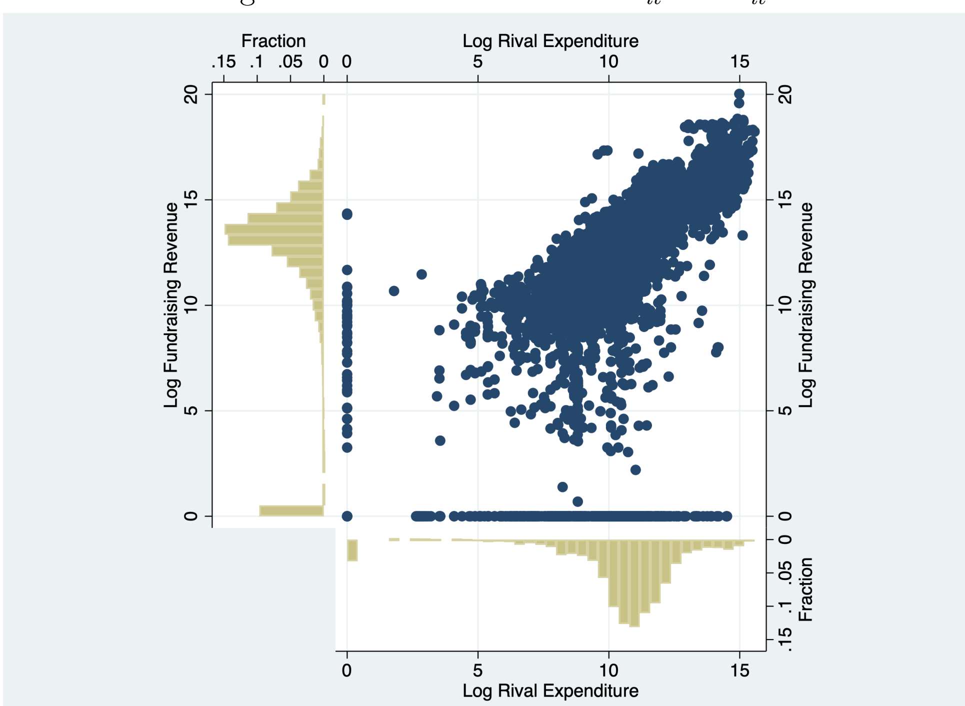

Figure 5: Correlation between f_{it} and a_{it}

Figure 3, Figure 4 and Figure 5 show the pairwise scatter-plots of the pooled variables of interest in logarithms[22] and the associated histograms. There is a clear positive correlation between all 3 variables, indicative of a **possible** causal link between them. In the next section, I present an empirical strategy to identify the causal relationships specified by the fundraising Production Function and fundraising Decision Function from section 3, using the data outlined in this section.

[22] I have in fact applied the $ln(1 + x)$ transformation to avoid generating unnecessary missing values at $x = 0$. $ln(1 + x) \approx ln(x)$ when x is large.

5 Identification Strategy

In this paper, I seek to consistently estimate the parameters in the system of equations that together characterise the fundraising process. Recall from section 3 the Fundraising Production Function and the Fundraising Decision Function respectively:

$$f_{it} = \rho_0 + \rho_1 f_{it-1} + \beta_0 b_{it} + \beta_1 b_{it-1} + \gamma_0 a_{it} + \varepsilon_{it}$$
$$b_{it} = \phi_0 + \phi_1 b_{it-1} + \psi_1 f_{it-1} + \omega_0 a_{it} + g_{it}$$

Where f_{it} is the log of fundraising revenue, b_{it} is the log of own fundraising expenditure, a_{it} is the average fundraising expenditure of rivals and ε_{it}, and its descendant g_{it}, are mean-zero error terms assumed to be uncorrelated across charities. Both error terms can be orthogonally decomposed further into a charity-specific (time-invariant) component η_i, a time-specific subsector shock that affects all charities equally m_t, and an idiosyncratic component potentially correlated across time v_{it}.

In this section, I first outline the threats to identification of the parameters in the models above and then provide an empirical strategy that best mitigates these threats.

In the ideal experiment, fundraising budgets each year are randomly allocated to different charities and revenue shocks are serially uncorrelated over time - this creates exogenous variation in b_{it}, f_{it} and a_{it} that can be used to identify all nine parameters. In reality, strict exogeneity is unlikely to hold due to endogeneity and the models above are potentially misspecified. I consider both of these threats in turn.

5.1 Threats to Identification: Endogeneity

The largest source of endogeneity arises from the fact that fundraising expenditure b_{it} is a choice variable, that is chosen to maximise net revenues in *response* to contemporaneous productivity shocks and donor preference shocks - captured by ε_{it} (and its descendant g_{it}); hence $cov(b_{it}, \varepsilon_{it}) \neq 0$ creating endogeneity in the Production Function [23]. For example, increased extreme weather events leading to wildfires and other environmental degradation in time t, could trigger an exogenous increase in donations to charity i, as donors become more aware and sensitised to the issue of climate change. Simultaneously, charity i could respond to the crisis by increasing fundraising efforts, attempting to capitalise on the heightened sensitivity. This creates a positive dependence between b_{it} and ε_{it} in the Production Function and potential upward (omitted variable) bias on β_0. To the extent that such shocks impact all charities homogeneously, they could be fully captured by m_t^{ε} (the time-specific component

[23] This can also be seen directly from the construction of B_{it}^{opt} in Equation 2, in which the optimal level of fundraising spending depends on $\mathbb{E}[E_{it}|X^P]$, where X^P is the matrix of regressors in the Production Function

of ε_{it}) [24]. Furthermore, to the extent that rivals respond to the donor preference shock in a similar way, this endogenises a_{it} in both the Production Function and the Decision Function.

Another source of endogeneity is due to serial correlation in the error terms ε_{it} and g_{it}. For example, if sensitivity to climate change is correlated over time, this could make b_{it-1} and f_{it-1} endogenous in both the Production Function and the Decision Function as they are related to the error terms through ε_{it-1} and g_{it-1} respectively. This creates so called Nickell Bias (1980) on the lagged dependent variables in the Production Function and the Decision Function.

A third source of endogeneity arises due to a reverse causal chain in the Production Function: higher fundraising revenue can *enable* higher fundraising expenditure in the same year, creating dependence between b_{it} and ε_{it} in the Production Function [25]. This can create very strong upward bias on the β coefficients.

Measurement error in F_{it} or B_{it} can also be a source of endogeneity. If we assume that each charity has a tendency to inflate (deflate) their revenues (expenditures) by a certain proportion, and this tendency is possibly related to the size of a charity, then this could create artificial dependence between regressors and error terms. To the extent that the propensity to misreport is time-invariant, then this measurement error may be captured fully by η_i^{ε} [26].

The final source of endogeneity arises from sample selection bias, caused by selecting observations on total income and by age.

The Charity Commission only requires charities with over £500k in annual income in a given year to report a detailed financial breakdown, hence the dataset only records such observations, allowing charities to drop in and out of the dataset depending on their total

[24] Variation in ε_{it} is not necessarily donor-driven; charity-driven variation such as changes in employee quality, managerial quality or governance structure can create endogeneity and positive omitted variable bias in a similar way.

[25] My specification of the Decision Function only allows for *lagged* enabling effects ψ_1, not *contemporaneous* enabling effects, with the latter creating a significant endogeneity concern

[26]

To see this, suppose we want to estimate the true fundraising production function:

$$\widetilde{F_{it}} = E_{it}\widetilde{F_{it-1}^{\rho_1}}\widetilde{B_{it}^{\beta_0}}\widetilde{B_{it-1}^{\beta_1}}A_{it}^{\gamma_0}$$

But we only observe the variables after they are artificially inflated or deflated relative to the true value by a constant fraction M_i - specific to each charity (e.g. observed $F_{it} = M_i^F\widetilde{F_{it}}$ and observed $f_{it} = \mu_i^F + \widetilde{f_{it}}$ in logs). Substituting the true variables for the observable counterparts and taking logs yields:

$$\frac{F_{it}}{M_i^F} = E_{it}\left(\frac{F_{it-1}}{M_i^F}\right)^{\rho_1}\left(\frac{B_{it}}{M_i^B}\right)^{\beta_0}\left(\frac{B_{it-1}}{M_i^B}\right)^{\beta_1}A_{it}^{\gamma_0}$$

$$f_{it} = \rho_0 + \rho_1 f_{it-1} + \beta_0 b_{it} + \beta_1 b_{it-1} + \gamma_0 a_{it} + \varepsilon_{it} - (\beta_1 + \beta_0)\mu_i^B + (1-\rho_1)\mu_i^F$$

Where the final unobservable term $-(\beta_1 + \beta_0)\mu_i^B + (1-\rho_1)\mu_i^F$ represents the measurement error. This creates a downward bias on the β coefficients and a positive bias on ρ_1, but the direction becomes indeterminate if the measurement errors μ_i are related to the true variables. The same reasoning can be applied to the Decision Function.

income level. Since only 7% of registered charities in the UK reported an annual income of over £500k in 2019 [27], the threshold for selection into the dataset is potentially very strict. Selecting on outcome creates a strong selection bias in the Production Function, and since g_{it} is strongly dependant on ε_{it}, amounts to selection-on-unobservables in the Decision Function, also creating selection bias. This results in a negative correlation between the error term and the regressors, since for lower values of the regressors, the error must be higher on average in order to be selected into the sample, creating a downward bias on the estimates[28].

Selection bias is also created by selecting into the sample only those charities above 14 years of age, hereon "older" charities [29]. To the extent that ε_{it} (and g_{it}) differ between older and younger charities on average perhaps due to higher managerial quality captured by η_i^{ε}, or larger donor preference shocks captured by m_t^{ε}, this could create dependence between error terms and regressors, creating upward bias in the estimated coefficients in the Decision Function and the Production Function. [30]

[27] BBC article

[28]

Downward bias in the coefficients in the Production Function can be represented mathematically using selection term that is omitted from the equation and is negatively correlated with the regressors. I use an indicator variable $s_{it} = 1$ to denote selection into the sample (Aguirregabiria (2012)):

$$f_{it} = \rho_0 + \rho_1 f_{it-1} + \beta_0 b_{it} + \beta_1 b_{it-1} + \gamma_0 a_{it} + \varepsilon_{it}|_{s_{it}=1}$$

Where $\mathbb{E}[\varepsilon_{it}|_{s_{it}=1}|X^P]$ is equivalent to $\mathbb{E}[\varepsilon_{it}|X^P, s_{it} = 1]$, where X^P is the matrix of regressors in the Production Function. The condition $s_{it} = 1$ can be rewritten as a condition on ε_{it}. Using NF_{it} to represent non-fundraised income to charity i in time t:

$$F_{it} + NF_{it} > 500,000 \Leftrightarrow F_{it} > 500,000 - NF_{it} \Leftrightarrow f_{it} > f(-NF_{it}) \Leftrightarrow \varepsilon_{it} > \varepsilon(-NF_{it}, -X^P)$$

Thus, the Production Function can be rewritten as:

$$f_{it} = \rho_0 + \rho_1 f_{it-1} + \beta_0 b_{it} + \beta_1 b_{it-1} + \gamma_0 a_{it} + \lambda(-NF_{it}, -X^P) + \widetilde{\varepsilon_{it}}$$

Where:

$$\lambda(-NF_{it}, -X^P) = \mathbb{E}[\varepsilon_{it}|\varepsilon_{it} > \varepsilon(-NF_{it}, -X^P), X^P]$$
$$\widetilde{\varepsilon_{it}} = \varepsilon_{it}|_{s_{it}=1} - \mathbb{E}[\varepsilon_{it}|\varepsilon_{it} > \varepsilon(-NF_{it}, -X^P), X^P]$$

$\lambda(-NF_{it}, -X^P)$ is a selection term omitted from the regression but negatively correlated with the regressors, creating downward bias in the estimated coefficients. I will refer back to the concept of a selection term throughout this paper.

Selecting on total income similarly creates selection bias in the Decision Function, since it introduces the selection term: $\lambda(-NF_{it}, -X^P) = \mathbb{E}[g_{it}|\varepsilon_{it} > \varepsilon(-NF_{it}, -X^P), X^D]$ which is non-zero since g_{it} and ε_{it} are highly correlated and decreasing in the regressors of the Decision Function(X^D) since X^D and X^P are very similar regressor matrices.

[29] See section 4 for the practical justification behind selecting these charities

[30] In a similar way to selecting on total income, selecting on age corresponds to selection on unobservables, creating the omitted selection term $\lambda(age_{it}) = \mathbb{E}[\varepsilon_{it}|age_{it} > 14, X^P]$ in the Production Function that is increasing in the regressors, since age is positively correlated with the regressors, hence creating upward bias in estimated coefficients.

5.2 Threats to Identification: Misspecfication

The second threat to identification arises due to potentially mis-specifying the functional form of the structural models. Largely for ease of interpretation, I have assumed the Production Function and the Decision Function are linear in parameters, which represent constant partial elasticities. This implicitly imposes continuity in the true Decision Function and Production Function.

However, there may exist discontinuities in the marginal cost of fundraising: if high sunk costs or entry barriers to fundraising exist, this could create large discontinuity at $b_{it} = 0$. For example, charities with lower brand recognition may face large sunk costs to fundraising compared to charities with high brand recognition, since large initial fundraising expenditures are required in order to generate a return. This forces these charities to spend either zero or some large amount on fundraising, creating discontinuity at $b_{it} = 0$ and resulting in non-linearity, biasing the estimated coefficients in the Decision Function [31]. There may also be discontinuity at $f_{it} = 0$ since increased public scrutiny may disincentivise charities to report positive fundraising revenues, causing charities that receive low fundraising revenues to report them as zero or to only rely on non-fundraised income, such as government grants or investment income. This creates non-linearity in the true Production Function.

A naive solution is to ignore the observations at $f_{it} = 0$ and $b_{it} = 0$ and estimate the models conditional-on-positive (COP) fundraising revenues in the Production Function and fundraising expenditure in the Decision Function. This creates selection bias similar but smaller in magnitude to the bias created by the £500k total income threshold, resulting in downward bias in the estimated parameters of both models. [32].

[31] This can be seen by examining the CEF under left-censoring at zero. Examining the CEF of the Decision Function:

$$\begin{aligned}
\mathbb{E}[b_{it}|X^D] &= 0.Pr(b_{it} = 0|X^D) + \mathbb{E}[b_{it}|X^D, b_{it} > 0].Pr(b_{it} > 0|X^D) \\
&= \mathbb{E}[b_{it}|X^D, b_{it} > 0].Pr(b_{it} > 0|X^D) \\
&= \mathbb{E}[b_{it}|X^D, b_{it} > 0].Pr(g_{it} > g(-X^D)|X^D) \\
&= \mathbb{E}[b_{it}|X^D, b_{it} > 0].f(-X^D)
\end{aligned}$$

Which is nonlinear in the regressors X^D of the Decision Function even if the underlying CEF is linear.

