ESSAYS ON THE NATURE, PURPOSE AND MEASUREMENT OF SOCIAL IMPACT

A thesis submitted to the Department of Geography and the Environment of the London School of Economics for the degree of Doctor of Philosophy

London, February 2017

Paolo Lucchino
London School of Economics

Dedication

To my father.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 53,000 words.

Statement of conjoint work

I confirm that Paper 2 was jointly co-authored with Richard Dorsett (National Institute of Economic and Social Research) and I contributed 50% of this work.

I confirm that Paper 3 was jointly co-authored with Fadi Hassan (Trinity College Dublin and Centre for Economic Performance) and I contributed 50% of this work.

Abstract

This thesis consists of a series of papers relating to the nature, purpose and measurement of social impact. Paper 1 makes an empirical contribution to the understanding of firm incentives to be perceived as *social*, and finds that even quite distinct variants of *social enterprises* can nevertheless benefit form a *strategic* use of social responsibility. Paper 2 relates to the theoretical prediction that non-profits will deliver higher standards of quality in markets where this quality is hard to observe, and tests this, possibly for the first time in the literature, in the case of unemployment support services. In contrast to existing literature, we do not find strong evidence of for-profit vs. non-profit differentials in product quality, be it observed or unobserved. The immediate focus of Paper 3 and Paper 4 is on the role of access to light in the socio-economic transformation and growth of rural communities in the developing world, which partly constitutes a second strand of this thesis. Using a randomised controlled trial, Paper 3 identifies strong evidence of a causal link between access to light and educational attainment. Through a further randomised controlled trial, Paper 4 finds that access to light contributes to a diversification in household livelihoods from agricultural to non-farm economic activities. To our knowledge, this constitutes the first robust evidence that small-scale lighting sources can help stimulate the very first steps in the direction of economic transformation. Importantly, Paper 3 and Paper 4 also speak to the questions of measurement of social impact evoked in Papers 1 and 2 and some of the practical challenges typically faced in this pursuit. Bringing these various threads together, these papers contribute to our understanding of contemporary social enterprise, with a particular focus on the role of information and its measurement in tightly linking *social enterprise* with genuine *social impact*.

Acknowledgements

I first and foremost would like to thank my supervisors for their support over these years. I am extremely grateful to Olmo Silva for his understanding and respect for my reasons to embark in doctoral research, for his insightful and disinterested guidance throughout, and for his unwavering sense of humour. I am also extremely grateful to Simona Iammarino for her passion for the study of social enterprise, and for her valuable support in my pursuit of this over these years.

I would like to thank my examiners, Oriana Bandiera and Simon Quinn, for accepting to take on this role and providing constructive and critical scrutiny of this work.

I am deeply indebted to all those with whom I worked together in the research projects forming part of this thesis.

I would like to extend my warmest appreciation to Richard Dorsett for the co-authorship of Paper 2, as well as for years of intellectual exchange and inspiration. As co-authors of Paper 2, we gratefully acknowledge the contributions of Department for Work and Pensions (DWP) colleagues; in particular, Janet Allaker and Matt Garlick. This paper draws on and extends research we carried out while at the National Institute of Economic and Social Research and forming part of a broader evaluation of the Work Programme commissioned by DWP.

I would like to thank the UK Secure Data Service, and all the team there, for facilitating access to the data used in Paper 1.

I would like to thank the team of Powering Impact for managing the projects behind Papers 3 and 4 and for their outstanding support. Particular regards go to Mauro Ometto. I would also like to thank GIVEWATTS for its precious support in the field, and extend special mentions to Jesper Hornberg, Michael Waiyaki, and David Orina. I am extremely grateful for the financial support received from Enel Foundation, STICERD, the Coca-Cola Company and the Private Enterprise Development in Low Income Countries initiative to make these projects happen. A special regard goes to Fadi Hassan for the years we worked together on these projects, including co-authoring Paper 3 and collaborating on Paper 4; the substantial reciprocal learning gained; and for the dusty adventures in the field.

I would like express my gratitude to the Economic and Social Research Council for financial support in the form of the Advanced Quantitative Methods Studentship.

I have also benefited from discussions with other faculty and colleagues at the LSE. In particular, I would like to thank Henry Overman, Steve Gibbons and the participants of the LSE Economic Geography seminar for the helpful comments. All errors and omissions remain my own.

I owe my good memories of student camaraderie to Alessandra Scandura, Alessandro Sforza, Alex Jaax, Andrea Ascani, David Nguyen, Davide Luca, Emanuele Campiglio, Luisa Gagliardi, Nicola Mastrorocco and Ted Pinchbeck,

I offer my final token of gratitude to my wife Ana, for her patience and support, and my children Niko and Esma, for sharing their radiance.

Preface

When businesses seek to conquer the hearts and minds of consumers, they often look beyond just offering a good product. The product must have a soul, and it must be a *good* soul. Many firms strive to communicate their voluntary actions to integrate social and environmental considerations into their operations. Such behaviour has been referred to as *corporate social responsibility* (CSR), and many firms go as far as presenting themselves as *social enterprises*.

This thesis is motivated by the desire to make sense of the above phenomenon, and consists of a series of papers on the nature, purpose and measurement of social impact. While each of these papers is designed to make a specific contribution in its relevant literature, they also relate to cross-cutting questions on social business and impact.

These papers are heavily influenced by work concerning the nature of *social enterprise* in Lucchino (2017). Through a critical review of different strands of relevant literature, that paper asks whether the notion of *social enterprise* captures a unitary phenomenon, defined either on the basis of the presence of institutional features common to all social enterprises, or on the basis of common empirical patterns of behaviour. It concludes that this common basis is in fact lacking, and questions whether the concept of *social enterprise* conveys any substantive meaning. It further argues that research efforts would be better defined if, rather than attempting to study all social enterprises, they focused on specific institutional features present among a subset of social enterprises.

In the course of this review, it also points to the possible decoupling between the social behaviour firms are *perceived* to uphold and that which they *actually* implement. Indeed, in the presence of asymmetric information, market forces are not able to close this gap and to force firms to tie their claims to actual behaviour. This is particularly worrisome given social enterprises are more likely to emerge in industries where the degree of asymmetric information is high. The implication of this postmodern disjuncture between perception and reality has important detrimental consequences for the growth of a vibrant socially responsible business sector. On the one hand, it implies the wasteful investment of firm resources into the marketing social achievements. Secondly, a strategic firm will first and foremost want to be *perceived* as social, as this is what ultimately affects their bottom line. Whether this requires them to actually *be* social becomes of second order. At the extreme, it would be possible to have *social responsibility* without *social impact*.

The papers in this thesis therefore tackle important follow-up questions: Why may firms want to appear socially responsible? Can we expect them to actually deliver social impact? Does the measurement of social impact have a role in enabling socially responsible business? Taken together, the papers build an understanding of the phenomenon of social enterprise, in particular highlighting the role of *information* and its *measurement*, which we argue is fundamental to the existence of a thriving socially responsible sector.

We make an empirical contribution to the understanding of the incentives to be perceived as social in Paper 1. In particular, the paper exploits the possibility to identify different 'variants' of social enterprise in the UK Small Business Survey 2010, and studies whether these are subject to different incentives to be perceived as social. Specifically, it compares firms making a broad claim to social enterprise by stating they are 'a business with primarily social or environmental aims' and the subset of these that also reinvest surpluses into the business or the community. Importantly, we fail to find strong evidence of differences in the patterns of behaviour between these two firm types. There results are consistent with a *strategic* use of social responsibility by all such firms. One interpretation of this result is that the actual reinvestment of surpluses toward the social purpose does not attenuate the need for a firm to invest in marketing themselves as social. This may be precisely because the ease with which competitors can make unsubstantiated claims obfuscates the actions of firms actually delivering social impact. On the other hand, an alternative interpretation is that it is perfectly rational for even strongly 'mission-driven' firms to capitalise on their social reputation among consumers. If so, whether this interferes or not with the actual achievement of social impact would remain an open question. In either case, these results do suggest that the definition of social enterprise as simply a 'business that works for the social good' may be too wide, capturing a large share of firms engaging in conventional (profit-maximising) behaviour, perhaps delivering little in the way of genuine social impact.

Pursuing the question of actual social impact further, Paper 2 relates to the theoretical prediction that non-profits will deliver higher standards of quality in markets where this quality is hard to observe. There is a rich empirical evidence on this topic in the case of health services and nursing homes, but none, to our knowledge, in the case of welfare services. This paper attempts to fill this gap by studying the context of the UK Work Programme. We exploit the fact that, in this context, explicit financial incentives are contracted ex-ante and based on objective and observable metrics, thereby eliminating the possibility of extracting financial benefits from manipulating how the social reputation of the firm is perceived. Under these conditions, the firm would then only deliver social impact beyond contracted levels if it is either intrinsically motivated to do so or if it is forbidden to distribute profits anyways. We address this broader research question by focussing on the extent to which for-profit and non-profit providers differ in the levels of quality they deliver on observable and contracted outcomes compared to less observable and non-contracted margins. Following the suggestion at the end of Lucchino (2017), this is an example of paper focussing on a specific feature of a subset of social enterprises (specifically, their non-profit status) and identifying its implications. While we estimate that non-profit providers supply higher levels of quality on unobserved or non-contracted margins, this difference is not statistically significant.[1] In contrast to predictions from the literature, therefore, we do not find strong evidence of for-profit vs. non-profit differentials in product quality, be it observed or unobserved.

The above papers highlight how the detachment between claims and actual social impact can have detrimental consequences for the growth and sustainability of a healthy socially responsible business sector. Indeed, the fuzziness around actual impact can generate much of the same drawbacks as those that emerge from unfair competition arising from, for example, the avoidance of regulation or brand infringement. The greater the opacity around actual social impact, the greater the scope for non-compliers to benefit from a (fictitious) social reputation, to the detriment of compliers. In this context, transparency is a crucial tool to ensure a level playing field to all. The rigorous and independent measurement of social impact therefore has an important role to play in enabling the viability of the sector.[2]

The immediate focus of Paper 3 and Paper 4 is on the role of access to light in the socio-economic transformation and growth of rural communities in the developing world, which partly constitutes a second strand of this thesis.

Paper 3 focusses on the link between access to light and education. We randomly distribute solar lamps to 7th grade pupils in rural Kenya and monitor their educational outcomes throughout the year at quarterly frequency. Our identification strategy rests on the variation in treatment at the pupil level and in treatment intensity across classes, and finds a positive and significant intention-to-treat effect as well as a positive and significant spillover effect on control students. We provide suggestive evidence that the mechanism through which lamps affect students is by increasing co-studying at school, especially after sunset.

Paper 4 aims to provide new evidence on the important question of whether interventions such as access to light, which relax constraints in relation to the number of productive hours available, can help initiate similar, if much more modest, socio-economic transformations as those that come along with large scale electrification. To understand and quantify these dynamics, we run a randomised controlled trial distributing solar lamps to households in rural Kenya. Our findings are that access to light contributes to a diversification in household livelihoods from agricultural to non-farm economic activities. This evidence is supported by a consistent set of results across time-use, the incidence

[1]Note, however, that given the small number of sampling clusters, there is a concern that this test may not be sufficiently powered.

[2]In this thesis, we focus exclusively on quantitative measurement. The role of qualitative measurement in the overall evaluation of social enterprise is unquestioned, but remains outside the scope of this thesis.

of different productive activities, and income levels. To our knowledge, this constitutes the first robust evidence that small-scale lighting source can help stimulate the very first steps in the direction of economic transformation.

Importantly, Paper 3 and Paper 4 also speak to the questions of measurement of social impact evoked in Papers 1 and 2. Indeed, they are the product of a 4-year collaboration as independent evaluators for a non-profit organisation working in East Africa. The experience documented in these papers highlights the importance of independent impact evaluation for the operations of social enterprises, and some of the challenges typically faced in these contexts.

GIVEWATTS is a registered non-profit enterprise operating in several countries in East Africa.[3] Its mission is to provide clean energy to rural off-grid communities, in the form of solar solutions like solar lamps and home systems. Importantly, GIVEWATTS operates as a social enterprise, specifically a *commercial non-profit*, in the sense that its revenues come primarily from their sales rather than donations. The lamps are not distributed for free, but sold in installments.[4] At the same time, GIVEWATTS also fundraises from donors, in particular to expand geographically.

For GIVEWATTS, and many organisations like it, the measurement of its social impact is central and instrumental to the sustainment of a business identity such as the above. Indeed, without some evidence of social impact, it would be difficult for GIVEWATTS to argue they are any different than a retail chain selling electrical goods. Additionally, as the literature discussed in Lucchino (2017) indicates, GIVEWATTS' non-profit status is essential to attracting donations from individuals that are unsatisfied by the level of provision of public goods provided by the State and gaining their trust that it will not pocket (part of) these for itself.[5] Credible measurement of social impact acts as a catalyst to sustaining both these mechanisms. Ultimately, therefore, measurement is what substantiates a social business identity, and makes the social business model viable.

On a more practical level, the evaluation work contained in Paper 3 and Paper 4, as well as the analysis in Paper 2, exemplify some very typical methodological challenges faced when seeking to measure social impact, and illustrate the use of recent solutions suggested by the literature. Firstly, many interventions can generate spillovers from the target population onto the non-target population, making causal impact measurement more difficult. In Paper 3, we deal with this challenge by mimicking an experimental design created to address precisely this issue.[6] Secondly, the scale of the interventions of many socially-minded firms is often relatively small, and implemented across an even

[3] See http://www.givewatts.org/.

[4] This is intended to avoid creating dependency. Note, however, that the lamps were given for free in the experiments in Paper 3 and Paper 4

[5] See Weisbrod (1975) on voluntary donations towards public goods and Hansmann (2008) on the role of asymmetric information and trust.

[6] We mimic a randomised saturation design, recently formalised in Baird et al. (2014)

smaller number of locations or social groups. This raises the challenge of correct statistical inference in the context of clustered data and only few sampled clusters. We face this challenge in Paper 2 and Paper 3, and adopt two alternative solutions to address it: in Paper 2 we adopt an approach that relies on assumptions on the structure of the error terms,[7] while in Paper 3 we use a more data-driven approach: permutation inference. A final challenge faced by organisations wanting to measure their impact is the operationally and financially costly nature of doing so. In Paper 4, we experiment with one of the many ways in which technology promises to offer novel solutions to collect data and measure impact. Specifically, we make use of Interactive Voice Response (IVR) calls for the collection of time use-data. While on this occasion the insights from this data were below expectations, the possibilities allowed by this technology are nevertheless notable. IVR allowed to have exact, indeed programmatic, control on the timing of surveys (essential for the experiential measurement of time-use); to operate a 2-month survey campaign from a desk thousands of kilometres from the field; and all at a fraction of the cost of alternative modes of data collection.

Bringing these various threads together, these papers contribute to our understanding of contemporary social enterprise, with a particular focus on the role of information and its measurement. While many organisations now falling under the recent and fashionable term *social enterprise* are institutions like co-operatives and non-profits which date back centuries, something genuinely new may be occurring. Consumers appear to be increasingly socially-minded, and this emerges as empirically relevant for the behaviour of all types of 'social enterprises' (Paper 1). In particular, a new wave of not-further-defined 'businesses working for the social good' are likely to have emerged in response to this (Lucchino, 2017). If consumer preferences are a driving force of this new phenomenon, then the battle for their hearts and minds becomes the central arena within which social enterprise will manifest itself. The credible measurement of social impact is therefore critical to actually keeping this manifestation tightly linked to genuine social impact (Papers 2, 3 and 4).

As a closing comment, this discussion also highlights the importance of the role of what we might call the *social data scientist* in enabling a thriving socially responsible business sector. Such a role would drive the implementation of rigorous quantitative evaluation methods that are nevertheless operationally light-weight and non-intrusive, in particular thanks to the possibilities offered by new technology, to address the increasing need for social business to credibly measure and communicate their social impact.

[7] Donald & Lang (2007)

Contents

Paper 1

Strategic use of Corporate Social Responsibility: Are social enterprises any different?

1.1 Introduction and motivation

Firms are increasingly keen to be seen as more than just producers of a good product. Many take voluntary actions to integrate social, environmental, ethical, human rights and consumer concerns into their operations. Such behaviour has been referred to as *corporate social responsibility* (CSR). While the term tends to be associated with large corporations, the essence of the concept is generally applicable to any firm engaging in the private provision of public goods, regardless of their size. Recent surveys of the economics and management disciplines document an emergent understanding of the phenomenon (Bnabou & Tirole, 2010; Kitzmueller & Shimshack, 2012; Crifo & Forget, 2014).

The literature has identified a number of circumstances that can give rise to CSR. Firstly, intrinsic motivation may substitute extrinsic, monetary rewards. If controllers of the firm value social objectives, they may support these by foregoing part of their surplus. This gives rise to what is referred to as *not for profit CSR* (Kitzmueller & Shimshack, 2012). However, strong evidence of this trade-off is hard to come by. The comprehensive meta-analysis in Margolis et al. (2009) finds that the relationship between social responsibility and financial performance is weak, but if anything, positive. Indeed, a second strand of research suggests that CSR is not incompatible with profit maximisation: quite the contrary, firms will engage in *strategic CSR* when the costs of doing so are outweighed by the benefits (Baron, 2001; McWilliams & Siegel, 2001; Kitzmueller & Shimshack, 2012). Such benefits can take the form of, for example, improved matching in the labour market; increased willingness to pay on behalf of socially-minded consumers; and pre-empting the detrimental consequences of public politics (regulation) and/or private politics (activists) on firm operations. Finally, if social performance is imperfectly observable, an *appearance* of CSR can be sufficient to improve profits (Calveras & Ganuza, 2014). We refer to this latter behaviour as *manipulative CSR*.

Beyond CSR, a distinct yet related strand of entrepreneurship literature has emphasised the notion of *social enterprise*. Social enterprise is generally intended as implying *more* than CSR. If for CSR the social element is a by-product, for social enterprise it is central to the mission, or at least is just as important as profit in the 'double bottom line' of social and financial performance. In line with this, such firms are often referred to as *mission-driven firms*.

The concept of social enterprise has been presented in a variety of ways, and a consensus on what actually defines a social enterprise is lacking. One definition of social enterprise that has gained currency in the literature and policy debates is simply that of a 'business that works for the social good' (Alter, 2007; Peattie & Morley, 2008). However, firms thus defined may fall short of fulfilling any transformative promise. Indeed, such definition could simply be capturing firms engaging in strategic CSR. Worst still, if social behaviour is hard to verify, such firms could easily be engaging in no more than manipulative CSR.

For these reasons, others within the social enterprise literature have argued that the mere presence of a social mission statement does not adequately characterise genuine *social enterprises*. Instead, harder governance features, such as forms of participatory governance and/or limitations on the distribution of profits, are necessary, and hold better promise of delivering truly alternative firm behaviour (see Paper **??** for a review).

This paper contributes to strengthening the empirical grounding of such conceptualisations of socially motivated firms. Indeed, due to the lack of a shared definition of *social enterprise* and general paucity of standardised data on the broader landscape of socially motivated firms, theoretical constructs such as the above have had only limited empirical validation. This paper makes two contributions in this field. The first one relates to the mechanisms that give rise to CSR. As noted in Kitzmueller & Shimshack (2012), the existing empirical evidence has tended to consider how CSR arises as the response to a single stakeholder relationship, but not several ones at the same time. For example, a study may consider how building a social reputation helps attract socially-minded consumers, but do so without considering its effects on employee recruitment and retention. Results may therefore be confounded by omitted variables. This paper adds to this literature by examining multiple CSR mechanisms simultaneously. The second contribution relates to the understanding of differing interpretations of social enterprise, and particularly in the extent to which they differ in their use of strategic CSR. Specifically, we exploit firm responses to the social enterprise module of the UK Small Business Survey 2010 to identify firms making a broad claim to social enterprise by stating they are 'a business with primarily social or environmental aims' and the subset of these that satisfy the more stringent UK Government definition of social enterprise. The latter states that "*a social enterprise is a business with primarily social objectives* but also requires that "*surpluses [be] principally reinvested for that purpose in the business or in the community* (DTI, 2002). Ultimately, this is a test of whether these two conceptual characterisations of social enterprise actually differ empirically in some substantive way. In particular, we test the hypothesis that firms satisfying the broad but not the narrower definition of social enterprise will engage more in strategic, or even manipulative, CSR, while firms satisfying the more stringent definition will exhibit more mission-driven, not for profit CSR behaviour.

The analysis proceeds by correlating the incidence of each type of socially motivated firm with key product attributes, elements of market structure, and worker and consumer characteristics previously considered in the literature. By doing so in a multi-variate context, we explore associations across a much broader set of potential drivers than has been previously done in the literature. Furthermore, by comparing the extent to which each potential driver is associated with broadly versus narrowly defined social enterprises, we gain an insight into the extent two which these two types of firms respond differently to the same market contexts.

Results are consistent with a strategic use of CSR by all firms claiming they operate for social aims. This is particularly the case in response to the consumer market context and demand. Additionally, while our results indicate that 'businesses that work for the social good' but do not satisfy the narrow definition of social enterprise show evidence of strategic use of CSR, but we also find at best only *tentative* evidence that 'official' *Social Enterprises* are any less prone to do so. Importantly, the results in this paper suggest that the definition of social enterprise as simply a 'business that works for the social good' may be too wide, capturing a large share of firms engaging in conventional (profit-maximising) behaviour, perhaps delivering little in the way of genuine social impact.

1.2 Identifying CSR behaviour and Social Enterprises

The core dataset used in this paper is the Small Business Survey (SBS) 2010. The SBS is a UK survey of firms with less than 250 employees. It was commissioned by the UK Department for Business Innovation and Skills, and was carried out in the summer of 2010. Survey respondents were randomly sampled from the Dun & Bradstreet database, stratifying by the four UK countries and by firm size bands. The survey covers key business metrics as well as a number of topics of particular interest in relation to small and medium enterprises.[1] Importantly, the survey includes a module on social aims and *Social Enterprises*. While this module is ultimately geared toward identifying firms that fully satisfy the UK Government definition of social enterprise, it provides all the

[1]See BIS (2010) for more information.

information necessary to identify the variants of firms of interest to the current study. Specifically, it reports whether firms receive no more than 75% of its income from grants and donations; distributes no more than 50% of its surplus to shareholders; think of themselves as a business that has primarily social or environmental aims; and feel they fit the UK Government definition of social enterprise.

To ensure we focus on the types of firms that are most relevant to the dynamics studied in this paper, we exclude firms with certain characteristics. Firstly, given our focus on trading firms and their response to market dynamics, we exclude any firm that receives more than 75% of its turnover form grants or donations. Survey respondents excluded by this criteria are predominantly those involved in social work without accommodation and religious membership organisations. Secondly, given our emphasis on the role of relationships vis a vis labour and consumer stakeholders, we focus on firms with at least one employee and firms selling final products. We identify firms selling final goods and services in two ways. As discussed in Section 1.3.2, we use the UK Living Cost and Food Survey 2010 to construct measures of customer characteristics by industry. As our first approach, we therefore drop all firms in the SBS found in industries for which no expenditure is recorded in Living Cost and Food Survey 2010. The rationale for exclusion is both because the lack of relevant records in the Living Cost and Food Survey 2010 suggests these firms are involved in the production of intermediate rather than final goods, but also because their consumer-related right-hand side variables used in the estimations would be missing. As a second step, we drop all firm in industries that cannot be classified under the Search/Experience/Credence classification (Nelson, 1974; Liebermann & Flint-Goor, 1996, see Section 1.3.1). These two operations drop all surveyed firms found in wholesale trade, other membership organisations, research and development, mining, manufacture of metals and public administration. It also excludes some firms in the insurance and pension fund industries.[2]

Finally, all firms are asked "do you pay more than 50% of your trading profit or surplus to owners/shareholders?". Approximately 20% of firms respond affirmatively and are asked no further questions on social performance. Because we do not observe their responses to the questions we are most interested in, we exclude them from the analysis.[3]

All remaining firms are asked questions on social aims and social enterprise. They are asked "do you think of your business as a social enterprise, by which I mean a business that has mainly social or environmental aims?" and "how well do you think your business fits the following definition of a social enterprise: A business with primarily social/environmental objectives, whose surpluses are principally reinvested for that purpose

[2]As the degree of asymmetric information is high in these sectors, we would ideally want to keep these firms in our sample. However, the failure to match these to expenditure data from the Living Cost and Food Survey 2010 means we cannot construct the relevant variables for these firms. Note also that not all firms in the insurance and pension fund industries are dropped.

[3]In Section 1.5, we re-run the estimations including these firms. Results do not change.

Table 1.1: Social and environmental aims

Do you think of your business as a social enterprise, by which I mean a business that has mainly social or environmental aims?			
	Freq.	Percent	Cum.
Yes	944	35.01	35.01
No	1,591	59.01	94.03
Don't know	161	5.970	100
Total	2,696	100	

Table 1.2: Fit with UK Government definition

How well do you think your business fits the following definition of a social enterprise: A business with primarily social/environmental objectives, whose surpluses are principally reinvested for that purpose in the business or community rather than mainly being paid to shareholders and owners?			
	Freq.	Percent	Cum.
Yes, it is a very good fit	442	16.39	16.39
Yes, it is quite a good fit	705	26.15	42.54
No, it is not a good fit	1,402	52.00	94.55
Don't know	147	5.45	100
Total	2,696	100	

in the business or community rather than mainly being paid to shareholders and owners?". The distribution across possible responses to these questions is presented in Table 1.1 and Table 1.2.

Our core sample therefore consists of 2,696 market-trading and consumer-facing firms with at least one employee. We use responses to the above questions to develop a classification of firms that lends itself to being used as dependent variable in the empirical tests carried out in this paper. The first empirical contribution of this paper relates to our understanding of the market circumstances and stakeholder relationships that are associated with an increased recourse to *any* form of corporate social behaviour. We identify all firms that are making any claim to some form of social behaviour as those that answer affirmatively to the question "do you think of your business as a social enterprise, by which I mean *a business that has mainly social or environmental aims*?" (see Table 1.1). Those responding 'don't know' are classified as missing.

A possible criticism of such measure is that it is self-reported and may, in principle, be unrelated to true corporate social behaviour. However, considering the objectives of this analysis, this may be an advantage. If firms are acting strategically, they will first and foremost want to *appear* as behaving socially, especially if social behaviour is observed imperfectly. Their self-reported perception of their aims will capture this.

We then further split firms that see themselves as 'a business that has mainly social or environmental aims' according to whether they fit the UK Government definition of social

Table 1.3: Presence of social aims

	Freq.	Percent	Cum.
No Social Aims	1,591	64.49	64.49
Declared Social Aims	555	22.50	86.99
Social Enterprise	321	13.01	100
Total	2,467	100	

enterprise ("A business with primarily social/environmental objectives, whose surpluses are principally reinvested for that purpose in the business or community rather than mainly being paid to shareholders and owners"). The SBS includes a derived variable indicating fully compliant Social Enterprises. This is primarily determined by how the firm responds to the question on the extent to which it feels it fits the official definition of social enterprise (see Table 1.2). Given these firms have already stated they have primarily social/environmental objectives, we would expect that their responses to this latter question will be ultimately driven by the extent to which their 'surplus is reinvested for this purpose or into the community'. Conditional on having primarily social aims, firms that respond affirmatively to this second question should be those that are either bound by a strict governance requirement to reinvest surpluses (for example, by taking a not-for-profit legal form) or do so voluntarily but systematically enough to feel a 'very good fit' with the official definition of social enterprise. The self-reported nature of this information here is somewhat problematic, as in this case we are indeed concerned with whether firms actually do what they say. However, in light of the scarcity of systematic and standardised data on corporate social behaviour, analysing this information can therefore provide valuable insights.

In line with this approach, firms can therefore be classified as presented in Table 1.3: firms that don't have a social purpose (for short, *No social aims*); those that declare a social purpose but do not reinvest surpluses for this purpose or into the community (*Declared Social Aims*); or those that do have a social purpose and also reinvest surpluses (*Social Enterprises*). To be clear, when we are interested in the associations with *any* form of corporate social behaviour, we are comparing firms with No Social Aims to all firms that are either *Social Enterprises* or have *Declared Social Aims.*

Unfortunately, not all firms fit neatly into the desired classification, as can be seen in Table 1.4. Some firms state they do not see themselves as a business with primarily social or environmental aims and yet later state they feel a very good fit (2.78% of firms) or quite a good fit (10.42% of firms) with the UK Government definition of social enterprise. This is a logical contradiction as the UK Government definition requires one to be a business with primarily social or environmental aims. We refer to this group as *confused.* In light of their negative response to the direct question about being a 'business with primarily social or environmental aims', these firms are classified as ones with No Social Aims in

Table 1.4: Unusual fit with the classification

	Fit with definition				
Social Aims	Yes, very	Yes, quite	no	Don't know	Total
Yes	13.06	14.35	6.230	1.370	35.01
No	2.780	10.42	42.84	2.970	59.01
Don't know	0.560	1.370	2.930	1.110	5.970
Total	16.39	26.15	52	5.450	100

our estimations. As discussed in Section 1.5, omitting this group of firms does not alter the results.

A second group of firms state they are a business with primarily social or environmental aims but only feel their business fits the UK Government definition 'quite well'. These firms are classified as firms with *Declared Social Aims* in our estimation. Indeed, in the context of the possible responses to the survey questions, specifically the availability of the option to answering 'very well', the option 'quite well' takes the meaning of 'to a considerable extent, but not wholly'. Considering the firms in question have already answered affirmatively that they are a business with primarily social or environmental aims, we deduce that it is something about what they do with surpluses that make them hesitate from declaring that they wholly fit the UK Government definition. Given our aim is precisely to discriminate these firms around their use of surplus, we interpret their answer as indicating that these firms do not reinvest surpluses in the business or the community, or at least not systematically. This is in line with the approach taken by the Small Business Survey, whereby the derived variable identifier for *Social Enterprises* does not include these firms. Section 1.5 carries out some robustness checks around the treatment of this group of firms, and results broadly hold.

We can attempt to map each of our three main firm classifications onto the different variants of CSR. As *Social Enterprises* state they pursue social aims and do so at the expense of shareholder profit, their operations may be seen as a form of mission-driven, not for profit CSR. Firms with *Declared Social Aims* state they pursue social aims but without constraints on the use of surpluses. Their use of CSR is possibly strategic, if not manipulative. Finally, firms with *No social aims* will contain firms that do not engage in any CSR, as well as firms that might undertake some strategic CSR but hesitate from making this their primary aim. The challenge of this paper will be to reveal if indeed these conceptual variants actually differ empirically.

1.3 Methodology

Our aim is to study the associations between market circumstances and/or stakeholder relationships and an increased identification with *any* or a *specific* form of 'corporate social responsibility' and/or 'social enterprise'. Indeed, where the firm's stakeholders hold social preferences that can affect demand, profit maximisation by a conventional firm can give rise to *strategic CSR* (Baron, 2001; McWilliams & Siegel, 2001). However, most empirical work on this topic considers only one stakeholder or one political mechanism at a time, and may therefore be subject to omitted variable bias. This paper contributes to this literature by focusing *simultaneously* on several such circumstances and relationships that may make a strategic use of CSR attractive.

We do so by carrying out a number of regressions to assess the strength of the relationship between the presence of firms we categorise in *any* or in a *specific* type of social firm and key product attributes, elements of market structure, and worker and consumer characteristics that we construct at 4-digit SIC level from other datasets.[4] The selection of potential right hand side variables is therefore driven by the desire to validate, in a multivariate context, a number of testable predictions explored in previous work. Specifically, we estimate a linear probability model (OLS) where the left-hand side variable is a dummy equal to one if the firm is, depending on the equation, of *any* or a *specific* type of social firm and zero if it is a firm with no social aims. In other words, firms of each 'social' type are compared to the same population of firms with no social aims. Importantly, this set-up allows to subsequently explore the extent to which these associations differ between social firms of different types by comparing coefficients across equations.

The analysis carried out therefore consists of the following steps, which are each discussed in dedicated subsections:

1. In Section 1.3.1 we enumerate a set of factors potentially correlated with corporate social behaviour (broadly defined).

2. In Section 1.3.2 we identify the data sources from which we construct the desire proxy variables used in the estimation.

3. For each type of socially motivated firm, identify which of these factors emerge as the strongest predictors of that type firm using the approach described in Section 1.3.3. Results from this step will provide some initial insight into how different firm types may be associated with different factors.

4. We select a suitable model specification to be estimated across all types of firms following the approach in Section 1.3.4. A statistical comparison of the coefficients then reveals differential associations across firm types to the same set of potential predictors.

[4]See Png & Reitman (1995) for an example of similar approach in the context of branded vs. unbranded petrol stations.

1.3.1 Potential drivers and testable predictions

Drawing on existing literature, this section presents the factors associated with corporate social responsibility and/or an identification with social aims, and the possible priors in relation to the dynamics underlying these relationships. It also details the proxy variables included in the set of candidate covariates for estimation. Details on the data sources for these variables is covered in Section 1.3.2.

Importantly, there is strong evidence that consumer behaviour is influenced by CSR (Kitzmueller & Shimshack, 2012). Such influence feeds into willingness to pay, implying at least some of the cost of CSR can be transferred to final consumers. However, the precise mechanism that determines this willingness to pay is less well understood. The literature points to three main functions of the CSR 'brand'.[5] Firstly, CSR may act as a signal of product quality in the context of asymmetric information. Alternatively, CSR may strengthen overall brand recognition and attachment. Finally, CSR may be a luxury good that allows consumers to express their own self-image and social commitment.