[32] For example, in the Decision Function, selecting only observations with positive fundraising expenditures b_{it} introduces the selection term $\lambda(-X^D) = \mathbb{E}[g_{it}|b_{it} > 0, X^D] = \mathbb{E}[g_{it}|g_{it} > g(-X^D), X^D]$ in a similar way to selecting on total income introduced in footnote 28. $\lambda(-X^D)$ is negatively correlated with the regressors in the Decision Function hence introduces a downward bias on the estimated coefficients in the Decision Function. Downward bias of a smaller magnitude is also created if the conditioning is probabilistic, since this introduces the selection term: $\lambda(-X^D) = \mathbb{E}[g_{it}|Pr(b_{it} > 0) > p, X^D] = \mathbb{E}[g_{it}|Pr(g_{it} > g(-X^D)) > p, X^D]$, which is also non-zero and decreasing in the regressors.

5.3 Identifying the Fundraising Production Function

After identifying donor preference trends and shocks, e.g. wildfires, that affect all charities equally to be a major source of omitted variable bias, the simplest solution is to include time dummies in the regression equation, thus partitioning out any (purely time-variant) variation that does not vary between charities. This transforms the error term ε_{it} such that m_t^{ε} is purged, reducing the dependence between regressors and errors.

However, the largest sources of omitted variable bias are most likely charity specific omitted variables captured by η_i^{ε}, such as managerial or employee quality[33] hence an obvious strategy would be to implement a within transformation or first difference transformation on the regression equation to purge the charity specific component η_i^{ε} from the error term. This may also solve the problems caused by measurement error, also plausibly captured by η_i^{ε} [34], and significantly reduce serial correlation in the error term, reducing the dependence between lagged variables and the error term. However, introducing a v_{it-1}^{ε} term into the error via such a transformation creates so called Nickell Bias (1980) in the estimated coefficients as it creates artificial correlation between the transformed error and the transformed lagged dependent variable containing f_{it-1}. Furthermore, in order to produce consistent estimates using OLS on the transformed model, I require strict exogeneity $\mathbb{E}[v_{it}^{\varepsilon}|\eta_i^{\varepsilon}, m_t^{\varepsilon}, X^P]$, which is not satisfied due to the existence of a lagged dependant variable.

Selecting charities based on age may also induce selection bias [35]. However, if age is only related to charity specific factors such as managerial quality captured by η_i^{ε} or donor preference shocks captured by m_t^{ε}, then purging m_t^{ε} and η_i^{ε} from the error term using time dummies and a first-difference transformation can mitigate this selection bias [36].

In order to mitigate the selection bias caused by charities dropping in and out of the sample based on total income, I have chosen to restrict the sample of charities further to only those that report an average non-fundraising income over the time period of over £1m; this effectively excludes charities from the sample that are close to the threshold for selection, and are thus more likely to be selecting into the sample based on fundraising revenues. The charities that remain in the sample have a larger (non-fundraised) income buffer to allow them to remain in the sample regardless of the level of fundraising revenue. Although non-fundraised income is not entirely independent of ε_{it}, selecting on non-fundraised income is likely to significantly reduce selection bias [37].

[33] Sieg and Zhang (2012) use a Olley-Pakes control-function approach to estimate a fundraising production function for Green Charities in the US, finding that managerial quality, measured by governance expenditure, is a strong determinant of donations. Hence, I control for governance costs in the regression to capture any confounding effect of managerial quality on fundraising revenues and fundraising expenditure.

[34] as outlined in footnote 26

[35] as outlined in footnote 30

[36] this requires that $\mathbb{E}[v_{it}^{\varepsilon}|age_{it}, X^P] = 0$

[37] Selecting on a larger average non-fundraised income $\overline{NF_{it}}$ attenuates the size of the selection term $\lambda(-NF_{it}, -X^P) = \mathbb{E}[\varepsilon_{it}|\varepsilon_{it} > \varepsilon(-NF_{it}, -X^P)]$ introduced in footnote 28, since it effectively forces $\varepsilon(-NF_{it}, X^P)$ towards negative infinity. Assuming errors are mean zero normally distributed, this attenuates the expected value of ε_{it} and hence the selection term.

In order to mitigate the non-linearity caused by discontinuity in fundraising revenues at $f_{it} = 0$, I choose to restrict the sample to only those charities that report at least one positive fundraising expenditure over the period $\overline{b_{it}} > 0$. This excludes charities from the sample that may be disincentivised to report positive fundraising revenues, since they are not expected to receive positive fundraising revenues with zero fundraising expenditure, and would therefore avoid public scrutiny by reporting $f_{it} = 0$. This also introduces selection bias [38] since positive fundraising expenditures are likely to be induced or enabled by a larger ε_{it}, but this poses a smaller threat to identification than a potential non-linear data-generating-process.

The final source of endogeneity not addressed is the existence of a reverse-causal chain: higher fundraising revenues *enabling* higher fundraising expenditure in the same year, creating dependence between ε_{it} and b_{it}. To address this, I propose using the minimum total labour cost to charity i in time t (c_{it} in logarithms) as instrumental-variable that creates exogenous variation in fundraising expenditure b_{it} [39]. c_{it} is the product of the total number of employees in charity i in time t and the national minimum/living wage in time t, hence is the smallest possible wage bill of charity i in time t. Relevance is provided by the fact that wage costs (associated with fundraising) often constitute a large proportion of fundraising expenditures and the living wage is often binding across the charity sector; hence $cov(b_{it}, c_{it}) \neq 0$[40]. The exclusion restriction is likely to be satisfied since minimum total wage costs are unlikely to affect fundraising revenues directly, only indirectly through changes to the cost of fundraising expenditure. Instrument validity $\mathbb{E}[c_{it}\varepsilon_{it}] = 0$ is harder to justify as, although the national minimum/living wage is plausibly exogenous and satisfies the exclusion restriction[41], the number of employees (the most variable component of c_{it}) may be responsive to the same donor preference shocks and productivity shocks captured by ε_{it}. For example, a man-made environmental disaster may trigger an exogenous increase in donations to green charities, captured by m_t^{ε}, enabling or inducing a charity to hire more employees. Another example is that charities with higher employee quality, captured by η_i^{ε} may hire fewer employees.

Hence, I choose to impose a weaker but more reasonable validity condition $\mathbb{E}[c_{it}v_{it}^{\varepsilon}] = 0$ thereby allowing the number of employees to vary with η_i^{ε} and m_t^{ε}, but maintaining that charities cannot contemporaneously adjust employees to idiosyncratic shocks captured by

However, it also introduces a new selection term $\mathbb{E}[\varepsilon_{it}|\ \overline{NF_{it}} > £1m, X^P]$, which is plausibly non-zero and related to the regressors, thus introducing bias. However, if $\overline{NF_{it}}$ only depends on the charity-specific productivity η_i^{ε}, then $\mathbb{E}[v_{it}^{\varepsilon}|\ \overline{NF_{it}} > £1m, X^P] = 0$ and purging η_i^{ε} from the error term through first differencing will remove the new selection term. This is a plausible assumption: if non-fundraising revenue and fundraising revenue come from different sources, then non-fundraised income may not be influenced by time-specific shocks v_{it}^{ε} and m_t^{ε} that influence fundraising revenue

[38] Conditioning on positive fundraising expenditures introduces the selection term $\mathbb{E}[\varepsilon_{it}|\bar{b}_{it} > 0, X^P]$ which is non-zero since $\bar{b}_{it}$ is endogenous.

[39]the use of a valid instrument for Δb_{it} also enables a relaxation of the assumption of strict exogeneity to the idiosyncratic error term v_{it}^{ε}, allowing omitted variables and measurement error to vary within charities over time.

[40]Relevance may be weakened by the fact that charities may reduce the non-labour fundraising expenditure or reduce the number of employees to compensate for higher labour costs.

[41]This assumes that changes to the minimum wage do not effect donations to green charities directly.

v^{ε}_{it}. This is plausible since it most likely takes longer than one year to adjust the number of employees to idiosyncratic shocks to fundraising revenue.

Implementing an instrumental variables strategy using the moment condition $\mathbb{E}[c_{it}v_{it}] = 0$ requires removing η^{ε}_i and m^{ε}_t from the error term. Performing a first-difference transformation and adding time dummies does exactly that, whilst additionally controlling for governance spending gov_{it} yields the preferred regression specification for estimating the Production Function:

$$\Delta f_{it} = \rho_0 + \rho_1 \Delta f_{it-1} + \beta_0 \Delta b_{it} + \beta_1 \Delta b_{it-1} + \gamma_0 \Delta a_{it} + \alpha_0 \Delta gov_{it} + \sum_{1}^{T} \alpha_t m^{\varepsilon}_t + \Delta v^{\varepsilon}_{it} \tag{6}$$

Transforming the Production Function in this way endogenises Δf_{it-1} and Δb_{it-1}[42] but also creates suitable candidate instruments for each endogenous variable. The transformation in Equation 6 enables past differences to be used as instruments for differenced endogenous variables (Anderson and Hsiao (1982)), presenting Δf_{it-2} as a suitable instrument for Δf_{it-1} and Δb_{it-2} as an instrument for Δb_{it-1}. Past differences may predict current differences with strong relevance even if the series is close to non-stationarity[43]. In addition, the instruments may be excluded if they do not influence fundraising revenues independently of the endogenous regressors, which requires that true brand effects and own fundraising effects are at most one period lagged [44].

This involves utilising two more moment conditions: $\mathbb{E}[\Delta f_{it-2} \Delta v^{\varepsilon}_{it}] = 0$ and $\mathbb{E}[\Delta b_{it-2} \Delta v^{\varepsilon}_{it}] = 0$, which are valid if f_{it-2} and b_{it-2} are uncorrelated with v^{ε}_{it-1} and v^{ε}_{it} is not serially correlated[45]; together with $\mathbb{E}[c_{it} \Delta v^{\varepsilon}_{it}] = 0$, [46] this creates a just-identified case with 3 endogenous variables and 3 instrumental variables, where Δa_{it} and gov_{it} are assumed to be exogenous variables.

[42]This is because Equation 6 reintroduces serial correlation in the error term (as $\Delta v^{\varepsilon}_{it}$ is correlated with $\Delta v^{\varepsilon}_{it-1}$ through the shared v^{ε}_{it-1}) making the lagged dependant variable Δf_{it} (containing f_{it-1}) endogenous to the transformed error, resulting in Nickell bias (1980). In addition, reverse causality results in contemporaneous correlation between the error and fundraising expenditure, making Δb_{it} and Δb_{it-1} endogenous to the transformed error (containing both v^{ε}_{it} and v^{ε}_{it-1}).

[43]I also test f_{it-2} and f_{it-3} as instruments for Δf_{it-1} and b_{it-2} and b_{it-3} for Δb_{it-1}, with levels as instruments improving relevance if the data is stationary and longer lags as instruments improving validity if there exists serial correlation in v^{ε}_{it}

[44]i.e. β_2 and ρ_2 do not exist in reality.

[45]This assumes that any dependence between lagged variables and the error term is eliminated once the fixed effects η^{ε}_i and m^{ε}_t are purged from the error.

[46]Note that this assumes v^{ε}_{it-1} and c_{it} are also uncorrelated, which is stronger than contemporaneous uncorrelatedness. Hence, I also test using c_{it-1} as an alternative instrument.

Table 4 presents a summary of the threats to identification and the strategies designed to mitigate them.

Table 4: Summary of Identification Strategy for estimating the Production Function

Threat to identification	Solutions
Omitted variable bias	1. Time FEs 2. First Differencing 3. Control for governance spending 4. Instrumental Variables
Reverse causality	1. Instrumental variables
Measurement error	1. First differencing 2. Instrumental variables
Serial correlation	1. First differencing 2. Instrumental variables
Selection bias	1. Restricting sample to only charities with an average non-fundraised income of over £1m 2. First differencing 3. Time FEs
Misspecification	1. Restricting sample to those reporting at least one positive fundraising expenditure

My strategy for identifying the Decision Function is similar, involving the implementation of a 2SLS procedure on the first-difference transformed model (employing lagged differences as instruments) in addition to introducing sample restrictions that mitigate selection bias and specification error. See subsection 10.1 for detailed outline of the strategy.

6 Results

In this section I present and describe the results of my preferred 2SLS estimation strategies alongside OLS estimates and summary statistics. In the next section I analyse and critique the results, subjecting the results to several robustness checks.

6.1 Estimating the Fundraising Production Function

Table 5: Selected Sample Summary Statistics (Production Function)

	Mean	SD	Min	Max	Number of observations
B_{it} (Expenditure)	1,782,576	5,467,322	0	57,938,000	3,013
F_{it} (Revenue)	4,973,977	14,592,970	0	152,343,008	3,013
A_{it} (Rival Exp.)	302,436	393,431	7,023	2,157,057	3,014
C_{it} (min. labour cost)	965	2,462	0	39,523	3,013

In Table 5 I summarise the four variables of interest, imposing the new sample restrictions introduced in the previous section[47] that attempt to mitigate selection bias and reduce specification error. Comparing Table 1 with Table 5, it is clear that imposing the new restrictions has produced a smaller sample of larger charities[48], spending on average around £ 1m more and raising £ 2.5m more on average. Imposing the fundraising expenditure restriction $\bar{b_{it}} > 0$ has reduced the discontinuity at $f_{it} = 0$, reducing the proportion of observations at $f_{it} = 0$ from 9.6% to 4.1%, which explains the higher average fundraising revenue and expenditure in the sample. Imposing the non-fundraised income threshold $\overline{NF_{it}} > £1m$ excludes charities likely to drop in and out of the dataset by providing a large enough income buffer, successfully improving balance in the panel: the average number of time periods reported per charity increases from 5.85 to 8.46 out of 9.

In Table 6 I present initial OLS estimates of the fundraising elasticities using the Production Function as a base model, where f_{it} and f_{it-1} are the current and previous year fundraising revenue respectively in logs whilst b_{it} and b_{it-1} are the current and previous year fundraising expenditure respectively in logs for a given charity i in time t. a_{it} represents the average fundraising expenditure of rivals to charity i in time t in logs. Comparing column 1 to column 2, the coefficients on the contemporaneous variables increase, whereas the coefficients on the lagged variables decrease as a result of the sample restrictions. The estimates appear to be robust to the inclusion of time fixed effects in column 3 and to the inclusion of governance expenditure as a control variable in column 4. Eliminating between variation in column 5 attenuates the magnitude of the estimated coefficients, whilst significantly reducing the R^2 from 0.745 to 0.236 and the F-statistic from 297.1 to 8.264. In addition, the

[47] $\overline{NF_{it}} > £1m$ and $\bar{b_{it}} > 0$

[48] the number of charities in the sample has decreased from 1671 to 354.

coefficient on a_{it} becomes insignificant, perhaps due to the lack of within-variation in average rival spending outlined earlier in Table 2.

Recall from section 5 my preferred regression specification for estimating the Production Function:

$$\Delta f_{it} = \rho_0 + \rho_1 \Delta f_{it-1} + \beta_0 \Delta b_{it} + \beta_1 \Delta b_{it-1} + \gamma_0 \Delta a_{it} + \alpha_0 \Delta gov_{it} + \sum_{1}^{T} \alpha_t m_t^{\varepsilon} + \Delta v_{it}^{\varepsilon} \tag{7}$$

Where Δf_{it-1}, Δb_{it-1}, Δb_{it} are endogenous regressors. Table 7 provides a summary of the first stage regressions of each endogenous variable, represented by different columns, on the instrument set $Z_{it} = [\Delta f_{it-2}, \Delta b_{it-2}, c_{it}]$. Whilst Δf_{it-2} and Δb_{it-2} appear to be relevant instruments for Δf_{it-1} and Δb_{it-1}, the external instrument minimum total labour cost in levels c_{it} appears to be weak. This results in a low F-statistic $F = 1.89$ and a low Sanderson-Windmeijer multivariate F-test statistic $SWF = 6.63$[49] in the first stage regression on the endogenous variable Δb_{it}. In addition, the Kleibergen-Paap (rk) Wald F-statistic is 1.764, hence we cannot reject the null hypothesis of weak identification [50].

Column 2 in Table 8 shows the 2SLS estimated coefficients of my preferred specification. The results provide evidence of a strong brand effect of fundraising revenues $\hat{\rho}_1 = 0.710$. The remaining coefficients are not statistically significant using clustered standard errors. The next largest determinant of fundraising revenue is contemporaneous fundraising expenditure: $\hat{\beta}_0 = 0.544\%$. This is followed by estimated spillover effects $\hat{\gamma}_0 = 0.157\%$ and the estimated lagged own fundraising effect $\hat{\beta}_1 = -0.269\%$.

See subsection 10.2 for a description of the estimation results for the Decision Function.

[49]The Sanderson-Windmeijer multivariate F test of excluded instruments is more robust test-statistic for the detection of weak instruments. See Sanderson and Windmeijer (2016) for details.

[50]the Kleibergen-Paap (rk) Wald F-statistic is chi-squared (1) distributed with a 10% critical value at 2.706.

Table 6: OLS Estimates

	(1) Full sample	(2) Selected sample	(3) Time FEs	(4) Controls	(5) TwowayFEs
f_{it-1}	0.800***	0.744***	0.745***	0.743***	0.364***
	(0.0125)	(0.0293)	(0.0292)	(0.0294)	(0.0629)
b_{it}	0.179***	0.266***	0.267***	0.267***	0.235***
	(0.0180)	(0.0307)	(0.0307)	(0.0307)	(0.0398)
b_{it-1}	-0.0959***	-0.162***	-0.164***	-0.164***	-0.0947***
	(0.0171)	(0.0268)	(0.0269)	(0.0269)	(0.0284)
a_{it}	0.0443	0.137***	0.138***	0.129***	-0.0409
	(0.0230)	(0.0365)	(0.0369)	(0.0378)	(0.127)
Observations	7759	2648	2648	2648	2648
R^2	0.761	0.744	0.745	0.745	0.236
Adjusted R^2	0.761	0.744	0.744	0.744	0.232
F	2626.8	767.9	288.4	297.1	8.264

Standard errors in parentheses.

All estimates are interpreted as elasticities.

The model specifications are cumulative, with each building on the model specification to its left.

Standard errors are robust to arbitrary heteroskedasticity and autocorrelation within charities.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: First Stage Summary

	(1) Δf_{it-1}	(2) Δb_{it}	(3) Δb_{it-1}
Δf_{it-2}	-0.245***	0.0745	0.0211
	(0.0628)	(0.0712)	(0.0642)
Δb_{it-2}	-0.0246	-0.0606	-0.261***
	(0.0324)	(0.0405)	(0.0440)
c_{it}	0.0694	0.0445	-0.00627
	(0.0439)	(0.0430)	(0.0414)
Observations	1926	1926	1926
R^2	0.073	0.015	0.074
Adjusted R^2	0.068	0.010	0.070
SWF	7.84	6.63	7.76
F	8.68	1.89	14.52

Standard errors in parentheses.

Each column corresponds to the first stage for the endogenous variable in the header.

Each row corresponds to a different instrument.

Standard errors are robust to arbitrary heteroskedasticity and autocorrelation within charities.

F is the F-statistic of the test for joint significance of the instruments.

SWF is the Sanderson-Windmeijer multivariate F test of excluded instruments.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: OLS and IV estimates of preferred regression specification for estimating the Production Function

	(1) OLS	(2) 2SLS (Preferred Strategy)
Δf_{it-1}	-0.263***	0.710***
	(0.0522)	(0.211)
Δb_{it}	0.244***	0.554
	(0.0333)	(0.543)
Δb_{it-1}	0.0423	-0.269
	(0.0262)	(0.179)
Δa_{it}	0.000403	0.157
	(0.123)	(0.263)
Observations	2284	1926
R^2	0.201	-0.748
Adjusted R^2	0.198	-0.757
F	6.965	2.988

Standard errors in parentheses.

All estimates are interpreted as elasticities.

The model specifications are cumulative, with each building on the model specification to its left.

Standard errors are robust to arbitrary heteroskedasticity and autocorrelation within charities.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7 Criticisms and Robustness Checks

7.1 Sample restrictions and selection bias

In this paper I impose two types of sample restrictions: the first set of restrictions introduced in section 4 are needed to produce a sufficiently coherent dataset, whilst the second set of restrictions introduced in section 5 are stronger and are aimed at mitigating selection bias (in particular the $\overline{NF_{it}} > £1m$ restriction) and misspecification bias (in particular the $\overline{b_{it}} > 0$ sample restriction used to estimate the Production Function and the $\overline{leg_{it}} > 0$ restriction used to estimate the Decision Function).

Examining column 1 and 2 of Table 6, and column 1 and 2 of Table 12, imposing the sample restrictions appears to have an indiscernible effect on the OLS estimated coefficients, which reflects the combined effect of both attenuating and introducing selection biases (which themselves differ in direction). In addition, none of the estimates are statistically significant, so it is impossible to separately identify the sources of selection bias even with further deconstructions of the sample restrictions. See subsubsection 10.3.1 for a deeper analysis of the effect of each restriction that utilises Table 19 and Table 18.

7.2 Mitigating endogeneity by first-differencing and including controls

The inclusion of charity-specific dummies in column 5 of Table 6 and Table 12 significantly attenuates the OLS estimated coefficients, which provides evidence that purging the η_i from the error term significantly mitigates the upward bias on the coefficients caused by time-invariant omitted variables, measurement error, serial correlation and selection bias. See subsubsection 10.3.2 for a discussion on why OLS on a first-differenced or within-transformed model will not yield consistent estimates.

The inclusion of governance spending does not appear to reduce positive omitted variable bias: the coefficients are stable across column 3 to column 4 of Table 6. Table 16 in section 10 indicates that governance spending, as a proxy for managerial quality, may not have a statistically or economically significant impact on fundraising revenues. This directly contradicts the findings of Sieg and Zhang (2012).

7.3 Mitigating endogeneity by including time-fixed effects

In section 5, I posited the existence of sector-wide donor preference shocks and trends that influence fundraising revenues, plausibly captured by m_t^{ε}, which charities and their rivals respond to by intensifying fundraising efforts- thereby confounding the effect of rival and own fundraising expenditure in the Production Function. This justifies the inclusion of time fixed effects in the Production Function. In addition, if these donor preference trends influence fundraising expenditures in the same way, then this confounds the effect of rival expenditure on own expenditure in the Decision Function. This justifies the inclusion of time fixed effects in the Decision Function. However, examining the results presented in Table 6, Table 16, Table 12, Table 17, Table 20, Table 21, Table 22 and Table 23, it is evident that trends and shocks may differ in magnitude and direction across charities, and thus are not likely to be captured by single year dummies. See subsubsection 10.3.3 for a deeper discussion as to the inclusion of time fixed effects in estimating both the Production Function and Decision Function.

7.4 Choice of instruments

The fact that I cannot reject the null hypothesis that the estimates are weakly identified in either the Decision Function nor Production Function models poses a significant problem for the internal validity of the estimates; in particular, IV estimators can be biased, t-tests may fail to control size and conventional IV confidence intervals may too often fail to cover the true parameter (Andrews et al. (2019)). In addition, the validity conditions depend on (arguably strong) assumptions about the decision-making process and behaviour of the error terms. Hence, I propose replacing instruments in my preferred set: $Z^P = [\Delta f_{it-2}\ \Delta b_{it-2}\ c_{it}]$ for estimating the Production Function and $Z^D = [\Delta f_{it-2}\ \Delta b_{it-2}]$ for estimating the Decision Function with suitable alternatives that may strengthen the relevance of the instrument set and provide valid alternatives if any instruments are deemed to be invalid. However, examining Table 28 to Table 35, it is evident that almost all alternative instrument sets provide a weaker identification of the parameters in the Production Function and Decision Function. See subsubsection 10.3.4 for a deeper analysis of the alternative instrument sets; included is a discussion as to why the estimates presented in column 1 of Table 33 and column 3 of Table 28 may be more reliable estimates of the causal effect than my initial estimates presented in Table 14 and presented in Table 8 respectively.

7.5 Different A_{it} average rival expenditure definitions

Another set of robustness checks pertain to the definition of average rival expenditure A_{it}, in particular whether the estimated spillover effects and strategic effects are consistently positive and relatively stable across slight variations in the weights given to other charities (of whom expenditure is averaged over). I find that estimated spillover and strategic effects are relatively stable and positive across different rival definitions, although most of the estimates are not statistically significant, so conclusive evidence for their existence is not guaranteed. In addition, accurately identifying which charities are considered to be rivals by charity i in time t may not be possible without further qualitative or survey data. See subsubsection 10.3.5 for a deeper analysis of the robustness of positive estimated spillover and strategic effects, that utilises Table 24 to Table 27.

7.6 Heterogeneity across subgroups

I also estimate the Production Function and Decision Function for different subgroups within my sample, to test the plausibility of the assumption of constant causal effects (elasticities) across different charities. I find that the causal estimates differ greatly between subgroups, which indicates that the estimates obtained in Table 8 and Table 14 may represent an average causal effect across substantially different subgroups, which in itself is arguably not a useful estimand. Heterogenous estimates could also indicate that the specification of the fundraising production function and decision function vary across different charities, perhaps due to different objective functions or fundraising techniques, which raises misspecification concerns. See subsubsection 10.3.6 for a detailed discussion on the heterogeneity across subgroups and an examination of Table 36 and Table 37.

7.7 Statistical significance and standard errors

Out of the seven estimated parameters of interest presented in Table 8 and Table 14, only the estimated brand effects $\hat{\rho}_1$ are statistically significant at the 95% level using my preferred 2SLS strategy. As shown in column 3 of Table 16 and Table 17, the results are also not statistically significant when HAC(1) standard errors are applied, which are robust to arbitrary AR(1) serial correlation across time and arbitrary heteroskedasticity[51]. This is likely due to the weak first stage of my chosen instrument sets, neither of which I can reject the null hypothesis of weak identification for. In addition, due to the nature of the non-profit sector, the factors that affect the choice of fundraising spending and level of donations are potentially infinitely dimensional, hence the large standard errors could reflect the amount of noise in the data. Misspecification and heterogeniety can also weaken the statistical significance of the estimated parameters.

[51] Observations between different charities are assumed to be independent

8 Discussion

In section 6, I use two instrumental variables strategies to estimate seven parameters of interest, represented graphically in Figure 1, that taken together characterise the fundraising process; in particular, the factors that determine how much a charity decides to spend on fundraising in a given year and how much is raised as a result. In this section I discuss the interpretation of the estimates and how they can be used to estimate the true fundraising efficiency.

8.1 Interpretation

Assuming my preferred empirical strategy allows for exact causal identification, estimated coefficients can be interpreted as a weighted average of partial elasticities[52].

The results presented in column 2 of Table 8 show that a 1% increase in fundraising expenditure by a given charity in a given year, causes a 0.554% increase in fundraising revenues in the same year, ceteris parabis[53]. This implies that for the green charity in the sample with the median fundraising ratio (2.62), £1 spent on fundraising generates an *individual* marginal return of £1.45 the same year[54]. Table 9 shows the contemporaneous individual marginal returns to fundraising, evaluated at the median charity in each year.

Table 9: Marginal Returns to fundraising for median charity (£) in each year

Year	**Marginal Returns to fundraising for median charity (£)**
2007	1.499
2008	1.445
2009	1.405
2010	1.521
2011	1.461
2012	1.388
2013	1.421
2014	1.499
2015	1.558

The negative coefficient on lagged fundraising expenditures b_{it-1} in Table 8 ($\hat{\beta}_1 = -0.269$) contradicts my initial hypothesis about the individual returns to fundraising being spread over at least a two-year period. Instead, it could plausibly indicate a differential causal effect

[52]Partial elasticities are averaged over different complier subpopulations corresponding to the different instruments used, and averaged along the length of a possibly nonlinear causal function (Angrist and Pischke (2009)). Assuming homogeneity and linearity in causal effects across complier populations, the estimates can be interpreted as partial elasticities.

[53]This is consistent with fundraising elasticity estimates from Arulampalam et al. (2015), Sieg and Zhang (2012), Khanna et al. (1995), Posnett and Sandler (1989) which range from 0.183 - 0.626.

[54]The parameter estimates are interpreted as elasticities, hence the individual marginal return to fundraising for charity i in time t: $\frac{\partial F_{it}}{\partial B_{it}} = \frac{F_{it}}{B_{it}}\hat{\beta}_0 = R_{it}\hat{\beta}_0$ is directly proportional to the fundraising ratio of charity i in time t.

of fundraising expenditure in the long run compared to the short run. For example, whilst revenues in the same year increase, revenues fall the year after due to a lagged signalling effect of fundraising spending: if donors conflate high fundraising expenditures with low charitable spending, then they may respond negatively to high fundraising expenditures in the long run. However, it is likely that the fundraising effects are incident over a period of less than a year, hence I would require higher frequency data to observe a true β_1 parameter.