In the first instance, the CSR 'brand' may be acting as a signal of product quality in the context of asymmetric information (Landes & Posner, 1987; Economides, 1998). When there is asymmetric information, profit-maximising firms may be tempted to reduce levels of unobservable product quality. In this context, CSR can be used to reassure consumers they won't be left short-changed. As such, the incentive to brand will be stronger for goods whose quality can only be ascertained after use compared with those whose attributes can be evaluated transparently before purchase. Such asymmetry levels can be classified using the Search/Experience/Credence taxonomy developed by Nelson (1974) and extended by Liebermann & Flint-Goor (1996). *Search goods* are defined as those whose quality can be evaluated fully before purchase - for example, a pencil. The quality of *experience goods and services*, on the other hand, may only be ascertained after consumption. An example of this is a restaurant meal. Finally, *credence services* refer to those services whose quality can be only imperfectly evaluated even after they have been used. These may be medical procedures or investment management. In line with the above hypothesis, Siegel & Vitaliano (2007) find that firms producing experience goods or credence services tend to invest more in CSR than firms selling search goods. We should expect to find the same results in our estimations, with possibly stronger correlations for 'strategic' than for 'mission-driven' firms. We therefore replicate their approach by including a Search/Experience/Credence classification of each 4-digit SIC industry into the set of potential covariates. Industries are classified, increasing in the degree of informational asymmetry, into search goods, experience goods and services, and credence services as presented in Table 1.5.[6] Furthermore, we expect that asymmetric information will pose a

[5] See Ramello (2006) for a review of the functions of brands and trademarks.

[6] Siegel & Vitaliano (2007) use the following, more detailed, classification: search goods, non-durable experience goods, durable experience goods, experience services and credence services. We omit the

Table 1.5: Search/Experience/Credence classification

Good Type	Industries
Search Goods	Clothing, Furniture, Footwear, Carpets, Mattresses
Experience Goods and Services	Health/Beauty, Cigarettes, Food, Cleaners, Newspapers, Office Supplies, Housing, Automobiles, Appliances, Hardware, Drugs, Glasses, Software, Signs, Books, Sporting Goods, Hobbies, Utilities, Advertising, Transportation, Vacations, Education, Training, Tours, Transportation, Banking, Car Rentals, Entertainment, Direct Mail, Real Estate, Cargo, Job Placement, Information, Nursing Homes, Sports Clubs, Hotels, Waste Collection, Landscaping
Credence Services	Investments, Trusts, Portfolio Management, Mutual Funds, Insurance, Health Care, Weight Control, Car Repairs

Categorisation developed by Nelson (1974) and extended by Liebermann & Flint-Goor (1996)

greater obstacle when purchase decisions are more financially demanding. For this reason, we include the average household budget share spent on goods in each industry in the set of variables considered. We expect this to be positively correlated with the incidence of CSR behaviour across industries.

A second influence of CSR on the consumer relationship relates to improving 'brand awareness' and 'salience'. Much like that which is sought through conventional advertising, this creates and maintains levels of brand differentiation and reduces competitive pressures. Investment in CSR is used to build customer loyalty, thereby raising barriers to entry. In line with this, Bagnoli & Watts (2003) show how the competition for socially responsible consumers leads to an inverse relationship between the provision of public goods and the level of competitiveness in an industry. We should therefore expect a negative correlation between the use of CSR and measures of competition in our estimations. To test this, our model allows for the potential inclusion of the Herfindahl index of turnover evaluated at 4-digit SIC level (as a measure of concentration, and therefore the absence of competition) and the annual entry rate, again at 4-digit SIC level, as a measure of the competitive threat on incumbents. We expect this is unlikely to be different between 'strategic' and 'mission-driven' firms.

Finally, the CSR brand be a *luxury good*, providing consumers a value in itself that becomes increasingly important as necessities are catered for. Under this view, individuals purchase socially responsible goods to a display their commitment to an issue or value and/or to signal adherence to a certain identity. This is different from the above two cases, as the demand for CSR is not driven by product or brand attributes, but relates to the consumers' tastes and demographic characteristics. Firms will seek to monetise

distinction between durable and non-durable goods and between goods and services to simplify the presentation and interpretation of the findings.

this demand, and one might expect that, controlling for other relevant demographic characteristics, there should be an increasing relationship between the provision of CSR and consumer incomes. Our model therefore allows for the inclusion of the average income level of families purchasing products by 4-digit SIC codes, normalised to have mean zero.

Importantly, however, the strength of this relationship may vary depending on the degree of asymmetric information about product attributes. The value to firms of using CSR to elicit willingness to pay from socially-minded consumers may be higher when asymmetric information on product attributes is high, as the former cannot be elicited transparently through the product itself. To test for this, our model also allows for the inclusion interacted slope terms for the consumer income variable for each of the above-mentioned Search/Experience/Credence taxonomy. We expect this will result in a higher incidence of 'strategic' firms in markets with high-income consumers and high information asymmetry (particularly if engaging in 'manipulative' CSR). On the other hand, this pattern should not arise among 'mission-driven' firms.

In light of the potential relevance of consumer preferences, we further include a number of consumer demographic characteristics in the set of potential right-hand-side variables. These are: the percentage of families who donate to charity; are headed by a female, someone under 40, a graduate; and that have children.

A second key stakeholder relationship where CSR can play a role is that with current and potential employees. CSR can be used as a signal for being a 'good employer' and attract better and/or more motivated applicants. It may allow to substitute monetary compensation with intrinsic rewards, thereby giving rise to a wage penalty and a 'labour donation' (Preston, 1989). Empirical evidence on this issue, however, indicates no such penalty once individual, job, and workplace characteristics are controlled for (Leete, 2001). Overall, existing evidence would suggest labour market considerations do not appear to be the primary driver behind observed CSR behaviour. Nevertheless, to account for a possible influence of worker intrinsic motivation on the behaviour of the firm, we include a number of variables describing worker demographics and preferences in the set of candidate right-hand-side variables. Specifically, we calculate the share of workers by 4-digit SIC who: donate to charity; engage in volunteer work; and state that their profession or political views are important to their identity. Additionally, we include the share of workers who are: female: under 30; graduates; and have children as further potential controls.

The literature on CSR has also focused on political stakeholder relationships as drivers of CSR. These include pre-empting the private politics of activists (Baron, 2001, as above) or the public politics of regulation (Lyon & Maxwell, 2011). Innes & Sam (2008) find evidence in support of these hypotheses in the context of the U.S. Environmental Protection Agency's 33/50 program. However, the lack of proxy data for the incidence of political and/or lobby pressures across industries places this aspect beyond the scope of this paper.[7]

Finally, we include firm size, age and region as potential controls. Furthermore, in light of the self-reported nature of the dependent variable, we also include the gender, age and qualification level of the survey respondent as possible controls.

1.3.2 Data Sources

The main dataset used in this paper is the Small Business Survey (SBS) 2010. We focus on the subsample of 2,696 market-trading and consumer-facing firms with at least one employee. Our estimation approach aims to relate our classification of 'social' firms to characteristics of the firm's product, market, workers and consumers. Much of this information is not available in the SBS. We therefore use other datasets to construct relevant variables at the 4-digit Standard Industrial Classification (SIC) 2003 and merge these onto the SBS.

We obtain information on workers from the first three waves of the UK Longitudinal Household Survey (also known as Understanding Society). Understanding Society is a panel dataset of over 30,000 households and 50,000 adults, and aims to be a multi-topic household survey to understand social and economic change in Britain. The industries where surveyed households work are classified according to 5-digit SIC 2007 codes. By converting these to 4-digit SIC 2003 codes, it is possible to construct variables characterising the workforce by industry (for example, the percentage female employees) and map these onto firms on the SBS. This operation is relatively straight-forward: 67% of 5-digit SIC 2007 codes each map fully onto a single 5-digit SIC 2003 code; 85% map fully onto at most two SIC 2003 codes; and only 2% map onto more than 5 SIC 2003 codes. Where SIC 2007 codes mapped onto more than one SIC 2003 code, the contribution of a given employee in Understanding Society to aggregate statistics by SIC 2003 was apportioned equally across the matching SIC 2003 industries. We used data from the first three survey waves, as some questions were only asked in a specific wave. Where the same questions are asked across multiple waves, pooling data improves the accuracy of the aggregate statistics by both increasing sample size as well as implicitly giving more weight to the characteristics of individuals who do not move in an out of an industry. This gives a closer approximation of the potential pool of employees faced by a firm.

Data on consumers is merged on to the SBS from the Living Costs and Food Survey 2010. This operation is more involved as the Living Costs and Food Survey 2010 classifies expenditure according to the Classification of Individual Consumption by Purpose (COICOP). Mapping the native SBS 4-digit SIC 2003 industry classification onto

[7] Note that the model in Baron (2001) identifies a mechanism predicting an inverse relationship between CSR and levels of competition as arising from the arena of private politics. Firms enjoying high levels of brand recognition (and therefore a lower level of competition, and higher rents) will be more sensitive to the threat that this image may be tarnished by activist actions, and will therefore respond strategically by pre-emptively delivering more CSR.

Table 1.6: Data sources

Data Source	Variables Constructed
Living Cost and Food Survey	% families giving to charity; % families headed by female; % families headed by under 40; % families who give to charity; % families with children; Mean disposable income of consumer; Mean budget share spent on good
Understanding Society	% does volunteering; % female; % gives to charity; % graduate; % has children; % saying political views are important to identity; % saying profession is important to identity; % under 30
Business Structure Database	Herfindal index at SIC 4 digit level; Annual firm entry rate at SIC 4 digit level
Small Business Survey	Age of firm; Age of respondent; Firm size by number of employees; Gender of respondent; Government Office Region; Qualifications of respondent
Nelson (1974) and Liebermann & Flint-Goor (1996)	Good type: Search / Experience / Credence

COICOP requires the use of correspondence tables to go through several classifications. Specifically, we map the 4-digit SIC 2003 codes through ISIC 3.1, ISIC 3, CPC 1.0 and finally COICOP codes. As before, when mapping classification A in terms of B, the latter classification was apportioned equally when its codes mapped onto more than one code in classification A. When mapping SIC 2003 in terms of ISIC 3.1, 93% of SIC 2003 codes mapped onto at most 3 ISIC 3.1 codes. This value was 90% when mapping ISIC 3.1 in terms of ISIC 3; 74% when mapping ISIC 3 in terms of CPC 1.0 ; and 86% when mapping CPC 1.0 in terms of COICOP. 368 SIC 2003 codes were successfully translated in terms of COICOP, each mapping onto a median of 5 and mean of 8 COICOP codes. Note that, given the Living Costs and Food Survey focused on final consumer products, the match with SBS data was not intended to be complete. Industries involved in the production of intermediate goods will not find a match on the Living Costs and Food Survey. Indeed, this is one way in which we are able to exclude firms not selling to final consumers. Family expenditure by COICOP was then aggregated in terms of SIC 2003 codes. Industry statistics where then calculated using both Living Costs and Food Survey weights and family expenditure so as to portray consumer profiles weighted by their contribution to demand, and hence their importance to firms.

Data on market structure is estimated from the Business Structure Database 2010 and merged on to the SBS via 4-digit SIC 2003 codes.

Table 1.6 provides a summary display of the data sources used to construct each of the variables discussed in Section 1.3.1. Table 1.7 and Table 1.8 present the mean value and standard deviation of all variables used in the estimations, respectively: a) across firms that do and do not state they are a ‘business with social or environmental aims’ and b) across firms that have No Social Aims, *Declared Social Aims* or are *Social Enterprises*.

Table 1.7: Descriptive statistics - By presence of social aims

	Not primarily social aims	Primarily social aims	Total
Market characteristics			
Search good	0.0857	0.0434	0.0699
	(0.280)	(0.204)	(0.255)
Experience good	0.791	0.880	0.824
	(0.407)	(0.325)	(0.381)
Credence good	0.124	0.0763	0.106
	(0.329)	(0.266)	(0.308)
Herfindahl index at SIC 4 digit level	0.0337	0.0296	0.0322
	(0.0729)	(0.0914)	(0.0803)
Annual firm entry rate at SIC 4 digit level	0.0952	0.0899	0.0933
	(0.0442)	(0.0412)	(0.0432)
Consumer characteristics			
(C) %Mean disposable income of consumer	0.0152	0.235	0.0972
	(0.816)	(0.921)	(0.863)
(C) %Mean budget share spent on good	1.367	2.745	1.881
	(2.054)	(2.768)	(2.438)
(C) % families who give to charity	2.157	2.132	2.148
	(1.137)	(1.247)	(1.179)
(C) % families headed by female	34.93	33.61	34.44
	(8.535)	(8.263)	(8.457)
(C) % families with children	38.85	47.12	41.93
	(13.38)	(22.39)	(17.75)
(C) % families headed by under 40	27.34	34.95	30.18

	Not primarily social aims	Primarily social aims	Total
	(14.03)	(19.84)	(16.84)
(C) % families giving to charity	30.16	35.26	32.06
	(9.910)	(13.58)	(11.68)
Worker characteristics			
(W) % female	36.79	51.00	42.09
	(23.68)	(25.73)	(25.40)
(W) % has children	33.09	31.89	32.64
	(7.751)	(7.654)	(7.735)
(W) % under 30	26.62	29.14	27.56
	(12.25)	(14.29)	(13.10)
(W) % graduate	36.18	37.02	36.50
	(21.49)	(18.96)	(20.58)
(W) % saying profession is important to identity	26.54	28.24	27.18
	(8.151)	(7.657)	(8.011)
(W) % saying political views are important to identity	5.207	5.554	5.336
	(3.858)	(3.283)	(3.657)
(W) % does volunteering	15.90	18.16	16.74
	(8.649)	(8.328)	(8.599)
(W) % gives to charity	69.07	67.66	68.55
	(10.68)	(9.805)	(10.39)

Table 1.8: Descriptive statistics - By type of firm

	No Social Aims	*Declared Social Aims*	Social Enter-prise	Total
Market characteristics				
Search good	0.0857	0.0505	0.0218	0.0694
	(0.280)	(0.219)	(0.146)	(0.254)
Experience good	0.791	0.858	0.935	0.825
	(0.407)	(0.350)	(0.248)	(0.380)
Credence good	0.124	0.0919	0.0436	0.106
	(0.329)	(0.289)	(0.205)	(0.308)
Herfindahl index at SIC 4 digit level	0.0337	0.0344	0.0242	0.0326
	(0.0729)	(0.102)	(0.0782)	(0.0812)
Annual firm entry rate at SIC 4 digit level	0.0952	0.0942	0.0840	0.0935
	(0.0442)	(0.0426)	(0.0379)	(0.0432)
Consumer characteristics				
(C) %Mean disposable income of consumer	0.0152	0.122	0.425	0.0927
	(0.816)	(0.905)	(0.921)	(0.861)
(C) %Mean budget share spent on good	1.367	2.266	3.494	1.847
	(2.054)	(2.440)	(3.111)	(2.420)
(C) % families who give to charity	2.157	2.189	2.043	2.150
	(1.137)	(1.367)	(1.073)	(1.185)
(C) % families headed by female	34.93	34.20	32.96	34.51
	(8.535)	(7.804)	(9.261)	(8.499)
(C) % families with children	38.85	42.60	54.32	41.71
	(13.38)	(19.65)	(24.65)	(17.53)
(C) % families headed by under 40	27.34	31.72	40.05	29.98
	(14.03)	(19.16)	(19.89)	(16.71)
(C) % families giving to charity	30.16	33.89	37.20	31.92

	No Social Aims	*Declared Social Aims*	Social Enterprise	Total
	(9.910)	(12.49)	(15.21)	(11.62)
Worker characteristics				
(W) % female	36.79	46.56	57.87	41.74
	(23.68)	(25.45)	(24.28)	(25.27)
(W) % has children	33.09	32.28	31.25	32.66
	(7.751)	(7.839)	(7.357)	(7.744)
(W) % under 30	26.62	29.09	28.88	27.47
	(12.25)	(14.63)	(13.33)	(13.01)
(W) % graduate	36.18	35.66	40.44	36.62
	(21.49)	(19.95)	(17.11)	(20.67)
(W) % saying profession is important to identity	26.54	27.61	29.55	27.17
	(8.151)	(7.844)	(7.293)	(8.037)
(W) % saying political views are important to identity	5.207	5.453	5.843	5.345
	(3.858)	(3.451)	(3.087)	(3.681)
(W) % does volunteering	15.90	17.26	20.14	16.76
	(8.649)	(8.531)	(7.615)	(8.610)
(W) % gives to charity	69.07	67.60	67.99	68.60
	(10.68)	(10.46)	(8.431)	(10.38)

1.3.3 Model selection

Having constructed the required dataset, we proceed to identify which factors emerge as the strongest predictors for each type of socially motivated firm. To be clear, the ambition is not to uncover any *causal* determinants of social responsibility. Instead, we aim to estimate the correlation between the incidence of firms identifying with some form of corporate social behaviour in a way that is at least conditional on other candidate features we can observe in our data. While certainly more modest than causal identification, this task nevertheless involves some challenges. Indeed, a naive approach would be to control for as many factors as possible in the estimation. This approach however suffers from some known pitfalls. For example, a larger number of predictors increases the risk that multicollinearity between variables leads to insignificant estimates. Furthermore, an excessive number of predictors can lead to overfitting the model to match the features of a specific sample, rather than provide credible inference about associations in the wider population.

To address these challenges, we adopt an approach to model selection that combines both insights from previous work as well as data-driven selection procedures. We define the set of candidate variables by drawing on existing literature, as detailed in Section 1.3.1. We then use a simulation approach to select the variables to be included in the model. We adopt a simulation approach to reduce arbitrariness and subjectivity in the model selection progress. In the absence of such approach, researchers normally tend to select only a limited set of candidate specifications. While these are typically selected on the basis of previous evidence, they may still suffer from a degree of subjectivity. Moreover, the goodness of fit is seldom evaluated out-of-sample, increasing the risk of overfitting.

The simulation approach adopted aims to generalise and systematise the intuition behind the conventional researcher-driven approach to model selection. As in that approach, we define the potential set of candidate predictors on the basis of previous work. However, instead of constraining the model exploration to a small number of specifications, we run several thousand iterations that each construct a model specification by selecting an independent random subsample from the set of candidate predictors. For each of these iterations, we evaluate the *out-of-sample* predictive accuracy of the specification, to address the risk of overfitting. This is evaluated by calculating the out-of-sample root mean squared error using 2-fold cross-validation (Stone, 1974). This involves randomly assigning each observation to one of two mutually exclusive subsets, also referred to as *folds*, of the data at the beginning of the simulation, stratifying by 4-digit SIC code.[8] Each specification is estimated on each fold in turn, and the predictive error is calculated using observations in the other fold. Running this simulation across many iterations produces a dataset relating out-of-sample predictive accuracy to randomly assigned covariates. Because covariates are independently and randomly sampled from the set of potential predictors, we can identify the causal effect of the inclusion of a given covariate on out-of-sample predictive accuracy with a simple regression of the out-of-sample root mean squared error on a full set of dummies indicating when a given covariate is included in the estimation.

The results from these simulations serve two purposes. Firstly, they provide evidence to inform the selection of variables in our estimations. More interestingly however, it

[8]Half of firms in each SIC industry are assigned to each fold. The fold is assigned by simple coin toss for the last firm in industries with an odd number of data points, or for firms in industries with only one data point

allows us to quantify the predictive power of a given covariate for each type of socially minded firm. This, in itself, already sheds some light on our research question. In particular, if predictors differ by firm type, we will have a first form of evidence that the incidence of these firms is differentially associated with market circumstances.

1.3.4 Model estimation

Drawing from the simulation results and previous literature, we define a common model specification to be estimated across all types of socially motivated firms. This will involve some trade-offs, as some covariates may improve the prediction for some firm types and worsen it for others. We resolve this trade-off by setting on a rule-based approach to the selection of variables. Specifically, as we are interested in comparing the association of a given covariate across firm types, we include all covariates that have been identified as improving the model fit for at least one type of social firm.

We estimate this model on each firm type and interpret the coefficients. The aim of these regressions is to provide a measure of the associations between each type of social firm and given stakeholder characteristics. We discuss how these results relate to the existing literature. Beyond this, we can then infer the relative association of each type of firm to these characteristics by taking the difference in the estimation coefficients across equations. As the samples from the estimations overlap, they are not statistically independent, and the coefficients from the two estimations will co-vary. To ensure the standard errors of the differences are computed correctly, the two models are estimated as Seemingly Unrelated Regressions (Zellner, 1962). We cluster standard errors at the 4-digit SIC level, as most regressors are defined at this level of hierarchy.

1.4 Results

1.4.1 Model selection

We start by interpreting the results from the model selection simulations. Table 1.9 presents the results from a regression of cross-validated out-of-sample root mean squared error, standardised to have mean zero and standard deviation of one, on a full set of dummy indicators for the inclusion of each potential covariate, estimated on data from 5,000 draws. The first column reports the results in relation to the prediction of firms that identify with any form of corporate social behaviour; the second column reports these for the prediction of *Social Enterprises*; and the third evaluates these for firms with only *Declared Social Aims*. Cells are colour-coded to give an immediate impression of the patterns of association across firm types. Green (red) cells indicate the covariate causes an improvement (deterioration) in the root mean squared error, statistically significant at the 10% level. Cells without highlighting indicate the effect of the covariate on the root mean squared error is insignificant.

Comparing the patterns across the columns already goes some way toward answering our questions of interest. For example, the top of the table reveals that a number of covariates are good predictors across all firm types. These include the percentage of workers who are female, who are under the age of 30, who engage in volunteering, and that donate to charity; the share of consumer female-headed families and the share that gives to charity. The financial importance of a purchase, as proxied by the mean budget share spend on a good, emerges as the strongest predictor across all firm types. To be clear, at this stage we can only identify whether a covariate is a good predictor of each firm type, not the direction of the association. Indeed, a given covariate could be a strong predictor of different firm types, but with the association being in opposite directions.

The Search/Experience/Credence classification, as well as its interaction with consumer income levels, has a mixed association across firm types. It is strongly predictive for *Social Enterprises*, but substantially reduces the predictive power of the models for *Declared Social Aims* firms or for firm identifying with any social behaviour. Additionally, consumer income on its own is largely irrelevant for predicting socially motivated firms. This mixed pattern of results suggests, in line with previous work by Siegel & Vitaliano (2007), that the informational asymmetries may play a role in whether firms present themselves as socially responsible. However, the mechanism may be more complex than believed previously.

More generally, the middle section of the table reveals the association of a number of covariates is heterogeneous across *Social Enterprises* and firms with *Declared Social Aims*. This result is strongly suggestive of the fact that these two types of firms may respond to quite a different set of external factors. It is to this task that we turn to next.

1.4.2 Model Estimation

The next step in the analysis consists of estimating the model to identify the direction and statistical significance of the associations between covariates and the incidence of types of socially motivated firms, with a view to interpreting both the coefficient estimates for each firm type as well as across firm types. This requires the selection of a common model specification to be estimated across all firm types. As discussed in Section 1.3.4, we include in our final model specification all variables that have emerged as improving the model fit for at least one type of social firm.

However, we make one ad hoc addition to the specification. Alongside the interaction of consumer income with the Search/Experience/Credence classification, we also include the classification indicators as covariates themselves, in light of the interest these have received in the literature. Because the consumer income variable is normalised to have mean

Table 1.9: Model selection simulation results

Covariate	Any corporate social behaviour	Social Enterprise	Declared Social Aims
(C) Mean budget share spent on good	-0.91128	-0.90573	-0.50237
(W) % female	-0.78885	-0.66785	-0.47039
(W) % does volunteering	-0.20555	-0.17673	-0.20214
(C) % families giving to charity	-0.04266	-0.10492	-0.12039
(C) % families headed by female	-0.09178	-0.05492	-0.08041
(W) % under 30	-0.08875	-0.06134	-0.06741
(W) % gives to charity	-0.06007	-0.07072	-0.09474
(C) % families headed by under 40	-0.08041	-0.30188	-0.0032
(C) % families with children	-0.19482	-0.95008	0.04116
Annual firm entry rate at SIC 4 digit level	-0.03687	-0.07012	0.08464
(C) % families who give to charity	0.00739	0.00233	-0.0657
Qualifications of respondent	-0.08654	0.2792	-0.3296
(W) % saying political views are important to identity	0.013	-0.00103	-0.03506
(W) % saying profession is important to identity	0.00468	-0.03863	0.07746
(W) % has children	0.03829	-0.08852	0.2526
(W) % graduate	0.13736	-0.00125	0.01435
Good type (SEC) x Consumer income	0.29141	-0.04017	0.20585
Firm size by number of employees	0.27533	-0.08544	0.40711
Good type: Search / Experience / Credence	0.23801	-0.20634	0.8941
Herfindal index at SIC 4 digit level	0.05725	0.01975	0.14408
Mean disposable income of consumer	0.14375	0.02573	0.12372
Age of firm	0.37665	0.02866	1.07009
Gender of respondent	0.54685	0.2965	0.49255
Age of respondent	0.74892	0.47208	0.90267
Government Office Region	0.91491	0.80927	0.99467

zero, the intercepts on the Search/Experience/Credence classification dummies represent the incidence of corporate social behaviour in an industry with customers on average incomes, while the slope coefficients on the interaction with income represent the conditional association between the presence of corporate social behaviour and a marginal increase in income levels within each classification group.

Despite the relatively inclusive approach to model selection, the out-of-sample root mean squared error for the chosen model is very close to the minimum value achieved across the 5,000 randomly selected specifications. This indicates the chosen variables are still likely to be among the most relevant for the prediction of the incidence of firms of all types, while at the same time avoiding overfitting the peculiarities of the data.

Table 1.10 presents the correlations between market circumstances and/or stakeholder relationships and an increased recourse to *any* or *specific* forms of corporate social behaviour as estimated with the chosen model. The first column reports the coefficients for the model where the left-hand side variable is a dummy for all firms identifying with any form of corporate social behaviour. The second column reports the same coefficients when predicting a dummy indicator for firms that fully fit the official definition of Social Enterprise, while the third column does so for firms that only *Declared Social Aims*. The final column is the difference in coefficients between the model for firms with *Declared Social Aims* and that for *Social Enterprises*.

Table 1.10: Main results

	Any social aims	*Social enterprise*	*Declared Social Aims*	*Declared Social Aims - Social enterprise*
Annual firm entry rate at SIC 4 digit level	-0.550	-0.446*	-0.218	0.228
	(0.16)	(0.08)	(0.56)	(0.54)
Search	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Experience	0.085	0.055	0.081*	0.026
	(0.23)	(0.25)	(0.09)	(0.55)
Credence	-0.029	-0.008	-0.003	0.005
	(0.70)	(0.87)	(0.95)	(0.91)
Search × Mean disposable income of consumer	-0.274*	-0.123	-0.150*	-0.027
	(0.06)	(0.18)	(0.06)	(0.68)
Experience × Mean disposable income of consumer	-0.050*	-0.004	-0.051*	-0.047*
	(0.05)	(0.80)	(0.05)	(0.04)
Credence × Mean disposable income of consumer	0.044	-0.075	0.059	0.134*
	(0.53)	(0.18)	(0.40)	(0.02)
(C) Mean budget share spent on good	0.019	0.021*	0.008	-0.013*
	(0.15)	(0.03)	(0.47)	(0.04)
(C) % families who give to charity	0.011	0.003	0.007	0.005
	(0.61)	(0.86)	(0.70)	(0.78)
(C) % families headed by female	0.001	0.001	-0.001	-0.002
Observations	2495	1884	2109	2427

Continued on next page

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.10 - continued from previous page

	Any social aims	*Social enterprise*	*Declared Social Aims*	*Declared Social Aims - Social enterprise*
	(0.78)	(0.47)	(0.81)	(0.37)
(C) % families headed by under 40	-0.002	-0.002	-0.002	0.000
	(0.39)	(0.38)	(0.52)	(0.98)
(C) % families with children	0.004*	0.006**	0.002	-0.004*
	(0.09)	(0.00)	(0.45)	(0.07)
(C) % families who give to charity	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
(W) % gives to charity	-0.003	-0.001	-0.003*	-0.003
	(0.12)	(0.56)	(0.06)	(0.11)
(W) % under 30	0.002	0.002	0.001	-0.001
	(0.47)	(0.42)	(0.66)	(0.70)
(W) % female	0.002*	0.001	0.002*	0.001
	(0.02)	(0.12)	(0.04)	(0.40)
(W) % does volunteering	0.001	0.002	0.001	-0.001
	(0.62)	(0.22)	(0.75)	(0.52)
(W) % saying political views are important to identity	0.006	0.001	0.007	0.006
	(0.19)	(0.84)	(0.14)	(0.19)
(W) % saying profession is important to identity	0.000	0.000	0.000	0.000
	(0.98)	(0.90)	(0.88)	(0.96)
(W) % has children	-0.002	-0.003*	0.000	0.004

Continued on next page

Observations	2495	1884	2109	2427

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.10 - continued from previous page

	Any social aims	*Social enterprise*	*Declared Social Aims*	*Declared Social Aims - Social enterprise*
	(0.39)	(0.06)	(0.83)	(0.11)
Observations	2495	1884	2109	2427

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

All models include the qualification of the respondent and firm size as controls

1.4.3 Correlates of any form of corporate social behaviour

We start by considering the associations with any form of social behaviour, as reported in the first column of Table 1.10. Firstly, we consider the finding of Siegel & Vitaliano (2007) and the idea that CSR is used to convey product quality and gain consumer trust in the presence of asymmetric information. As our model allows for an interaction of the degree of asymmetric information with consumer income levels, our results are somewhat more complex to interpret. The pattern across coefficients on the interaction with income suggests that the association between consumer income levels and CSR branding strengthens as the degree of informational asymmetry increases. Compared to the omitted category of firms offering search goods, the marginal effect of income on the probability of being a firm with any social aims is .224 for firms offering experience goods and services and .318 for firms selling credence services.[9] The left-most panel in Figure 1.1 conveys this more directly. In plots the differences in the probabilities of being a firm of a given social type between either firms selling experience goods and services or selling credence services and the baseline category of firms selling credence good, calculated at different levels of consumer income.[10] The increasing relationship is clear and the difference becomes statistically significant in the upper half of the income distribution.

[9]The former is the difference between the coefficient 'Experience $\times$ Mean disposable income of consumer' minus the coefficient 'Search $\times$ Mean disposable income of consumer', while the latter is the difference between the coefficient 'Credence $\times$ Mean disposable income of consumer' minus the coefficient 'Search $\times$ Mean disposable income of consumer'

[10]For firms with any social aims and for the case of experience goods and services, the plotted line is: $.085 + (-.050 - -.274) \times Meandisposableincomeofconsumer.$

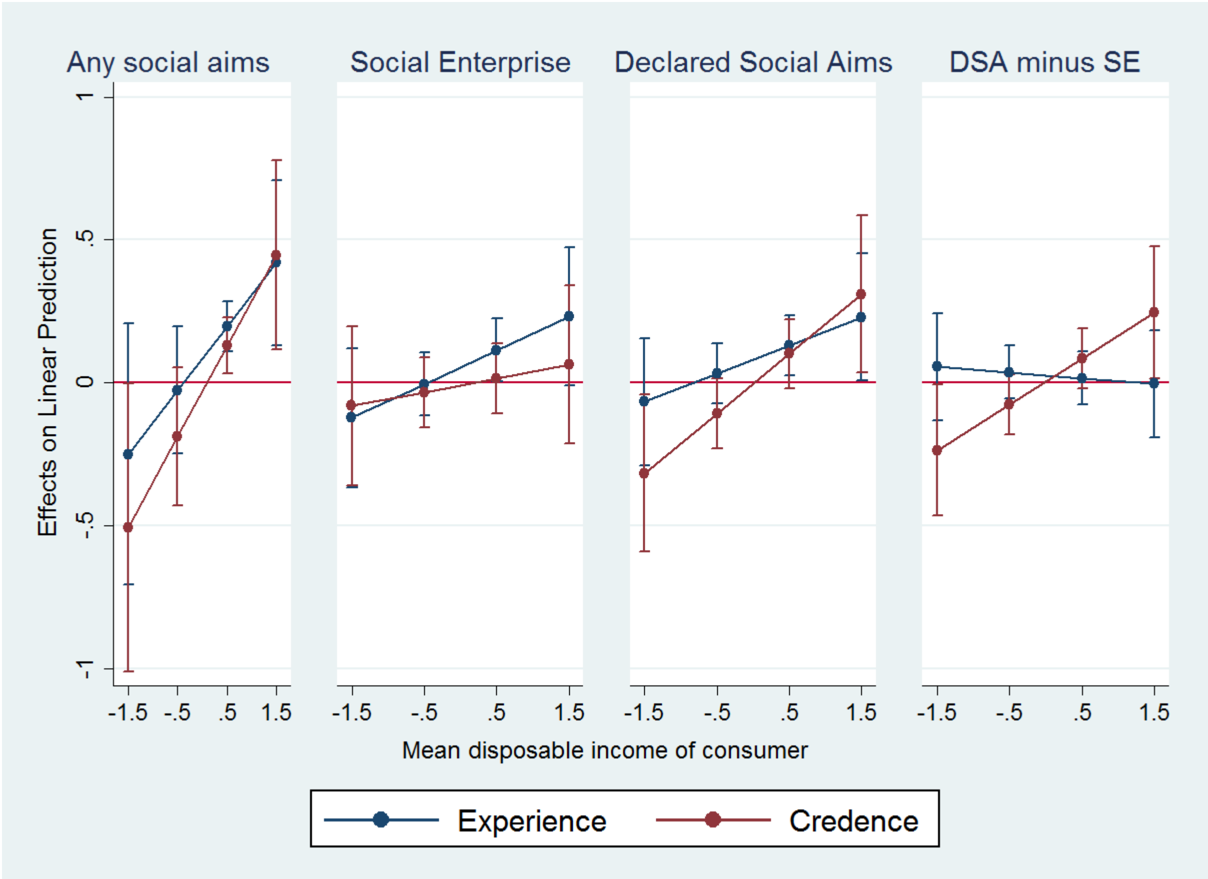


Figure 1.1: Information asymmetry and income

These results are consistent with the view that the demand for socially responsible production is indeed a luxury good. However, when informational asymmetry is low, firms may cater to this demand by communicating their socially responsible processes through the product itself rather than by branding themselves as socially responsible. On the other hand, when information is strongly asymmetric, firms wanting to cater towards socially-minded high-income consumers will be constrained to investing in their CSR image.[11] Overall, the estimations suggest that the use of CSR branding to attract socially minded affluent consumers intensifies as the level of informational asymmetry increases.[12]

The above result is robust to the inclusion of budget shares as a control. The latter emerges as positively associated with CSR, though it is not statistically significant. This lends some weak support to the view that the value of CSR branding in the eyes of the consumer becomes increasingly relevant the larger the financial commitment of a purchase.