The estimated spillover effects $\hat{\gamma}_0$ and estimated strategic effects $\hat{\omega}_0$ taken together characterise the two types of interactions between rival charities in the fundraising process. The results indicate that a 1% increase in the average expenditure of rivals to charity i in time t leads to a 0.157% increase in fundraising revenues to charity i and induces a 0.0467% increase in fundraising expenditure by charity i in the same year, ceteris parabis. The total elasticity of rival spending on fundraising revenues is equal to the direct (spillover effect) plus the indirect effect through increased own fundraising efforts[55] equal to $\hat{\gamma}_0 + \hat{\omega}_0 \hat{\beta}_0 = 0.157 + 0.0467 * 0.554 = 0.183$[56], in which the direct effect accounts for 86% of the total elasticity. Positive spillover effects and small positive strategic effects together imply that although competition for donors is driving charities to compete on fundraising expenditure, fundraising techniques are inclusive[57] generating *collective* returns that are higher than the *individual* returns to fundraising[58].

The results also show a strong enabling effect $\hat{\psi}_1 = 0.180$ of revenue received by charity i in the previous year on the fundraising expenditure of charity i, which could suggest that charities are financially constrained, such that their available funds (which are a function of revenue received in previous years) are a binding constraint on the optimal fundraising expenditure, leading to a strong dependence between past fundraising revenue and current expenditure. In addition, relative to own fundraising effects, high estimated brand effects ($\hat{\rho}_1 = 0.710$) and persistence effects ($\hat{\phi}_1 = 0.305$) suggest that the fundraising decision and production processes are relatively stable over time. Persistence in expenditure may be largely due to rigidity in labour costs, whilst strong brand effects may be explained by donor inertia[59] or a signalling effect[60].

The discussion thus far has assumed that the estimates are internally valid, i.e. consistent and estimated to a reasonable degree of uncertainty. However, there are several reasons to doubt the accuracy and precision of the estimates. Firstly, 6 of the estimated coefficients are not (statistically) significantly different from zero, and this result is robust to different estimates of the standard error. Hence, the existence of the true parameters and their

[55] See Equation 4

[56] This point estimate is only mentioned qualitatively as its uncertainty is difficult to quantify, and it is likely very imprecisely estimated. In addition, the estimates are taken from two different samples hence may not be valid for the same subset of charities.

[57] In that they engage new donors rather than steal their rivals existing donors

[58] Recall Equation 3

[59] donors reluctance to switch charities or cancel charitable standing orders

[60] If donors associate high existing or past income with reliability, then they may be encouraged to donate in the future; thus, past revenues influence present revenues via a brand effect

magnitude is very uncertain[61]. Secondly, sample selection bias, heterogeneity and weak IV issues represent the largest threats to identification that persist even after implementing my preferred empirical strategy[62].

In addition, it is likely that the results would differ significantly when estimated over a more recent time period, different subsector, and a sample of younger and smaller charities. For example, younger and smaller charities may experience much smaller individual returns to fundraising β as they lack the brand recognition of older charities. Charities in more mature subsectors, such as health and housing, may experience smaller or negative spillover γ effects as the total donor population (or potential donable funds) are not growing over time. More recent data may yield higher returns to fundraising β, due to the growth in online and social media fundraising campaigns, which can generate much larger returns, and returns over a longer time-span. Furthermore, government austerity measures post-2015 and the COVID19 Pandemic in 2020 have caused charities to become more financially constrained, which may accentuate enabling effects ψ in the data, in addition to ushering in a selection effect, in which the charities that can afford to fundraise (and stay alive), become more frugal and selective with the fundraising projects they engage in, leading to higher observable returns to fundraising β.

8.2 Estimating the *true* fundraising efficiency

Recall from section 3 the derivation of the bias in the fundraising ratio r_{it}, the ratio of fundraising revenue to expenditure, in logarithms as a measure of the true fundraising efficiency ε^c_{it}:

$$r_{it} - \varepsilon^c_{it} = \varepsilon^d_{it} + (\beta_0 - 1)b_{it} + (\beta_1 + \rho_1)b_{it-1} + \gamma_0 a_{it} + \rho_1 r_{t-1}$$

Which assumes that ε_{it} - the unobservable shocks and determinants of fundraising revenue - can be decomposed linearly into a charity-driven component ε^c_{it} and a donor-driven component ε^d_{it}. ε^d_{it} represents the factors taken to be exogenous[63] by the charity, such as donor preferences and external events, whereas ε^c_{it} represents factors under the control of the charity, which are assumed to capture the true fundraising efficiency of a charity.

First, using the estimates of the fundraising Production Function parameters obtained in the previous section, I obtain the residuals $\hat{\varepsilon_{it}}$ using:

$$\hat{\varepsilon_{it}} = r_{it} - (\hat{\beta}_0 - 1)b_{it} - (\hat{\beta}_1 + \hat{\rho}_1)b_{it-1} - \hat{\gamma}_0 a_{it} - \hat{\rho}_1 r_{it-1}$$

[61] See section 7 for a deeper discussion on the lack of statistical significance

[62] See section 10 for a deeper discussion on selection bias, heterogeneity and weak IV issues.

[63] Exogenous in the economic sense, not econometric sense. Charities may still respond to ε^d_{it} making it a source of endogeneity in the causal model.

Next, I regress these residuals $\hat{\varepsilon_{it}}$ on year dummies and beneficiary dummies[64] which are assumed to capture all unobservable variation in fundraising revenue purely attributable to donor preferences and external shocks, to obtain the new residuals that capture remaining variation in $\hat{\varepsilon_{it}}$, unexplained by and uncorrelated with these exogenous regressors, assumed to be determined by the charity themselves[65]. This allows me to decompose the original residuals $\hat{\varepsilon_{it}}$ into the (fitted) donor driven unobservables $\hat{\varepsilon_{it}^d}$ and the (residual) charity driven unobservables $\hat{\varepsilon_{it}^c}$, which provide a best estimate of the true fundraising efficiency of a charity[66].

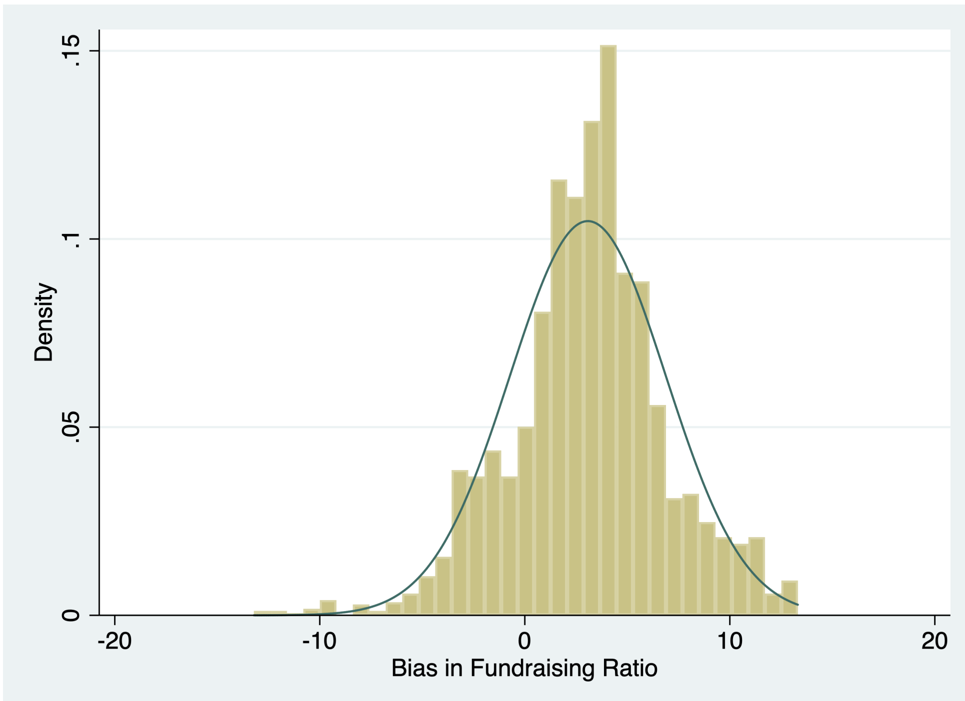

Figure 6: Bias in the fundraising ratio as a measure of true fundraising efficiency $(r_{it} - \varepsilon_{it}^c)$ in logs

[64]which indicate the other listed beneficiaries of the charity e.g. the advancement of health, eradication of poverty, support for arts/heritage/culture, advancement of education etc

[65]The exogeneity of sector-wide shocks or trends and a charity's listed beneficiaries is a relatively strong assumption. It assumes that charity managers take their listed beneficiaries as given - determined by the charitable objectives set forth in their mission statement - unable to be changed in a way that may be correlated with the inherent efficiency of the charity. Hence any variation in donations that is correlated with a particular beneficiary is assumed to represent a stronger donor preference or public sensitivity towards that cause and is hence captured by ε_{it}^d (and is independent of ε_{it}^c). For example, if green charities that also work towards the eradication of poverty experience higher unexplained fundraising revenue ε_{it} on average, then this due to a stronger donor preference for green charities that also work to eradicate poverty, and not because green charities that also work towards the eradication of poverty are any more or less inherently fundraising efficient than charities that do not. In addition, any shock or trend in unexplained fundraising revenue that is common to all charities in a particular year is assumed to be the result of an exogenous donor preference shock or trend captured by ε_{it}^d, and not the result of a sector-wide shock or trend in inherent fundraising efficiency (hence is independent of ε_{it}^c). This means that on average, charities cannot become inherently more or less efficient over time by assumption

[66]The regression of $\hat{\varepsilon_{it}}$ on year dummies and beneficiary dummies yields an adjusted-R^2 of 0.0505, with an F-statistic of 65.09, implying that the donor-driven influence on the residual $\hat{\varepsilon_{it}}$ is non-negligible

Figure 6 provides a density plot of the estimated bias in logarithms ($r_{it} - \hat{\varepsilon^c_{it}}$), which suggests that on average, the fundraising ratio overestimates the true fundraising efficiency for the charities in the sample[67]. This has important implications for the use of the fundraising ratio as a measure of financial performance, advocated for by the NCVO and Charity Navigator. From Equation 5, due to the high estimated brand effect $\hat{\rho}_1$ and positive estimated spillover effect $\hat{\gamma_0}$, the fundraising ratio may be positively biased by high fundraising expenditure in the year before, high fundraising expenditure of rivals and a higher ratio the year before, whereas diminishing returns to current fundraising expenditure $\hat{\beta_0} < 1$ implies that high fundraising expenditure can negatively bias the fundraising ratio in the same year.

However, the fundraising ratio appears to remain a relatively strong predictor[68] of the fundraising efficiency ε^c_{it}, hence I argue that despite the biases, charity managers and evaluators should continue to use the fundraising ratio as an (imperfect) measure of fundraising efficiency. In fact, with a greater understanding of the main limitations (biases) in the fundraising ratio, it may become a more useful and informative tool for evaluating the efficiency of fundraising performance. However, fundraising efficiency itself may be of secondary importance to fundraising optimality, which considers the extent to which charities are committing an optimal amount on fundraising. See subsection 10.4 for a deeper discussion on optimality.

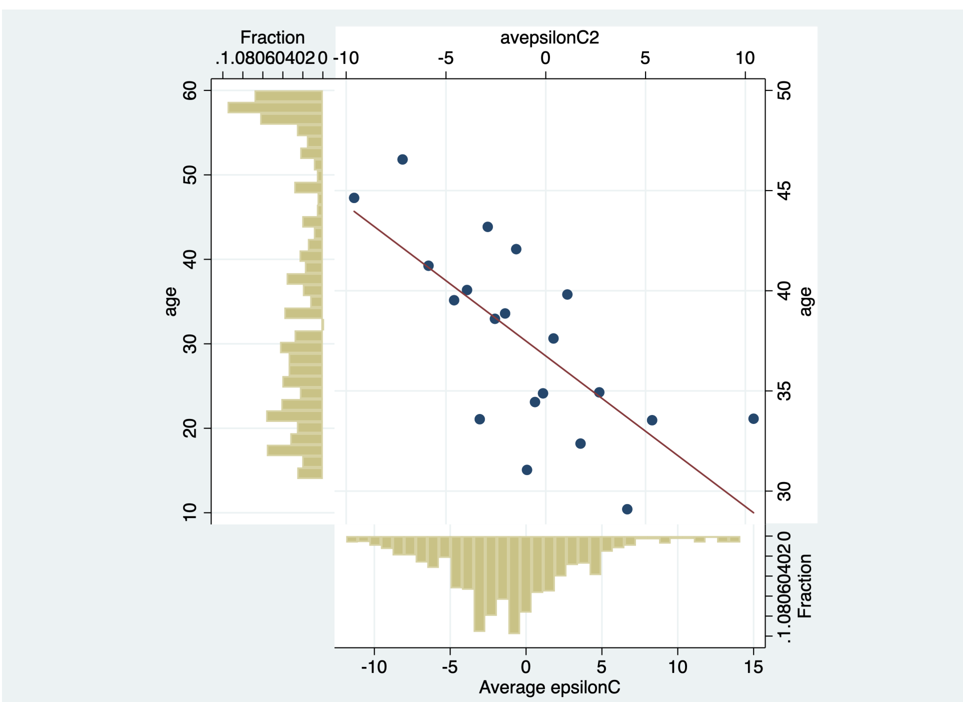

Figure 7: Correlation between charity *age* and $\overline{\varepsilon^c_{it}}$

Figure 7 shows a negative correlation between the age of a charity and the average

[67]Note that the bias is expressed in logarithms, hence the units do not reflect the magnitude of bias in raw terms. A more rigorous statistical analysis would be needed to estimate the magnitude and direction of the bias in absolute terms.

[68]The Pearson correlation coefficient is 0.605 which is significant at the 5% level.

fundraising efficiency. This could be because older charities may be less likely innovate in the fundraising techniques they use to capture donors; for example, younger charities may be more likely to utilise social media. Another explanation could be that younger charities have more agile governance structures and are more transparent, which is more attractive to donors. The phenomenon of older charities being less fundraising efficient could explain the positive bias in the fundraising ratio for charities in the sample presented in Figure 6, since sampled charities are on average 7 years older than the average green charity in the population.