Thirdly, the results are consistent with the evidence of higher CSR efforts in the context of imperfect competition, though they are not statistically significant (Bagnoli & Watts, 2003; Baron, 2001). The coefficient on the industry entry rate is large and negative, indicating that industries where there is higher firm entry (a proxy for high competition) are associated with a lower incidence of social firms.

The coefficients on the controls for worker and consumer characteristics reveal only a few patterns. The share of female workers in an industry is positively and statistically significantly associated with the incidence of social responsibility. The share of female-headed households is not associated with CSR, but firms selling to families with children tend to invest in more CSR. One could speculate this may be because these families have a higher threshold for product quality, and CSR is a means for the firm to communicate this.

Overall, the results are at least consistent with a strategic use of CSR. Our results suggest that CSR branding may be used to reassure consumers of product quality when this is imperfectly observed; to increase product and brand differentiation and maintain monopolistic power; and to capture demand for socially responsible production by affluent consumers, in particular when asymmetric information limits the extent to which this can

[11] For example, both beauty products and office supplies are experience goods. Firms producing beauty products will probably cater to a more affluent clientele, which we posit will also be more able to afford socially responsible production. Beauty product firms may do so by building a positive social brand image of themselves, but may be better off investing directly in the social responsibility of, for example, their ingredients. Consumer income levels will be associated with investment in CSR branding, but only weakly so. On the other hand, both car repairs and mutual fund management are credence services, though the former will have a wealthier customer base. Fund managers wanting to cater to this market will need to invest in a CSR image for the fund, as the high levels of information asymmetry make it difficult to do so convincingly through the attributes of the service itself. Within this category of goods, a strong correlation between consumer income and CSR image will emerge.

[12] We note the large negative coefficients in the bottom of the income distribution would instead suggest the opposite relationship: among industries catering towards low income consumers, firms operating in the context of asymmetric information will be *less* likely to present themselves as socially responsible. However, these coefficients are not statistically different from the base category of search goods.

be done directly through product attributes. Importantly, as an extension to existing empirical evidence, these results emerge when simultaneously including characteristics of the firm, workers, consumers and the market in the estimation.

1.4.4 Variants of social enterprise

Columns 2-4 in Table 1.10 aim to shed light on how correlations between social behaviour and stakeholder and market characteristics may be different between firms that only 'Declare social aims' and those that 'reinvest surpluses for this purpose or into the community', thereby fully satisfying the UK Government definition of 'Social enterprise'. The main interest here is testing the prior that firms that only 'declare they have social aims' will emerge as more responsive to circumstances where there are higher returns to such branding. Conversely, we want to test whether official *Social Enterprises* may be just as likely to exploiting this element of their brand. The second column reports the model coefficients for *Social Enterprises*, the third column does so for firms that only *Declared Social Aims*, and the final column is the difference in coefficients between firms with *Declared Social Aims* and *Social Enterprises*.

The results confirm and extend the insights discussed in the previous subsection. Firstly, even when taking firms with *Declared Social Aims* and *Social Enterprises* separately, the results broadly replicate the findings in the first part of the estimation. Coefficients are generally similar in direction and magnitude. In particular, the coefficient on the industry entry rates is still negative for both types of firms, and becomes statistically significant for *Social Enterprises*.

This similarity in results implies that, apart from very few exemptions, the differences in the two sets of estimation coefficients are not statistically different from zero. In other words, the estimations fail to reject the hypothesis that the two types of social firms show the same association with worker, consumer and market circumstances.[13]

However, a couple of exceptions point to some suggestive evidence of differential behaviour between the two firm types. Firstly, while the estimations for firms with *Declared Social Aims* replicates the overall finding of an increasing correlation with consumer income as levels of asymmetric information increase, this pattern is weaker for official *Social Enterprises*. Again, we can use Figure 1.1 to interpret the results more immediately. The second and third panels plot the marginal effect of offering experience goods and services or credence services at different levels of consumer incomes for *Social Enterprises* and firms with *Declared Social Aims* respectively. The difference between the two is in the rightmost panel. The positive relationship is evident in both, but is somewhat weaker for *Social Enterprises*. In particular, while the difference in the marginal effects is not statistically significant, the rightmost panel lends support to the view that firms with

[13]This interpretation must be considered alongside the possibility that high measurement error in our data may be what is ultimately determining the lack in statistical significance across coefficients.

Declared Social Aims may respond more to consumer incomes in markets for credence services. In other words, relative to *Social Enterprises*, firms with only *Declared Social Aims* may show a more aggressive use of CSR branding to capitalise on the demand by socially-minded high-income consumers in markets with asymmetric information.

Secondly, the coefficient on the consumer budget share and the share of familes with children is positive for both types of firms, but statistically significant for *Social Enterprises* only. Moreover, the difference in these coefficients between firms with *Declared Social Aims* and *Social Enterprises* is statistically significant. This suggests that it may be official *Social Enterprises* more than other types of social firms to use forms of CSR branding in relation to specific types of consumers. On the other hand, where statistically significant associations emerge in relation to worker characteristics, these tend to be for firms with only *Declared Social Aims*, suggesting these firms more than other types may be making use of their social image to attract and retain workers.

1.5 Robustness checks

The main results presented in the preceding sections are broadly robust to a number of alterations to the estimation approach. Summary results of these robustness tests are presented in Table 1.11 and Table 1.12. Table 1.11 focuses on the associations with any form of corporate social behaviour. The first column replicates the first column in Table 1.10. Each of the remaining columns display the results under different robustness checks. Similarly, Table 1.12 looks at the difference in the estimated coefficients for firms with *Declared Social Aims* and *Social Enterprises*. The first column replicates the last column in Table 1.10, and again each of the remaining columns display the results under each robustness check.

Table 1.11: Robustness checks - Correlates of CSR

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
Annual firm entry rate at SIC 4 digit level	-0.550	-0.669	-0.714*	-0.310	-0.513	-0.508
	(0.393)	(0.414)	(0.409)	(0.370)	(0.429)	(0.380)
Search	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Experience	0.085	0.071	0.046	0.068	0.096	0.054
	(0.071)	(0.071)	(0.070)	(0.059)	(0.080)	(0.070)
Credence	-0.029	-0.025	-0.047	-0.034	-0.043	-0.053
	(0.075)	(0.074)	(0.074)	(0.063)	(0.085)	(0.074)
Search × Mean disposable income of consumer	-0.274*	-0.265*	-0.253*	-0.210*	-0.307*	-0.292*
	(0.144)	(0.139)	(0.135)	(0.117)	(0.156)	(0.144)
Observations	2495	2247	2091	3202	2147	2330

Continued on next page

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.11 - continued from previous page

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
Experience × Mean disposable income of consumer	-0.050*	-0.065*	-0.048*	-0.045*	-0.036	-0.033
	(0.026)	(0.028)	(0.029)	(0.023)	(0.026)	(0.024)
Credence × Mean disposable income of consumer	0.044	0.140*	0.352***	0.001	0.044	-0.007
	(0.070)	(0.076)	(0.105)	(0.056)	(0.077)	(0.067)
(C) Mean budget share spent on good	0.019	0.000	-0.002	0.016	0.025*	0.018
	(0.013)	(0.014)	(0.014)	(0.013)	(0.015)	(0.013)
(C) % families who give to charity	0.011	0.033	0.039*	0.000	0.010	0.011
	(0.021)	(0.020)	(0.019)	(0.019)	(0.021)	(0.020)
(C) % families headed by female	0.001	0.001	0.002	-0.000	0.001	0.000
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
(C) % families headed by under 40	-0.002	0.000	0.005*	-0.003	-0.001	-0.001

Continued on next page

Observations	2495	2247	2091	3202	2147	2330

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.11 - continued from previous page

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(C) % families with children	0.004*	-0.002	-0.009**	0.005*	0.004*	0.005*
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
(C) % families who give to charity	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
(W) % gives to charity	-0.003	-0.002	-0.003	-0.003*	-0.004*	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(W) % under 30	0.002	0.003	0.004	0.001	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(W) % female	0.002*	0.002*	0.001	0.002*	0.002*	0.002*

Continued on next page

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
Observations	2495	2247	2091	3202	2147	2330

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.11 - continued from previous page

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(W) % does volunteering	0.001	0.001	0.001	0.001	0.001	0.002
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
(W) % saying political views are important to identity	0.006	0.005	0.005	0.003	0.007	0.007
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
(W) % saying profession is important to identity	0.000	-0.001	-0.002	-0.000	0.001	-0.000
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
(W) % has children	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	2495	2247	2091	3202	2147	2330

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

All models include the qualification of the respondent and firm size as controls

Firstly, the omission of suspected outliers does not alter the results. Outliers were identified as follows. A flag was created for each of the independent variables used in the estimation indicating observations with values above or below 2.5 standard deviations from the mean. Next, an outlier risk variable identifying the percentage of regressors that was flagged as possible outliers for each firm was created. We then re-estimate the model omitting firms with an outlier risk greater than 20%. Results are unchanged across all specifications.

A second set of robustness tests were carried out to address the concern that some industries may be perceived as being intrinsically 'social'. The concern here is that if CSR is the joint provision of a private and a public good, then it may not make sense to talk about CSR by a firm that is already in the business of producing a public good. Examples of such activities are nurseries and hospitals, whose product inevitably bundles a public good in them (e.g. an educated and healthy citizenry). Additionally, their cost may be subsidised, such that the pure private market mechanism is weakened. To address this concern, we re-run the estimation dropping all firms the Health, Social Services and Education sectors.[14] The results are unchanged. Indeed, they actually strengthen (in terms of magnitude and statistical significance). This may be readily explained. Firms that are already in a business that is widely perceived as 'obviously social' or that can point to their legal form as proof of their social commitment, may not need to do much more to assert their CSR. Instead, when the sample is restricted to 'normal' private firms in 'normal' markets, the market-based incentives that are associated with a strategic use of CSR emerge even more strongly precisely because these incentives allow firms to use CSR to differentiate themselves from other competitors in the same industry.

[14]We also re-run the estimation on samples that: a) exclude Health & Social Services, b) drop any SIC code at 4-digit level that had an incidence of firms with social aims of 90% or higher and c) drop all firms with legal forms that suggested they may be in a business of providing partly or fully public goods (e.g. charities, non-profits, industrial and provident societies, friendly societies). Results are robust to all such tests.

Table 1.12: Robustness checks - *Declared Social Aims* minus social enterprise

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
Annual firm entry rate at SIC 4 digit level	0.228	0.262	0.270	0.228	0.298	0.358
	(0.372)	(0.387)	(0.370)	(0.372)	(0.417)	(0.315)
Search	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Experience	0.026	0.017	0.033	0.026	0.024	-0.020
	(0.044)	(0.045)	(0.044)	(0.044)	(0.055)	(0.039)
Credence	0.005	-0.001	0.002	0.005	-0.000	-0.026
	(0.047)	(0.047)	(0.048)	(0.047)	(0.056)	(0.049)
Search × Mean disposable income of consumer	-0.027	-0.029	-0.033	-0.027	-0.046	-0.045
	(0.066)	(0.065)	(0.064)	(0.066)	(0.084)	(0.064)

Continued on next page

Observations	2427	2196	2046	2427	2079	2262

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.12 - continued from previous page

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
Experience × Mean disposable income of consumer	-0.047*	-0.043*	-0.053*	-0.047*	-0.043*	-0.024
	(0.023)	(0.024)	(0.022)	(0.023)	(0.025)	(0.019)
Credence × Mean disposable income of consumer	0.134*	0.154*	0.056	0.134*	0.137*	0.076
	(0.058)	(0.070)	(0.099)	(0.058)	(0.060)	(0.057)
(C) Mean budget share spent on good	-0.013*	-0.015*	-0.012*	-0.013*	-0.011	-0.016**
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)
(C) % families who give to charity	0.005	0.004	-0.002	0.005	0.003	0.004
	(0.016)	(0.017)	(0.016)	(0.016)	(0.018)	(0.015)
(C) % families headed by female	-0.002	-0.002	-0.002	-0.002	-0.002	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(C) % families headed by under 40	0.000	0.001	-0.002	0.000	-0.000	0.001

Continued on next page

Observations	2427	2196	2046	2427	2079	2262

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.12 - continued from previous page

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
(C) % families with children	-0.004*	-0.005	-0.001	-0.004*	-0.004*	-0.003
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
(C) % families who give to charity	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
(W) % gives to charity	-0.003	-0.003	-0.003	-0.003	-0.003*	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
(W) % under 30	-0.001	-0.000	-0.001	-0.001	-0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(W) % female	0.001	0.001	0.001	0.001	0.001	-0.000
Observations	2427	2196	2046	2427	2079	2262

Continued on next page

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

Table 1.12 - continued from previous page

	Main	Outliers	Drop Education, health and social work	Include not asked	Exclude confused	Exclude stark
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(W) % does volunteering	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(W) % saying political views are important to identity	0.006	0.007	0.006	0.006	0.007	0.007
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
(W) % saying profession is important to identity	0.000	0.000	0.001	0.000	0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(W) % has children	0.004	0.004	0.004	0.004	0.004	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Observations	2427	2196	2046	2427	2079	2262

p-values in parentheses

* $p < .10$, ** $p < .01$, *** $p < .001$

All models include the qualification of the respondent and firm size as controls

The final set of robustness checks relates to the treatment of firms whose responses imply they do not fit neatly into our classification. As discussed in Section 1.2, a minority of *confused* firms state they do not see themselves as a business with primarily social or environmental aims and yet later state they feel a good fit with the UK Government definition of social enterprise. We re-run the estimations omitting these firms, and find the results are unchanged.

Finally, we consider the issue that our group of firms with only *Declared Social Aims* consists of two distinct subgroups. One of these captures firms that state they are a business with primarily social or environmental aims but only feel their business fits the UK Government definition 'quite well'. We refer to these firms as *hesitants*. The other subgroup is firms that state they have social aims but also state that they do not identify themselves at all with the definition of Social Enterprise. As the UK definition requires having social aims, these responses are in stark contrast with each other, and indeed the number of firms in our sample that fits this response pattern is rather small. We refer to these firms as *starks*. To rule out that the main results may be driven by the inclusion of such a peculiar group of firms, we re-run the analysis by including *hesitants* only in our *Declared Social Aims* group. Precisely because hesitants differ from official *Social Enterprises* in their responses only ever so slightly, any difference could potentially be considered even more noteworthy. Overall, as can be seen in the right-most columns in Table 1.11 and Table 1.12, we confirm the broad pattern of results is replicated. It is particularly interesting to see that firms that hesitate to claim they reinvest surpluses in the business or community but do state they are a business with social aims are more likely to operate in industries with high-income consumers and asymmetric information. This evidence lends support to the idea of a strategic use of CSR.

1.6 Discussion and conclusion

Recent work in the economics and management disciplines has consolidated our understanding of corporate social responsibility. However, empirical validation of this understanding has tended to lag somewhat behind, greatly hampered by the lack of suitable data. This shortcoming is even more severe in the literature on social enterprise, which faces important additional challenges in defining the object of study. The aim of this paper is to make an empirical contribution on these topics by exploiting valuable information on social responsibility and behaviour available in the UK Small Business Survey 2010.

Our findings indicate that insights from previous work generally remain valid when explored together in a multivariate framework. In particular, although rarely statistically significant, the results are at least consistent with the use of CSR as a means to: elicit consumer engagement, especially in the presence of asymmetric information; improve matching in the labour market; and to generate a degree of market power. Overall, this analysis brings additional evidence in support of the view that CSR is not a defining feature of (some) firms, but rather a behaviour that is set in motion by conventional profit-maximising firms in response to (consumer) market incentives.

The paper then proceeds to ask whether being a 'social enterprise', which is indeed put forward as a defining feature by some firms, is associated with any different behaviour in this respect. Overall, social enterprises are found to respond somewhat more weakly, but ultimately no differently, than strategic firms to just about all margins considered. Therefore, the results at best tentatively suggest that 'official' social enterprises may be somewhat less responsive to strategic use of their social image.

It may very well be that it is perfectly rational for even strongly 'mission-driven' firms to capitalise on their social reputation among consumers, and that doing so brings no detriment to the achievement of social impact. If this view is correct, rather than focus on branding strategy, further research might best concentrate on how different variants of social enterprise differ in the extent to which they actually deliver social impact.

Developing a comprehensive evidence base on the empirical difference in behaviour and outcomes between enterprises that merely 'trade for a social purpose' and others that substantiate this claim by imposing limitations on surplus distribution is therefore still very much open to further research. Nevertheless, the evidence presented in this paper strengthens the view that 'a business that works for the social good' is likely to be too wide a characterisation of social enterprise, capturing a large share of firms engaging in conventional (profit-maximising) behaviour.

Paper 2

Non-profit status and service quality: evidence from unemployment support services in the UK

2.1 Motivation

In a seminal contribution to the theory of non-profit organisations, Hansmann (1980) argues that the absence of the profit motive implies that non-profits will deliver higher standards of quality in markets where this quality is hard to observe. In such situations of asymmetric information, for-profit firms have an incentive to reduce levels of hard-to-verify quality, while the non-distribution constraint in non-profit firms ensures that these motives will not come into play. A rich theoretical literature builds and extends on this intuition (Bilodeau & Slivinski, 1998; Ben-Ner & Van Hoomissen, 1991; Glaeser & Shleifer, 2001)

As the degree of competition intensifies, the two firm types may be pushed to become increasingly similar (Brown & Slivinski, 2006). Specifically, under competitive pressures, firms will be pressed to provide similar levels of observable quality at a given price. However, for-profit firms will still have an incentive, compared to non-profits, to undersupply quality on product features that are difficult to observe or verify.

There is a rich empirical evidence on this topic in the case of health services and nursing homes. Earlier reviews highlighted the absence of evidence of differences in quality levels between for profit and non-profits (Malani et al., 2003). However, more recent work generally concludes that non-profits do outperform for-profit firms in the provision of quality levels. Schlesinger & Gray (2006) offers a review, and specific examples include Ben-Ner et al. (2012); Brickley et al. (2011); Chou (2002); Grabowski & Hirth (2003); O'Neill et al. (2003); Grabowski & Stevenson (2008). Schlesinger & Gray (2006) also review a body of work analysing this research question in the context of the childcare industry. While the welfare and employment services industry is, in many countries, also populated by both non-profit and for-profit firms, there does not appear to be previous research testing the above hypotheses in this industry. This paper attempts to fill this gap by studying the context of the UK Work Programme.

The Work Programme is the UK's major welfare-to-work programme, providing personalised support to welfare recipients who have been receiving benefits long-term or who are at risk of doing so. The Programme is being delivered by 18 'prime providers' working under 5-year contract to the UK Department for Work and Pensions (DWP), and who in turn can subcontract the provision to other providers in their supply chain. Importantly, both for-profit and non-profit providers are represented at all levels of delivery, both as prime providers and within their supply chains.

Providers are paid by DWP on the basis of observable and carefully monitored Job Outcomes. The payment structure is set to incentivise the sustainability of these employment outcomes. Notably, the contracts do not specify what support services should be supplied by the provider to achieve these outcomes; their choice is left to the experience and initiative of the provider. As payments are made exclusively for the achievement of the specific set of outcomes stipulated in contracts, all other possible outcomes do not affect the financial incentives of the provider. For example, the level of earnings achieved by individuals who find work or the quality of individuals' experience of the programme are not directly relevant to the providers' bottom-line. Therefore, if the above hypothesis on the provision of quality levels is correct, we should expect non-profits and for-profits to achieve similar outcome levels on contracted and fully observed measures, but that non-profit providers should offer higher levels of quality in relation to non-contracted and less closely monitored measures. In this paper, we focus on measures of customer satisfaction as examples of the latter.

The context of the Work Programme offers a valuable opportunity to address this paper's research question, as its institutional features allow for the identification of quality differences in a way that is robust to a number of relevant potential confounders. Importantly, individuals are randomly assigned to prime providers within each geographical Contract Package Area (CPA). This allows estimating quality differentials that are free from any biases that might otherwise arise from customer characteristics, local area circumstances and the interaction between the two. Secondly, we can estimate quality differentials while limiting the confounding effect of price competition. Indeed, most payment parameters are set by DWP in each CPA. While providers were encouraged, at the time of bidding for contracts, to offer year-by-year discounts on the headline Job Outcome payment at the contracting stage, these then remain fixed for the five-year duration of the contract. Additionally, all our results are robust to including contracted price levels in the estimation. The institutional set-up of the Work Programme therefore allows to identify causal measures of relative effectiveness across prime providers in the same CPA.

However, it is important to keep in mind that when we relate these differences in outcomes to the providers' profit status, the estimated associations may still be confounded by unobservable or omitted provider service characteristics. As such, we cannot identify the causal effect of profit status.

Our estimations indicate that there are no differences in quality outcomes, whether closely monitored and observed or not, between for-profit and non-profit primes. We also fail to find any statistically significant association between quality differentials and the share of customers dealt by non-profit providers across a prime's supply chain. In this latter case, however, the estimated associations between non-profit status and non-contracted outcomes are consistently positive and sizeable. As we discuss, the interpretation of this result is closely linked to whether one deems the analysis to be sufficiently powered. Overall, however, in contrast to predictions from the literature, we do not find strong evidence suggesting the sector of the provider is associated with differentials in product quality, be it observed or unobserved.

The paper is structured as follows. Section 2.2 provides a overview of the Work Programme. Section 2.3 describes the data sources used in the analysis, and how quality outcome measures and explanatory variables are constructed. Our estimation strategy is set out in Section 2.4. After presenting some preliminary analyses in Section 2.5, we outline our results in Section 2.6. Section 2.7 concludes.

2.2 The Work Programme

The Work Programme is the UK's major welfare-to-work programme. It was launched throughout Great Britain in June 2011, replacing much of the employment support previously on offer. It provides personalised support to welfare recipients who have been receiving benefits long-term or who are at risk of doing so.

Everyone receiving Jobseeker's Allowance (JSA)[1] or Employment Support Allowance (ESA)[2] access the Work Programme at a specified point in their claim, depending on their circumstances. Depending on their benefit history, some claimants must join the Work Programme while for others participation is voluntary.[3]

The Work Programme is delivered by organisations known as prime providers, or 'primes', working under a 5-year contract with the Department for Work and Pensions (DWP). For the purposes of the Programme, the country is divided into 18 geographical areas called Contract Package Areas (CPAs). The DWP tendered to contract two prime providers in 14 of these CPAs, and for three prime providers in the remaining 4 CPAs.

[1] Jobseeker's Allowance (JSA) is the national unemployment benefits

[2] Employment Support Allowance is the national benefit in support of individuals who are unable to work due to illness or disability.

[3] See DWP (2016) for a detailed overview.

In what follows, we refer to the 'contract level' as the operations of a given provider in a given CPA. There are 40 such contracts.

These contracts were awarded to 18 prime providers, with some providers winning contracts in more than one CPA. Of these 18 prime providers, 15 are for-profit, 2 are non-profit, and 1 is a mixed public-private-voluntary organisation. Primes can subcontract the provision of the Programme to other providers in their supply chain, but vary in the extent they do so. Non-profit providers are represented at all levels of Work Programme delivery, i.e. as prime providers and within the prime providers' supply chains.

Prime providers are contracted by DWP to hold the ultimate responsibility for the provision of employment support for the caseload assigned to them. Within each CPA, individuals referred to the Work Programme are assigned to primes on a random basis.

Following referral to a Work Programme provider, the provider is responsible for contacting that person to discuss the programme and begin planning the steps needed to support them into sustained employment. Once this activity has taken place, the provider registers an 'attachment' to the Work Programme. Individuals remain on the Work Programme for two years and can only be referred and attached once during this period.

Primes are paid according to the employment outcomes achieved by individuals assigned to them (see Table 2.1). Payments are primarily for getting participants into sustained employment; there are no payments for job entry. The amount and eligibility for payments depends on the Payment Group of each individual, reflecting differences in the anticipated difficulty in achieving positive outcomes.

At most one Job Outcome payment can be paid per individual. This is due after either 13 or 26 weeks of continuous or cumulative employment, depending on which Payment Group the individual belongs to. Following a Job Outcome payment, Sustainment Payments are paid for every subsequent four week period in continuous employment, for up to two years. Participants complete the Work Programme after 104 weeks or when the maximum number of Sustainment payments has been made.

Table 2.1: Work Programme parameters by Payment Group

Payment group:	Customer group	Weeks in employment to achieve Job Outcome:	Sustainment: maximum number of 4 weekly payments:	Attachment fee, Year 1 only (£)	Job Outcome fee (£, maximum)	Sustainment Payment (£)
1	JSA 18-24	26	13	400	1200	170
2	JSA 25+	26	13	400	1200	215
3	JSA early access	13	20	400	1200	250
4	JSA ex-IB	13	20	400	1200	250
5	ESA volunteers	13	20	400	1000	115
6	New ESA claimants	13	20	600	1200	235
7	ESA ex-IB	13	26	600	3500	370
8	IB/IS (England only)	13	13	400	1000	145
9	JSA prison leavers	26	20	400	1200	200

Table 2.1 also shows the maximum payment amounts for each outcome. At the time of bidding, prime providers were invited to offer discounts on the Job Outcome payments, leading to some degree of variation across providers in the agreed size of these payments.

2.3 Data and measures of outcomes and service delivery

2.3.1 Datasets

This section provides an overview of the data sources used in the analysis and the measures constructed from each. The analysis draws on the following data sources:

- Relevant variables from the administrative National Benefit Database (NBD), detailing the full benefit history for the population of individuals who participated in the Work Programme, followed forward until December 2014.
- Relevant variables from the HM Revenue and Customs (HMRC) administrative dataset of tax returns, detailing the employment spells for all Work Programme participants.
- A survey of Work Programme participants, which asks respondents about their experience of the Work Programme, carried out as part of the external evaluation of the Work Programme.
- Data on referrals from prime providers to subcontractors, provided by DWP.
- Data on discount on Job Outcome fees offered by providers, taken from the contracts.

Administrative data on individuals participating in the Work Programme were provided by DWP. These data detail the circumstances for each of the 1,655,097 individuals referred to the Work Programme between June 2011 and September 2014, and track their outcomes until December 2014. This includes CPA and provider information and dates of referral, attachment, and any payments made to the provider.

This dataset also includes information on the claimant's characteristics, such as the claimant type, Payment Group, gender, age, health and disabilities, qualification levels, ethnicity and previous occupation. Data are complete for most characteristics. Qualification levels are an exception to this, as these are either missing or not known for 56% of the sample. Around 6% of individuals in the sample do not report a previous occupation. This is predominantly among individuals on disability benefits. Around 5% of individuals do not report their ethnicity.[4]

[4]A small number of data observations contain possibly contradictory information. Specifically, 2,251

The Work Programme administrative data contained a unique and anonynimised identifier allowing us to merge this data with the National Benefit Dataset and HMRC data. Only four pairs of individuals with a duplicate identifier were found in the 1,655,097 Work Programme records, and one arbitrarily selected case from each pair was dropped. A further 367 records showed a unique identifier of 1 and were dropped.

The full benefit history for the population of individuals who participated in the Work Programme was merged in from the National Benefit Database. These datasets were matched with a very high success rate. Only 6,143 individuals (0.37% of the population) were not matched.[5] Drawing on the information in the National Benefit Database, we construct benefits histories over the 24 months before referral, and for the 24 months after referral (or to the latest available time period in the data).

We also matched Work Programme participants to administrative data on tax returns held by HM Revenue and Customs (HMRC). 1,497,976 Work Programme participants were successfully matched to the HMRC data, delivering a match rate of 90.5%. Individuals not matched either had no employment records or had employment spells only before their referral to the Work Programme. Their failure to be matched is therefore likely to be a correct representation of their circumstances. We use the HMRC dataset to construct employment status histories for the 24 months after referral (or to the latest available time period in the data).

It was necessary to carry out a number of data cleaning operations on the employment data. A large number of end dates were missing (indicated by an end date in December 2099). Around 82% of these spells were followed by spells with a later start date. We interpret this as implying that the end date is missing rather than the spell being ongoing. We set these to the end of the tax year during which the employment spell started. The remaining 18% of spells were not followed by any subsequent spell. Almost half of these had a start date of 2014, suggesting they are likely to be ongoing spells. Their end date was set to a date in the near future (specifically, the end the 2014/15 tax year). The remaining spells had earlier start dates and were assumed to have genuine missing end dates. These were set to the end of the relevant tax year. Other minor cleaning operations include dropping spells with dates in the far future (but not December 2099) or the distant past. These amendments inevitably make assumptions that may not be entirely correct. However, it is important to note that, as individuals are randomly assigned to providers, measurement error will be balanced across these, and will not bias inter-provider comparisons.

individuals are reported as being in the JSA 18-24 year old Payment Group, but are aged 25 or over; and 2,038 individuals aged 25 or over are reported as being in the JSA 18-24 Payment Group. These individuals represent 0.25% of the sample and were dropped.

[5]Possible reasons for this very slight failure to match to NBD data include: a time lag in logging ex-offenders onto the NBD; having been in prison long enough to not have any benefit record since 1999; lower engagement with Jobcentre Plus; and errors in the data.

The DWP provided data detailing the referrals made by each prime to each of their subcontractors on a monthly basis. There are some limitations to these data. For instance, there is no distinction by Payment Group, nor do we know how many participants the referrals relate to, since individuals may be referred on by primes more than once. Nevertheless, it is the best available information to characterise how primes tend to distribute work across their supply chain.

We also use responses to a survey of Work Programme participants, which asks respondents about their experience of the Work Programme, carried out as part of the external evaluation of the Work Programme. The survey interviewed around 2500 individuals typically around 3 to 6 months from referral to the Programme. From the survey of participants, it was possible to characterise different aspects of how individuals experienced the Work Programme. The responses from participants do not distinguish between prime providers and subcontractors. However, since each surveyed individual's prime provider is known, the data can be used to construct contract-level summary measures of individuals' experiences. Note, therefore, that our estimates of the outcomes on non-contracted and less observed outcomes are calculated from a much smaller underlying sample than for measures of contracted and/or more observable outcomes. We would therefore expect the former measure to be more noisy, due to the underlying sampling variation.

2.3.2 Outcomes

To address the paper's research question, we use the above datasets to construct a number of outcome measures at the individual level. We further posit a hierarchy between these, ranging from the most salient and easily observed to the ones that are most subject to the limitations of asymmetric information.

We argue that the outcomes on which providers receive payments for will be the most observable and carefully monitored ones. These are whether there has been a Job Outcome payment within 12 months of referral and whether there has been a Sustainment Payment within 12 months of referral. Both are constructed from the Work Programme Administrative Database.

We place important and visible labour market outcomes that are, however, not directly contracted upon as part of the Work Programme slightly lower down the scale of observability. Falling within this group are whether an individual is off benefit at 12 months after referral (source: NBD) and whether they are employed 12 months after referral (source: HMRC)

Finally, the we argue that customer satisfaction outcomes are the ones that are most subject to asymmetric information and least relevant contractually, and we place these at the lowest level in the scale of observability. From the Work Programme Customer Survey, we can construct the following measures: the percentage saying support was better than that of the state provider of support for the unemployed (Job Centre Plus); the percentage saying support was useful for moving closer to a job; the percentage saying support was matched to need and circumstances; and the percentage saying the level of contact was about right.