9 Conclusion

In this paper, I have introduced a system of equations that capture the decision-making process and the revenue generation process of fundraising. Taken together, the fundraising Production Function and Decision Function allow me to separately estimate the individual returns to fundraising and the collective returns, which account for the interplay between rival charities. Exploiting plausibly exogenous variation in minimum total labour costs, driven by changes in the national minimum wage, as well as lagged revenues and expenditures, I am able to use an instrumental-variables strategy to identify seven parameters of interest, including own fundraising effects (β_0 and β_1), spillover effects (γ_0) and strategic effects (ω_0) depicted graphically in Figure 8 [69].

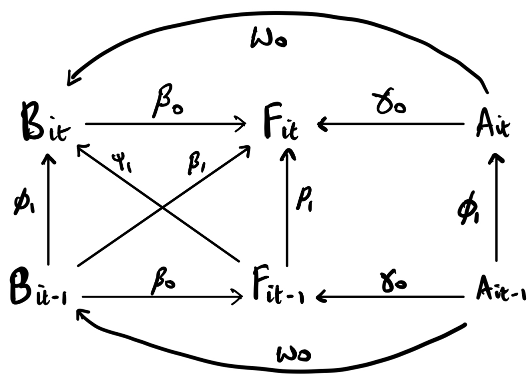

Figure 8: Complete model

I find that a 1% increase in fundraising expenditure by a given charity in a given year leads to a 0.554% ($\hat{\beta}_0$) increase in fundraising revenues in the same year, ceteris parabis, which corresponds to a £1.45 individual marginal return - for the charity in the sample with the median fundraising ratio (ratio of fundraising revenue to expenditure). This point estimate is consistent with other estimates of fundraising elasticity in the literature[70].

I also find that a 1% increase in the average expenditure of rivals to charity i in time t leads to a 0.157% ($\hat{\gamma}_0$) increase in fundraising revenues to charity i, and induces a 0.0467% ($\hat{\omega}_0$) increase in fundraising expenditure by charity i in the same year. This corresponds to a total effect of charity i's *rivals* fundraising expenditure on charity i's own fundraising revenues in time t equal to the direct effect $\hat{\gamma}_0$ alongside the indirect effect ($\hat{\beta}_0\hat{\omega}_0$) through a change in the own fundraising efforts of charity i: 0.157+0.0467x0.554= 0.183%. The direct effect and indirect effects are depicted as the two different paths from A_{it} to F_{it} in Figure 8. By accounting for the positive strategic and spillover influences of rival charities, it is evident that the collective returns to fundraising are larger than the returns an individual charity could achieve by fundraising in isolation.

[69] F_{it} represents the fundraising revenue (donations) of a charity i in time t, B_{it} represents the fundraising expenditure (budget) of charity i in time t and A_{it} represents the average fundraising expenditure of charity i's rivals in time t.

[70] See Arulampalam et al. (2015), Sieg and Zhang (2012), Khanna et al. (1995), Posnett and Sandler (1989)

Finally, I develop a framework for characterising the magnitude and direction of bias in the fundraising ratio (the ratio of fundraising revenue to expenditure) as a measure of the true fundraising efficiency of a charity, and use the results to estimate the fundraising efficiency for the charities in my sample. I find that higher fundraising expenditure the year before, a higher ratio in the year before, higher rival fundraising expenditure and more "popular" beneficiaries can positively bias the fundraising ratio, whilst higher current fundraising expenditure can negatively bias the fundraising ratio, as a measure of the fundraising efficiency of a charity. For the charities in my sample, the fundraising ratio overestimates the true efficiency on average, and older charities tend to have lower average fundraising efficiency than younger charities. This could be explained by the different fundraising techniques employed by older and younger charities, with younger charities being more likely to engage in digital fundraising campaigns and utilise social media, which can generate much higher returns due to its reach.

However, there are several reasons to doubt the accuracy and precision of the estimates. Firstly, 6 of the estimated coefficients are not (statistically) significantly different from zero, and this result is robust to different estimates of the standard error. Hence, the existence and magnitude of the true parameters, in addition to the direction of average bias in the fundraising ratio, is very uncertain. Secondly, sample selection bias, heterogeneous causal effects and weak IV issues represent large threats to consistency that persist after implementing my preferred identification strategy. Hence, the estimated coefficients are likely asymptotically biased.

Furthermore, the results likely have limited external validity, as my selected sample of charities differ from the population of charities in several ways. Using a sample of younger charities may yield much smaller individual returns to fundraising as they lack the brand recognition of older charities whilst using a sample of charities in a more mature subsectors such as health may yield smaller or negative spillover effects, as fundraising simply results in donor stealing. Using more recent data may yield higher returns to fundraising, due to the growth in online and social media fundraising campaigns, which can generate larger returns over a longer time-span. Furthermore, the COVID19 Pandemic in 2020 may have ushered in a selection effect, in which only the most fundraising efficient charities stay afloat and can afford to fundraise, leading to higher estimated returns to fundraising across the population.

Nonetheless, the novel theoretical framework, identification strategy and results provide a useful benchmark for future researchers, as well as charity managers and regulators that seek a deeper understanding of the fundraising process, the role of rivalry in the non-profit sector, and the caveats to employing the fundraising ratio as a measure of the efficiency of fundraising. The fundraising ratio is just one of several indicators of financial performance used by regulators and charity evaluators, which is arguably a level of performance secondary to the level of social impact performance of the charity itself.

10 Appendix

10.1 Identifying the Fundraising Decision Function

Selecting into the sample charities with over £500k in total income also creates selection bias in the Decision Function, since it amounts to selecting on ε_{it} which is highly correlated with the error term g_{it} [71]. However, restricting the sample of charities to only those reporting over £ 1m in average non-fundraised income over the period may not be as effective in mitigating selection bias in the Decision Function as in the Production Function. This is because average non-fundraising revenue may constitute a relevant omitted variable in the Decision Function, hence is strongly correlated with g_{it}. This is because average non-fundraising income is strongly correlated with the regressors f_{it-1} and b_{it-1} and can enable higher fundraising expenditure b_{it}. Although this may reduce the selection bias caused by the total income threshold, it may introduce selection bias of a similar magnitude, effectively replacing one non-random selection criteria with another [72] However, I choose to impose the sample restrictions for the sake of continuity in the external validity between the estimated coefficients of the Production Function and the Decision Function.

The bias caused by attempting to fit a non-linear data-generating process into a linear model, due to discontinuity at $b_{it} = 0$, cannot be solved by dropping charities from the sample that report zero fundraising expenditures, since this introduces selection bias [73]. Hence I choose to restrict the sample to charities that report at least one positive legacy income over the period $\overline{leg_{it}} > 0$, which is largely driven by high brand recognition. This is because charities with higher brand recognition should face lower entry barriers to fundraising, hence any amount of fundraising expenditure can expect to generate a return [74]. This would ensure fundraising expenditures are more continuous at $b_{it} = 0$. Although this introduces some selection bias [75], this poses a smaller threat to identification than a potential non-linear data-generating-process[76].

[71]This introduces the selection term: $\lambda(-NF_{it}, -X^P) = \mathbb{E}[g_{it}|\varepsilon_{it} > \varepsilon(-NF_{it}, -X^P), X^D]$ which is non-zero since g_{it} and ε_{it} are highly correlated and decreasing in the regressors of the Decision Function(X^D) since X^D and X^P are very similar regressor matrices.

[72]

Selecting on average non-fundraised income attenuates the selection term introduced by the total income threshold $\mathbb{E}[g_{it}|\ \varepsilon_{it} > \varepsilon(-NF_{it}, -X^P), X^D]$ whilst simultaneously introducing a new selection term $\mathbb{E}[g_{it}|\overline{NF_{it}} > £1m, X^D]$, which is non-zero and likely increasing in the regressors. A naive solution is to control for $\overline{NF_{it}}$ in the Decision Function, thereby capturing any confounding effect of $\overline{NF_{it}}$ on the regressors, and purging $\overline{NF_{it}}$ from the error term g_{it}. However, it is not clear if $\overline{NF_{it}}$ is itself also consequence of the regressors, in which case it constitutes a bad control (Angrist and Pischke (2009)) and conditioning on it would reintroduce selection bias.

[73]see footnote 31

[74]This cannot be said for charities with lower brand recognition, since a larger amount of fundraising may be required to generate a return, forcing these charities to spend either zero or some large amount on fundraising

[75] conditioning on positive legacy income introduces the selection term $\mathbb{E}[g_{it}|\overline{leg_{it}} > 0, X^D]$ which is non-zero since positive legacy income can enable higher g_{it}, and is increasing in the regressors.

[76]see footnote 31

In a similar way to estimating the Production Function, including time dummies in the Decision Function can partition out any variation that does not vary between charities, which include sector-wide fundraising expenditure trends captured by m_t^g, the time-specific component of the error term g_{it}. This can mitigate endogeneity since the fundraising trends may be driven by the trend in public sensitivity to climate change (as charities seek to capitalise on the increased sensitivity) and hence can act as an omitted variable, confounding the effect of rivals charity fundraising expenditure a_{it} on own charity fundraising expenditure b_{it} [77].

The charity-specific component of the error term η_i^g is likely to capture time-invariant characteristics such as area of operation and governance structure; these variables may directly influence fundraising expenditure, and also be correlated with fundraising revenue, confounding the effects of f_{it-1} and b_{it-1} on b_{it}. Hence, implementing a Within or First-Difference transformation on the Decision Function would purge η_i^g from the error term, mitigating the bias caused by these time-invariant omitted variables [78].This would also mitigate endogeneity caused by measurement error, likely captured by η_i^g [79] and significantly reduce serial correlation in the error term. Purging η_i^g from the error term also reduces the selection bias caused by selecting charities on age [80].

First-differencing and including time dummies yields my preferred regression specification for estimating the parameters of the Decision Function:

$$\Delta b_{it} = \phi_0 + \phi_1 \Delta b_{it-1} + \psi_1 \Delta f_{it-1} + \omega_0 \Delta a_{it} + \sum_{t=1}^{T} \alpha_t m_t^g + \Delta v_{it}^g \tag{8}$$

However, Equation 8 results in so-called Nickell Bias (1980), as it creates artificial correlation between the transformed error, containing v_{it-1}^g, and the transformed lagged dependent variable containing b_{it-1}. Hence I propose using an instrumental variables strategy that instruments the endogenous Δb_{it-1} with Δb_{it-2} and the plausibly endogenous Δf_{it-1} with Δf_{it-2} [81], taking a_{it} to be exogenous to the error Δv_{it}^g. This utilises two moment conditions $\mathbb{E}[\Delta f_{it-2} \Delta v_{it}^g] = 0$ and $\mathbb{E}[\Delta b_{it-2} \Delta v_{it}^g] = 0$ which are valid if f_{it-2} and b_{it-2} are uncorrelated

[77] However, it is also likely that sector-wide fundraising expenditure trends captured by m_t^g are the consequence of rising rival fundraising expenditure not just the cause, due to vicious cycle of fundraising (Rose-Ackerman (1982)) outlined in section 2. If the trend in fundraising expenditure that emerges is more the consequence of rival spending than the cause, then controlling for time fixed effects may constitute a bad-control (Angrist and Pischke (2009)) conditioning on which introduces bias

[78] It is very likely that such omitted variables also vary across time, hence are captured by v_{it}^g. If the omitted variables are serially correlated across time, then f_{it-1} and b_{it-1} become endogenous (related to the error term via v_{it-1}^g).

[79] as outlined in 26

[80] This is only valid if we assume that whilst $\mathbb{E}[\eta_i^g | age_{it}, X^D] \neq 0$, $\mathbb{E}[v_{it}^g | age_{it}, X^D] = 0$ i.e. idiosyncratic fundraising expenditure shocks do not vary with age

[81] I also test f_{it-2} and f_{it-3} as instruments for Δf_{it-1} and b_{it-2} and b_{it-3} for Δb_{it-1}, with levels as instruments improving relevance if the data is stationary and longer lags as instruments improving validity if there exists serial correlation in v_{it}^g

with v^g_{it-1} and v^g_{it} is not serially correlated[82]. In addition, the instruments may be excluded if they do not influence fundraising expenditure independently of the endogenous regressors, which requires that true persistence effects and enabling effects are at most one period lagged [83].

Table 10 presents a summary of the threats to identification for the Decision Function and the strategies designed to mitigate them.

Table 10: Summary of Identification Strategy for estimating the Decision Function

Threat to identification	Strategies to mitigate threat to identification
Omitted variable bias	1. Time FEs 2. First Differencing 3. Instrumental Variables
Measurement error	1. First differencing
Serial correlation	1. First differencing 2. Instrumental variables
Selection bias	1. Restricting sample to only charities with an average non-fundraised income of over £1m
Misspecification	1. Restricting sample to those reporting at least one positive legacy income over the period

[82] This assumes that any dependence between lagged variables and the error term is eliminated once the fixed effects η^g_i and m^g_t are purged from the error.

[83] i.e. ϕ_2 and ψ_2 do not exist in reality.

10.2 Estimating the Fundraising Decision Function

Table 11: Selected Sample Summary Statistics (Decision Function)

	Mean	SD	Min	Max	Number of observations
B_{it} (Expenditure)	2,410,363	6,475,238	0	57,938,000	1,673
F_{it} (Revenue)	7,621,303	19,102,482	0	152,343,008	1,674
A_{it} (Rival Exp.)	404,318	483,964	7,023	2,157,057	1,675

In Table 11, I summarise the 3 variables of interest, imposing the sample restrictions introduced in the previous section[84] that attempt to mitigate selection bias and reduce specification error. Comparing Table 1 with Table 11, it is evident that imposing the new restrictions has produced a smaller sample of larger charities[85], spending on average around £ 2m more and raising £ 5m more on average. This is largely due to the legacy income restriction $\bar{leg}_{it} > 0$ that restricts the sample of charities to those with large enough brand recognition to face continuous marginal costs of fundraising[86]. This reduces the proportion of observations at $b_{it} = 0$ from 35% to 15%. Imposing the non-fundraised income threshold $\overline{NF_{it}} > £1m$ excludes charities likely to drop in and out of the dataset by providing a large enough income buffer, successfully improving panel balance: increasing the average number of time periods reported per charity in the sample from 5.85 to 8.62 out of 9.

Table 12 provides OLS estimates of the parameters in the Decision Function, using the Decision Function as a baseline model. Imposing the sample restrictions in column 2 attenuates the estimated coefficients on f_{it-1} and a_{it}, whilst increasing the estimated coefficient on b_{it-1}. Including time fixed effects in column 3 makes a negligible difference to the estimates. Eliminating between variation in column 5 attenuates the magnitude of the estimated coefficients, whilst significantly reducing the R^2 from 0.892 to 0.160 and the F-statistic from 1087.4 to 4.060. In addition, the coefficients on a_{it} and f_{it-1} become insignificant.

Recall my preferred regression specification for estimating the Decision Function from section 5:

$$\Delta b_{it} = \phi_0 + \phi_1 \Delta b_{it-1} + \psi_1 \Delta f_{it-1} + \omega_0 \Delta a_{it} + \sum_{t=1}^{T} \alpha_t m_t^g + \Delta v_{it}^g \qquad (9)$$

Where Δb_{it-1} and Δf_{it-1} are endogenous regressors. Table 13 provides a summary of the first stage regressions of each endogenous variable, represented by different columns, on the instrument set $Z_{it} = [\Delta f_{it-2}, \Delta b_{it-2}]$. Δf_{it-2} and Δb_{it-2} appear to be weakly relevant instru-

[84] $\overline{NF_{it}} > £1m$ and $\bar{leg}_{it} > 0$

[85] the number of charities in the sample has decreased from 1671 to 193.

[86] This restriction replaces the weaker $\overline{b_{it}}$ restriction used to model the Production Function and hence the Decision Function is estimated over a smaller sample of larger charities compared to the estimated Production Function.

ments for Δf_{it-1} and Δb_{it-1}, since the F-statistics and Sanderson-Windmeijer multivariate F-test statistics are low in the first stages of both endogenous variables. In addition, the Kleinbergen-Paap Wald (rk) F-statistic is 1.76, hence we cannot reject the null hypothesis of weak identification at the 10% level of significance[87]

Column 2 in Table 14 shows the 2SLS estimated coefficients of my preferred specification. None of the coefficients are statistically significant using clustered standard errors. The results show a strong persistence effect of fundraising expenditures ($\hat{\phi}_1 = 0.305\%$), and a strong enabling effect of lagged fundraising revenues ($\hat{\psi}_1 = 0.180\%$). The results also show a small and positive estimated strategic effect: $\hat{\omega_0} = 0.0467\%$.

[87] the Kleinbergen-Paap Wald (rk) F-statistic is chi-squared (1) distributed with a 10% critical value at 2.706.

Table 12: OLS Estimates of Elasticities

	(1) Full sample	(2) Selected sample	(3) Time FEs	(4) Twoway FEs
f_{it-1}	0.0789***	0.0497*	0.0469*	0.0392
	(0.00950)	(0.0233)	(0.0227)	(0.0438)
b_{it-1}	0.873***	0.929***	0.931***	0.386***
	(0.00756)	(0.0195)	(0.0192)	(0.0947)
a_{it}	0.0657**	0.0296	0.0306	0.296
	(0.0224)	(0.0462)	(0.0444)	(0.179)
Observations	7759	1472	1472	1472
R^2	0.820	0.892	0.892	0.160
Adjusted R^2	0.820	0.891	0.892	0.154
F	12613.6	2750.6	1087.4	4.060

Standard errors in parentheses.

All estimates are interpreted as elasticities.

The model specifications are cumulative, with each building on the model specification to its left.

Standard errors are robust to arbitrary heteroskedasticity and autocorrelation within charities.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: First Stage Summary

	(1) Δf_{it-1}	(2) Δb_{it-1}
Δf_{it-2}	-0.351***	0.0369
	(0.104)	(0.0307)
Δb_{it-2}	0.0852**	-0.261**
	(0.0311)	(0.0839)
Observations	1078	1077
R^2	0.129	0.096
Adjusted R^2	0.122	0.089
SWF	8.63	10.46
F	5.75	6.48

Standard errors in parentheses.

Each column corresponds to the first stage for the endogenous variable in the header.

Each row corresponds to a different instrument.

Standard errors are robust to arbitrary heteroskedasticity and autocorrelation within charities.

F is the F-statistic of the test for joint significance of the instruments.

SWF is the Sanderson-Windmeijer multivariate F test of excluded instruments.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: OLS and 2SLS estimates of preferred regression specification for estimating the Decision Function

	(1) OLS	(2) 2SLS (preferred strategy)
Δf_{it-1}	0.0444	0.180
	(0.0325)	(0.194)
Δb_{it-1}	-0.261**	0.305
	(0.0828)	(0.384)
Δa_{it}	0.0297	0.0467
	(0.0978)	(0.106)
Observations	1272	1076
R^2	0.072	-0.251
Adjusted R^2	0.065	-0.261
F	1.853	1.263

Standard errors in parentheses.

All estimates are interpreted as elasticities.

The model specifications are cumulative, with each building on the model specification to its left.

Standard errors are robust to arbitrary heteroskedasticity and autocorrelation within charities.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

10.3 Robustness checks continued

10.3.1 Sample restrictions and selection bias

Columns 3 and 4 of Table 18 and Table 19 separate the effect of the two restrictions and report the estimated 2SLS coefficients. For both Table 18 and Table 19, comparing column 3 with the preferred results in Table 14 and Table 8 reveals no clear effect of the restriction designed to mitigate the downward selection bias caused by the total income threshold[88]. This is due to the fact that the non-fundraised income restriction also introduces positive selection bias in both cases, although more so in the Decision Function coefficients [89]. In addition, for both Table 18 and Table 19, comparing column 4 with the preferred results in Table 14 and Table 8 reveals no clear effect of the restrictions designed to mitigate the misspecification bias. This is because the two restrictions $\overline{b_{it}} > 0$ and $\overline{leg_{it}} > 0$ introduce upward selection bias, since $\overline{b_{it}}$ is positively correlated with the error term ε_{it} and $\overline{leg_{it}}$ is positively correlated with the error term g_{it}[90], in addition to introducing a downward selection bias, since they effectively impose a conditional-on-positive interpretation of the parameters[91]creating a similar bias to that caused by the total income threshold.

Column 2 in both Table 18 and Table 19 report the Tobit estimates on the full sample that account for potential left-censoring at zero - which is evident from the histograms in Figure 3 and Figure 4. However, these estimates are most likely inconsistent as the Tobit estimator requires that the models are specified with all relevant explanatory variables - which is unlikely to hold.

10.3.2 Mitigating endogeneity by first-differencing and including controls

The inclusion of charity-specific dummies in column 5 of Table 6 and Table 12 significantly attenuates the OLS estimated coefficients, which provides evidence that purging the η_i from the error term significantly mitigates the upward bias on the coefficients caused by time-invariant omitted variables, measurement error, serial correlation and selection bias. However, a within transformation does not entirely mitigate Nickel bias as the existence of a lagged dependent variable f_{it-1} violates strict exogeneity by construction. Hence, the estimates in column 5 are invalid. Nickel Bias is perhaps more prevalent in column 1 of Table 8 and Table 14 which present the OLS estimates of Equation 6 and Equation 8 respectively. The negative coefficients on the lagged dependent variables indicate that a first-difference transformation may in fact accentuate Nickel bias[92] providing further justification for the use of an instrumental-variables strategy.

[88]See footnote 28

[89]See footnote 37 and 72 for more details.

[90]See footnotes 38 and 75

[91]See the naive solution and a probabilistic selection threshold outlined in footnote 31.

[92]For example, in estimating the Production Function, differencing constructs negative correlation between Δf_{it-1} and $\Delta v^{\varepsilon}_{it}$, via the $-v^{\varepsilon}_{it-1}$ term, creating significant downward bias on the Δf_{it-1} coefficient.

10.3.3 Mitigating endogeneity by including time-fixed effects

Including time dummies into the Production Function base model in column 3 of Table 6 appears to marginally increase the coefficients, whilst including time dummies in my preferred specification presented in column 2 of Table 16 only appears to mitigate positive omitted variable bias on the estimated spillover effects $\hat{\gamma_0}$ and are not individually statistically significant. This contradicts my predictions regarding the direction of omitted variable bias caused by sector-wide shocks captured by m_t^{ε}. Assuming that such an omitted variable exists, i.e. an upward trend in public sensitivity toward climate change driving a trend in fundraising revenues and expenditures describes a plausible confounding effect, the bigger question arises as to whether this can be mitigated by including time dummies. Table 20 and Table 21 present the magnitude of time trends and shocks to fundraising revenue for different sized charities, measured by quintile of the average fundraising revenue distribution. It appears that whilst the smallest and largest quintiles are experiencing a downwards trend in f_{it}, the middle quintiles experience an upward trend[93]. In addition, relative to post-2015 levels, sector-wide changes in fundraising revenue appear to differ widely across quintiles. Thus, it is likely whilst the donor preference shocks may be sector-wide, the effect on fundraising revenue differs widely across charities, hence purging m_t^{ε} from the error will not fully mitigate omitted variable bias.

The theoretical justification for including time dummies in the estimation of the Decision Function is weaker than in the Production Function, since the type of variation captured by m_t^g is not clear[94]. However, column 1 and 2 of Table 17 show that including time dummies in the Decision Function decreases the estimated strategic effect $\hat{\omega_0}$, plausibly mitigating positive omitted variable bias. In addition, trends in fundraising expenditures appear to be more homogenous, with all quintiles except the lowest showing a large upward trend, as shown in Table 22 and Table 23. Hence, although the theoretical justification is weaker, the results appear to justify the inclusion of time fixed effects in the Decision Function.

10.3.4 Choice of instruments

First, I propose replacing Δf_{it-2} and Δb_{it-2} in the preferred set of instruments for the twice lagged levels f_{it-2} and b_{it-2} as instruments for the endogenous variables Δb_{it-1} and Δf_{it-1} in both the Production Function and Decision Function. Examining Table 29, it is clear that despite higher F-statistics, the lower SWF-statistics indicate that twice lagged levels do not improve the relevance of the twice differenced instruments[95] in Production Function estimation. However, Table 34 shows that twice lagged levels are far better performing instruments for the endogenous Δb_{it-1} and Δf_{it-1} variables in the Decision Function: I can

[93] although, only for the 2nd largest quintile is the trend statistically significant. Hence there is much uncertainty as to the exact fundraising revenue trend in the data.

[94] See footnote 77

[95] in addition, the Kleinberg-Paap F statistic is 0.01, which approaches perfect irrelevance.

reject the null that the system is weakly identified[96]. Hence the 2SLS results in column 1 of Table 33 may be more reliable than my initial estimates presented in Table 14. I also propose replacing c_{it} in my preferred instrument set with Δc_{it} in the 2SLS estimation of the Production Function. I hypothesised that the existence of a reinforcing reverse causal chain (higher revenues enabling higher expenditure) would create upward bias on the estimated coefficients in the Production Function; however, comparing column 1 to column 2 in Table 8, my preferred instrumental-variables approach has not appeared to mitigate this bias, since the estimated coefficients have increased. Examining column 3 of Table 28, it appears that using Δc_{it} in place of c_{it} does appear to mitigate the bias caused by reverse causality, as the estimate of own fundraising elasticity falls compared to the OLS estimates of Equation 6 presented in Table 8. In addition, examining the first stage results presented column 3 of Table 31, this appears to greatly increase the SWF-statistics compared to the preferred instrument set, however I am still unable to reject the null hypothesis of weak identification[97]. Nonetheless, the 2SLS results in column 3 of Table 28 may be more reliable than my initial estimates presented in Table 8.

Further lagged instruments offer valid alternatives if any instruments in the preferred set are deemed to be invalid or plausibly violate the exclusion restriction. For example, twice differenced instruments Δf_{it-2} and Δb_{it-2} are invalid if (1) there exists serial correlation in either v^g_{it} or v^ε_{it}[98] or (2) past values of fundraising revenue or expenditure influence future errors[99]. Hence, I conduct the Arellano and Bond test for autocorrelation in second differences using the residuals obtained from estimating Equation 6 and Equation 8 via OLS. In both cases I am able to reject the null of AR(2) serial correlation in differences, and equivalently AR(1) serial correlation in levels at the 10% level of significance[100] hence it is likely that criterion (1) is not violated. However, criterion (2) is not easily testable, hence I propose replacing Δf_{it-2} and Δb_{it-2} in the preferred set of instruments with the thrice lagged levels f_{it-3} and b_{it-3} as instruments for the endogenous variables Δb_{it-1} and Δf_{it-1} in both the Production Function and Decision Function. However, the first stage results in Table 30 and Table 35 show that in both cases thrice lagged levels are too weak to consider valuable. I also propose replacing c_{it} in my preferred instrument set with c_{it-1} in estimating the Production Function since the existing moment condition: $\mathbb{E}[c_{it}(v^\varepsilon_{it} - v^\varepsilon_{it-1})] = 0$ requires that c_{it} and v^ε_{it-1} are uncorrelated. This may not be plausible since charities may adjust their employee numbers, and hence c_{it}, to idiosyncratic fundraising revenue shocks in the previous year v^ε_{it-1}[101]. However, upon examining the first stage in Table 32, the new instrument set

[96] Since the Kleinberg-Paap Wald rk F-Statistic is 5.31 which exceeds the 5% critical value at 3.841.

[97] the Kleinberg-Paap Wald rk F-Statistic is 1.26 which does not exceed the 10% critical value at 2.706

[98] For example, if v^g_{it} is AR(1) serially correlated such that $v^g_{it} = kv^g_{it-1} + e_{it}$, the moment condition: $\mathbb{E}[\Delta f_{it-2}\Delta v^g_{it}] = \mathbb{E}[(f_{it-2} - f_{it-3})(v^g_{it} - (k-1)v^g_{it-2} - e_{it-1})] \neq 0$ since f_{it-2} and v^g_{it-2} are correlated by construction.

[99] For example, if f_{it-2} affects v_{it-1}, then the two are correlated and hence the moment condition: $\mathbb{E}[\Delta f_{it-2}\Delta v^g_{it}] = \mathbb{E}[(f_{it-2} - f_{it-3})(v^g_{it} - v^g_{it-1})] \neq 0$.

[100] the test statistics are -2.53 and -1.78, with the OLS residuals from Equation 8 showing slightly larger autocorrelation.

[101] I propose the more plausible moment condition: $[c_{it-1}(v^\varepsilon_{it} - v^\varepsilon_{it-1})] = 0$ that only requires that charities

appears too weak to consider.