2.3.3 Characteristics of service delivery

In line with the focus of this paper, the characteristics of service delivery we are most interested in relate to the for-profit vs. non-profit nature of the provider and their supply chain. The Work Programme is being delivered by 15 private, 2 non-profit, and 1 mixed public-private-voluntary organisation working under contract to DWP. The sector of each prime was identified by looking up company information in the official company register at Companies House. The 2 non-profit organisations are working in 3 of the 18 CPAs, while the mixed organisation works in a further, distinct CPA. We therefore construct a dummy variable indicating the primes where the 'Provider is not 100% private sector', which is the case in 8 out of 40 contracts.[6]

We further draw on information provided by DWP on the volume of referrals to each subcontractor in each primes' supply chain to construct the percentage of referrals dealt by providers that are not 100% private. This measure adds the cases self-delivered by the prime if this is a non-profit to the number of cases subcontracted to non-profit providers. As with the primes, this is possible as the sector of the subcontractor can be identified from the database at Companies House. While this data is provided on a monthly basis, we construct the measures at an annual basis.

Finally we construct a measure of price from details on the financial discounts offered by prime providers at the time of bidding to DWP. These discounts could vary by year and by Payment Group. The consequence of the discounts was that providers differed in the amount of revenue each referral could potentially attract. We assess the relevance of discounts by considering a variable for the total potential revenue for each individual. This is calculated as the Job Outcome fee (where providers can offer discounts) plus the Sustainment Payment multiplied by the maximum number of Sustainment Payments possible (which is fixed by Payment Group).

2.4 Research design

The aim of the paper is to identify quality differentials between for-profit and non-profit providers in the absence of confounding factors relating to customer and local characteristics and while controlling for differences in prices. In this section, we present our

[6]As we show in Section 2.6, all results are robust to the omission of the 1 mixed prime provider.

methodological approach, highlighting specific advantages and challenges present in the context of the Work Programme.

As a general framework, we can think of the outcome for individual i in CPA c referred to provider p, is determined by:

$$y_{icp} = c_c'\gamma + l_{cp}'\pi + x_{icp}'\beta + \phi_c + v_{cp} + \epsilon_{icp} \tag{2.1}$$

Outcomes will depend on the circumstances in each CPA, c_c. Examples of these could be the strength of the local labour market or the sectoral composition of the local economy. Individual characteristics, x_{icp}, such as labour market history or qualification levels, may also influence outcomes. Finally, the above equation allows for providers in each contract to affect outcomes, through their service characteristics (l_{cp}). Outcomes are also influenced by unobserved factors, which we can decompose into a CPA component (ϕ_c), a contract component (v_{cp}), and an individual component (ϵ_{icp}).

Our aim is to identify the association of specific elements of the service characteristics vector l_{cp}, namely the profit status of the provider or its supply chain, and quality outcomes. Importantly, we want to ensure this association is not driven by omitted characteristics that are associated both with the providers' profit status and the outcomes.

Individuals are randomly assigned to providers within each CPA by Job Centre Plus. On a daily basis and for each Payment Group, referrals are assigned to the provider which had the fewest referrals to date, relative to their contracted market share.[7] This institutional set-up allows to fully account for the possible effect of several potential confounding factors.

Firstly, random assignment of individuals allows experimental identification of relative provider effectiveness as within-CPA provider differences in outcomes. This can be estimated using a pooled model across all CPAs, and introducing CPA fixed effects to capture the influence of both observed (c_c) and unobserved (ϕ_c) factors which are common at CPA level. In other words, for example, differing customer composition across providers in different CPAs and/or prevailing conditions across CPAs cannot bias our estimates of relative provider effectiveness.

Secondly, as confirmed in Section 2.5, randomisation ensures the sets of individuals assigned to each provider exhibit balanced profiles across observable characteristics. Importantly, such balanced profiles also lend support to the (untestable) assertion that the groups are balanced over unobservable characteristics. Once CPA fixed effects are netted

[7] Other than in a limited number of clerical cases and mandatory ESA (IR) WRAG 12 month prognosis cases, opportunities for human intervention are minimal. Prior to 30 January 2012, the information systems only displayed allocations to providers by Claimant Group (rather than Payment Group) and at the Job Centre Plus district level (rather than the CPA). The same principle as above was then applied to the allocation to providers but was carried out by advisers and was calculated using data at these lower levels. The tests discussed in Section 2.5 confirmed that randomisation delivered balanced customer profiles across providers under both allocation mechanisms.

out, randomisation guarantees that the residual variance in x_{icp} and ϵ_{icp} is uncorrelated with provider service characteristics l_{cp}. Indeed, controlling for the effect of individual characteristics is not strictly necessary to estimate relative provider effectiveness, though doing so will tend to increase the precision of such estimates. Within-CPA comparisons across providers will therefore also not be biased by compositional differences in customers across these providers. Additionally, they will not be biased should the effect of individual characteristics vary by CPA.

The above framework allows the causal identification of the overall provider effect on outcomes. However, identifying the causal effect of a given service characteristic on outcomes is more difficult. On the one hand, by comparing within-CPA differences in outcomes across providers, this setup is robust to endogenous correlations of service provision in response to CPA circumstances or the profile of individuals in each CPA, or indeed the interaction between the two. For example, we need not be concerned if basic Maths and English training is more likely to be administered in CPAs with strong service sectors. However, Equation 2.1 highlights the fundamental limitation to identifying causal effects of service characteristics. Essentially, this would require having a correctly specified model in which unobserved provider service characteristics (v_{cp}) are not correlated with the observed service characteristics that we include in our estimations (l_{cp}). This is the challenge of any observational study.

In the specific context under study, omission of relevant service characteristics from the model may bias the coefficient on variables included in the specification. This is an important caveat worth keeping in mind given the limitations in the available data on service provision, as well as the need to keep the model parsimonious.

Aside from the issue of bias, the estimation of correct standard errors raises an additional challenge in this context. This consists in the fact that despite drawing on a very large number of individual observations, the data is clustered in a small number of groups. Specifically, the variation we are most interested in exploiting only occurs across the 40 provider contracts across the 18 CPAs.

When the number of clusters is small, conventional approaches to the calculation of standard errors are inadequate. With a larger number of contracts, commonly used clustered standard errors would have been sufficient to address the issue (Liang & Zeger, 1986) However, as these standard error calculations rely on asymptotic results, they are known to be biased when the number of clusters is small (Bertrand et al., 2004; Cameron et al., 2008; Cameron & Miller, 2013).

A number of alternative approaches to correct the bias in the test size exist (Bertrand et al., 2004; Cameron et al., 2008; Donald & Lang, 2007; Bester et al., 2011). For example, Brewer et al. (2013) show that applying bias-reducing corrections to the formula for clustered standard errors (Bell & McCaffrey, 2003) combined with using the t_{G-1} distribution with G equal to the number of groups rather than a standard $N(0, 1)$ distribution as the reference distribution for statistical tests (Bester et al., 2011) provides unbiased rejection probabilities. However, they also show this is no longer the case when there is large imbalance in the number of treated and control units, or, in as in our case, non-profits and for-profits. The wild cluster bootstrap-t procedure (Cameron et al., 2008) delivers unbiased estimates of the standard errors, including when groups are unbalanced (Mackinnon & Webb, 2016). However, this solution is too computationally demanding in the context of our large dataset.

To ensure we estimate standard errors correctly, we adopt an approach which draws on the two-step estimation procedure proposed by Donald & Lang (2007). This strategy consists of a first stage were covariate-adjusted group level means in outcomes are estimated on the full individual-level dataset. These are then regressed onto group level variables. Donald & Lang (2007) show theoretically and with MonteCarlo simulations that the second stage estimate of the coefficients and rejection probabilities will be unbiased even when the number of groups is small, so long as the number of observations per group is large. These are the conditions in the context of the Work Programme.[8]

It should be noted however, that the Work Programme context departs from the specific circumstance considered in the above stream of literature, including Donald & Lang (2007). That literature considers the case where units are clustered in groups. In our case, however, the data is structured at several different hierarchies. Apart from the trivial case of characteristics varying at the individual level, most provider characteristics (including profit status) will vary at the level of the 40 provider contracts. Importantly, however, we want to be able to control for price, which is defined at the contract-Payment Group-year level. Similarly, the we construct the share of referrals going to non-profit subcontractors in each supply chain at the level of the contract and year. Finally, our Work Programme and labour market outcomes are estimated from the large population of individual data and can hence be aggregated to any level, however those estimated from the more limited survey data can only be aggregated at the contract level. The implication of this more detailed hierarchical structure is that, while we replicate the Donald & Lang (2007) approach to the extent possible, the existing literature does not provide precise recommendations for our specific case.

Our specific estimation strategy is as follows. Drawing on data from these different hierarchies, the general equation we wish to estimate is the following:

$$y_{icpst} = a_{cst} + l'_{cpst}\pi + a_{cst} \cdot x'_{icpst}\beta + \phi_c + v_{cp} + \epsilon_{icpst} \tag{2.2}$$

where the quality outcome y_{icpst} for individual i of Payment Group s in CPA c assigned to contract p in year t is regressed on a fixed-effect absorbing outcome levels across the CPA(c)-Payment Group(s)-year(t) cells, a_{cst}. Individual characteristics, x_{icpst}, are also

[8] The Donald & Lang (2007) two-step estimator is also less computationally demanding, as only the first stage is estimated on the full dataset.

allowed to influence outcomes, and their effect is allowed to vary across CPA-Payment Group-year groups through an interaction with all cell dummies, a_{cst}.[9]

Following Donald & Lang (2007), Equation 2.2 is estimated in two stages. The first stage regresses, one CPA-Payment Group-year group at a time, individual outcomes on individual characteristics and dummies for the contracts with each group cell. Together, this set of first stage regressions identifies the parameters of the following equation:

$$y_{icp} = \delta_{cpst} + a_{cst} \cdot x'_{icpst}\beta + \epsilon_{icpst} \tag{2.3}$$

From this set of equations, we extract the mean and standard error of the 40 provider contract dummies (δ_{cpst}) for each Payment Group-year cell.[10] The coefficient on each dummy is the mean quality outcome for that contract, adjusted for individual characteristics. Because randomisation was carried out daily within CPAs and Payment Groups, comparisons across these outcomes at provider k and provider j within-CPA can be interpreted causally.

The data extracted from the first stage allows estimating a second stage at the CPA-Payment Group-year level. This involves regressing these estimated covariate-adjusted group means on service level variables as follows:

$$\hat{\delta_{cpst}} = a_{cst} + NP'_{cpst}\pi + p'_{cpst}\psi + \phi_c + v_{cp} \tag{2.4}$$

In this second stage, we decompose the vector of overall service characteristics, l_{cpst} in Equation 2.2 into $NP_{cpst} + p_{cpst}$. NP_{cpst} is our service characteristic of interest, representing either non-profit status of the prime provider or the share of referrals to non-profit subcontractors. The former will vary at the contract level, while the latter will vary at the contract-year level. p_{cpst} is the price, defined at contract-Payment Group-year level and introduced as a control. Any omitted but confounding service characteristic in l_{cpst} not accounted by $NP_{cpst} + p_{cpst}$ will enter the error term v_{cp} and, as discussed, is the main threat to causal identification.

A full set of CPA-Payment Group-year dummies (a_{cst}) is included to absorb any group level effects and errors. Since the ϕ_c error components are CPA wide, we cluster the second stage standard errors at CPA level. This is a cautious approach, in line with best practice of clustering at the highest available hierarchy (Cameron & Miller, 2013).

In Section 2.6, we estimate 3 specifications of Equation 2.4. Specification 1 estimates Equation 2.4 without the price variable as control, while Specification 2 includes it. Specification 3 includes price as a control and omits the one mixed public-private-voluntary prime provider as a robustness check. Observations are weighted by the inverse of the standard error of the estimate in the first stage (Angrist & Pischke, 2009).

[9]Monte Carlo simulations in Donald & Lang (2007) indicate that allowing the effect of individual characteristics to vary across groups allows for a cautious approach to inference.

[10]There are 9 Payment Groups and 4 years of data, implying a maximum of 1440 observations.

2.5 Preliminary analysis

As a precondition to carrying out the estimation, this section presents evidence ascertaining that the randomisation process assigning customers to providers within each CPA delivered a balanced profile in observable characteristics across providers, and that there is a statistically significant difference in provider outcomes over and above that which one might expect by pure chance.

We conduct these statistical tests by estimating a series of regression equations. Formally, we can modify Equation 2.1 to combine the effects of all provider service characteristics, l_{cp}, and of unobserved provider service characteristics, v_{cp}, into a single provider indicator, P_{icp}:

$$y_{icp} = a_c'\gamma + \delta_{cp}P_{icp} + x_{icp}'\beta + \phi_c + v_{cp} + \epsilon_{icp} \tag{2.5}$$

To avoid perfect multi-collinearity between the set of provider indicators in each CPA, we first randomly define one provider in each CPA to be the 'reference' provider. The other provider, or providers, in each CPA are each identified as 'treatment' providers. This assignment is purely arbitrary, does not affect the results and involves no loss of generality. The set of indicators, P_{icp}, is therefore only included for 'treatment' providers.[11]

Importantly, because individuals within a CPA are assigned at random to providers, P_{icp} is independent of ϵ_{icp} and so δ_{cp} has a causal interpretation as the mean effect on individuals in CPA c of being assigned to provider p rather than the reference provider. Each two-prime CPA provides one estimate of relative effectiveness. Each three-prime CPA provides two estimates of relative effectiveness. In total, this gives us 26 experimental estimates of relative effectiveness for each outcome, of which 22 are independent.[12]

When y_{icp} is a background characteristic, an F-test of joint significance of the δ_{cp} terms across all treatment providers and CPAs provides a test of whether that characteristic is balanced in the sense of being similar across providers within a CPA. Failure to reject the null hypothesis of no significant coefficients would indicate that background characteristics are statistically similar across providers in the same CPA, while a rejection of the null hypothesis would indicate differences in the customer profiles across providers in the same CPA. Analogously, with y_{icp} as an outcome, the null hypothesis is of no significant variation in effectiveness. In this case, rejection of the null hypothesis indicates that there is significant variation in within-CPA provider effectiveness, meaning that the question of which provider an individual is assigned to has an important bearing on their outcomes. When testing for variation in outcomes, the model allows for the effect of individual background characteristics, X_{icp}. These are omitted when testing for balance. This

[11] Following this operation, $a_c'\gamma + \phi_c$ captures both the CPA level effect and the effect of the reference provider. The two cannot be separately identified.

[12] In 3-provider areas, 3 pair-wise comparisons can be made. However, the third estimate can be derived from the other two: $A - C = (A - B) - (C - B)$. Therefore, only two are independent.

implies that our tests for variation in outcomes are more stringent.

We run the tests over a set of 45 customer characteristics.[13] This delivers a large number of estimates. We therefore do not present these here, but only discuss the overall patterns across the results.[14]

The first thing to note is that we find more evidence of imbalance over observable characteristics than we should expect purely by chance, if all tests were statistically independent. Given our lowest threshold of statistical significance is 10%, we might expect 4 or 5 of the F-tests on the 45 balancing variables to come up as statistically significant by chance. Instead, 14 of the tests emerge as statistically significant in the pooled sample. This suggests that there might be differences in observable customer characteristics across providers that, from a statistical point of view, depart from zero. However, it is also important to emphasise that the expected number of significant results is on the assumption of independence; it may be that the higher number found reflects a correlation between characteristics.

However, in socio-economic terms the magnitude of the coefficients is very small. For example, the typical difference in the share of lone parents in the whole sample of customers across any two providers is 0.003 or, equivalently, 0.3 percentage points. Therefore, while statistically significant, it is nevertheless materially very small. The largest typical difference estimated related to the share of customers with No Qualifications and is 0.6 percentage points.

The fundamental point at stake is whether these balancing test results should alter our view on the ability of our approach to provide causal estimates of the impact of providers on outcomes. This would be the case if we suspected that differences in outcomes across providers could be driven by these small differences in the profile of customers at different providers. Our view is that is most likely not the case as, while there is evidence of statistically significant imbalance with regard to (say) gender, the estimated impacts presented in the next section are sometimes 5 or even 10 times larger than the observed imbalances among background characteristics, suggesting these would have to be implausibly influential on outcomes in order to argue that the impact is not attributable to the provider. We therefore conclude that there is evidence of sufficient balancing to proceed with our proposed approach.

On the other hand, the F-tests for provider variation in outcomes generally reject the hypothesis that the provider dummies are jointly irrelevant to the outcome levels achieved. Furthermore, the mean absolute deviation between pair-wise provider comparisons is sev-

[13]These are: 6 age bands; gender; disability; being an ex-offender or with a history of alcohol or substance abuse; being a lone parent; being on benefit 12 months before referral; 7 indicators of mental or physical health conditions; 5 Work Programme Customer Groups; 6 ethnic groups; 7 indicators for educational qualification levels; and 9 indicators of previous occupational classification. We do this for the whole sample as well as by high level-Payment Group aggregates (JSA 18-24, JSA 25+, Other JSA, New ESA, Other ESA/IB).

[14]Subject to approval from the DWP, these can be made available by the authors upon request.

eral times larger than was typically the case for background characteristics. The mean absolute values of the differences in the outcomes across providers are found to be in the order of 1-2 percentage points. These are sizeable if one considers, for example, that the grand mean incidence of Job Outcome within 12 months of referral is 11 percentage points. Providers therefore appear to have a meaningful and measurable effect on outcomes.

To overcome the difficulty interpreting such a large number of estimates, we make use of visualisations, which can perhaps give a more immediate impression of what the data are telling us. Specifically, in Figure 2.1 we display the 26 point estimates and 95% confidence intervals for a selection of characteristics (top) and outcomes (bottom). Importantly, all charts use the same x-axis scale, allowing an immediate comparison of the relative magnitudes.

Such visual exploration reveals two important messages. Firstly, despite the statical significance of some of the coefficients on pair-wise comparisons across background characteristics, their magnitude almost always essentially zero or negligible. On the other hand, the variation in provider effectiveness is much more pronounced, and in the majority of cases statistically significant. On this basis, we interpret the results as confirming that background characteristics are generally balanced and that variation in provider effectiveness is present, statistically significant and of material magnitude.

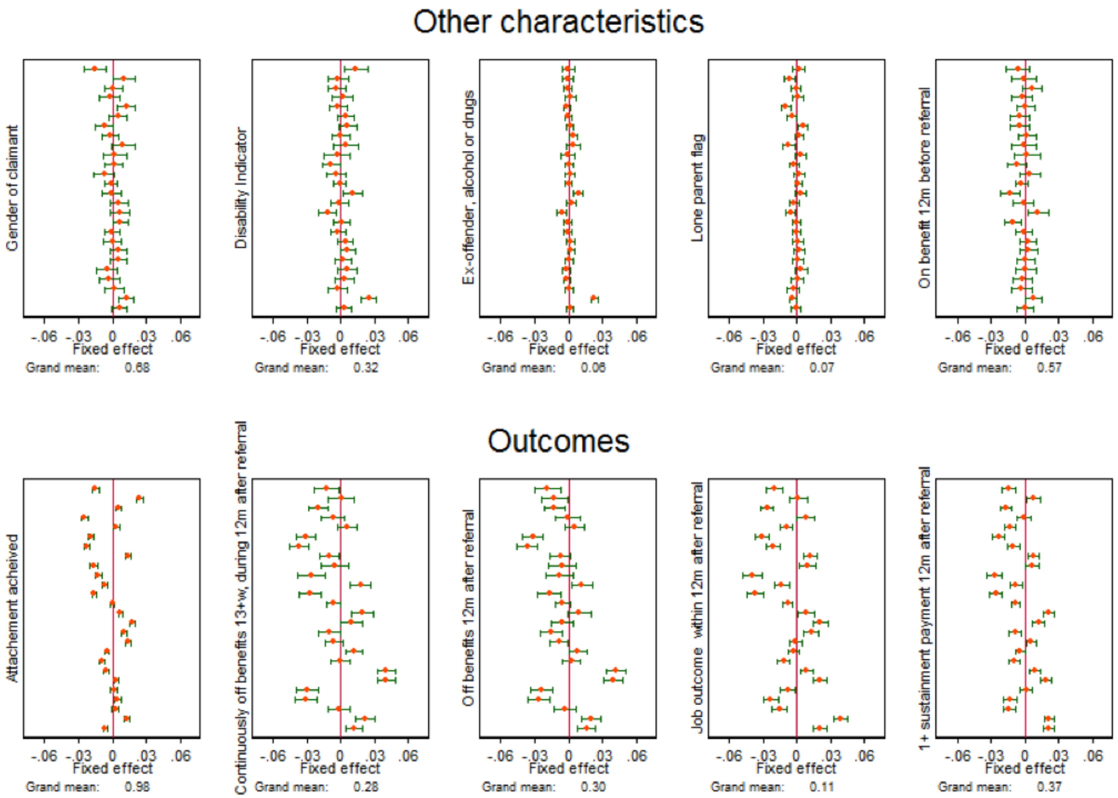


Figure 2.1: Visualisation of balancing tests (top row) and impacts (bottom row)

2.6 Results

The results of the regressions relating outcome measures to non-profit status are presented in the Table 2.2 below. Each cell reports the main coefficient of interest for the variable identified in the rows from a regression with the outcome identified in the column as the dependent variable. Standard errors and number of observations are reported below each coefficient. Outcomes across columns are ordered in broadly increasing degree of observability. The first row of the table reports the standard deviation of the residual variance in each outcome once CPA-Payment Group-year effects (a_{cst}) are absorbed. This parameter provides a benchmark against which to compare the magnitude of the coefficient estimates. The top panel of the table reports the results for estimations relating outcomes to the profit status of the prime. The bottom panel shows the results on the share of referrals handled by non-profit providers. Specification 1 does not control for price. This is done in Specification 2. Specification 3 controls for price and omits the one mixed public-private-voluntary prime provider.

Table 2.2: Ownership and service quality

Outcome Specification	Unobserved quality Useful	 Matched	 Better than JCP	 Contact is right	Observed quality Off benefits	 Employed	 Any Sustainment Payments	 Job Outcome
Std dev in residual variance	0.046	0.042	0.047	0.028	0.028	0.041	0.025	0.020
Provider not 100% private sector								
Spec 1	0.013	-0.007	-0.016	0.012	-0.000	-0.001	-0.003	-0.003
	(0.037)	(0.035)	(0.037)	(0.034)	(0.008)	(0.011)	(0.012)	(0.010)
	1394	1394	1394	1394	1013	1011	1002	998
Spec 2	0.014	-0.008	-0.016	0.012	-0.000	-0.001	-0.003	-0.003
	(0.037)	(0.034)	(0.038)	(0.033)	(0.007)	(0.010)	(0.010)	(0.009)
	1394	1394	1394	1394	1013	1011	1002	998
Spec 3	-0.003	-0.035	-0.020	0.031	0.001	-0.002	0.000	-0.001
	(0.010)	(0.031)	(0.057)	(0.041)	(0.010)	(0.015)	(0.012)	(0.010)
	1295	1295	1295	1295	939	937	928	924
% of referrals to non-profit providers								
Spec 1	0.085	0.057	0.040	0.027	0.001	0.003	0.005	0.005
	(0.060)	(0.062)	(0.060)	(0.036)	(0.012)	(0.015)	(0.016)	(0.013)
	1059	1059	1059	1059	1013	1011	1002	998
Spec 2	0.086	0.051	0.038	0.028	-0.003	0.001	0.000	0.002
	(0.061)	(0.062)	(0.062)	(0.036)	(0.010)	(0.014)	(0.014)	(0.012)
	1059	1059	1059	1059	1013	1011	1002	998
Spec 3	0.065	0.028	0.050	0.030	-0.001	0.001	0.003	0.004
	(0.059)	(0.074)	(0.086)	(0.043)	(0.014)	(0.019)	(0.014)	(0.014)
	984	984	984	984	939	937	928	924

*Standard errors in parentheses. Number of observations below each coefficient. * p<0.1, ** p<0.01, *** p<0.001.*

Overall, the results provide little evidence that the non-profit status of providers is systematically associated with negative or positive outcomes.

The profit status of the prime in particular does not appear to have any association with quality outcomes. The coefficients in relation to observable quality are not only statistically insignificant, but materially very small compared to the standard deviation of the variation in the outcome. Coefficients on the less observable quality outcomes are somewhat larger relative to the variation to be explained, but are all statistically insignificant. They also straddle above and below zero, therefore revealing no consistent pattern. Consistently across these, the take-away message is that we do not find any evidence that non-profit prime providers achieve contracted outcomes or customer satisfaction that is any different from that achieved by fully for-profit providers.

The percentage of cases handled by non-profit providers in the supply chain is also largely unrelated to all outcomes. Again, the coefficients on observable quality are all materially small, as well as statistically insignificant. The coefficients in relation to unobservable quality are also all statistically insignificant. They are however consistently all positive and of sizeable magnitude.[15] This is the closest we come to any evidence that non-profit provision may be associated with higher levels of unobservable quality.

All results are robust across specification. This means they are unaffected by including price levels as controls and the omission of the one mixed prime provider.

2.7 Conclusion

This paper provides novel evidence on the relationship between non-profit status and the provision of levels of service quality, in particular where these differ in their degree of observability. While several papers have found evidence of a higher provision of quality, particularly on features that are hard to observe, in the context of health, social care and child-care services, to our knowledge, this hypothesis has been seldom, if ever, studied in the context of unemployment support services. This is where this paper makes its contribution.

Overall, and contrary to existing evidence from other industries, our evidence broadly indicates that non-profit and for-profit providers do not differ in the levels of service quality offered. This is found to be the case both in relation to observed measures of quality, where the absence of a difference was expected, as well in relation to less observed measures of quality, where the literature would suggest we should expect non-profits to outperform. One interpretation of these results is that the two firm types do not ultimately differ in their operational objectives or mission, at least in the context of unemployment support

[15]The coefficient can be interpreted as a marginal effect of moving the supply chain from using 0% to 100% non-profit subcontractors. The coefficient size is typically around 1 standard deviation of the residual variation in outcomes.

services.

The results in this paper point to an aspect worth considering in further research as well as policy design. While we could rely on data for a very large number of individuals, when estimating variation across provider outcomes only a small number of effective observations were available. Drawing on recommendations from the literature, we have taken care to avoid misleading inference by adopting a 2-stage approach. Following this approach makes the structure of the estimation problem much more evident, and the constraint of small sample size readily apparent. As discussed in Brewer et al. (2013), however, the appropriate corrections to the calculations of standard errors give rise to an issue of statistical power. In other words, some correlations may go undetected. In other words, our approach to statistical inference has prioritised avoiding Type 1 errors, at the price a higher risk of Type 2 errors. It is important to highlight that balancing the risks of the two types of errors is a ultimately a subjective choice, where no hard and fast rules exists. This is particularly salient given that we do find consistently positive and sizeable estimates relating the share of referrals handles by non-profit subcontractors and less observable quality outcomes, but fail to detect a statistically significant effect. Further research would benefit from analysing institutional contexts allowing a larger effective sample size. In the case of the Work Programme, this could be easily facilitated by the policy-maker by increasing the number of primes within each CPA.[16]

[16]Two primes within a CPA give a single experimental estimate of relative effectiveness. Having five primes would increase it to 10.

Paper 3

Powering Education

3.1 Introduction

More than 1.3 billion people worldwide lack access to electricity and 40% of them live in Sub-Saharan Africa (IEA, 2013). This means that roughly a quarter of humanity lives without lights at home in the evening, without power at the workplace during the day, and without the possibility of reading and studying after dark. Energy poverty implies that most people are strongly constrained in their standards of living.

In Africa, the electrical power grid reaches only about 400 million of the continent's 1 billion people. In urban and semi-urban areas, over 30% of people have access to grid electricity. However, this figure drops to under 2% in rural areas. The electrical power grid is expanding slowly and unevenly. Governments and the private sector are working to reach deeper into remote areas, but financial, political and logistical barriers have proven to be significant obstacles to overcome. Households in off-grid areas rely heavily on kerosene lamps, on which they spend 20-25% of their income, despite the fact that the cost of equivalent useful lighting can be 150-times higher than of incandescent bulbs (IEA, 2014).

If we look at aggregate data we can see a strong correlation between electrification rate and completion of primary schooling (see Figure 3.1). Lack of access to light limits the possibility of studying after sunset and puts constraints on the time distribution of activities by students. In developing countries, it is not uncommon to find students of all ages gathering to read at night under the lights of a gas station or a shop (see Figure 3.2).[1] However, in rural areas, the lack of such basic infrastructure means that even this may not be possible. Electrifying rural areas in developing countries is a long and costly process. By the time this occurs, generations of students risk being affected by the lack of lighting, undermining the process of human capital accumulation in these countries.

[1] The first picture, which made the headlines of major newspapers, refers to Daniel Cabrera, a nine-year-old boy from the Philippines, who is studying under the lights of a McDonald. The second picture is taken in Guinea and has been reported by the New York Times and BBC.

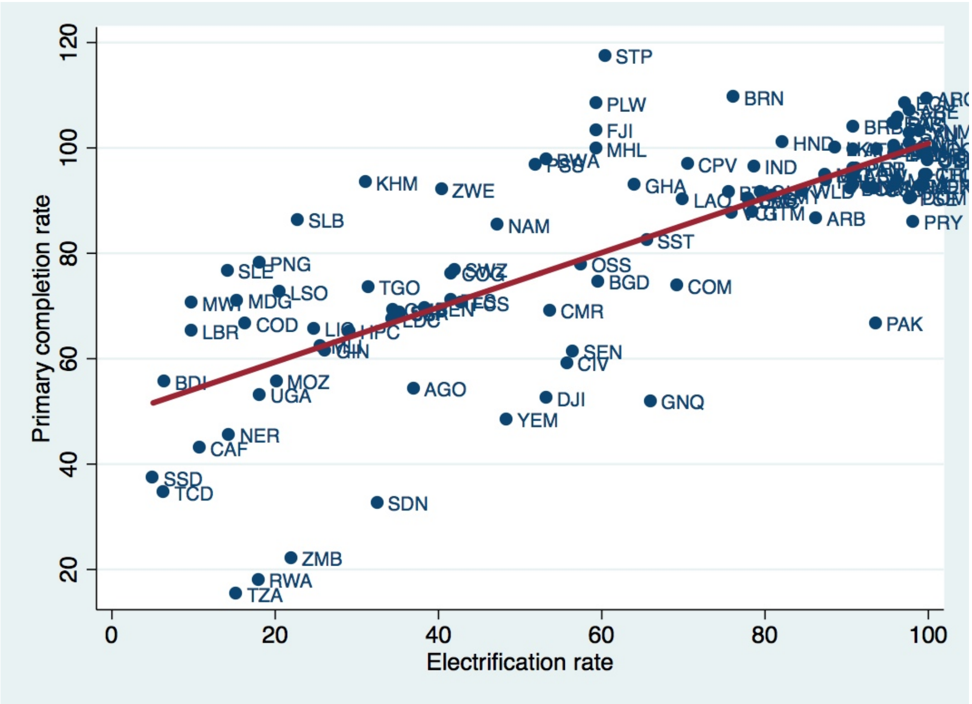


Figure 3.1: Electricity access and primary schooling (WDI data, electricity<100%)

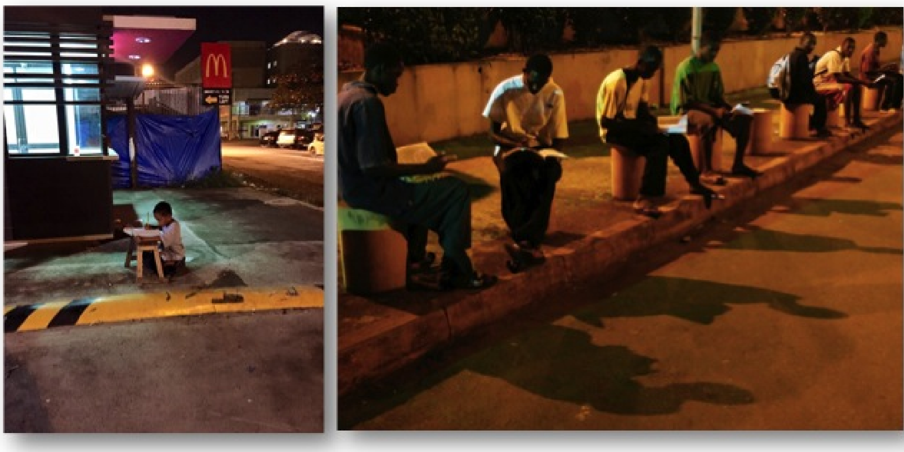


Figure 3.2: Students and lack of electrification

In this project, we evaluate the impact that solar lamps, which are a readily available source of lighting, can have on education. This is a novel experiment on a potentially key educational input for development. We distribute solar lamps to students in off-grid areas, randomising treatment at the pupil level. During the course of the project, however, both survey data and evidence from fieldwork revealed that a substantial degree of lamp sharing was occurring. The presence of possible spillovers between treated and control students invalidates the Stable Unit of Treatment Assumption on which our individual randomisation rested.

In order to account for these spillovers, we are able to exploit the natural variation in treatment intensity across classes that arose from name mismatch between school transcripts and survey at baseline. We argue that the variation in the match rate is random and we show that variation in treatment intensity is balanced across key observable characteristics of students, teachers, classes, and the local setting. Therefore, we use an identification strategy based on the econometric specifications of a randomised saturation design as proposed by Baird et al. (2014). We are able to find a positive and significant intention-to-treat effect such that treated students, in a class where 50% of the pupils get a lamp, experience an increase in math grades of 0.97 standard deviations.[2] Moreover, the lamp affects control students too, such that increasing treatment intensity by 10% increases their grades by 0.22 standard deviations. Finally, we collect data on the geolocation households and exploit the variation of treatment intensity at the geographical level induced by the randomisation. We do not find robust evidence of geographical spillovers. We conclude that within school interaction constitutes the the main source of spillovers.