10.3.5 Different A_{it} average rival expenditure definitions

Table 15: Different A_{it} definitions

Variable	Grouping by area of operation and beneficiaries	Distance Measure	Minimum length of time charities can be rivals (stability)	Maximum distance between rivals (exclusivity)	Top 10 largest
80	yes	Charitable exp.	1 year	30	no
81	yes	N/A	9 years	N/A	no
82	yes	Fundraising rev.	1 year	N/A	yes
83	yes	Fundraising rev.	1 year	30	no
84	yes	Charitable exp.	1 year	10	no
85	yes	Charitable exp.	1 year	40	no
86	yes	Charitable exp.	5 years	30	no
87	yes	Charitable exp.	9 years	30	no
88	yes	Charitable exp.	3 years	30	no

Table 15 summarises 8 other algorithms used to construct the weights; all other charities that satisfy the criteria in the table are given a weight equal to 1. For example, 80 gives my preferred definition, 83 uses fundraising revenue as a measure of distance, 86 only allows rivals to change at least every 5 years. In addition, 81 does not differentiate charities based on distance, hence is the most inclusive, and 82 only averages the fundraising expenditure of the top 10 largest charities by fundraising revenue in each rival group- defined by their beneficiaries and area of operation- which makes it conceptually distinct from the other definitions of a rival charity[102]. I have implicitly assumed that it is possible to accurately infer which charities are considered to be rivals by charity i in time t using observable financial data, however, greater accuracy may not be possible without further qualitative or survey data.

Table 24 and Table 25 show that estimated spillover effects are positive across all definitions, ranging from of 0.00395 (which corresponds to the definition of rival as the top 10 largest charities in a particular rival group) to 0.723 (which corresponds to the use of fundraising revenue as a measure of distance[103]). Furthermore, the estimates appear to be more stable across the small time and distance adjustments in 84-88 ranging in point elasticity estimates from 0.151 to 0.344, providing evidence of robustness with respect to the choice

cannot adjust employee numbers to fundraising shocks in the same year

[102]This is the definition of A_{it} used in Arulampalam et al. (2015) to ascertain the direction of spillover effects amongst international development charities

[103]This estimate is abnormally high, mostly likely because dependence between f_{it} and a_{it} is baked into the definition of a rival, hence the correlation, and 95% significance, is largely artificial

of time stability and exclusivity. However, the estimates are not statistically significant so conclusive evidence for their existence is not guaranteed.

Table 27 and Table 26 show that estimated strategic effects are positive and relatively stable across all definitions except 84 (which corresponds to a point elasticity estimate of -0.106) otherwise ranging from 0.349 to 0.0615. 84 differs from my preferred specification in that it is more exclusive, restricting the maximum distance between rivals to 10; a negative coefficient on 84 could indicate that strategic effects become negative when rivals reach a certain proximity to each other - the distance between rivals falls below a certain threshold. However, the estimates are not statistically significant, so we are unable to identify a causal relationship between exclusivity and the direction of strategic effects.

10.3.6 Heterogeneity across subgroups

Table 37 and Table 36 present the estimated coefficients for certain subgroups in the selected sample, in particular for exclusively animal focused charities (1), exclusively environment related charities (2), international charities (3) and exclusively national charities(4)[104]. The estimates appear to differ significantly across subgroups, with National charities reporting strongly positive estimated own fundraising effects, and International charities reporting negative own fundraising effects as shown in column 3 and 4 of Table 36. In addition, Environment charities report weakly positive estimated strategic effects and Animal charities report weakly negative estimated strategic effects, as shown in column 1 and 2 of Table 37. This suggests that the causal estimates obtained in Table 8 and Table 14 may represent an average causal effect across substantially different subgroups, which in itself is not a useful estimand.

The heterogeneity in causal estimates across subgroups is an oft encountered problem in the non-profit literature largely due to divergent organisational objectives between charities. I hypothesised that by focusing on fundraising, an inherently economic phenomenon, I would be able to uncover a set of behavioural equations common to all charities, albeit within a particular subsector. However, this assumes that charities are able to separate their fundraising objective, static net revenue ($\pi_{it} = F_{it} - B_{it}$) maximisation, from their main charitable objectives, such that charities are indifferent between different fundraising methods that generate the same net revenue. This is unlikely to hold in reality. For example, some charities may be unable or unwilling to engage in fundraising techniques such as street-chugging whilst others may only be able to raise money at certain times or events during the year leading to differently specified fundraising production functions and decision functions across charities. Furthermore, the fundraising objective itself may differ across charities; for example, it may be the case that smaller charities choose the fundraising expenditure budget B_{it} to maximise (gross) fundraising revenues F_{it}, in order to maximise their brand presence, whilst larger charities choose the fundraising budget to maximise long-run net

[104]The pairs, [(1) and (2)] and [(3) and (4)] are together mutually exclusive but not collectively exhaustive.

revenues ($V_i = \pi_{it} + \delta\pi_{it+1} + \delta^2\pi_{it+2} + \ldots$).

Pooling together charities with differently specified production and decision functions creates identification problems pertaining to misspecification bias[105].

10.4 Optimality

Optimal fundraising expenditures are defined as the level of fundraising expenditure that maximises a charity's net revenues $\pi_{it} = F_{it} - B_{it}$, such that an additional pound of fundraising expenditure generates one pound in revenue. Optimality ensures that charities are maximising their available funds, and by extension, maximising the welfare of their beneficiaries. Therefore, a case could be made that fundraising optimality is a more important measure of financial performance than efficiency, since a charity can become more fundraising efficient and spend a sub-optimal amount on fundraising. However, fundraising *optimality* is harder to determine for charity managers and evaluators, since it requires knowledge the true fundraising production function of each charity.

[105] I decide not to address this threat to identification in the main section, by distinguishing between charities with different fundraising objectives and techniques, mainly due to a lack of data and a concrete theoretical framework.

10.5 Tables

Key:

f_it_1	$= f_{it-1}$
f_it_2	$= f_{it-2}$
f_it_3	$= f_{it-3}$
b_it_1	$= b_{it-1}$
b_it_2	$= b_{it-2}$
b_it_3	$= b_{it-3}$
a_it_1	$= a_{it-1}$
a_it_2	$= a_{it-2}$
a_it_3	$= a_{it-3}$
f_it_d	$= \Delta f_{it}$
b_it_d	$= \Delta b_{it}$
a_it_d	$= \Delta a_{it}$
f_it_d_1	$= \Delta f_{it-1}$
f_it_d_2	$= \Delta f_{it-2}$
f_it_d_3	$= \Delta f_{it-3}$
b_it_d_1	$= \Delta b_{it-1}$
b_it_d_2	$= \Delta b_{it-2}$
b_it_d_3	$= \Delta b_{it-3}$
a_it_d_1	$= \Delta a_{it-1}$
a_it_d_2	$= \Delta a_{it-2}$
a_it_d_3	$= \Delta a_{it-3}$

Specifications are not cumulative. Cluster (within charity) standard errors are reported unless stated otherwise. Results presented are estimated using my preferred 2SLS strategy unless stated otherwise. F-Statistics are reported alongside SWF-Statistics where weak instruments are a concern. The Sanderson-Windmeijer multivariate F test (SWF) of excluded instruments is more robust test-statistic for the detection of weak instruments. See Sanderson and Windmeijer (2016) for details.

Table 16: Full output for the Production Function

	(1) Preferred specification	(2) no Time FEs	(3) HAC(1) SEs
f_it_1_d	0.710***	0.691**	0.710*
	(0.211)	(0.211)	(0.341)
b_it_d	0.554	0.522	0.554
	(0.543)	(0.542)	(1.087)
b_it_1_d	-0.269	-0.242	-0.269
	(0.179)	(0.177)	(0.277)
a_it_d	0.157	0.201	0.157
	(0.263)	(0.261)	(0.440)
y2007	0		0
	(.)		(.)
y2008	0		0
	(.)		(.)
y2009	0		0
	(.)		(.)
y2010	0.276		0.276
	(0.190)		(0.201)
y2011	-0.249		-0.249
	(0.268)		(0.540)
y2012	0.134		0.134
	(0.271)		(0.491)
y2013	0.00773		0.00773
	(0.204)		(0.206)
y2014	-0.0789		-0.0789
	(0.242)		(0.542)
exp_gov_d	5.51e-08	7.76e-08	5.51e-08
	(0.00000138)	(0.00000137)	(0.00000193)
_cons	-0.00219	0.00894	-0.00219
	(0.135)	(0.0233)	(0.263)
N	1926	1926	1926
R^2	-0.748	-0.679	-0.748
adj. R^2	-0.757	-0.683	-0.757
F	2.988	3.285	0.937

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: Full output for the Decision Function

	(1) Preferred specification	(2) no Time FEs	(3) HAC(1) SEs
f_it_1_d	0.180	0.191	0.180
	(0.194)	(0.198)	(0.194)
b_it_1_d	0.305	0.302	0.305
	(0.384)	(0.385)	(0.364)
a_it_d	0.0467	0.0495	0.0467
	(0.106)	(0.109)	(0.109)
y2007	0		0
	(.)		(.)
y2008	0		0
	(.)		(.)
y2009	0		0
	(.)		(.)
y2010	0.177		0.177
	(0.212)		(0.209)
y2011	0.273		0.273
	(0.183)		(0.182)
y2012	0.114		0.114
	(0.172)		(0.172)
y2013	-0.0288		-0.0288
	(0.226)		(0.204)
y2014	0.215		0.215
	(0.168)		(0.174)
_cons	-0.144	-0.0186	-0.144
	(0.158)	(0.0438)	(0.158)
N	1076	1076	1076
R^2	-0.251	-0.259	-0.251
adj. R^2	-0.261	-0.262	-0.261
F	1.263	0.558	0.721

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Changing the sample for the Production Function

	(1) Full sample	(2) Tobit	(3) No $\overline{NF_{it}}$ restriction	(4) No $\overline{b_{it}}$ restriction
main				
Delta f_it-1	0.319	0.200***	0.767*	0.251
	(0.227)	(0.0263)	(0.313)	(0.152)
Delta b_it	1.356	0.142***	0.329	1.065
	(1.532)	(0.0234)	(0.744)	(1.106)
Delta b_it-1	-0.438	0.0876***	-0.219	-0.206
	(0.510)	(0.0234)	(0.279)	(0.244)
Delta a_it	0.00416	0.146	0.0345	0.161
	(0.192)	(0.125)	(0.0981)	(0.339)
Observations	5165	6838	4062	2557
R^2	-2.542		-0.855	-0.784
Adjusted R^2	-2.549		-0.860	-0.791
Pseudo R^2		0.003		
F	0.324		1.441	0.719

Standard errors in parentheses
cluster standard errors are used
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Changing sample for estimating the Decision Function

	(1) Full sample	(2) Tobit	(3) No $\overline{NF_{it}}$ restriction	(4) No $\overline{leg_{it}}$ restriction
main				
Delta f_it-1	-0.146	-0.0503	0.117	-0.105
	(0.106)	(0.0507)	(0.138)	(0.134)
Delta b_it-1	0.321*	0.470***	0.361	0.196
	(0.134)	(0.0437)	(0.330)	(0.158)
Delta a_it	-0.0132	0.351	-0.0228	-0.261
	(0.144)	(0.232)	(0.118)	(0.232)
Observations	5165	6839	2240	2557
R^2	-0.242		-0.242	-0.110
Adjusted R^2	-0.244		-0.246	-0.114
Pseudo R^2		0.003		
F	2.205		1.027	1.427

Standard errors in parentheses
Cluster standard errors are used
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Differintial fundraising revenue time trends between $\overline{F_{it}}$ quintiles

	(1) Lowest quintile	(2) 2nd quntile	(3) 3rd quintile	(4) 4th quntile	(5) highest quintile
year	-0.0618	0.0679**	0.0138	0.0184	-0.0109
	(0.146)	(0.0239)	(0.0149)	(0.0369)	(0.0433)
Constant	132.1	-123.7*	-13.96	-22.67	37.99
	(293.4)	(48.03)	(29.95)	(74.29)	(87.15)
Observations	200	281	325	426	442
R^2	0.001	0.028	0.003	0.001	0.000
F	0.179	8.075	0.863	0.248	0.0627

Standard errors in parentheses

Each column corresponds to different quintile of average fundraising revenue.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: Differintial fundraising expenditure time trends between $\overline{B_{it}}$ quintiles

	(1) Lowest quintile	(2) 2nd quntile	(3) 3rd quintile	(4) 4th quntile	(5) highest quintile
year	-0.0906	0.117*	0.0307	0.0658	0.0561**
	(0.0938)	(0.0496)	(0.0293)	(0.0349)	(0.0207)
Constant	185.0	-224.7*	-48.92	-118.7	-97.07*
	(188.7)	(99.81)	(58.98)	(70.18)	(41.69)
Observations	340	334	331	340	328
R^2	0.003	0.017	0.003	0.010	0.022
F	0.932	5.591	1.095	3.553	7.313

Standard errors in parentheses

Each coloumn corresponds to different quintile of average fundraising expenditure.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: Differintial fundraising revenue time shocks between $\overline{F_{it}}$ quintiles

	(1) Lowest quintile	(2) 2nd quntile	(3) 3rd quintile	(4) 4th quntile	(5) highest quintile
y2007	0.110	0.394	-0.0889	0	0
	(0.736)	(0.362)	(0.419)	(.)	(.)
y2008	0	0.389	-0.140	0.0565	0.259
	(.)	(0.355)	(0.418)	(0.273)	(0.452)
y2009	-0.415	0.0403	-0.153	0.131	0.375
	(0.719)	(0.358)	(0.415)	(0.274)	(0.453)
y2010	0.582	-0.0899	-0.0777	-0.0978	0.479
	(0.721)	(0.356)	(0.418)	(0.274)	(0.457)
y2011	0.168	0	0.149	-0.193	0.568
	(0.719)	(.)	(0.419)	(0.273)	(0.453)
y2012	0.193	0.253	0.0749	0.0558	0.958*
	(0.721)	(0.362)	(0.421)	(0.272)	(0.453)
y2013	0.288	0.236	0	0.140	0.574
	(0.711)	(0.358)	(.)	(0.272)	(0.452)
y2014	0.261	0.347	-0.163	0.173	0.781
	(0.716)	(0.362)	(0.419)	(0.271)	(0.452)
y2015	-0.187	0.451	-0.0121	-0.101	0.611
	(0.719)	(0.360)	(0.424)	(0.273)	(0.453)
Constant	9.511***	12.33***	13.54***	14.42***	15.29***
	(0.514)	(0.256)	(0.299)	(0.194)	(0.324)
Observations	605	601	611	595	601
R^2	0.004	0.008	0.002	0.006	0.010
F	0.328	0.607	0.134	0.431	0.778

Standard errors in parentheses

Each coloumn corresponds to different quintile of average fundraising revenue.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: Differintial fundraising expenditure time shocks between $\overline{B_{it}}$ quintiles