Our paper is related to the large literature on education in developing countries. Promoting human capital accumulation is one of the key steps in the development process. The literature shows the importance of building school infrastructure (Duflo, 2001; Burde & Linden, 2013); the relevance of providing free primary education (Lucas & Mbiti, 2012); the effect of subsidies to households and pupils on enrolment (Schultz, 2004; Angrist & Lavy, 2009; Ambler et al., 2015); and the impact of monetary incentives on teachers' performance (Muralidharan & Sundararaman, 2011). Soft inputs like information on schooling returns (Jensen, 2010) and involving parents in school management (Gertler et al., 2012) have also be found to have a positive impact on educational outcomes. Our paper is closer in spirit to the literature that analyses the role of complementary inputs to education, like deworming programs (Miguel & Kremer, 2004) and flip-charts (Glewwe, 2002). Given the lack of electrification in developing countries, our contribution is to investigate the importance of access to light as an additional input for education and to measure the spillover effects that access to light can have on students.

Our paper is also related to the literature on energy access and development, such as Dinkelman (2011), Rud (2012), and Lipscomb et al. (2013). These papers are con-

[2] We find no effects on English, Swahili, Science and Social Science.

cerned with the effects of energy access on employment, industrialisation, and human development and housing values. Our study complements this field by having a focus on education. However, these studies examine the impact of electrification, which is a large region-wide technology shock, whereas we evaluate the effect of providing solar lamps, which is a smaller and individual technology shock that relates to a more easily available and cheaper source of energy access.

Finally, our paper speaks to the broader literature on randomised control trials. Many experiments are likely to fail or have biased results because of the presence of spillovers. Our paper provides methodological guidance on how to use a randomised saturation approach, as described in Baird et al. (2014), in order to account for spillovers even if the experiment was not initially designed for that. This requires a variation of treatment intensity that is as good as random and being able to approximate the functional form between treatment intensity and the dependent variable.

The paper is structured as follows: Section 3.2 describes the experiment structure; Section 3.3 discusses lamp usage and attrition; Section 3.4 analyses the average treatment effect on education; Section 3.5 assesses the presence of spillover effects arising within the class and from geographical proximity; and Section 3.6 concludes.

3.2 Project structure and randomisation

The experiment involves about 300 7$^{\text{th}}$ grade students across 13 classes in the Loitokitok and Nzaui districts in Kenya, relatively close to the Tanzanian border and mount Kilimanjaro (see Figure 3.3). We focus on schools in off-grid rural areas where household electrification is below 2.6%.

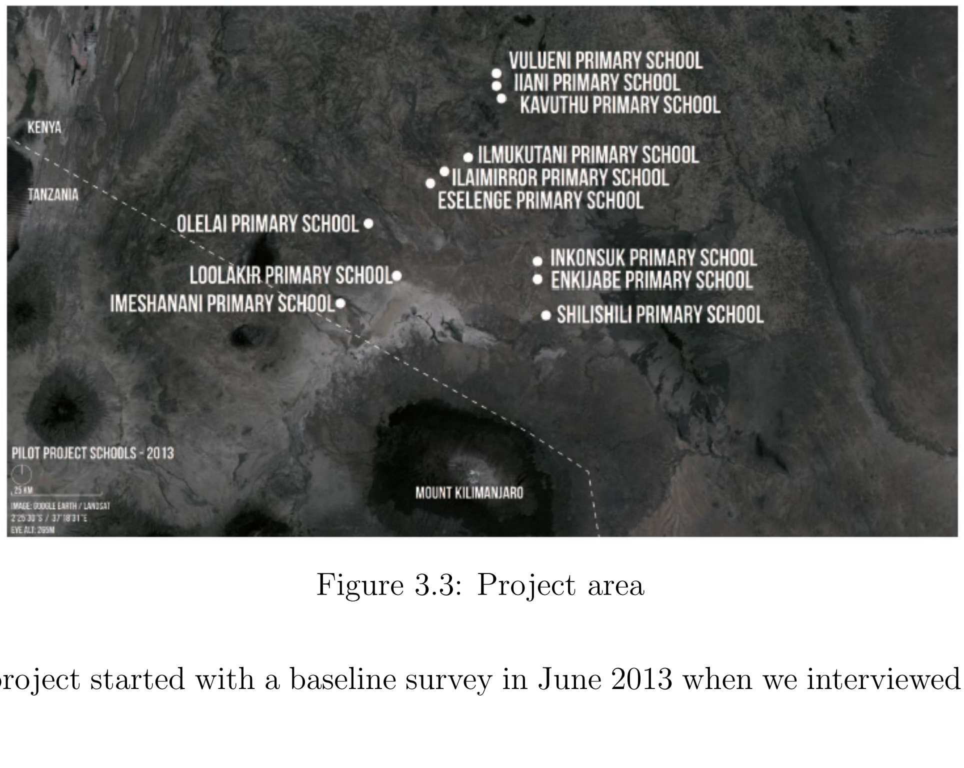


Figure 3.3: Project area

The project started with a baseline survey in June 2013 when we interviewed students and collected end of term grades from school transcripts. Lamps were distributed to the treatment group in September 2013, at the beginning of a new school term.[3] We then collected end of term grades for the treatment and control groups in November 2013, March 2014, and June 2014. We also run extensive face-to-face surveys of students in November 2013 and March 2014.

The baseline survey covered 341 pupils. We were able to match 286 of these with transcript data and they constitute our core sample, over which we conducted the randomisation. We distributed a solar lamp to 143 pupils and the remaining students were promised a lamp after one year, at the end of the experiment.[4] We randomise assignment to treatment at the pupil level so that within each class some students are in the treatment and some in the control group. We choose this level of randomisation to maximise statistical power, given the budget and the size of our sample. In our randomisation strategy, we seek balance between treatment and control groups on grades, which is our dependent variable, gender, classes, and a proxy for wealth.[5]

Given our sample size and the number of variables that we want to balance, we follow Bruhn & McKenzie (2009) and use a re-randomisation method where we select the allocation of lamps that minimises the statistical difference in means between control and treatment out of 10,000 draws (the so called *MinMax t-stat* method). We prefer this method to stratification, because the number of variables we would have been able to stratify on would have been severely constrained by our sample size. In this way, we avoid strong imbalance on several variables without forcing close balance on each. Moreover, we choose re-randomisation rather than pairwise matching, because attrition would have posed the risk of loosing too many observations, and potentially invalidating the experiment. We follow recommendations in Bruhn & McKenzie (2009) in the econometric analysis, and account for our randomisation method by including balancing variables in the regression and also run permutation tests to validate our inference.[6]

[3]New academic years start in January. This implies that our sample started in 7th grade and finished in 8th grade. This contributed to attrition, as some students in our sample did not graduate to 8th grade or changed schools. As we discuss in Section 3.3, attrition is unrelated to treatment status.

[4]Students in the control group received the lamp in September 2014.

[5]We construct a wealth index through a principal component analysis based on house characteristics (e.g. type of walls, water, and toilet facilities) and a set of goods owned (e.g. radio, telephone, bicycle, etc.).

[6]Bugni et al. (2015) shows that a in simple regression with strata or balancing variables fixed effects, the t-test on the treatment indicator has an unbiased probability of rejection. On that basis, our permutation test may actually be unnecessary.

Table 3.1: Balance between treatment and control on variables at baseline

Explanatory variable: treatment	**Initial randomisation**		**End of Experiment**	
Balanced variables	Coefficient	p-value	Coefficient	p-value
Mathematics	2.26	0.20	3.1	0.19
English	-0.59	0.45	-091	0.60
Kiswahili	-0.29	0.81	-0.91	0.55
Science	0.51	0.78	-0.20	0.93
Social Studies	-1.44	0.29	-1.75	0.33
Gender	0.00	1.00	0.00	0.99
Wealth index	0.02	0.77	0.02	0.82
School 1	-0.06	0.23	-0.03	0.64
School 2	0.01	0.47	0.03	0.26
School 3	0.04	0.28	0.03	0.58
School 4	0.05*	0.09	0.08*	0.08
School 5	0.02	0.70	-0.04	0.37
School 6	-0.03	0.42	0.00	0.97
School 7	-0.03	0.12	-0.04	0.17
School 8	0.02	0.49	0.02	0.61
School 9	0.00	0.99	-0.02	0.49
School 10	0.02	0.57	0.02	0.68
School 11	-0.02	0.40	-0.04	0.29
School 12	-0.04	0.22	-0.02	0.49
Additional variables				
Hours of study	0.11	0.79	0.16	0.76
Missed days of schools (previous month)	0.06	0.78	0.5*	0.07
Source of studying light: wood/candle	0.00	0.97	0.05	0.55
Source of studying light: kerosene	-0.04	0.69	-0.05	0.60
Mother's education	0.04	0.44	0.06	0.38

***. **, *, significant at the 1%level, 5% level, 10% level

Table 3.1 reports regressions of the baseline values of the balancing variables on treatment over the sample at the beginning and at the end of our project. The balance between treatment and control was well maintained throughout our study, reflecting attrition at random across terms. Moreover, we show that our sample is balanced also on other relevant variables for which we did not explicitly seek balance, like hours of studying at baseline, light source, school attendance, and mother's education.

3.3 Treatment compliance and attrition

We run two student surveys, 3 and 6 months after treatment. During these, we asked specific questions about lamp usage and appropriation. In terms of appropriation, 84% of treated students reported that the lamp stayed in their household when they were sleeping; the remaining 16% said that the lamp stayed at school at night. The lamp was resistant and broke in only three cases; in all the other cases the lamp was reported to be in good condition or with only minor problems. Moreover, in all cases students declared that the solar charge was sufficient for either all or most of the activities they wanted to carry out with the lamp. All these elements suggest that compliance was high, implying that intention-to-treat will be very close to the average treatment effect.[7] Additionally, 97% of students declared that studying was the main activity they used the lamp for.[8]

Beyond compliance with the experiment in terms of lamp appropriation and usage for studying, our experiment exhibits a degree of attrition. Grades are our main dependent variable of interest, but this is not always available for all students in our initial sample. This could be, for example, because they did not sit the end of term exam or left the school. Specifically, grades data is missing for 13% of our initial sample in Term 1. This increases to 23% in Term 2, and to 39% in Term 3. After the exam in Term 1, students were promoted from 7th to 8th grade. Unfortunately 16% of students in our sample did not pass the exam and had to repeat 7th grade. This explains a large share of attrition in Term 2 and 3, but not all of it. We regress a dummy indicator for those repeating 7th grade on treatment and find an insignificant value.[9] Moreover, in Table 3.2 we regress a dummy indicator for students with missing grades on treatment and find no statistically significant relation. Notice also that Table 3.1 shows that balance between treatment and control over balancing and additional baseline variables has been preserved across all terms among students sitting the exam. Therefore, we conclude that attrition is unrelated to treatment and that our results are unlikely to be affected by attrition bias.

3.4 Intention-to-treat effect

In this section, we run a series of reduced form regressions to identify the impact of treatment on educational outcomes. Given randomisation, the coefficients of the regressions can have a causal interpretation.

[7] We could not systematically check if students sold the lamp. In the first survey, we asked students to bring the lamp at the interview. About 55% of them did, but many declared that the lamp was installed at home in a way that was not easily removable. Indeed, during our field visit, we saw many cases in which the lamp was wired in the house and used as a proper lighting fixture. During the fieldwork, we visited households at random and with no notice; in all cases the lamp was in the house. In light of this, we believe lamp resale was minimal, if it happened at all.

[8] 40% of the students reported using the lamp to study all subjects equally, 25% to study mainly mathematics, and 20% mainly science.

[9] The coefficient is 0.04 with a p-value of 0.34.

Table 3.2: Attrition

Y: Grades data available	Term 1	Term 2	Term 3
Treatment	0.01	-0.03	-0.17
	(0.04)	(0.05)	(0.25)
Observations	286	286	286

***significant at the 1%level; ** significant at the 5% level; * significant at the 1% level.
Clustered standard errors at the school level in parentheses.

We start our analysis by running a single difference equation with an OLS estimation for each round of grades that followed our treatment. Our basic specification is the following:

$$y_{ij} = \beta_0 + \beta_1 \; Treatment_{ij} + \mathbf{Z_{ij}}\boldsymbol{\gamma} + \lambda_j + \epsilon_{ij} \tag{3.1}$$

where y_{ij} is the grade of student i in class j; λ_j captures class fixed effects; and $\mathbf{Z_{ij}}$ is a vector of controls that includes student's age, mother's education, and number of siblings. We also include the balancing variables used in the re-randomisation as controls.

We then extend our analysis to a lagged dependent variable specification. This allows controlling for past grades that, given the cumulative process of education and learning, might influence current grades. We use grades at baseline as the lagged dependent variable of reference. Therefore, we estimate the following regression:

$$y_{ijt} = \beta_0 + \beta_1 \; y_{ij0} + \beta_2 \; Treatment_{ij} + \mathbf{Z_{ij}}\boldsymbol{\gamma} + \lambda_j + \epsilon_{ijt} \tag{3.2}$$

We run this regression also with the controls specified for the cross section estimate.

Finally, we run a first difference estimation that allows us to control for individual fixed effects. Despite the randomisation, this specification offers an important robustness check. The first difference is taken with respect to grades at baseline, so all time-invariant variables between the two periods drop.[10] Therefore, we estimate:

$$\Delta y_{ijt} = \beta_0 + \beta_1 \; Treatment_{ij} + \epsilon_{ijt} \tag{3.3}$$

Table 3.3 summarises the main findings of these specifications. We report p-values for clustered standard errors in parentheses and those from permutation test in brackets. We are unable to detect any treatment effect in the short, medium, and longer run independently from the specification used and the controls that are added.[11]

[10]The controls used in the other specifications are all time-invariant, so they are not included in this case. Controls can be added to account for differential trends, but lack of intention-to-treat effect remains.

[11]We present the effect on the average grade across all subjects. We have also run these specifications on each subject separately, but results do not change.

Table 3.3: Intention-to-treat effect - Pooled regressions

	Cross section		Lagged dependent variable		First difference
Y: Grades	(1)	(2)	(1)	(2)	(1)
Treatment	0.047 (0.60) [0.3]	0.048 (0.54) [0.38]	-0.024 (0.77) [0.58]	-0.008 (0.91) [0.48]	-0.079 (0.47) [0.8]
Age		-0.057 (0.26) [0.65]		-0.061 (0.10)* [0.46]	
Mother's education		0.038 (0.88) [0.59]		-0.1 (0.27) [0.44]	
Number of siblings		0.025 (0.15) [0.55]		0.027 (0.18) [0.55]	
Grades at baseline			0.61 (0.00)*** [0.25]	0.63 (0.00)*** [0.48]	
Observations	646	582	639	575	641

***significant at the 1%level; ** significant at the 5% level; * significant at the 10% level. P-Values from clustered standard errors at the school level in parentheses () and p-values from permutation testing in brackets []. The dependent variable is the standardised grade in mathematics. All specifications account for class fixed effects and balancing variables.

3.5 Accounting for spillover effects

The lack of intention-to-treat effect may be due to the presence of spillovers that we need to account for. Given the partially non-rival nature of the lamp and the randomisation structure at pupil level, we cannot rule out spillovers. The consequence of spillovers is that the lamp affects both treated and control students. This can explain why we do not find evidence of treatment effects by directly comparing the performance of students in the two groups. However, this does not imply that the lamps have no effect on grades at all. It means that our estimation strategy should account for the presence of spillovers.

Spillovers can arise from interaction within the classroom or between treated and control students living close by. In principle, the spillover mechanisms can be i) lamp sharing increasing the quantity and/or quality of study time; ii) improved learning of treated students that then share their knowledge with control students; and iii) competition from control students who feel disadvantaged and increase their study effort. We are able to disentangle within-class and geographical spillovers, but can only provide suggestive evidence of the different mechanisms of spillovers.

3.5.1 Within-class spillovers

We identify within-class spillovers by exploiting variation in class treatment intensity. Variation in the treatment intensity between classes arose during the process of matching survey responses with school transcripts at baseline. Starting from the full sample of 341 students surveyed at baseline, a match with transcripts was achieved only for 286 students. The match rate differed across classes, leading to a variation of treatment intensity ranging between 14% and 62% (see Table 3.4). We argue that the variation in the match rate is random. As we discovered at a later stage, this was due to casual issues: like major misspellings of names in the survey; the use of Baptismal names in the survey and traditional names in the transcript or vice versa; and inverting name and surname in the transcripts. Our enumerators did not use the transcripts as a reference for names when interviewing students, but asked students their names directly. So, for example, a mismatch was created when the student "Wambua Kyalo", as reported in the transcript, used his baptismal name "Jonathan Kyalo" in the survey.

Table 3.5 shows that matched and unmatched students are not statistically different across key observable characteristics like hours of study, wealth, mother's education, source of light etc. Given the random nature of being matched or unmatched, we argue that the two groups are balanced also on unobservable characteristics, so that we have natural variation in the data. Table 3.6 reinforces this point, and shows that treatment intensity across classes is balanced over gender, teacher experience, wealth, and most grades. A student in a class whose treatment intensity is 10% higher than another class tends to have a statistically insignificant 1.9 extra points in mathematics. There is imbalance in grades for science and social studies. The content of the Kenyan Primary School syllabus generates little complementarity between these subjects and mathematics.[12] Furthermore, our results are robust to the inclusion of baseline science grades, baseline social studies grades, or indeed the average of all baseline grades as controls.

Econometric specification and results

We use the econometric specification of a randomised saturation design as proposed by Baird et al. (2014), where saturation is defined as the percentage of students treated within a class (treatment intensity). This methodology allows to identify different components of the experimental effect of treatment: spillovers on the control group, spillovers on the treated group, and treatment on the uniquely treated. This methodology involves a two-step randomisation process: treatment intensity is firstly randomised across clusters;

[12] Note that science covers topics like vegetation, how to create compost, human diseases and similar issues and not fields like physics or chemistry, which would have complementarities with math.

Table 3.4: Treatment intensity variation

	Treatment intensity	Class size
Class 1	14.20%	7
Class 2	33.3%	18
Class 3	33.3%	36
Class 4	37.0%	27
Class 5	38.4%	13
Class 6	40.0%	15
Class 7	42.8%	28
Class 8	47.5%	40
Class 9	48.0%	25
Class 10	51.4%	35
Class 11	55.5%	9
Class 12	57.8%	38
Class 13	61.9%	21

Table 3.5: Mean difference between matched and unmatched students (t-test)

Explanatory variable: treatment intensity	Coefficient	p-value
Hours of study	-0.35	0.48
Missed school days	-0.16	0.55
N. of people in the household	0.31	0.63
Wealth index	0.06	0.12
Source of study light: kerosene	0.02	0.79
Mother education	0.05	0.41

Table 3.6: Balance of treatment intensity

Explanatory variable: treatment intensity	Coefficient	p-value
Mathematics	19.01	0.41
English	13.04	0.64
Swahili	1.67	0.91
Science	55.61***	0.00
Social studies	-27.7*	0.07
Gender	0.06	0.88
Wealth index	-0.11	0.87
Teacher experience	-5.86	0.38
Teacher education	-0.17	0.81
Hours of study	-1.06	0.71
N. of people in the household	-1.86	0.67
Source of study light: kerosene	-0.12	0.79
Mother education	0.66	0.52

***significant at the 1%level; * significant at the 10% level.

then, individual treatment is randomised within clusters. As argued above, our first step is as good as random and the second step was randomised explicitly.

However, contrary to the original design of Baird et al. (2014) we do not have a pure control group. Therefore, we follow an identification strategy that addresses this limitation as in McIntosh et al. (2014). This involves estimating the pure control outcome by imposing a functional form assumption for the effect of treatment intensity on control students. Hence, our econometric model is:

$$y_{ijt} = \beta Treatment_{ij} + \mu(TI_j \ * \ \delta_t) + \gamma(TI_j \ * \ Treatment_{ij} \ * \ \delta_t) + \delta_t + s_{ij} + \epsilon_{ijt} \quad (3.4)$$

where TI_j captures treatment intensity in class j; δ_t is a time dummy for post-treatment period and s_{ij} are individual fixed effects.

Estimating this regression as a difference in difference model between a specific term date and grades at baseline is equivalent to estimating this simplified version in first difference:

$$\Delta y_{ijt} = \alpha + \beta \ Treatment_{ij} + \mu \ TI_j + \gamma(TI_j \ * \ Treatment_{ij}) + \epsilon_{ijt} \quad (3.5)$$

β is the treatment effect on the uniquely treated (TUT) and captures the theoretical intention-to-treat effect at the point of zero saturation. Defining π_j as the share of treated students in class j, $TUT = E(\Delta y_{ijt}|T_{ij} = 1, \pi_j = 0) - E(\Delta y_{ijt}|T_{ij} = 0, \pi_j = 0)$, where T_{ij} indicates if a student i in class j is treated or not. The coefficient μ is the saturation slope for the control group and captures spillovers on the control group: $SC(\pi) = E(\Delta y_{ijt}|T_{ij} = 0, \pi_j = \pi) - E(\Delta y_{ijt}|T_{ij} = 0, \pi_j = 0)$. γ is the differential of the saturation slope for the treated and measures the effect of changing saturation on the treated compared to control, so that $\mu + \gamma$ captures the spillover on treated, defined as $ST(\pi) = E(\Delta y_{ijt}|T_{ij} = 1, \pi_j = \pi) - E(\Delta y_{ijt}|T_{ij} = 1, \pi_j = 0)$. This methodology allows to compute an intention-to-treat measure as the sum of the treatment on uniquely treated and of spillovers on treated such that $ITT(\pi) = E(\Delta y_{ijt}|T_{ij} = 1, \pi_j = \pi) - E(\Delta y_{ijt}|T_{ij} = 0, \pi_j = 0)$.

The results of this regression are presented in Table 3.7. We account for the small number of clusters by i) calculating statistical significance relative to the small sample t-distribution with eleven degrees of freedom while clustering standard errors at the school level; and ii) re-calculating the p-values using randomisation inference (Rosenbaum, 2002).[13]

[13] This permutation exercise constructs the distribution of the parameter of interest under the null hypothesis of no effect had the randomisation procedure been carried out as a randomised saturation design. This is, of course, not the procedure we carried out. The substantive question, however, is whether we obtain a valid p-value when comparing our actual coefficient to this distribution. The literature does not provide us with an answer as no one has considered our very specific case. However, Bugni et al. (2015) provide reason to believe that our permuted distribution will most likely have an equal or even greater variance than the one that would emerge if one could replicate the more stringent conditions of

Table 3.7: Spillover effect estimates - Pooled sample

Y: Grades in Mathematics	
Treatment, β	0.37
(treatment effect at 0 saturation)	(0.25)
	[0.15]
Class Treatment Intensity, μ	2.21
(saturation slope in control)	(0.026)**
	[0.09]*
Treatment * Class Treat. Intensity, γ	-1.03
(differential saturation slope in treatment)	(0.18)
	[0.83]
Intention-to-treat at saturation=0.5	0.97
	(0.03)**
	[0.1]*
Marginal effect of 10% higher	0.11
treatment intensity on treated students	(0.18)
	[0.3]
Observations	641

***significant at the 1%level; ** significant at the 5% level; * significant at the 1% level. P-Values from clustered adjusted standard errors at the school level in parentheses () and p-values from randomisation inference in brackets[]. The dependent variable is the standardised grade in mathematics.

We can see that there is a positive and significant spillover effect on the control group. The estimates of μ are positive, significant, and large in magnitude such that a 10% increase in saturation raises math grades of the control group by 0.22 standard deviations. The effect on treated students is more difficult to read straight from the table. The treatment effect at zero saturation (β) and marginal effect of treatment intensity on the treated, ($\mu + \gamma$) are both positive but not statistically different from zero. However, the overall intention-to-treat is positive and statistically significant. Table 3.7 reports the intention-to-treat effect at an example saturation level of 50%.[14] The size of the coefficient is such that on average treated students improve their grades in mathematics by 0.97 standard deviations if they are in class where half of students receive a lamp.

To aid with interpretation, we provide an overall picture of the estimated effects in Figure 3.4. The Figure shows the estimated intention-to-treat and spillover effects on control students at different levels of class treatment intensity. Dashed lines represent 90% confidence bands. Figure 3.4 clearly shows as sizeable effects on grades, which are statistically significant across the entire range of treatment intensities in our project. It also makes clear how the overall effect is very similar between treated and control students, especially at higher treatment intensities, suggesting a large part of the this is working through the class rather than the individual level.

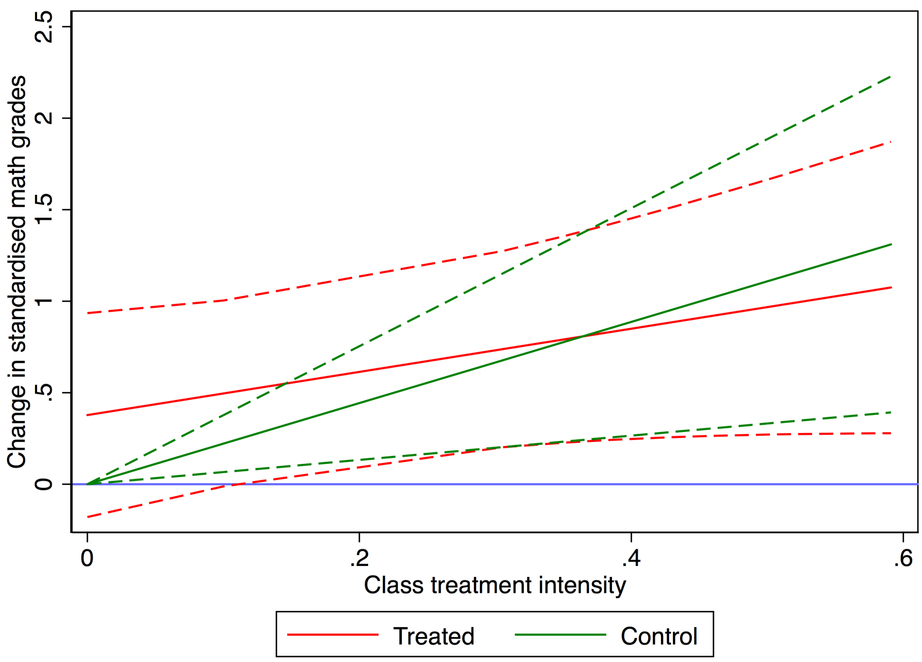


Figure 3.4: Estimated effects by class treatment intensity

As in McIntosh et al. (2014), we do not have a pure control group. Therefore, our ITT estimates rely on an out-of-sample prediction that hinges on the linear specification

our randomisation procedure. Our permuted p-values are, therefore. conservative.

[14] This is calculated as $\beta + 0.5 \times (\mu + \gamma)$

of the model. In Figure 3.5, we let the data speak for itself and use a local polynomial smoother to analyse the relationship between grade first difference of control students and treatment intensity. We find a positive relation, which is what we would expect in the presence of spillovers, and a linear functional form seems to be appropriate for the interval of our data. Moreover, adding a squared term on treatment intensity to Equation 3.5 delivers insignificant results. Our estimates rest on the assumption that the linear specification extends also between 0% and 14% saturation. However, given evidence of positive spillover effects and treatment effects, our ITT estimates are more likely to provide a lower bound of the true effect. In fact, one might expect a concave function of treatment intensity on grades. Providing a few lamps to a class at zero saturation is likely to have a stronger effect than adding lamps at a 90% saturation. If that's the case a linear specification provides a lower bound estimate by Jensen's inequality.

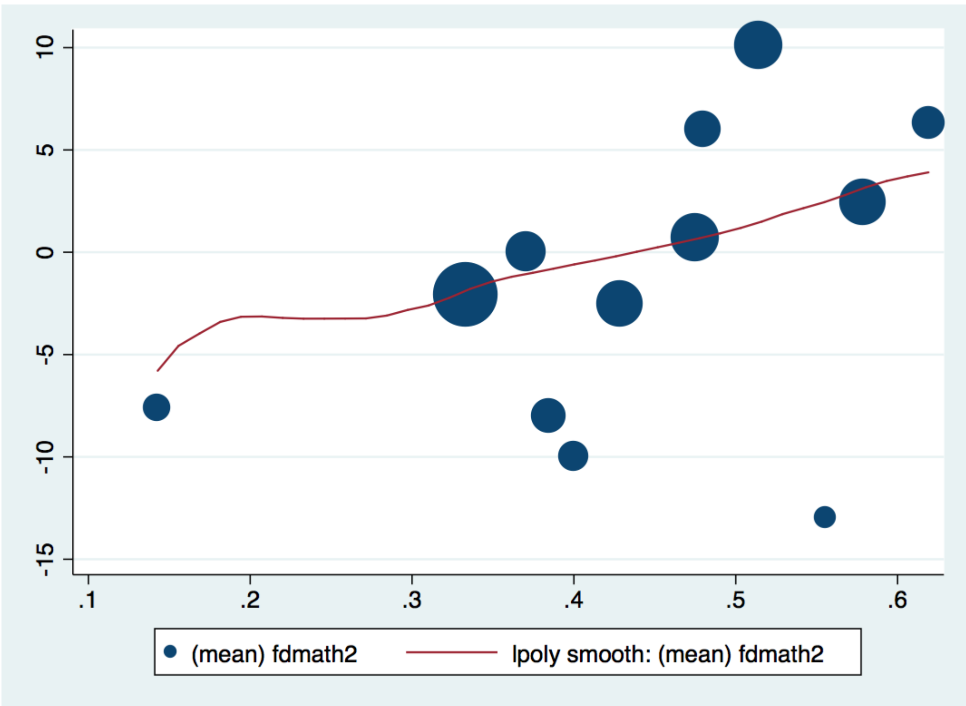


Figure 3.5: Local polynomial smoother of control groups' grades and class treatment intensity

3.5.2 Geographical spillovers

Externalities may not only take place within the classroom, but also at the homes of treated students. Students live in clusters of houses called *bomas*. There are no roads or illumination to connect bomas, so pupils are unlikely to move between them at night because of the danger of wild animals. However, students can interact around the house before and just after sunset or on their way to/from school. Therefore, Equation 3.5 needs to account for the fact that some of the spillovers that we attribute to class-level interaction, may actually be due to geographical proximity between treated and control students.

In order to do that, we the exploit exogenous variation in the geographical density of treatment across pupils generated by the experiment. We collected the geographical coordinates of the houses where students live and we use this information to construct a measure of the geographical treatment intensity. For each student, we compute the percentage of treated students within a radius of 500 meters, one kilometre, and 1.5 kilometres of their home. This is going to include both students in the same and in a different class, with the latter accounting for about 23% of the variation in the data. Finally, we rely on the following specification to identify the overall experimental effect accounting for both within-class and geographical externalities, thereby disentangling the two effects:

$$\begin{aligned}\Delta y_{ijt} = \alpha + \beta \ Treatment_{ij} + \mu \ TI_j + \gamma(TI_j \ * \ Treatment_{ij}) + \\ \sigma \ GTI_{ik} + \phi(GTI_{ik} \ * \ Treatment_{ij}) + \epsilon_{ijt}\end{aligned} \tag{3.6}$$

where GTI_{ik} is the geographical treatment intensity around student i within a radius $k = 0.5, 1, 1.5 \ km$.

Table 3.8 reports the results of Equation 3.6.[15] We can see that there is a positive spillover effect on control students arising from geographical proximity to treated pupils (coefficient σ). A 10% increase of geographical treatment intensity within 1km leads to an increase of grades of control students by 0.047 standard deviations, but the result is not robust to randomisation inference. Similarly, geographical spillovers on treated students are also not robust to permutation inference. Finally, results in Table 3.8 show that intention-to-treat effect does not change significantly once we account for geography, and spillovers on control students driven by class treatment intensity remain stable. Overall, we interpret these results as indicating that within-class, rather than geographical spillovers, account for the bulk of the spillovers identified.

3.5.3 Uncovering the causal chain

The preceding sections identify positive, significant and sizeable intention-to-treat effects, as well as spillovers on control students. In this section, we provide some suggestive evidence on how these effects may be materialising. In particular, the lamp can affect educational attainment by increasing the quantity and quality of study effort. To explore this hypothesis, we re-estimate Equation 3.5 with aggregate time-use data as the depen-

[15] We report only the results for geographical treatment intensity within 1km as it turns to be the relevant distance. Measures based on a distance of 0.5 or 1.5 kms do not deliver significant results. Details available upon request.

Table 3.8: Spillover effect with geographical estimates - Pooled sample

Y: Grades in Mathematics	
Treatment, β	0.52
(treatment effect at 0 saturation)	(0.12)
	[0.23]
Class Treatment Intensity, μ	1.86
(saturation slope in control)	(0.002)***
	[0.09]*
Treatment * Class Treat. Intensity, γ	-1.33
(differential saturation slope in treatment)	(0.089)*
	[0.81]
Geo Treatment Intensity, σ	0.47
	(0.019)**
	[0.23]
Treatment * Geo Treat. Intensity, ϕ	-0.05
	(0.83)
	[0.62]
Intention-to-treat at class saturation=0.5	1.00
& geo intensity = 0.5	(0.000)***
	[0.09]*
Marginal effect of 10% higher	0.05
class treatment intensity on treated students	(0.27)
	[0.30]
Marginal effect of 10% higher	0.04
geo treatment intensity on treated students	(0.015)**
	[0.33]
Observations	521

***significant at the 1%level; ** significant at the 5% level; * significant at the 1% level. In the first parentheses p-values from robust standard errors. In the second parentheses permutation testing. The dependent variable is the standardised grade in mathematics.

Table 3.9: Impact on aggregate time use

Outcome	ITT at 50% saturation	Saturation slope on control
Share of time at school	0.05	0.01
	(0.13)	(0.12)
Time at school (hrs)	0.34	0.02
	(0.38)	(0.49)
Time at school or studying (hrs)	-0.49	-0.16
	(0.68)	(0.82)
Time playing (hrs)	0.16	0.03
	(0.36)	(0.38)
Time sleeping (hrs)	0.71	0.18*
	(0.17)	(0.08)
Time studying (hrs)	-0.83	-0.18
	(0.88)	(0.89)
Time working at home (hrs)	-0.58	-0.11
	(0.83)	(0.83)
Time working away form home (hrs)	-0.03	0.02
	(0.59)	(0.37)

*Permutation inference p-values in parentheses. * p<0.1, ** p<0.01, ***p<0.001.*

dent variable, and calculate the intention-to-treat at 50% saturation and the marginal effect of saturation on control students. Results are presented in Table 3.9, with p-values estimated using a 1000-permutation inference. Overall, the lamp does not appear to be generating any changes in the aggregate allocation of time for either treated or control students. Specifically, it does not appear to be affecting the *quantity* of study effort, as we are unable to find any significant impacts on total time at school, total time studying, or the sum of the two.

Analysing the distribution of activities across the hours of the day provides some, albeit still tentative, insight into how the lamp may be effecting study time. Firstly, we re-estimate Equation 3.5 on the incidence of specific activities at specific hours over the course of the day. Results are visualised in Figure 3.6 and Figure 3.7. The lines show the intention-to-treat at 50% saturation and the marginal effect of saturation on control students on the incidence of each activity as each hour of the day. The shaded area captures the 90% confidence interval under the null hypothesis of no effect constructed from a 1000-permutation inference, such that the effect is statistically significant if the line falls outside the shaded area. The vertical line marks the time of sunset. In general, we fail to identify any clear effects of the lamp on the allocation of activities across hours of the day. However, we see some tentative indication that both treatment and control students are more likely to be in school in the hours immediately after sunset.

Additional evidence from student surveys supports the claim that lamp saturation may be bringing students to study together at school in the early evening. 48% of treated

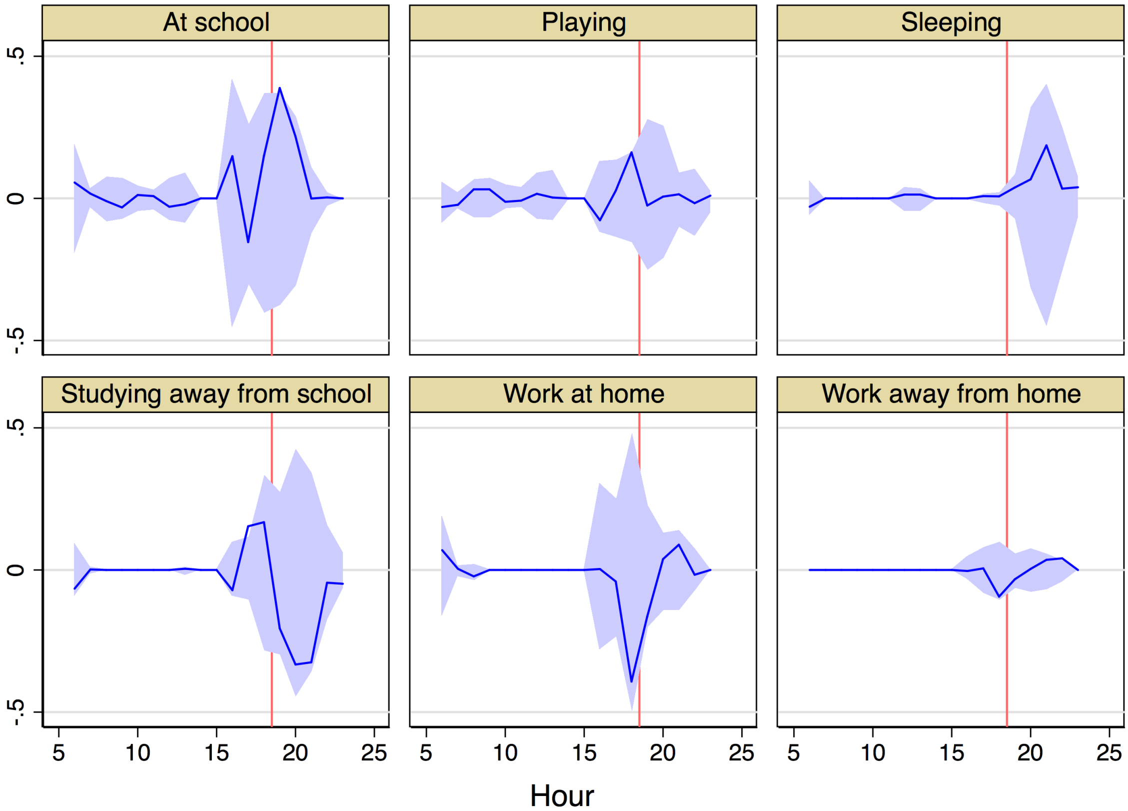


Figure 3.6: Intention-to-treat by time of day

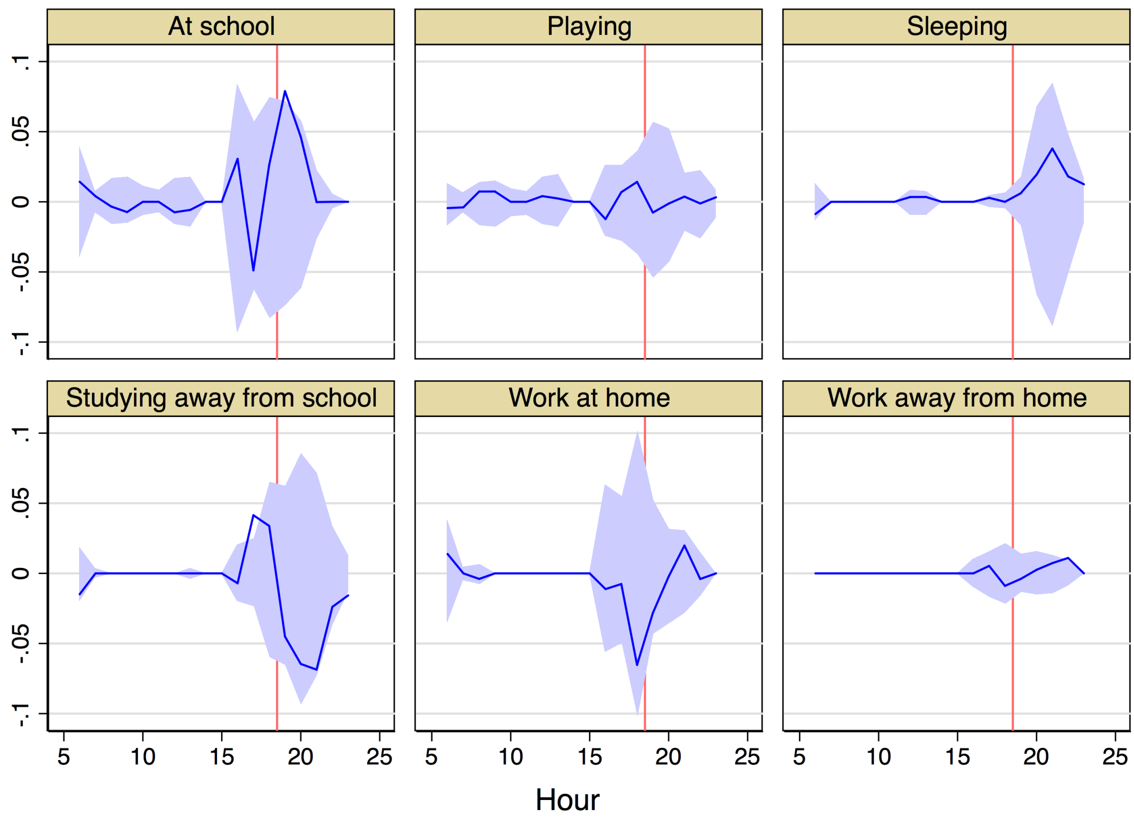


Figure 3.7: Saturation slope on control students by time of day

Table 3.10: Impact on study habits

Outcome	ITT at 50% saturation	Saturation slope on control
Preferred study location: School	0.26	0.06
	(0.14)	(0.13)
Preferred study time: Evening	-0.16	-0.04
	(0.72)	(0.81)
Studies with others	0.45*	0.10*
	(0.02)	(0.01)
Studies with others at home after sunset	0.13	0.03
	(0.23)	(0.20)
Studies with others at home before sunset	0.01	0.00
	(0.41)	(0.43)
Studies with others at school after sunset	0.31*	0.05
	(0.05)	(0.10)
Studies with others at school before sunset	0.04	0.03
	(0.45)	(0.26)

*Permutation inference p-values in parentheses. * p<0.1, ** p<0.01, ***p<0.001.*

respondents declared they shared the lamp with other people when studying; 60% of these shared the lamps with students of the same class and the remaining shared primarily with siblings. When studying with other students, about 90% of the pupils reported to do so at school. Additionally, running Equation 3.5 on responses to the survey's study habits module, we find a statistically significant 30 percentage point intention-to-treat at 50% saturation effect on the incidence of students indicating 'school after sunset' as their preferred study arrangement, and a marginally insignificant spillover effect control students (see Table 3.10). Importantly, we also find that treatment saturation is associated with an increased incidence of co-studying among both treated and control students. Possible reasons why the lamp allows students to spend more time at school in the evening could be that students (or their parents for them) can now carry out chores more efficiently at home at night, or because the lamps allow students to walk home more safely later in the day, when sunset and darkness are approaching.[16]

Therefore, while the lamp is not affecting the aggregate quantity of time dedicated to studying, we find evidence it may be influencing student study habits. Specifically, lamps appear to be triggering increased co-studying at school during the early evenings, thereby affecting control students as well. As total study time does not change, we hypothesise that our estimated impacts on grades are being generated by an increased *quality* of studying. However, we are unable to separate the extent to which this quality

[16]From fieldwork experience both alternatives are plausible; some students need to walk for as much as two hours to get to school.

improvement is determined by better lighting or by the benefits of studying together. Moreover, given that only 48% of treated students stated they shared the lamp with other students, lamp sharing is unlikely to account for all the spillovers. Sharing of knowledge due to interaction both at school and around home, as well as a competition effect are plausible possible candidates. Further investigation using network data on study partners could help identify the different sources of spillovers.

A possible alternative causal chain could be related to how the lamps affect the behaviour of household members, and how this in turn affects students' education. During the surveys, we asked students about lamp use at home. The family members who most frequently used the lamps are mothers (47% of cases), siblings (35%), and fathers (15%). Mothers used the lamp mainly for chores, siblings for studying, and fathers for other working activities at home. Time use analysis on parents shows that the lamp allows mothers to do chores more effectively at night, freeing time for other activities, especially paid work, during the day and in the evenings.[17] Additionally, we find evidence that the lamp helps generate savings on other lighting fuels, kerosene in particular. Evidence from student surveys and household expenditure surveys indicates that families with treated students experience a reduction of fuel expenditure of about 60-90 Ksh ($0.66-$1) per week. This is equivalent to around 10-15% of the median weekly income in our sample. Both these dynamics therefore suggest the lamp generates an income effect on households, which could potentially affect grades. We are unable to distinguish this income effect from the direct effect of the lamp itself, and this issue deserves further investigation. Das et al. (2013) show that increasing school inputs may affect household spending responses and, in turn, learning outcomes. We ran a household expenditure survey after one year at the end of our experiment and we did not find significant differences across expenditure categories between treatment and control group. This is not inconsistent with the results of Das et al. (2013) given that they find an effect on household expenditure in the second year and if the input was anticipated, which is not our case. Additionally, income effects would only be able to able to explain spillover effects on grades if the income effect itself spills over onto control households. These considerations strengthen our confidence that the income effect on grades is likely to be of second order compared to that of the lamp itself.

3.6 Conclusions

This study presents a novel experiment to assess the effect of access to light on education. Through a randomised control trial, we document an overall positive effect of solar lamps on education in rural Kenya. Once our identification strategy takes into account the potential presence of spillovers, we are able to find a positive and significant intention-

[17]Results available upon request. We build on these initial findings in Paper 4

to-treat effect and a positive and significant spillover effect on controls. Most of these spillovers arise from within-class interaction, while geographical proximity between treated and control students does not have a robust effect.

Given the small size of the technology shock that our experiment provides, all our estimates are likely to be a lower bound of the true effect of lighting, and more in general energy access, on education. Moreover, any experimental issues like lamp appropriation by teachers and lamp sharing with students in different classes, are likely to bias our estimates downwards.

The mechanisms through which spillovers arise deserve additional investigation. We have been able to disentangle within-class and geographical spillovers. Exploiting data on co-studying networks could help us verify whether spillovers arise primarily from the sharing of lamps or the sharing of knowledge.

Further research into this topic with larger samples and in different settings may help improve our understanding of the effects of access to light on education, and the mechanisms that can enhance or limit such links. Moreover, additional studies on household effects, especially on employment, would also provide a relevant and interesting line of research.

Paper 4

Entrepreneurship, gender and the constraints of time: a randomised experiment on the role of access to light

4.1 Introduction

In the absence of artificial light sources, families must rely on the limited hours of daylight to carry out their activities. This constraint affects the more than 1.3 billion people worldwide who lack access to electricity, 40% of whom live in Sub-Saharan Africa (IEA, 2013), predominantly in rural areas.

The constraint on productive hours is particularly relevant to women, who overwhelmingly carry the burden of having to fit housework into the time available to them (World Bank, 2012). Necessary but less productive activities typically relegated to women, such as cooking or cleaning the house, may crowd-out more remunerative uses of this limited time remaining. Moreover, an important fraction of daylight hours are often devoted to collecting firewood or sourcing kerosene to ensure some basic level of lighting during the hours of dark, thereby further reducing the time available for other uses.

This paper aims to provide evidence about whether access to light, which relaxes the time constraint in relation to the number of productive hours available, can stimulate the emergence of currently pent-up productive potential, particularly of women. There is evidence that grid electrification increases labour supply and employment, both at the local area level (Dinkelman, 2011; Rud, 2012) and at the household level (van de Walle et al., 2013), as well as household incomes (Khandker et al., 2013, 2012). There is considerably less evidence, however, on the extent to which a cheap and renewable source of energy used exclusively for *lightling*, like a solar lamp, allows to reap (some of) the above benefits of full scale electrification. To the extent this evidence exists, it is ultimately mixed and inconclusive. This is a policy relevant issue, as electrification is a long and costly process. Hence, assessing the effectiveness of ready-to-use small-scale off-grid solar technologies is where this paper makes its contribution.

Firstly, we dedicate considerable attention to time use (and indirectly labour supply and employment) effects. We do so despite an emerging set of recent evidence indicating small-scale solar energy products do not appear to influence time allocations. The RCTs in Grimm et al. (2014) and Aklin et al. (2015) find no change in hours worked.[1] Similarly, IDinsight (2015) finds no statistically significant change in the amount of time spent on productive activities (income generating activities, chores, and study). However, we believe there is a compelling case for studying the possible effects of light on time use in greater detail.

Specifically, as the studies above focus on aggregate time dedicated to activities, they cannot identify whether the are changes in the allocation of activities *within* a day. To the extent that the productivity or effectiveness of a task depends on the time of day it is carried out, 'task-shifting' may have economic consequences even in the absence of changes in total time use. Indeed, there is extensive anecdotal evidence suggesting access to light allows tasks to be moved from daytime to night-time and vice-versa. For example, albeit based on a relatively crude matched comparison, Harsdorff & Bamanyaki (2009) find evidence that women owning a solar home system were more likely to carry out domestic work in the evenings after sunset, and did so for more hours. Similarly, some exploratory proxy data on parental time use collected as part of the experiment in Hassan & Lucchino (2014) suggests lamps allowed mothers and fathers of treated students to substitute housework for paid work during the day. To address the limitation of previous studies, therefore, we adopt a diary approach to the collection of time use. Not only this is generally considered more accurate than the stylised aggregate approach (Kan, 2007; Bonke, 2005), but it also provides valuable information on the timing of activities across the day. We further complement the diary approach with data collected via real-time *experience* sampling implemented using interactive voice response calls, thereby eliminating recall bias. In line with the paper's focus on entrepreneurship, we pay particular attention on quantifying the effect access to light may have on the distribution of time between housework, farm work and entrepreneurial activity.

Building on this, this paper contributes to the very limited evidence on the role of light in promoting entrepreneurship and income diversification. We use the term entrepreneurship in the broadest sense of starting or expanding an economic activity, or generally making changes in how one earns their livelihood. Considerable anecdotal or observational evidence suggests this is a mechanism worth studying. For example, light

[1] Grimm et al. (2014) find a positive coefficient on hours of house work for spouses, but this not statistically significant.

is often identified as allowing micro-enterprises to extend their working hours (Harsdorff & Bamanyaki, 2009). Furthermore, qualitative evidence from the fieldwork for Hassan & Lucchino (2014) revealed that many mothers in treated households started to produce bags, baskets, and jewellery to be sold in local markets as the availability of light allowed them to engage in such informal entrepreneurial activity in the evenings. This points to the possibility that light my facilitate the emergence of new productive activities, thereby enabling diversification at the household and local level. To address this question, we place a considerable emphasis on collecting data on informal and micro non-farm productive activities. Additionally, considering the heavy predominance of agricultural activities in the project region, we also consider whether access to light is associated with changes in the types of farm and livestock activities carried out. To our knowledge, the only econometric evidence in relation to this broad topic are very recent emergent findings in Aklin et al. (2015), suggesting no association between solar micro-grids and the incidence of business ownership. This paper therefore aims to contribute to the currently very scarce evidence on the possible effects of access to light on small-scale entrepreneurship and economic diversification.

Importantly, access to solar-powered light may change household economic activities and circumstances without changing time use. An alternative causal mechanism runs via the effect of lamp ownership on fuel expenditure. Several studies identify a significant negative impact of access to solar-powered light on expenditure on alternative light sources,notably kerosene (see for example, Grimm et al., 2014; Hassan & Lucchino, 2014; IDinsight, 2015). These financial resources could be invested in improving or expanding the household's productive capital, thereby potentially affecting income and wealth. While this area does not constitute the main focus of the paper, we nevertheless include some references to this topic in our data collection to ensure we do not omit this potential mechanism.

To understand and quantify these dynamics, we run a randomised controlled trial (RCT) distributing solar lamps to households in rural Kenya. Our analysis exploits random variation in solar lamp ownership among the parents participating in the companion randomised controlled trail on the effects of access to light on education described in Hassan & Lucchino (2016). The latter randomly distributed solar lamps free of charge to a pool of 2,229 7th grade students across 60 schools in the Gucha South district in Kenya. The results in this paper are drawn from the 806 parents of these students surveyed via mobile phone, and across whom lamp allocation is indirectly randomly assigned.

Our findings are that access to light contributes to a diversification in household livelihoods from agricultural to non-farm economic activities. This evidence is supported by a consistent set of results across time use, the incidence of different productive activities, and incomes levels. To our knowledge, this constitutes the first robust evidence that small-scale lighting sources can help stimulate the very first steps in the direction of economic transformation. At the same time, the paper delivers some sobering evidence on the gender dimension of the effect of access to light. While we find evidence that access to light does indeed relax time constraints, and those of women in particular, we find that a large part of the benefits of this additional time ultimately flows to men.

The paper is structured as follows. Section 4.2 presents an overview of the experimental design and the project context. Section 4.3 details the data collection carried out, how this related to our priors of the relevant causal chain, and presents descriptive statistics of the data. Section 4.4 presents our identification strategy in detail, including a discussion of the nature of our sample and the outcome of the randomisation. Finally, Section 4.5 presents the results of the analysis and Section 4.6 concludes.

4.2 Project context and research design

This project, alongside its companion RCT in Hassan & Lucchino (2016), is concerned with the impact access to modern forms of lighting compared to traditional fuels (such as kerosene), or indeed no lighting at all. This requires identifying a target area exhibiting both low penetration of the electricity grid and limited presence of off-grid energy providers.

The project was implemented in partnership with Givewatts, a non-profit NGO providing clean energy to school children through schools and other institutions. Drawing on their local knowledge, we identified the Kisii County as a suitable candidate region for the project. Givewatts further agreed to not carry out their own operations in the target area for the duration of the project.

We use existing data to cross-validate this recommendation and further define our target area. The Kisii county is divided into 5 districts, and within 3 of these (Gucha, Gucha South and Masaba) more than 95% of the population is reported as lacking access to electricity in the Kenya Population and Housing Census 2009.[2] We complement this information with satellite night light data (see Lowe, 2014, for an overview of the data). Night light data offers a more up-to-date snapshot of energy access in the region as well as accurate measurement of light intensity for areas as small as 1 square kilometer. Including Stable Lights 3 from the latest satellite image available (2013), we identify Gucha South as the district with the lowest current levels of electricity assess. We select Gucha South as the target district for the project.

A striking feature of the project area is its geographical and socio-economic homogeneity. Dwellings tend to be constructed with the same similar materials and technique, and are broadly similar in size. These are typically built within the family plot of land, which is also invariably cultivated. The most common local amenity is the primary school. Families also source their basic goods from local 'shopping centres', which amount to little

[2]This is the latest official statistical source.

more than a handful of shops/stalls selling basic goods (e.g. vegetables, soap, kerosene) and services (e.g. mobile phone charging). Figure 4.1 shows a typical landscape in the project region and illustrates its homogeneity.

This paper bolts onto our companion experiment on the effect of solar lamps on the educational attainment of Grade 7 pupils in the Gucha South district. The latter adopts a randomised saturation design (Baird et al., 2014), whereby 2,229 students across 60 schools are assigned to one of three possible treatment statuses: all students in the class receiving lamp; half the students in the class receiving lamps; and no students receiving lamps. This paper exploits the exogeneity of treatment to establish the causal effect of access to light on the time use and economic activity responses of the mothers and fathers of participating students.

We argue this experimental design mimics individual randomisation across parents despite it resting on a randomised saturation design across classes and students. Indeed, the only departure from a direct individual randomisation consists in the fact that the randomised saturation design will introduce a degree of geographical clustering in the treatment. This introduces two methodological issues, which, however, we believe can be easily addressed.

Firstly, geographical clustering in treatment can bias estimates if influential and omitted factors share a similar spatial distribution. The socio-economic homogeneity of the region already goes some way towards addressing this aspect. More fundamentally though, we collect geolocation data on the homes of the project participants, allowing us to assess the robustness of results to the inclusion of geographical fixed effects (see Section 4.4.2 for more details).

Secondly, different treatment intensities across locations can be a problem to our identification strategy if we believe the individual intention-to-treat effect varies with the treatment intensity in the individual's neighbourhood. This could be the case if economic responses by a large number of parents triggers local general equilibrium effects. This is unlikely to be the case in our study, as our project participants are ultimately only a fraction of the economically active population in a locality.

Based on these considerations, as well as further evidence presented in Section 4.4.2, we argue this paper's research design allows for the causal identification of the impact of access to light on the economic livelihoods of the project families.

4.3 Data collection

The project involved a number of data collection operations. Some of these were designed primarily for the companion randomised controlled trail on the effects of access to light on education described in Hassan & Lucchino (2016) but nevertheless include data pertinent the the focus of this paper. Other surveys were designed specifically to be used to address this paper's research questions. This section starts by enumerating the type, outcome and timeline of the operations carried out. It then proceeds to detail the topics covered across the survey instruments.

4.3.1 Overview of the operations

Our broadest relevant sampling frame is the population of 2,229 7th grade students across the 60 schools in the Gucha South district in Kenya participating in our companion RCT. Treatment assignment was allocated using this sampling frame. Drawing on this frame, we carried out three data collection efforts relevant to this paper:

- A paper-based collection of mobile phone numbers targeting the parents of all 2,229 students. This operation identified 1,292 unique phone numbers relating to 1,375 students.
- A tablet-assisted face-to-face survey targeting a 55% random sample of 1,181 students selected from the 2,159 we had full baseline data for. The sampling was stratified by school, treatment assignment, gender and high and low baseline math grades. This operation was conducted in January 2016 and reached 876 students (a response rate of 74%).
- A geographical mapping of the homesteads targeting all 2,229 students, which was able to successfully identify the homes of 1775 (80%) of the students. The mapping was conducted in February 2016, recording residential locations at May 2015.

The population of parents who accepted to share their mobile phone number consists of our main sample of interest for the purposes of this paper. Drawing on the sample of parents who offered their contact details, we carried out two main data collection operations:

- A single wave of computer assisted telephone survey targeting all 1,292 unique mobile numbers, which was able to obtain full responses for 806 adults (a 62% response rate). This operation was conducted during November and December 2015.
- A repeated experiential sampling time use survey of the the 1,292 unique mobile numbers using Interactive Voice Response calls, conducted at random times of day over the 17 Tuesdays or Thursdays over the period between the 4$^{\text{th}}$ February 2016 and 31$^{\text{st}}$ March 2016 inclusive. The calls successfully got through to the respondent in 55% of cases, and 23% of these respondents completed at least part of the survey. A total of 2,817 person-time observations were collected. The average number of entries per person was 2.18

We also conducted a field visit in November 2015 aimed at gaining a qualitative understanding of the project context and refining our theory of change ahead of the preparation of the survey instruments and operations. This was also an occasion to run some cognitive testing of survey questions. During the fieldwork, we interviewed 13 families, 6 from the control group and 7 from among the treated. We interviewed 4 fathers, 7 mothers, 2 step-parents, and had an impromptu focus group with a group of mothers (some of whom where project participants) while they were selling vegetables on the side of the road. A topic guide was used, but not all topics were discussed in each interview.

4.3.2 Topics covered

We proceed to present an overview of the topics we collected data on, and descriptive statistics for the sample obtained. We collected a number of background characteristics, but we focus our attention here primarily on metrics concerning the outcomes of interest.

The core of the data analysed in this paper was collected through a mobile phone survey. This mode of data collection imposes some constraints to the scope of the survey, as the engagement and availability of the respondent is typically significantly lower than with conventional face-to-face approaches. The range of data collected should therefore be considered with this limitation in mind.

Gender and other background characteristics

This paper places an important emphasis on how the effects of access to light may differ by gender. As such, we took a number of steps to increase our reach to students' mothers. For example, when collecting mobile contact details, we specifically asked for the mother's own phone number, if she owned her own phone. However, in the majority of cases mobile phones were either shared among the adult members of the household, or primarily of the students' fathers.

To maximise our reach to mothers, we asked if mothers were available in the early stages of the questionnaire. Specifically, after having confirmed we had reached the parents or guardians of the student in question, we asked to speak to their mother if she was available. If so, we asked for the 'initial respondent' to hand the phone over to what we call the 'ultimate respondent'.[3] Women composed 38% of the initial respondents, and this figure rises to 44% of the 'ultimate' respondents. The sample of ultimate respondents consisted of 452 men and 348 women.

Despite the attempts to maximise the participation of women into our survey, they nevertheless accounted for less than half of the participants. To be able to use of data from the full sample, our first step in the analysis is therefore gender neutral. In subsequent

[3] As shown in Section 4.4.2, the change in respondents was equally likely across treatment and control groups

steps, we also allow for heterogenous effects across gender. However, this inevitably implies a reduction in the sample used to identify effects for a given gender.

Other background characteristics we collected are the age and highest level of education of the ultimate respondent, and the number of adults and children in the household.

Time use

In light of the focus of our research, we dedicate a considerable amount of space in our surveys to the documentation of time use. As discussed in Section 4.1, to improve on existing work on the effects of access to light on time use, we adopt a combination of diary based and experience sampling methods.

The time use module consisted of the largest section of our phone survey of parents. Respondents were asked: *Please tell me what you did yesterday, starting from when you woke up to when you went to sleep for the night, indicating start and end time of each activity and whether this was carried out at home or away from home.* Respondents were also asked to provide a proxy response for the same information for their spouse. All but 2 survey participants responded to the time use model, while 298 (37%) participants were able to provide proxy responses for their spouses. Enumerators were tasked to code the responses as open-ended. However, they were provided with a list of the most commonly occurring activities to be able to address the bulk of the task with ease. The list was drawn from most frequently activities in the 2005 Tanzania Time Use Survey Pilot (Rugaimukamu, 2005).

The specific activities reported were later aggregated into the following main activities: agriculture; livestock; non-farm work; house chores; family care; shopping and sourcing goods; personal care; social engagements; sleep and rest. A further category called 'in transit' covered time spent transferring between places and activities. Figure 4.8 and Figure 4.9 display word clouds of the most frequently occurring terms used by the survey respondents to describe the activities grouped under each heading. The clouds reveal a high degree of homogeneity within each high-level activity group. Activities were further grouped to create the broad time aggregates of: productive activities (agriculture; livestock; non-farm work); informal work (house chores; family care; shopping and sourcing goods) and leisure (personal care; social engagements; sleep and rest).

It is worth noting that, contrary to most other groupings, the activities falling within the categories 'non-farm work' and 'social engagements' are relatively heterogenous. For example, social engagements include both having tea with friends as well as economically relevant activities such as attending a meeting at a rotating saving group or helping building a local church. Similarly, 'non-farm work' includes any income-generating activity not directly related to the agricultural or livestock production. These can range from selling produce at the market to running a shop or being a teacher. Respectively, these may arguably be seen as incidences of social and economic diversification, and are therefore

Table 4.1: Time use descriptives statistics

Activity	Men		Women	
	As respondents	As spouses (proxy)	As respondents	As spouses (proxy)
Productive Activities	8.12	6.29	5.75	4.45
Agriculture	2.98	2.79	3.27	2.93
Livestock	1.00	0.41	0.72	0.28
Non-Farm Work	4.13	3.10	1.77	1.24
Informal Work	1.11	2.75	4.30	6.62
Family Care	0.25	0.40	0.30	0.44
House Chores	0.79	2.22	3.63	5.82
Shopping or Sourcing Goods	0.08	0.13	0.37	0.35
Leisure Time	14.21	14.28	13.34	12.15
Personal Care	2.23	2.09	2.14	1.84
Sleeping or Resting	10.54	11.19	10.32	9.58
Social Activities	1.44	1.01	0.87	0.72
Weighted sample size	452	109	346	189

particularly relevant to this paper's focus.

Table 4.1 shows the average number of hours per person over the day spent carrying out each of the main activities and the broad aggregates. The statistics are split by gender and by whether we restrict the sample to main respondents or to proxy information on spouses. Overall, the statistics are plausible and in line with what one would expect. The gender difference in the allocation of activities emerges clearly. Women respondents report 4 hours of informal work per day compared to 1 hour reported by men. These additional hours of work come at the expense of a reduction in productive and leisure activities by 2 and 1 hour respectively. The context portrayed by this data corroborates the hypothesis that access to light could have the potential to reduce constraints on productive activities by extending the total hours available, for instance by allowing women to do chores more quickly especially after sunset.

A number of arguments support the view that the previous day diary approach used in our survey delivers more accurate measurements of time allocations than the use of stylised direct questions (see Budlender, 2007, for a review). Indeed, the evidence indicates that answers to stylised questions exhibit systemic error compared to diary approaches (Kan, 2007; Bonke, 2005). Time spent on socially undesirable activities tends to be underreported and vice versa, leading to social desirability bias (United Nations Statistics Division, 2005). Respondent subjectivity may also affect which specific activities are deemed to fall within the broad activity type being asked for. For example, respondents may differ in whether they include unpaid or domestic responsibilities in their estimate of the time they spend 'working'. Our diary approach resolves this by post-coding the specific activities mentioned by the respondent in a way that is consistent across individuals.

Stylised questions can also be more demanding for the respondent. The estimation of total time across a given type of activity is computationally intensive, particularly when activities occur frequently and intermittently. This is in line with impressions gained from our cognitive testing of questions. We invariably found that parents were most comfortable accounting for time by describing atypical day, or even better, the previous day. They were able to recount the sequence of events, including start and end times.[4] This was particularly the case for agricultural, livestock, selling and domestic activities. Asking for the number of hours per day was found to be quite complex. Moreover, families never responded in terms of hours per week, even if the question was framed that way.[5] By using a diary approach, all computation of time aggregates is carried out at the analysis stage.

The above biases and errors are exacerbated as the recall period lengthens (Paull, 2002). In light of this, our time diary is limited to a one day recall method, whereby respondents narrate the event of the previous day. This reduces recall bias.[6] Surveys where conducted between Tuesdays and Saturdays, to ensure the previous day was always a working day.

We complement the diary approach with experience sampling via Interactive Voice Response (IVR). We do so for two reasons. Firstly, even a previous day diary by still be subject to some recall bias. Additionally, if the day involves many different spells, the diary approach may generate survey fatigue. These concerns are less relevant in the case of experience sampling. More fundamentally however, another main objective of using IVR methods was to experiment with what is still an innovative and largely unexplored approach to survey data collection and assess its viability for research.