	(1) Lowest quintile	(2) 2nd quntile	(3) 3rd quintile	(4) 4th quntile	(5) highest quintile
y2007	0.0177	-0.915	0.0355	-0.104	0
	(1.061)	(0.551)	(0.322)	(0.384)	(.)
y2008	-0.408	-0.821	-0.287	-0.381	0.109
	(1.047)	(0.551)	(0.324)	(0.384)	(0.230)
y2009	0	-0.575	0.157	-0.727	0.156
	(.)	(0.547)	(0.326)	(0.384)	(0.231)
y2010	-0.258	-0.249	0.180	0	0.304
	(1.047)	(0.551)	(0.324)	(.)	(0.233)
y2011	-0.375	0.142	0.332	0.0659	0.288
	(1.040)	(0.554)	(0.322)	(0.384)	(0.231)
y2012	-0.452	0.188	0.315	0.102	0.297
	(1.047)	(0.562)	(0.322)	(0.384)	(0.230)
y2013	-0.918	-0.241	0.0355	0.0897	0.374
	(1.027)	(0.554)	(0.324)	(0.384)	(0.230)
y2014	-0.578	-0.0807	0.400	0.281	0.413
	(1.027)	(0.554)	(0.328)	(0.384)	(0.230)
y2015	-0.706	0	0	-0.0546	0.506*
	(1.033)	(.)	(.)	(0.386)	(0.230)
Constant	3.254***	11.57***	12.68***	13.63***	15.41***
	(0.740)	(0.397)	(0.232)	(0.273)	(0.164)
Observations	340	334	331	340	328
R^2	0.004	0.026	0.021	0.029	0.023
F	0.175	1.094	0.877	1.257	0.952

Standard errors in parentheses

Each column corresponds to different quitile of average fundraising expenditure.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 24: Using different rival definitions in my preferred specification for the Production Function

	(1) a_it81_d	(2) a_it82_d	(3) a_it83_d	(4) a_it84_d
f_it_1_d	0.713***	0.714***	0.490*	0.710***
	(0.209)	(0.209)	(0.191)	(0.211)
b_it_d	0.563	0.561	0.523	0.535
	(0.551)	(0.563)	(0.509)	(0.519)
b_it_1_d	-0.272	-0.273	-0.255	-0.264
	(0.183)	(0.186)	(0.169)	(0.174)
a_it81_d	0.343			
	(0.402)			
a_it82_d		0.00395		
		(0.271)		
a_it83_d			0.723*	
			(0.366)	
a_it84_d				0.151
				(0.215)
N	1926	1926	1924	1926
R^2	-0.766	-0.766	-0.250	-0.723
adj. R^2	-0.775	-0.775	-0.257	-0.732
F	3.021	3.069	10.42	3.042

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 25: Using different rival definitions in my preferred specification for the Production Function continued

	(1) a_it85_d	(2) a_it86_d	(3) a_it87_d	(4) a_it88_d
f_it_1_d	0.703***	0.716***	0.714***	0.716***
	(0.210)	(0.210)	(0.209)	(0.209)
b_it_d	0.519	0.565	0.556	0.555
	(0.518)	(0.550)	(0.559)	(0.554)
b_it_1_d	-0.261	-0.273	-0.272	-0.273
	(0.173)	(0.184)	(0.183)	(0.183)
a_it85_d	0.344			
	(0.321)			
a_it86_d		0.293		
		(0.235)		
a_it87_d			0.136	
			(0.251)	
a_it88_d				0.261
				(0.230)
N	1926	1926	1926	1926
R^2	-0.692	-0.773	-0.759	-0.760
adj. R^2	-0.701	-0.782	-0.768	-0.769
F	3.193	2.947	2.936	3.190

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 26: Using different rival definitions in my preferred specification for the Decision Function

	(1) a_it81_d	(2) a_it82_d	(3) a_it83_d	(4) a_it84_d
f_it_1_d	0.183	0.186	0.122	0.185
	(0.195)	(0.195)	(0.193)	(0.196)
b_it_1_d	0.305	0.318	0.273	0.306
	(0.385)	(0.385)	(0.366)	(0.386)
a_it81_d	0.335			
	(0.182)			
a_it82_d		0.349*		
		(0.173)		
a_it83_d			0.199	
			(0.131)	
a_it84_d				-0.106
				(0.0989)
N	1076	1076	1076	1076
R^2	-0.252	-0.265	-0.179	-0.253
adj. R^2	-0.261	-0.274	-0.188	-0.262
F	1.410	1.375	1.352	1.503

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 27: Using different rival definitions in my preferred specification for Decision Function continued

	(1)	(2)	(3)	(4)
	a_it85_d	a_it86_d	a_it87_d	a_it88_d
f_it_1_d	0.180	0.179	0.179	0.179
	(0.194)	(0.194)	(0.194)	(0.194)
b_it_1_d	0.304	0.305	0.305	0.304
	(0.384)	(0.384)	(0.385)	(0.384)
a_it85_d	0.0615			
	(0.273)			
a_it86_d		0.125		
		(0.161)		
a_it87_d			0.116	
			(0.147)	
a_it88_d				0.257
				(0.184)
N	1076	1076	1076	1076
R^2	-0.251	-0.251	-0.251	-0.249
adj. R^2	-0.261	-0.260	-0.261	-0.258
F	1.440	1.318	1.281	1.363

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 28: Using different instruments in my preferred specification for Production Function

	(1) twice lagged levels	(2) thrice lagged levels	(3) Delta C_it	(4) Once lagged C_it
f_it_1_d	0.620	0.111	0.584	0.923
	(0.853)	(0.417)	(0.334)	(0.505)
b_it_d	4.561	1.058	0.126	1.281
	(21.93)	(0.900)	(0.153)	(1.972)
b_it_1_d	-2.406	-0.723	-0.157	-0.458
	(11.54)	(0.663)	(0.151)	(0.500)
a_it_d	1.444	0.362	-0.0111	0.443
	(7.322)	(0.493)	(0.181)	(0.843)
N	2284	1926	1926	1926
R^2	-62.381	-2.507	-0.412	-3.229
adj. R^2	-62.688	-2.525	-0.419	-3.251
F	0.0985	0.824	2.093	0.781

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 29: Production Function: First Stage Summary using twice lagged levels (1) as instruments

	(1) f_it_1_d	(2) b_it_d	(3) b_it_1_d
f_it_2	-0.221***	0.0891*	0.156***
	(0.0321)	(0.0412)	(0.0379)
b_it_2	0.0263	-0.135***	-0.254***
	(0.0136)	(0.0290)	(0.0254)
c_it	0.183***	0.116*	0.219***
	(0.0450)	(0.0497)	(0.0459)
Observations	2284	2284	2284
R^2	0.115	0.040	0.123
Adjusted R^2	0.111	0.035	0.119
SWF	0.04	0.04	0.04
F	16.85	18.61	34.31

Standard errors in parentheses

Each coloumn corresponds to an endogenous variable. Each row corresponds to an instrument.

F and SWF statistics are as before

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 30: Production Function: First Stage Summary using thrice lagged levels (2) as instruments.

	(1) f_it_1_d	(2) b_it_d	(3) b_it_1_d
f_it_3	-0.125**	0.0297	0.0993*
	(0.0381)	(0.0453)	(0.0457)
b_it_3	0.0180	-0.0902**	-0.136***
	(0.0196)	(0.0287)	(0.0326)
c_it	0.109*	0.135**	0.0831
	(0.0534)	(0.0504)	(0.0477)
Observations	1926	1938	1926
R^2	0.038	0.027	0.037
Adjusted R^2	0.033	0.022	0.032
SWF.	1.30	1.57	1.53
F	8.86	9.16	12.21

Standard errors in parentheses

Each coloumn corresponds to an endogenous variable. F and SWF are as before.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 31: Production Function: First Stage Summary using differenced C_it (3) as an instrument

	(1) f_it_1_d	(2) b_it_d	(3) b_it_1_d
Delta f_it-2	-0.242***	0.0768	0.0208
	(0.0629)	(0.0714)	(0.0643)
Delta b_it-2	-0.0249	-0.0600	-0.261***
	(0.0325)	(0.0403)	(0.0440)
c_it_d	0.0752	0.766*	0.0194
	(0.0901)	(0.314)	(0.189)
Observations	1926	1926	1926
R^2	0.071	0.023	0.074
Adjusted R^2	0.066	0.018	0.070
SWF	10.44	3.76	8.14
F	6.81	2.20	14.53

Standard errors in parentheses

Each coloumn corresponds to an endogenous variable. F and SWF are as before.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 32: Production Function:First Stage Summary using lagged C_it (4) as an instruments.

	(1)	(2)	(3)
	f_it_1_d	b_it_d	b_it_1_d
Delta f_it-2	-0.245***	0.0763	0.0211
	(0.0628)	(0.0714)	(0.0642)
Delta b_it-2	-0.0247	-0.0608	-0.261***
	(0.0324)	(0.0405)	(0.0440)
c_it_1	0.0662	0.00358	-0.00740
	(0.0438)	(0.0428)	(0.0410)
Observations	1926	1926	1926
R^2	0.073	0.014	0.074
Adjusted R^2	0.068	0.009	0.070
SWF	0.82	0.77	0.80
F	8.35	0.91	14.52

Standard errors in parentheses

Each coloumn corresponds to an endogenous variable. F and SWF are as before.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 33: Using different instruments in my preferred specification for Decision Function

	(1)	(2)
	twice lagged levels	thrice lagged levels
f_it_1_d	0.0592	-0.191
	(0.0924)	(0.852)
b_it_1_d	0.532*	0.119
	(0.215)	(0.670)
a_it_d	0.0100	0.0394
	(0.116)	(0.111)
N	1272	1076
R^2	-0.547	-0.084
adj. R^2	-0.558	-0.092
F	2.722	0.824

Standard errors in parentheses

All estimates are interpreted as elasticities

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 34: Decision Function: First Stage summary using lagged twice levels (1) as instruments

	(1)	(2)
	f_it_1_d	b_it_1_d
f_it_2	-0.203***	0.0450
	(0.0436)	(0.0242)
b_it_2	0.0865***	-0.0654**
	(0.0246)	(0.0211)
Observations	1275	1274
R^2	0.103	0.035
Adjusted R^2	0.097	0.028
SWF	10.92	11.65
F	6.31	4.95

Standard errors in parentheses

Each coloumn corresponds to the first stage for each endogenous variable.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 35: Decision Function: First Stage summary using lagged thrice levels (2) as instruments

	(1)	(2)
	f_it_1_d	b_it_1_d
f_it_3	-0.00495	0.00858
	(0.0476)	(0.0265)
b_it_3	0.0226	-0.0236
	(0.0255)	(0.0202)
Observations	1079	1077
R^2	0.009	0.012
Adjusted R^2	0.002	0.004
SWF	0.15	0.18
F	1.06	1.66

Standard errors in parentheses

Each coloumn corresponds to the first stage for each endogenous variable.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 36: Production Function:Heterogeneous treatment effects across subgroups

	(1) Animals	(2) Environment	(3) International	(4) National
Delta f_it-1	-0.502	0.821*	0.652	0.983**
	(0.575)	(0.365)	(0.635)	(0.313)
Delta b_it	0.0478	0.000786	-0.176	0.825
	(0.0728)	(0.393)	(0.462)	(0.800)
Delta b_it-1	-0.258	-0.168	-0.299	-0.395
	(0.527)	(0.177)	(0.327)	(0.359)
Delta a_it	-0.424	-0.0152	0.0909	0.205
	(0.231)	(0.300)	(0.693)	(0.379)
Observations	99	1544	683	1243
R^2	0.010	-0.846	-0.699	-2.227
Adjusted R^2	-0.103	-0.858	-0.724	-2.253
F	1432.3	2.379	0.701	1.599

Standard errors in parentheses
subgroups are mutually exclusive
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 37: Decision Function: Heterogeneous treatment effects across subgroups

	(1) Animals	(2) Environment	(3) International	(4) National
Delta f_it-1	-0.00618	0.578	-0.0604	0.367
	(0.00823)	(0.713)	(0.0874)	(0.401)
Delta b_it-1	0.229*	0.328	-0.226	0.676
	(0.0951)	(0.419)	(0.131)	(0.693)
Delta a_it	-0.0645	0.0956	0.0948	0.0730
	(0.0445)	(0.145)	(0.167)	(0.141)
Observations	95	812	452	624
R^2	0.175	-0.574	-0.004	-0.912
Adjusted R^2	0.098	-0.590	-0.022	-0.937
F	5.538	0.776	1.375	0.871

Standard errors in parentheses
subgroups are mutually exclusive
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Aguirregabiria, V. (2012): "Empirical Industrial Organization: Models, Methods, and Applications," .

Anderson, T. and C. Hsiao (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47–82.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Arulampalam, W., P. G. Backus, and J. Micklewright (2015): "Unofficial Development Assistance: A Model of Development Charitiesâ Donation Income," *Oxford Bulletin of Economics and Statistics*, 77, 191–209.

Bagwell, K. (2007): "The Economic Analysis of Advertising," 3, 1701–1844.

Dang, C. T. and T. Owens (2020): "Does transparency come at the cost of charitable services? Evidence from investigating British charities," *Journal of Economic Behavior Organization*, 172, 314–343.

Jacobs, F. A. and N. P. Marudas (2006): "Excessive, optimal, and insufficient fundraising among the Nonprofit Times 100," *International Journal of Nonprofit and Voluntary Sector Marketing*, 11, 105–114.

Khanna, J., J. Posnett, and T. Sandler (1995): "Charity donations in the UK: New evidence based on panel data," *Journal of Public Economics*, 56, 257–272.

Okten, C. and B. A. Weisbrod (2000): "Determinants of donations in private nonprofit markets," *Journal of Public Economics*, 75, 255–272.

Posnett, J. and T. Sandler (1989): "Demand for charity donations in private non-profit markets: The case of the U.K." *Journal of Public Economics*, 40, 187–200.

Rose-Ackerman, S. (1982): "Charitable Giving and Excessive Fundraising," *The Quarterly Journal of Economics*, 97, 193–212.

Sanderson, E. and F. Windmeijer (2016): "A weak instrument F-test in linear IV models with multiple endogenous variables," *Journal of Econometrics*, 190, 212–221, endogeneity Problems in Econometrics.

Sieg, H. and J. Zhang (2012): "The Importance of Managerial Capacity in Fundraising: Evidence from Land Conservation Charities," *International Journal of Industrial Organization*, 30, 724–734.

Steinberg, R. (1994): "Economics and philanthropy: A marriage of necessity for nonprofit organizations," *New Directions for Philanthropic Fundraising*, 1994, 7–25.

Vardanyan, M. and V. J. Tremblay (2006): "The measurement of marketing efficiency in the presence of spillovers: theory and evidence," *Managerial and Decision Economics*, 27, 319–331.