Overall, stylised questions suffer from a number of biases, which can particularly affect the measurement of activities occurring in short and unstructured spells. It is arguably these sorts of activities that may be most influenced by the availability of light. By using a combination of previous day recall diaries and experience sampling we can gear our data collection to maximise the ability to detect activities of this sort.

Furthermore, diaries and experience sampling also allow for the collection of more granular and detailed information on time use that is possible via stylised questions. Critically, these approaches uncover the timing of activities across the time of day and night. This allows the exploration of possible 'task-shifting' effects of access to light, which were beyond the scope of previous work on the issue. Secondly, these approaches allow for the inclusion of contextual variables. Typical examples of these are where and

[4]All individuals in our sample consists of mobile phone owners, and therefore have access to clock.

[5]The only exception to the above was the incidence of casual work, which tended to be reported in number of days this week.

[6]Note, however, that this increases the variation in the data, which reduces the statistical power to detect differences across treatment and control groups.

with whom the activity is taking place, and whether the activity is paid. In our survey instruments, we capture whether the activity is carried out at home or away from home. This aims to corroborate the hypothesis that access to light may be particularly relevant for home production.

Economic activity and productive assets

The second main objective of our data collection was to document and measure the household's productive capacity. The intention here is to detect whether a possible change in time use and/or savings on alternative fuel expenditure (see Section 4.3.2) could trigger an expansion and/or change in the economic activities the household is involved in. Considering the near universal involvement in agriculture, we paid considerable attention to measuring crop and livestock activity. At the same time, we included a detailed module on non-farm activities in line with the focus of the paper.

During our field visit, we gained an immediate impression of how virtually all families relied on agriculture for a relevant share of their living. Similarly, most, though not all, families owned animals. An accurate measurement of the activities and income flows of rural households is notoriously difficult (see The Wye Group, 2007, for a review). It involves dealing with issues such as distinct and highly seasonal income and expenditure flows, and high degree of measurement error on key parameters such as plot size.[7] We were therefore faced with the challenge of identifying a small number of questions that could act as strong proxies for overall agricultural activities. We addressed this by noting that each crop type is clearly associated with a distinct purpose: maize is planted for subsistence; vegetables can be sold; tea and sugar cane are exclusively for cash. Similarly, poultry is the main income-generating animal, while a typically small number of cows is kept primarily for family consumption. A family's productive capacity can therefore be proxied by simple questions on the type of crops they cultivate and animals they own, and we include these in our survey.[8] We also included questions on the number of each type of animal owned. In light of the high expected measurement error, we opted to not ask about the size of the plot for each crop.[9]

Descriptive statistics on the responses to each of these questions are presented in Table 4.2. These confirm many of the indications from the Kenya Integrated Household Budget Survey 2006 and the field visit described above. They also evidence the reasonably high level of casual agricultural work. Indeed, around 55% of project families engage in

[7]We tested questions on these topics during the field visit and confirmed the presence of these difficulties.

[8]We coded possible crop types by selecting the most frequently occurring responses among Gucha South respondents to the Kenya Integrated Household Budget Survey 2006 survey

[9]We tested a direct question asking about the size of the land owned by the household during our fieldwork and found a majority were unable to provide an confident answer.

Table 4.2: Farm-based productive activities

Variable	Mean	Standard deviation	10th percentile	90th percentile
Agriculture	97.25			
Livestock	82.81			
Casual agricultural work	55.30			
N# of crops types	3.41	1.65	2.00	6.00
Grows Maize	91.06			
Grows Beans	47.55			
Grows FingerMillet	11.38			
Grows Vegetables	53.30			
Grows Bananas	38.92			
Grows SugarCane	54.55			
Grows Tea	11.44			
Grows Coffee	12.29			
Grows Other	20.69			
N# of livestock types	1.45	0.90	0.00	2.00
Owns Cattle	63.01			
Owns Poultry	65.12			
Owns Goats	15.07			
Owns Rabbits	0.63			
Owns Others	1.38			
N# of Cattle	1.19	1.31	0.00	3.00
N# of Poultry	5.61	10.00	0.00	12.00
N# of Goats	0.41	1.20	0.00	2.00
N# of Rabbits	0.03	0.41	0.00	0.00
N# of Other animals	0.27			
Weighted sample size	800			

Table 4.3: Non-farm productive activities

Variable	All respondents Mean	Respondents with non-farm activities Mean
Non-farm activities	38.63	100.00
Number of non-farm activities	0.44	1.15
Started less than 12 months ago	1.88	4.85
Woman involved in non-farm work	13.75	35.60
Man involved in non-farm work	28.38	73.46
Carried out during hours of darkness	3.50	9.06
Carried out at home	3.50	9.06
Carried out at fixed place away from home	32.07	83.01
Activity is mobile	5.19	13.43
Weighted sample size	800	311

agricultural activities for pay on other people's land.[10]

Around 39% of families relied on incomes other than agriculture and livestock, and these constitute a prime interest for this paper. During the survey, we therefore asked respondents to enumerate all non-farm income-generating activities where an adult member of the household was involved. The activities were coded as open-ended responses. The most frequently occurring terms are displayed in the Figure 4.2. Similar examples were quoted during the fieldwork, and included being: a teacher, a security guard at a plantation, a motorcycle taxi driver (*boda boda* driver), a soapstone carver, a repairs tailor and running small cafe.

The non-farm activity module replicated the core questions from the corresponding module Kenya Integrated Household Budget Survey 2006.[11] These included questions identifying who was the main adult responsible for the economic activity, the time of day and location this was typically carried out, and whether the household was involved in the activity 12 months ago. Descriptive statistics on the responses to each of these questions are presented in Table 4.3. We see that if families engage in any non-farm activities, it will typically be only one. Men are twice as likely than women to be the main person involved in the activity, and the activity is typically carried out at a fixed location away from home.

Household income and saving

Our survey also included a small number of questions relating to the financial circumstances of the household. The motivation for this was twofold. On the one hand, we

[10] The most common example of these is weeding.

[11] The Kenya Integrated Household Budget Survey 20015/16 survey fieldwork was being carried out during the same period as our project. Responses to that survey may allow opportunities for further analysis.

Figure 4.1: A typical landscape in the Gucha South district

Figure 4.2: Non-farm activites word cloud

wanted to collect a measure, even if highly approximate, of the families' income flows to attempt to detect any causal links running from time use to productive capacity and ultimately income. Again, the complex financial circumstances of rural families make this a difficult task. Indeed, during our field visit, we confirmed that while families were generally able to estimate revenues for specific activities or events (for example, the sale of vegetables per day or the price of chicken if sold), they had strong difficulties combining all income sources into an estimate for a homogenous time periods. Despite this limitation, the context of a phone survey simply did not afford asking a detailed set of questions on this issue. Therefore, so as to at least collect some measure, we included a question asking respondents to estimate their typical weekly income, separately from farm and non-farm sources.

The second motivation is to document, again in a stylised fashion, what may be happening in terms of the household's savings. Previous research indicates that off-grid solar products deliver significant savings on expenditure on alternative fuels (notably kerosene). We therefore include a question asking the respondent to report their weekly expenditure on lighting fuels. We hypothesise that these savings may trigger an investment dynamic in the household, either by reinvesting these savings directly or by using this saving stream to support an increased credit capacity. Fully mapping these channels is beyond the scope of this project. However, we seek to gain some proxy signal in relation to these dynamics by asking respondents whether they feel they are setting aside savings on a regular basis and whether they are involved with any financial institutions.[12] Descriptive statistics on the responses to each of these questions are presented in Table 4.4. The income data includes some very high observations, in the order of several multiples of the sample mean. In the analysis, we trim the top 1% in the total income distribution.

4.4 Experiment validity and estimation strategy

This paper is concerned with identifying and quantifying the impact, if any, of access to modern light sources on household economic activities in rural developing country contexts. In this section, we discuss and present evidence on the extent to which our project and data can adequately support causal inference, as well as the extent to which any results should be deemed externally valid. Drawing on the insights from this analysis, we proceed to define our estimation strategy and robustness checks.

[12]The exact wording of the questions was 1) *Do you feel you are able to set a side part of your income as savings on a regular basis?* and 2) *Are you depositing your savings into any form of saving institution?* Enumerators were instructed to *probe for institutions like 'chamas', Savings and Credit Cooperative Organization (SACCOs), microfinance institutions, mobile money etc.*

Table 4.4: Income and savings

Variable	Mean	Standard deviation	10^{th} percentile	90^{th} percentile
Farm income	783.22	969.27	0.00	2000.00
Non-farm income	677.59	1374.02	0.00	2200.00
Total income	1460.81	1744.03	200.00	3200.00
Equivalised total income	122.26	180.04	14.81	280.95
Weeky expenditure on lighting fuel (Kshs)	103.44	101.47	0.00	200.00
Do you feel you are able to set a side part of your income as savings?	38.88			
Savings set in a savings institution?	42.26			
Weighted sample size	800			

4.4.1 Sampling frame and external validity

The population of parents that accepted to share their mobile phone number consists of our main sample of interest for the purposes of this paper. The nature of this sample has implications for the external validity of the results of this study.

Firstly, the adults in our sample will by definition be parents or guardians of a 7th grader, as well as owners of a mobile phone. The selection of these observable characteristics sets them apart from the broader population of adults in the target district. While this will potentially limit the generalisability of any results, we argue this should be a minor concern as both requirements are very common among the target population. Indeed, the Kenyan Housing and Population Census 2009[13] indicates that 89% of household heads and their spouses of working age in Gucha South have dependent children, and 58% of these own a mobile phone.[14] Our sample is therefore still likely to be relevant to broad sections of the rural population.

Nevertheless, it is worth being aware of how such a subset of the population is likely to be different from the overall population. In Table 4.5 we use the 2009 Census data to report the extent to which some key characteristics differ when comparing all adults of working age, those who are also parents, and those parents who also own a mobile phone. Stars refer to the difference with the 'All working age' group. While the large Census sample allows for a high degree of statistical precision across all comparisons, the differences are in some cases of marginal practical significance. However, parents with a mobile phone clearly emerge as being more affluent and better educated than the broader working age population. They are noticeably more likely to have completed at least primary education, and to own a radio. Their homes are somewhat less likely to

[13] At the time of writing, this is the most recent census in Kenya.

[14] Mobile penetration has continued to rise rapidly since, and stood at 88% in 2015 at a national level.

Table 4.5: Sample selection - Census 2009 data

Characteristic	All working age	Working age parents	Working age parents with a mobile phone
Home Owner	0.929	0.939***	0.927
Has earth floor at home	0.898	0.909***	0.865***
Has mud walls at home	0.818	0.829***	0.795***
Connected to electricity grid	0.019	0.016**	0.026***
Use wood fuel for cooking	0.942	0.950***	0.940
Completed primary education	0.688	0.703***	0.779***
Owns a radio	0.746	0.754***	0.893***
Weekly hours worked	35.942	36.311***	36.904***
	4393	3908	2270

** p<0.1, ** p<0.01, *** p<0.001. Stars refer to the difference with the 'All working age' group*

Table 4.6: Sample selection - Student survey data

Characteristic	Parents without contact details	Parents with contact details	p-value of the difference
Father completed primary education	0.80	0.83	0.33
Mother completed primary education	0.76	0.73	0.32
Father completed secondary education	0.48	0.55*	0.05
Mother completed secondary education	0.38	0.40	0.65
Has covered latrine walls at home	0.65	0.70	0.12
Has mud walls at home	0.71	0.69	0.51
Weighted sample size	296	580	

** p<0.1, ** p<0.01, *** p<0.001.*

use traditional materials, in favour of using more expensive materials such as cement. As such, while still strongly relevant to rural populations in developing countries, our sample is likely to be less representative of the poorest and more marginalised population groups in these communities.

However, data from our own sample suggest the selection on having consented to share mobile phone contacts is not as strong as suggested by the Census. Table 4.6 reports relevant responses to the student survey, which included both students whose parents had shared mobile contacts and others than did not. As with the census, the subsample of parents for whom we obtained mobile numbers appears to be somewhat more educated. Contrary to the 2009 Census, the quality of their home amenities do not indicate they are any wealthier. This is likely attributable to the increased popularisation in mobile phone ownership in recent years.

Finally, as parents are voluntarily offering their contact details, there may be an element of self-selection on the basis of both observable and unobservable characteristics. Ultimately, as self-selection on unobservables cannot be tested for, this is just a threat to external validity that we must keep in mind.

4.4.2 Balancing and internal validity

We proceed to discuss the extent to which the data supports the internal validity of the experimental design. Most notably, to substantiate a claim for causal attribution we would want to show that the randomisation has successfully split the sample in two groups differing only in their allocation to treatment.

Table 4.7 provides important evidence to that effect. Firstly, it shows that while the likelihood of offering mobile contact details is higher among the treated than the control group, this difference is not statistically significant. Furthermore, participation rates into the phone survey (conditional on having given contact details) are identical across treatment and control groups. This suggests that unobserved factors that determine participation into the survey are not differentially distributed across the two groups, thereby attenuating the concern that these might be driving differences in outcomes we might observe.

Table 4.7 then proceeds to present a standard set of balancing statistics to evaluate the success of the randomisation. The results in the table show that the distribution of gender, age, education and household structure are very similar across treatment and control groups. On the basis of this evidence, we have reason to believe that the randomisation has been successful in constructing two comparable groups.

As discussed in Section 4.3, space constraints limited the number of background variables we could collect. This can potentially limit the confidence in the randomisation outcome in two respects. Firstly, a notable omission to the background characteristics considered is a proxy of wealth. Differential wealth status across treatment and control households would be important confounders in the context of this study. The achieved balance in educational status should partly address this concern, to the extent that wealth and education tend to be correlated. Nevertheless, results in this paper should be considered with this caveat in mind.

Secondly, some aspects of the results we discuss in Section 4.5.3 may raise concerns over the balancing of the incidence of non-farm income streams at baseline. If present, baseline differences in the incidence of non-farm activities would be an important threat to causal inference. Firstly, these are likely to explain a large part of end-line differences otherwise attributable to a treatment effect. More generally, however, causal inference on other outcomes would also be called into question to the extent that the incidence of non-farm activities is a relevant confounding factor. To address this latter concern, we

Table 4.7: Balancing statistics

Variable	Control	Treated	P-value of difference
Offered contact details	0.58	0.66	0.13
Weighted sample size	1442	787	
Participated in phone survey	0.63	0.61	0.53
Weighted sample size	791	501	
Initial respondent is female	0.36	0.41	0.16
Ultimate respondent is female	0.43	0.45	0.45
Age of respondent	42.08	41.36	0.31
Completed primary education	0.40	0.44	0.24
Completed secondary education	0.27	0.30	0.33
Number of Adults in the household. Adult is over 18.	2.72	2.84	0.22
Number of Children in the household. Children is under 18	4.05	4.08	0.79
Weighted sample size	495	305	

** p<0.1, ** p<0.01, *** p<0.001.*

test the robustness of all our results to a specification that controls (among other things) for the reconstructed incidence of non-farm activities at baseline (see Section 4.4.4 for more details).[15] Reassuringly, the estimated effects on all other margins are confirmed, albeit sometimes attenuated, when estimating such a specification. Overall, therefore, we argue this shows that baseline differences in non-farm work, even *if* present, do not appear to invalidate estimation of treatment effects on other margins.

A final aspect concerning the internal validity of the experiment relates to the geographical distribution of treatment. As discussed in Section 4.2, identification rests on the exogenous allocation of solar lamps to households induced by a clustered experiment operating at school level. As such, we expect treatment to not be uniformly distributed across space. Figure 4.3 confirms this. It plots a geographical kernel density estimate of treatment intensity, calculated over areas with radius of 7.5×10^{-3} degrees (800m ca.) around the centroids of hexagonal grid cells of 2.5×10^{-3} degrees in diameter (300m ca.). Actual sample data points are overlaid on the map, colour-coded in green for treated units and red for control units.

The lack of geographical homogeneity in treatment poses a threat to identification to the extent that influential and omitted factors share a similar geographical distribution. The strong economic and cultural homogeneity across the project region observed during

[15] Assuming an equal hazard of activities ending over the previous 12 months across treated and control groups, the dummy indicator for the reconstructed incidence of non-farm activities at baseline is set as equal to one for all families reporting a non-farm income stream and reporting this was not started in the last 12 months.

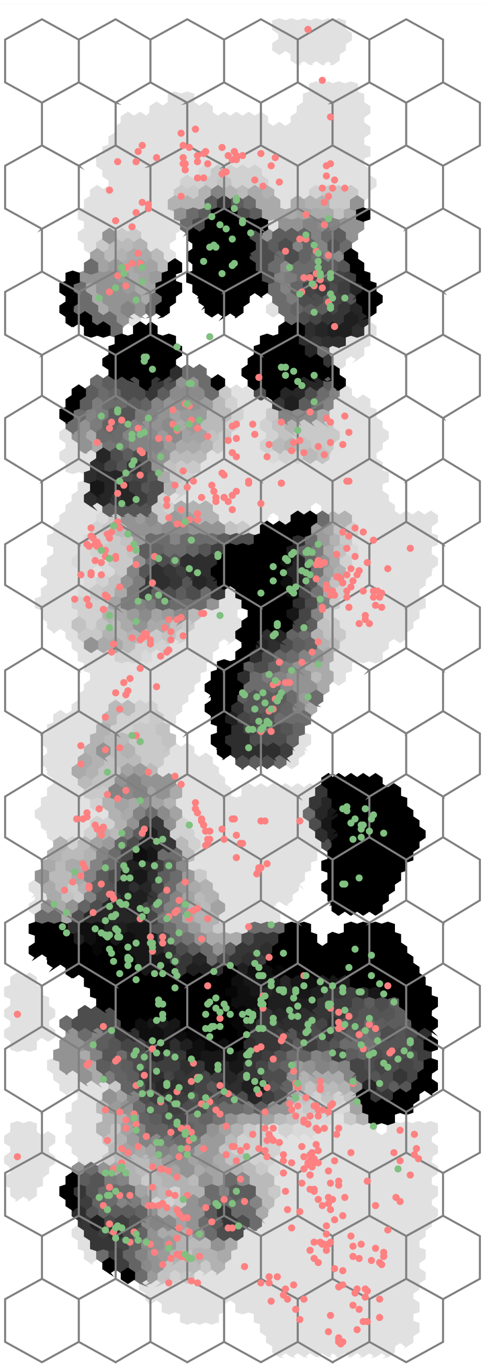


Figure 4.3: Geographical distribution of treatment

Table 4.8: Balancing statistics - with geographical fixed effects

Variable	Mean Untreated	Mean Treated	P-value of diff
Initial respondent is female	0.34	0.40	0.19
Ultimate respondent is female	0.38	0.42	0.40
Age of respondent	44.97	43.68	0.20
Completed primary education	0.19	0.25	0.23
Completed secondary education	0.44	0.44	0.94
Number of Adults in the household. Adult is over 18.	2.87	2.97	0.41
Number of Children in the household. Children is under 18	3.82	3.97	0.41
Weighted sample size	324	249	

** p<0.1, ** p<0.01, *** p<0.001.*

the field visit somewhat reduces this concern. Nevertheless, we seek an identification strategy that is robust to this possible bias. Specifically, as presented formally in Section 4.4.4, we include geographical fixed effects for arbitrarily constructed hexagonal grid cells 1.5×10^{-2} degrees in diameter (1550m ca.), as displayed in Figure 4.3. In this variant of our estimation strategy, identification hinges on the comparison of treated and control units within the same grid cell, thereby netting out the effect of any local area factors. As such, it rests only on observations falling within cells that include both treated and control units. In our case, this consists of 71% of our sample. Table 4.8 shows how the balancing of background variables is maintained when including geographical fixed effects and hence restricting the effective sample used.

Overall, the data available to us does not point to unsurmountable threats to internal validity. While a more extensive set of background characteristics, or indeed a baseline survey, would have provided a more solid basis to evaluate the outcome of the randomisation, the evidence and estimation approach presented in this section should provide a reasonable degree of confidence for causal inference.

4.4.3 Treatment adoption and compliance

Identification rests on the assumption that the treatment is administered as expected and participants comply with it. This may not always be the case. In the context of this study, lamps could be left unused or break, or treated families may sell or lend them to others. Control families may also take initiatives that might invalidate the experiment. In particular, they might choose to purchase a lamp.

Our data provides strong evidence indicating that the treatment was adopted by a majority of the treatment group and that compliance was high. Evidence from the student survey indicates that about 90% of respondents reported that the lamp was working well

Table 4.9: Compliance with treatment

Variable	Mean Control	Mean Treated	P-value of difference
Solar is main light source	0.20	0.57***	0.00
Kerosene is main light source	0.60	0.28***	0.00
Solar is among light sources	0.24	0.70***	0.00
Has any electrict light source	0.28	0.74***	0.00
Weighted sample size	525	328	

** p<0.1, ** p<0.01, *** p<0.001.*

or with minor problems only; in more than 94% of cases, the solar charge of the lamp was sufficient for the required activities; and in more than 90% of cases the lamp stayed at home during the night. Responses from the parent survey also indicate that being assigned to treatment has a real effect on the light sources available to the household. Table 4.9 shows how that families assigned to treatment are three times more likely to indicate *solar* as their main light source, or as one among the family's light sources. Equally, they are half as likely to indicate kerosene as their main light source.

Nevertheless, Table 4.9 also shows that compliance was not complete. As we discuss in Section 4.4.4, this suggests an instrumental variable estimation strategy would provide important insights on the treatment effect on those that actually comply.

4.4.4 Estimation strategy

Household level outcomes

In consideration of the evidence presented in the preceding sections, we here set out our estimation strategy. Firstly, in light of the random allocation of the lamp across households, our core specification is a reduced form OLS regression of economic outcomes on treatment. The Intention-To-Treat (ITT) on household-level outcomes is estimated using the following specification:

$$y_{hj} = \beta_0 + \beta_1 \, Treatment_{hj} + \epsilon_{ij} \tag{4.1}$$

where y_{hj} is the outcome of household h in grid cell j. This specification is referred to as *CX* in the tables.

In light of randomisation, our core specification can already be interpreted causally. However, to maintain a cautious approach to identification and inference, we also run two robustness specification. These are intended to assess the robustness to imbalances in individual characteristics and geographical location respectively.

The first specification replicates Equation 4.1 with additional controls for gender, age and educational level of the respondent and the number of adults and children in the household. In light of the discussion in Section 4.4.2, we also include the reconstructed baseline incidence of non-farm income-generating activities. We estimate:

$$y_{hj} = \beta_0 + \beta_1\ Treatment_{hj} + X_{hj} + \epsilon_{ij} \tag{4.2}$$

where X_{hj} is the vector of controls. This specification is referred to as *CTL* in the tables.

Next, we re-estimate the main specification adding geographical fixed effects over arbitrarily defined grid cells. By estimating the following equation with grid cell fixed effects, we can account for the effect of any local omitted factors λ_j:

$$y_{hj} = \beta_0 + \beta_1\ Treatment_{hj} + \lambda_j + \epsilon_{ij} \tag{4.3}$$

This specification aims to address the concern that treatment may be correlated with omitted variables across space. If this is the case, estimated coefficients will be different from the main specification. On the contrary, if coefficients do not change significantly, we would conclude that local omitted factors should not be a major concern. We cluster standard errors across the 97 grid cells. The standard error will therefore increase as treatment is known to be geographically uneven.[16] This specification is referred to as *GEO* in the tables.

Additionally, we adopt an instrumental variable approach to estimate the effect of treatment in the context of partial compliance. Specifically, we use the random allocation to treatment to instrument for distribution of access to any source of electric light as observed in the data. Formally, the first and second stages for household level outcomes are:

$$Elec_{hj} = \beta_0 + \beta_1\ Treatment_{hj} + \mu_{ij} \tag{4.4}$$

$$y_{hj} = \beta_0 + \beta_1\ E\hat{l}ec_{hj} + \epsilon_{ij} \tag{4.5}$$

The above approach allows for the estimation of the effect of treatment among those who make use of it (Local Average Treatment Effect). This is identified if the treatment does not affect those who do not comply. There is reason to believe this assumption may be inviolated in our context, as, for example, non-compliers will draw a monetary benefit if they sell the lamp. This specification is referred to as *IV* in the tables.

Note that we do not estimate heterogenous treatment effects by gender in these specifications as the outcomes are for the household as a whole.

Using a plurality of specifications allows to test the robustness of the results under differing assumptions on what would allow for causal identification. While results for any single specification will be relevant in their own right, those that remain robust to these

[16] The intra-cluster correlation of treatment across the grid cells in 0.44

differing assumptions will constitute the strongest findings of this paper.

Individual aggregate time use outcomes

The estimation equations require some minor modifications when analysing aggregate time use. This is due to the fact that time use data is collected at the individual, rather than household, level. The respondent reports the activities carried out over the course of the previous day, and then offers a proxy response for their spouse. The latter are likely to be of significantly lower quality due to measurement errors and response biases.[17] As such, we analyse time use data both including and excluding spouse data. Specifically, our core specification becomes:

$$y_{ihj} = \beta_0 + \beta_1 \ Treatment_{ihj} + \epsilon_{ij} \tag{4.6}$$

where y_{ihj} is the outcome of adult i in household h in grid cell j. Estimation is initially restricted to the time use data on main respondents provided by the respondents themselves. The equation is then estimated on data including proxy information on spouses provided by the respondents. When spouses are included, standard errors are clustered at the household level. We adapt the above specifications in the same way to allow them to be estimated on individual level data.

Given the importance of the gender dimension in relation to our research questions, we allow for the possibility of heterogeneous effects by gender in our estimation of time use. We do so by allowing full interactions of the treatment variable with a gender indicator. Our core specification therefore becomes as follows, and all other specifications are altered in the same way:

$$y_{ihj} = \beta_0 + \beta_1 \ Treatment_{ihj} + \beta_2 \ Female_{ihj} + \beta_3 \ Treatment_{ihj} \times Female_{ihj} + \epsilon_{ij} \tag{4.7}$$

Individual continuous time use outcomes

As discussed in Section 4.3.2, the experiential time use data varies by respondent and time of day. Similarly, we transform the previous day diaries to identify the activity being undertaken by the respondent and their spouse at 15 minute snapshots across the 24 hours of the day. We use this data to estimate the effect of treatment over the different times of day.

We do so by separately estimating non-parametric specifications of the incidence of each activity type across the course of the day by treatment and control group, and calculating the difference between the two. Specifically, we estimate a kernel-weighted

[17] During the phone survey, a number of respondents reported they only had imperfect knowledge of the activities carried out by their partner.

local mean smoothing regression (Nadaraya, 1964; Watson, 1964; Gasser & Mller, 1979) of dummy indicators of the incidence of activity A_{iht} for adult i in household h at time of day t. We estimate the mean and standard errors of the incidence of each activity across the time of day separately for treatment and control group. We then calculate the difference between treatment and control group means. The standard errors of the difference is calculated from the standard errors of the two group means on the assumption of zero covariance between the two. We perform this estimation for the pooled sample, as well as for males and females separately.

4.5 Results

4.5.1 Aggregate time Use

We find evidence of treatment effects on the time dedicated to some activities. We report the treatment effects for broad time aggregates (productive activities, informal work, and leisure), as well as the main activities that compose these, in Tables 4.10, 4.11 and 4.12. Table 4.11 display results for men, Table 4.12 for women and pooled results are in Table 4.10. The left panel of each table restricts the sample to time use data on respondents, while the right panel pools information on respondents and proxy information in relation to spouses. Each panel then displays the 4 main specifications discussed in Section 4.4.4.

The overall picture portrayed by the results across specifications is that the lamp favours a reduction in time spent on agricultural activities in favour of activities of a broadly social nature, particularly for men. We find an intention-to-treat effect indicating a statistically significant reduction in time allocated to agricultural activities by 30 minutes per day.[18] We also find some evidence indicating that men reduce their involvement in informal work (primarily house chores) by around 20-30 minutes per day.

Men appear to reallocate this time primarily to leisure time, in the order of a statistically significant 45-60 minutes per day. This is mainly spent on personal care and social engagements, though the coefficients on the latter are statistically insignificant. Men in treated household also report sleeping in by an additional 10-15 minutes per day. This leads to a, possibly counter-intuitive, *reduction* in total time awake, albeit statistically insignificant. We can speculate that the postponing of wake-up times could be related to men's reduced involvement on the farm.

There are no other statistically significant results for women, most likely due to the lower sample size. The pattern across coefficients, however, provides some indication that women transfer time saved on agriculture to informal work (primarily house chores by 15 minutes) and social engagements by 10 minutes per day. There also is some indication that women in treated households do postpone going to sleep by some 10-15 minutes on

[18]Note this is robust to the inclusion of imputed baseline non-farm activities in Specification CTL

average, and report a corresponding increase in time awake. To reiterate, non of these results are statistically significant.

Some results indicate the lamp may trigger an increased engagement beyond the farm context. Beyond statistically insignificant but positive coefficients on social engagements for both men and women, we also generally find positive (though insignificant) coefficients on non-farm work, and positive and significant effects on time spent in transit. Again, these effects are larger work men than for women.

Overall, access to light appears to trigger a move away from the farm and toward increased leisure and social participation. It may also be associated with an increased involvement in non-farm work, though the evidence is weak. These effects are stronger for men than for women, with some indication that this is because the extension of productive hours leads women to take on house chores previously carried out by men. We are not aware of any other studying identifying a causal effect of access to light on aggregate time use.

4.5.2 Time use over the course of the day

As discussed in Section 4.1, our time use data allows us to explore any treatment effects on the timing of different activities. These treatment effects are visualised in Figure 4.4, and Figures 4.5 and 4.6 for men and women respectively. Each panel displays the estimated treatment effect (blue line) and 90% confidence interval (light blue shaded area) across all hours of the day. The panels in the top row of each Figure display the treatment effects across the broad time aggregates, with the panels below each representing their component parts. The red horizontal line runs at zero, and the vertical lines represent approximate sunrise and sunset times.[19]

[19] As Kenya lies on the Equator, sunrise and sunset times are broadly constant across the year.

Table 4.10: Aggregate time use - Pooled men and women

Outcome	Main respondents only				Respondents and spouses			
Specification	CX	CTL	GEO	IV	CX	CTL	GEO	IV
Productive Activities	-0.46	-0.53*	-0.64	-1.00	-0.21	-0.28	-0.45	-0.43
	(0.28)	(0.26)	(0.45)	(0.62)	(0.25)	(0.24)	(0.38)	(0.52)
Agriculture	-0.58*	-0.44*	-0.60*	-1.27*	-0.36*	-0.26	-0.38	-0.75*
	(0.23)	(0.22)	(0.33)	(0.51)	(0.21)	(0.20)	(0.28)	(0.44)
Livestock	-0.11	-0.06	-0.32*	-0.25	-0.07	-0.04	-0.25*	-0.14
	(0.11)	(0.11)	(0.17)	(0.24)	(0.09)	(0.09)	(0.12)	(0.18)
Non-Farm Work	0.23	-0.02	0.28	0.51	0.22	0.01	0.19	0.46
	(0.32)	(0.28)	(0.53)	(0.70)	(0.27)	(0.24)	(0.41)	(0.57)
Informal Work	0.05	-0.01	-0.20	0.11	-0.10	-0.16	-0.08	-0.21
	(0.23)	(0.19)	(0.30)	(0.50)	(0.21)	(0.20)	(0.24)	(0.43)
Family Care	0.04	0.02	-0.05	0.08	-0.02	-0.03	-0.05	-0.05
	(0.05)	(0.05)	(0.08)	(0.11)	(0.06)	(0.05)	(0.07)	(0.12)
House Chores	0.03	-0.03	-0.11	0.06	-0.07	-0.14	-0.02	-0.14
	(0.21)	(0.18)	(0.27)	(0.45)	(0.19)	(0.19)	(0.24)	(0.41)
Shopping or Sourcing Goods	-0.02	-0.00	-0.04	-0.04	-0.01	-0.00	-0.01	-0.02
	(0.06)	(0.05)	(0.10)	(0.12)	(0.05)	(0.05)	(0.07)	(0.10)
Leisure Time	0.30	0.43	0.63	0.66	0.20	0.31	0.25	0.42
	(0.27)	(0.27)	(0.45)	(0.58)	(0.25)	(0.25)	(0.43)	(0.52)
Personal Care	0.19*	0.20*	0.21	0.42*	0.11	0.10	0.09	0.24
	(0.11)	(0.11)	(0.14)	(0.23)	(0.10)	(0.11)	(0.14)	(0.22)
Sleeping or Resting	-0.12	-0.02	0.11	-0.27	-0.11	0.02	0.04	-0.22
	(0.21)	(0.21)	(0.31)	(0.45)	(0.19)	(0.19)	(0.28)	(0.40)
Social Engagements	0.23	0.25	0.31	0.51	0.19	0.19	0.12	0.40
	(0.18)	(0.18)	(0.22)	(0.40)	(0.16)	(0.16)	(0.20)	(0.33)
In Transit	0.17*	0.16*	0.04	0.36*	0.13*	0.13*	0.05	0.27*
	(0.09)	(0.10)	(0.08)	(0.21)	(0.08)	(0.08)	(0.07)	(0.16)
Total time awake	-0.15	-0.17	-0.06	-0.34	-0.14	-0.17	-0.13	-0.29
	(0.18)	(0.19)	(0.26)	(0.40)	(0.16)	(0.16)	(0.25)	(0.33)
Time wake up	0.21*	0.23*	0.23*	0.45*	0.18*	0.20*	0.23*	0.37*
	(0.10)	(0.11)	(0.13)	(0.23)	(0.09)	(0.09)	(0.12)	(0.20)
Time go to sleep	0.05	0.06	0.17	0.12	0.03	0.03	0.09	0.07
	(0.13)	(0.14)	(0.18)	(0.29)	(0.11)	(0.11)	(0.17)	(0.23)
At Home	0.33	0.43	0.14	0.72	-0.06	-0.00	-0.09	-0.13
	(0.33)	(0.32)	(0.50)	(0.73)	(0.30)	(0.29)	(0.42)	(0.62)

*Standard errors in parentheses. * p<0.1, ** p<0.01, *** p<0.001.*

Table 4.11: Aggregate time use - Men

Outcome	Main respondents only				Respondents and spouses			
Specification	CX	CTL	GEO	IV	CX	CTL	GEO	IV
Productive Activities	-0.41	-0.47	-0.56	-0.87	-0.25	-0.38	-0.45	-0.51
	(0.37)	(0.35)	(0.50)	(0.80)	(0.35)	(0.34)	(0.47)	(0.72)
Agriculture	-0.61*	-0.49	-0.72*	-1.31*	-0.58*	-0.44	-0.69*	-1.20*
	(0.32)	(0.30)	(0.40)	(0.72)	(0.30)	(0.28)	(0.36)	(0.62)
Livestock	-0.16	-0.14	-0.39*	-0.35	-0.07	-0.04	-0.25	-0.14
	(0.17)	(0.16)	(0.23)	(0.36)	(0.14)	(0.14)	(0.18)	(0.29)
Non-Farm Work	0.37	0.15	0.54	0.80	0.40	0.09	0.48	0.83
	(0.48)	(0.40)	(0.61)	(1.03)	(0.43)	(0.38)	(0.53)	(0.89)
Informal Work	-0.29	-0.33*	-0.55*	-0.62	-0.36	-0.35	-0.33	-0.74
	(0.19)	(0.19)	(0.29)	(0.42)	(0.23)	(0.24)	(0.31)	(0.49)
Family Care	0.04	0.03	-0.02	0.09	0.01	0.02	-0.01	0.02
	(0.07)	(0.07)	(0.08)	(0.14)	(0.08)	(0.08)	(0.10)	(0.16)
House Chores	-0.25	-0.29*	-0.42	-0.55	-0.31	-0.32	-0.24	-0.64
	(0.18)	(0.17)	(0.26)	(0.38)	(0.21)	(0.22)	(0.27)	(0.44)
Shopping or Sourcing Goods	-0.08	-0.07	-0.12	-0.16	-0.06	-0.05	-0.07	-0.12
	(0.05)	(0.05)	(0.10)	(0.11)	(0.05)	(0.05)	(0.07)	(0.10)
Leisure Time	0.73*	0.83*	1.07*	1.57*	0.61*	0.72*	0.61	1.26*
	(0.37)	(0.36)	(0.50)	(0.82)	(0.35)	(0.34)	(0.51)	(0.73)
Personal Care	0.39**	0.40**	0.38*	0.85**	0.22	0.21	0.18	0.45
	(0.15)	(0.15)	(0.19)	(0.33)	(0.14)	(0.14)	(0.17)	(0.29)
Sleeping or Resting	0.04	0.13	0.26	0.09	0.04	0.15	0.13	0.08
	(0.29)	(0.29)	(0.37)	(0.63)	(0.27)	(0.26)	(0.35)	(0.55)
Social Engagements	0.29	0.30	0.44	0.62	0.35	0.35	0.30	0.73
	(0.27)	(0.26)	(0.31)	(0.58)	(0.24)	(0.23)	(0.27)	(0.49)
In Transit	0.24	0.23	0.10	0.51	0.16	0.16	0.07	0.34
	(0.15)	(0.16)	(0.13)	(0.34)	(0.13)	(0.13)	(0.09)	(0.27)
Total time awake	-0.42	-0.43	-0.31	-0.91	-0.45*	-0.47*	-0.42	-0.93*
	(0.28)	(0.28)	(0.35)	(0.61)	(0.26)	(0.26)	(0.37)	(0.55)
Time wake up	0.37*	0.38*	0.40*	0.80*	0.36*	0.38*	0.41*	0.75*
	(0.17)	(0.17)	(0.19)	(0.38)	(0.15)	(0.15)	(0.18)	(0.33)
Time go to sleep	-0.05	-0.05	0.09	-0.11	-0.09	-0.09	-0.01	-0.19
	(0.19)	(0.19)	(0.26)	(0.40)	(0.17)	(0.17)	(0.25)	(0.35)
At Home	0.53	0.66	0.21	1.14	0.15	0.31	0.08	0.30
	(0.45)	(0.44)	(0.50)	(0.98)	(0.42)	(0.41)	(0.51)	(0.86)

*Standard errors in parentheses. * $p<0.1$, ** $p<0.01$, *** $p<0.001$.*

Table 4.12: Aggregate time use - Women

Outcome	Main respondents only				Respondents and spouses			
Specification	CX	CTL	GEO	IV	CX	CTL	GEO	IV
Productive Activities	-0.40	-0.60	-0.59	-0.89	-0.05	-0.17	-0.33	-0.10
	(0.39)	(0.40)	(0.55)	(0.86)	(0.33)	(0.33)	(0.46)	(0.69)
Agriculture	-0.55*	-0.38	-0.46	-1.23*	-0.14	-0.07	-0.08	-0.30
	(0.32)	(0.32)	(0.45)	(0.71)	(0.27)	(0.26)	(0.35)	(0.56)
Livestock	-0.03	0.03	-0.22	-0.07	-0.04	-0.03	-0.23*	-0.09
	(0.13)	(0.14)	(0.19)	(0.29)	(0.10)	(0.10)	(0.14)	(0.21)
Non-Farm Work	0.18	-0.25	0.10	0.41	0.14	-0.07	-0.01	0.29
	(0.36)	(0.37)	(0.49)	(0.80)	(0.28)	(0.28)	(0.38)	(0.58)
Informal Work	0.31	0.40	0.04	0.70	-0.00	0.03	-0.00	-0.01
	(0.36)	(0.36)	(0.44)	(0.81)	(0.34)	(0.35)	(0.41)	(0.72)
Family Care	0.03	0.01	-0.09	0.07	-0.06	-0.07	-0.09	-0.13
	(0.08)	(0.08)	(0.12)	(0.19)	(0.08)	(0.08)	(0.10)	(0.17)
House Chores	0.24	0.31	0.10	0.53	0.03	0.05	0.05	0.06
	(0.34)	(0.34)	(0.40)	(0.75)	(0.32)	(0.33)	(0.40)	(0.68)
Shopping or Sourcing Goods	0.04	0.08	0.03	0.10	0.03	0.05	0.04	0.07
	(0.10)	(0.10)	(0.14)	(0.23)	(0.08)	(0.08)	(0.10)	(0.18)
Leisure Time	-0.19	-0.08	0.11	-0.42	-0.16	-0.11	-0.06	-0.34
	(0.37)	(0.38)	(0.61)	(0.83)	(0.31)	(0.32)	(0.48)	(0.66)
Personal Care	-0.06	-0.06	-0.00	-0.13	0.01	-0.01	0.01	0.03
	(0.15)	(0.16)	(0.17)	(0.34)	(0.13)	(0.14)	(0.15)	(0.28)
Sleeping or Resting	-0.33	-0.20	-0.06	-0.72	-0.23	-0.11	-0.03	-0.47
	(0.29)	(0.30)	(0.36)	(0.66)	(0.24)	(0.24)	(0.31)	(0.51)
Social Engagements	0.19	0.18	0.18	0.43	0.05	0.02	-0.04	0.11
	(0.23)	(0.23)	(0.29)	(0.50)	(0.18)	(0.18)	(0.21)	(0.37)
In Transit	0.09	0.08	-0.02	0.20	0.11	0.10	0.05	0.22
	(0.08)	(0.08)	(0.08)	(0.18)	(0.07)	(0.07)	(0.09)	(0.15)
Total time awake	0.20	0.17	0.26	0.45	0.18	0.15	0.16	0.39
	(0.22)	(0.22)	(0.28)	(0.48)	(0.16)	(0.16)	(0.20)	(0.33)
Time wake up	0.00	0.03	0.02	0.00	-0.01	0.02	0.04	-0.03
	(0.10)	(0.11)	(0.12)	(0.23)	(0.08)	(0.09)	(0.10)	(0.18)
Time go to sleep	0.20	0.19	0.28	0.44	0.17	0.16	0.20	0.35
	(0.18)	(0.19)	(0.21)	(0.41)	(0.12)	(0.13)	(0.15)	(0.26)
At Home	-0.05	0.12	-0.11	-0.12	-0.39	-0.32	-0.38	-0.82
	(0.46)	(0.47)	(0.63)	(1.02)	(0.38)	(0.39)	(0.53)	(0.80)

*Standard errors in parentheses. * $p<0.1$, ** $p<0.01$, *** $p<0.001$.*

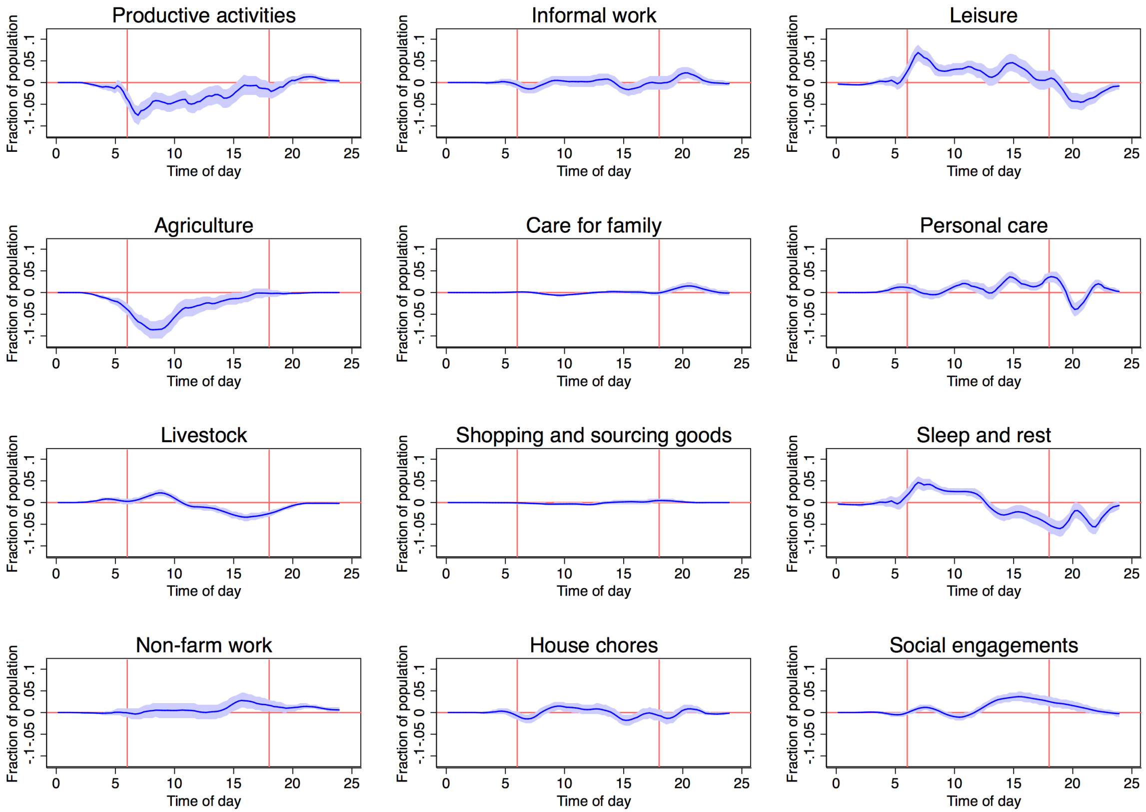


Figure 4.4: Change in time use - Men and women

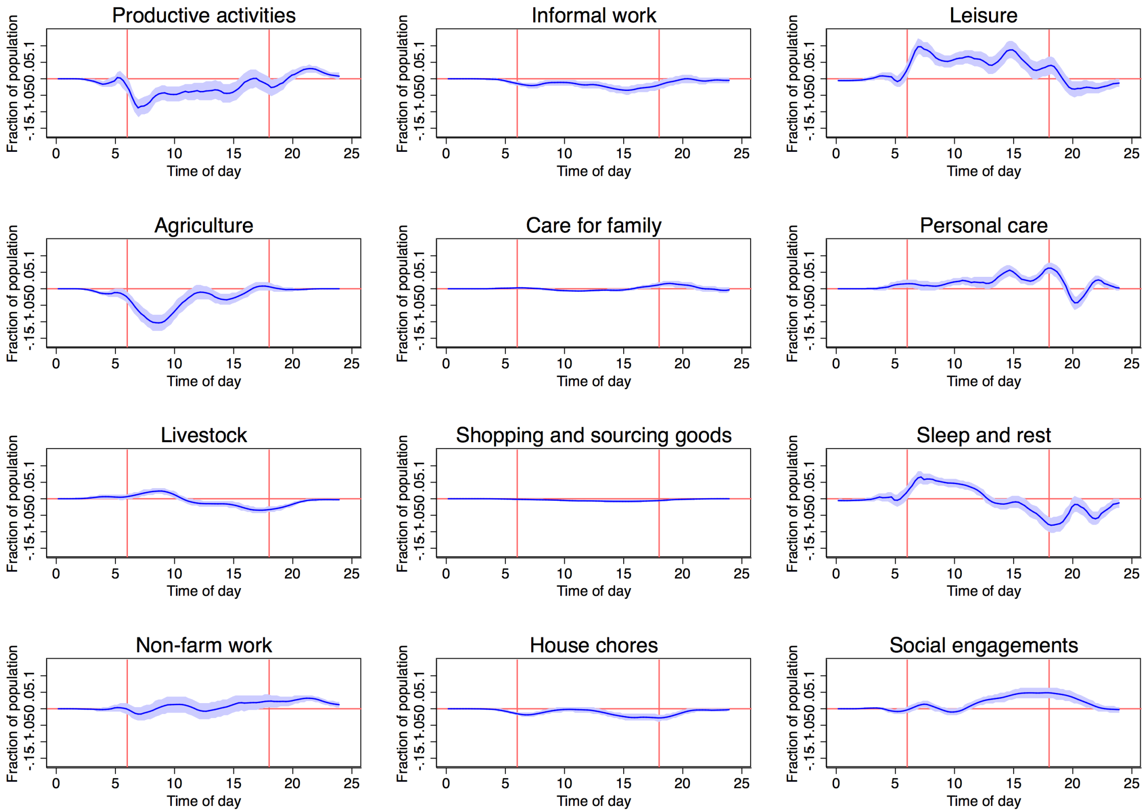


Figure 4.5: Change in time use - Men. Respondents only.

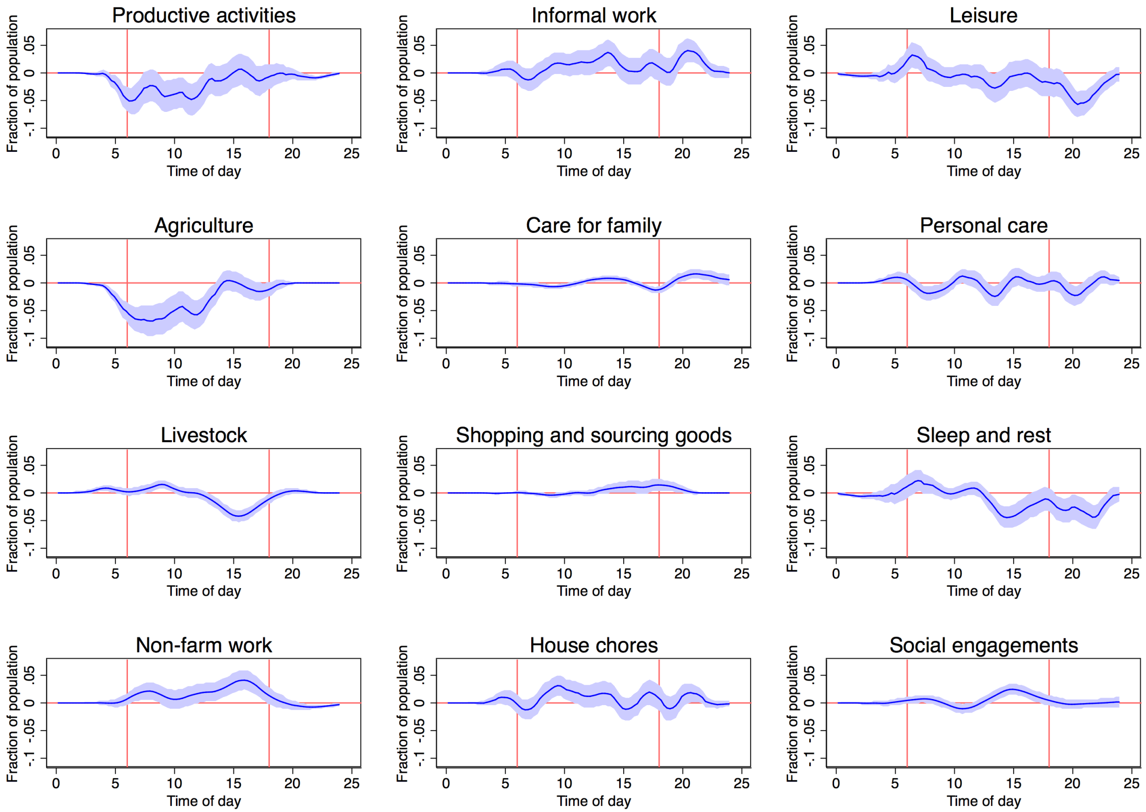


Figure 4.6: Change in time use - Women. Respondents only.

The results displayed in these Figures provide interesting insights that go beyond what can be inferred from aggregate data. In particular, many of the panels exhibit a 'waveform' shape, whereby peaks are compensated by throughs at different times of day and evidence 'task shifting'. For example, this is particularly clear in the trade-off between productive activities, informal work and leisure. While we have already shown in Section 4.5.1 a treatment effect leading to a net shift from productive to leisure activities for men, Figure 4.4 and 4.5 paint a more detailed picture on how and when this materialises. Specifically, we see that productive work gives way to leisure (primarily rest) in the morning, but the reverse occurs during the evenings as men in treated households increase their involvement in non-farm work and in social engagements, as they sleep later. Similarly, women also partly increase leisure time (and informal work) in the morning at the expense of productive activities, but reduce leisure time in the evenings to make time for an increase in informal work (split equally between family care and house chores). Women in treated households also see a notable increase in non-farm work and social engagements in the late afternoon, mainly at the expense of livestock activities and resting.

The Figures also provide further evidence that lamps influence wake-up and sleep times. As with responses to the direct questions on wake-up times, the pattern of reported activities indicates treated individuals, and men in particular, delay their wake-up time. In contrast to the answers to direct questions, we find stronger evidence indicating that treated household delay going to sleep, both for men and women.

The results in Figures 4.4, 4.5 and 4.6 support some important points. Firstly, the evidence of 'task shifting', especially when moving from one side of sunset to the other strongly substantiates the belief that the differences we identify between treated and control households are driven by access to light, and ultimately the treatment.

Secondly, 'task shifting' is economically important even in the absence of changes to aggregate time allocations if the productivity of a task depends on the timing when this is carried out. Social engagements are a case in point. As discussed in Section 4.3.2, these include socially and economically relevant activities such as listening to the radio or paying a visit to members of the community. Being able to reap the social benefits of participating in these also depends on the ability to attend at the most propitious moment. The same applies, for example, for the timing of when to sell at the market.

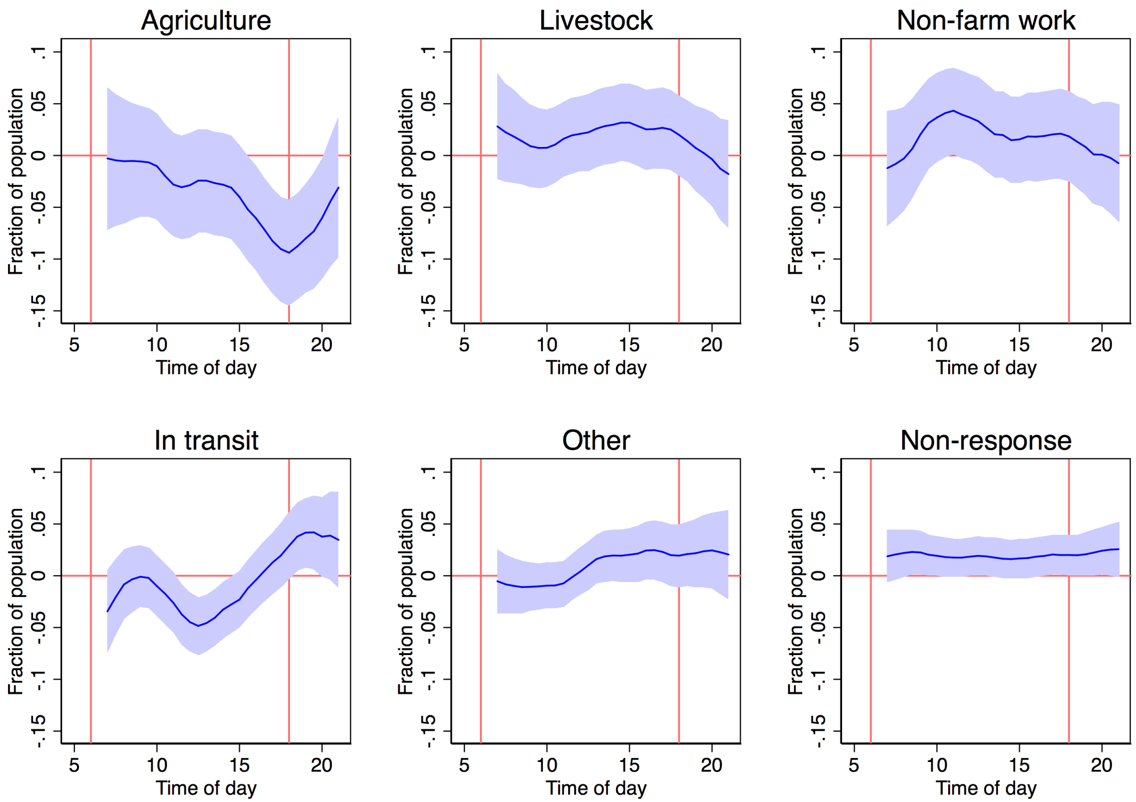


Figure 4.7: Change in time use via experience sampling

Table 4.13: IncomeSources

Outcome	CX	CTL	GEO	IV
Agriculture	0.20	0.50	0.98	0.44
	(1.18)	(1.14)	(1.11)	(2.57)
Livestock	3.39	2.31	1.50	7.40
	(2.65)	(2.65)	(3.95)	(5.79)
Casual agricultural work	-3.31	-1.37	-7.69	-7.22
	(3.56)	(3.53)	(6.11)	(7.79)
Non-farm activities	7.28*	0.95	8.88	15.86*
	(3.51)	(1.03)	(5.77)	(7.67)

*Standard errors in parentheses. * p<0.1, ** p<0.01, *** p<0.001.*

Figure 4.7 shows the equivalent results using data collected via experience sampling. Overall, the results are much less clear-cut, not least because the more limited sample implies estimates are subject to a higher degree of uncertainty. We can nevertheless identify some of the same patterns we have identified so far: statistically significant reductions in agricultural activities and increases in non-farm work, albeit statistically insignificant. To the extent that social engagements are proxied by 'In transit' and 'Other', we again see an increase in social participation in the afternoon and evening among the treated.

4.5.3 Productive activities and income

In this section, we report estimates of treatment effects on household economic activities and productive capacity. In particular, the hypothesis is that changes in time use and 'task shifting' made possible by the lamp may lead to a change in the household's involvement in productive activities both at the extensive and intensive margins. Ultimately, we also test whether these changes lead to a change in incomes.

Table 4.13 gives an initial overview of the possible effects of access to light. As one might expect given the highly agricultural project region, we find that access to light does not alter the household's propensity to engage in agriculture or livestock activities. However, we find some evidence indicating that access to light may lead to an economically significant increase in the incidence of non-farm income-generating activities, in the order of 7 percentage points. However, as discussed in Section 4.4.2, we may have reason to believe this difference may have been present at baseline. We discuss the effects on farm and non-farm activities in more detail in the next two subsections, before presenting results on incomes.

Farm activities

Despite the lack of aggregate effects on the extensive margin in relation to agricultural and livestock income streams, the lamp may nevertheless bring about a change in the

Table 4.14: Crops

Outcome	CX	CTL	GEO	IV
N# of crops types	-0.07	-0.09	-0.14	-0.16
	(0.12)	(0.12)	(0.18)	(0.26)
Grows Maize	0.40	0.72	2.82	0.87
	(2.05)	(2.08)	(3.03)	(4.47)
Grows Beans	-4.30	-5.14	-2.07	-9.38
	(3.57)	(3.63)	(5.00)	(7.76)
Grows FingerMillet	-1.95	-1.99	0.78	-4.25
	(2.25)	(2.27)	(3.27)	(4.90)
Grows Vegetables	2.31	1.58	-0.44	5.02
	(3.57)	(3.62)	(5.50)	(7.78)
Grows Bananas	-0.97	-1.04	-6.54	-2.10
	(3.49)	(3.50)	(4.28)	(7.59)
Grows SugarCane	1.08	1.68	-1.54	2.35
	(3.57)	(3.62)	(5.14)	(7.76)
Grows Tea	-1.52	-2.20	-6.25*	-3.32
	(2.24)	(2.23)	(2.99)	(4.88)
Grows Coffee	-4.85*	-4.76*	-8.43*	-10.56*
	(2.24)	(2.25)	(3.75)	(4.91)
Grows Other	2.61	2.09	7.36*	5.69
	(2.93)	(2.99)	(4.13)	(6.39)

*Standard errors in parentheses. * p<0.1, ** p<0.01, *** p<0.001.*

extent or composition these activities. This section therefore reports results in relation to crop composition (Table 4.14), and the ownership of livestock (Tables 4.15 and 4.16). Overall, we do not find strong evidence of treatment effects on the specific nature of the agricultural and livestock activities activities carried out by the household. There are a couple notable exceptions, however, that fit well with the results described in other sections and anecdotal evidence from the fieldwork.[20]

Firstly, assignment to treatment is associated with a 5 percentage point reduction in the incidence of coffee cultivation. Coffee is exclusively a cash crop, and this result brings further support to evidence suggesting that access to light favours a move away from farm-based livelihoods. At the same time, we find some evidence of an intention-to-treat effect on the number of poultry owned. At around 1 chicken, compared to an average of 5.6 in the sample, the economic magnitude of the result not insubstantial. This result chimes very well with qualitative evidence collected in the field (see Box 1).

Non-farm activities

Table 4.17 summarises results in relation to non-farm activities. We find a potentially sizeable treatment effect on the incidence of non-farm income generating-activities. Our

[20] Naturally, we would expect a fraction of all coefficients estimated to be statistically significant by pure chance.

Table 4.15: Livestock - Extensive

Outcome	CX	CTL	GEO	IV
N# of livestock types	0.07	0.05	0.02	0.15
	(0.06)	(0.06)	(0.09)	(0.14)
Owns Cattle	4.20	3.17	-0.83	9.14
	(3.44)	(3.45)	(5.00)	(7.53)
Owns Poultry	2.56	1.64	6.13	5.58
	(3.39)	(3.41)	(4.67)	(7.40)
Owns Goats	1.10	1.60	-1.28	2.40
	(2.57)	(2.53)	(4.32)	(5.59)
Owns Rabbits	-0.48	-0.57	-0.37	-1.05
	(0.50)	(0.50)	(0.30)	(1.09)
Owns Others	-0.63	-0.57	-1.37	-1.38
	(0.80)	(0.80)	(1.20)	(1.74)

*Standard errors in parentheses. * p<0.1, ** p<0.01, *** p<0.001.*

Table 4.16: Livestock - Intensive

Outcome	CX	CTL	GEO	IV
N# of Cattle	0.09	0.05	0.01	0.19
	(0.09)	(0.10)	(0.11)	(0.20)
N# of Poultry	1.23*	1.01	2.15**	2.68*
	(0.68)	(0.70)	(0.76)	(1.49)
N# of Goats	-0.02	-0.02	-0.08	-0.05
	(0.08)	(0.08)	(0.13)	(0.18)
N# of Rabbits	-0.03	-0.04	-0.02	-0.07
	(0.02)	(0.02)	(0.01)	(0.05)
N# of Other animals	0.11	0.15	-0.48	0.25
	(0.39)	(0.43)	(0.98)	(0.84)

*Standard errors in parentheses. * p<0.1, ** p<0.01, *** p<0.001.*

Box 1 A very entrepreneurial family

While undoubtedly an outlier compared to the other families in the target community, the stories collected from this family strongly illustrate the potential economic benefits of the lamp.

The family was in the treatment group, although it turns out they were one of the few families that had already bought a lamp before the project. Interestingly, this purchase was part of a clear strategy. The father had estimated that with the savings on kerosene generated by having a solar lamp, he could quickly save enough to buy a 'hen in a box' kit. The kit costs around 1,500 KSh (USD $15) and consists of a hen and a metal cage (to protect it from predatory birds). It is designed to provide the basic inputs to start a small poultry farm. When we interviewed the family, less than a year from the purchase of the first lamp, we saw three such cages and about a dozen adult chickens.

The mother is a teacher, but runs a retail activity on the side. She buys wholesale in the town where she teaches and sells retail every afternoon in her community. While this activity predated the purchase of a solar lamp, the lamp permits her to stay at the market a little longer into the twilight hours, thereby extending her working hours.

Table 4.17: Non-farm Activities

Outcome	CX	CTL	GEO	IV
Non-farm activities	7.28*	0.95	8.88	15.86*
	(3.51)	(1.03)	(5.77)	(7.67)
Number of non-farm activities	0.13**	0.06*	0.16*	0.28**
	(0.04)	(0.02)	(0.07)	(0.10)
Started less than 12 months ago	0.68	0.95	3.35	1.48
	(1.00)	(1.03)	(2.04)	(2.19)
Weighted sample size	800			
Woman involved in non-farm work	8.19	4.21	0.93	20.71
	(5.47)	(4.79)	(7.87)	(13.97)
Man involved in non-farm work	-0.14	4.08	11.48	-0.35
	(5.02)	(4.15)	(7.68)	(12.66)
Carried out during hours of darkness	6.73*	5.75*	7.18	17.02*
	(3.39)	(3.43)	(4.87)	(8.93)
Carried out at home	-2.54	-3.75	-4.36	-6.41
	(3.24)	(3.20)	(5.06)	(8.21)
Carried out at fixed place away from home	1.78	2.51	-1.98	4.49
	(4.22)	(4.25)	(5.88)	(10.61)
Activity is mobile	0.45	-0.08	9.01	1.13
	(3.83)	(3.83)	(5.55)	(9.65)
Weighted sample size	309			

*Standard errors in parentheses. * $p<0.1$, ** $p<0.01$, *** $p<0.001$.*

core cross-sectional specification identifies a 7 percentage point intention-to-treat effect. The inclusion of geographical fixed effects does not meaningfully alter the estimated size of the effect (indicating this is not due to the omission of local factors), though clustering at the geographical grid level makes the estimate statistically insignificant. We also identify a statistically significant treatment effect on the number of non-farm income-generating activities, in the order of 0.13 units compared to a mean of 0.44 in the sample. At the same time, however, Table 4.17 also shows that the incidence of activities started less than 12 months ago is only an insignificant 1 percentage point higher among treated households.

One's overall interpretation of these coefficients ultimately depends on one's belief on the accuracy and consistency of respondents' answers. If we believe the responses are always fully consistent, we would suspect that a difference in the mean incidence between treated and control groups was already present at baseline, in the order of 6 percentage points. As one would expect, the intention-to-treat effect disappears when we control for this imputed baseline incidence of non-farm activities, as shown in Specification

CTL.[21] Given randomisation, the ex-ante likelihood of the realisation of such imbalance is very low. The p-value for a standard t-test on the reconstructed baseline variable is 0.05. Nevertheless, this is little consolation if one takes the responses to the two above questions as suggesting that that is indeed the state of the world this project has been implemented in.

On the other hand, if one believes that the responses to the above two survey questions need not be fully consistent, then it is no longer implied that baseline differences were present. In particular, there is reason to be skeptical of the data quality in relation to whether activities were started less than 12 months before. Activities were reported as being recent in only 15 households (8 in the control group and 7 in the treatment group). Even just a handful of cases of incorrect reporting would therefore have a large effect on the estimated treatment coefficient. On the contracry, the variable identifying whether households are involved in non-farm income-generating activities is derived from whether the individual responded to the non-farm module of the survey. This involved 5 questions for each non-farm activity, making it less likely that entire activities would be reported incorrectly.

In summary, a cautious interpretation of the estimates would indicate that the treatment effect on the incidence of non-farm income streams is bracketed between a statistically insignificant 1 percentage point and a statistically significant 7 percentage points.

Qualitative interviews reveal some anecdotes on the types of activities that may have emerged thanks to having access to light. Indeed, on a few occasions, albeit often after direct prompting, the families identified ways in which the lamp had allowed them to increase their productive activity. Examples of this include carrying out clothing repairs or soapstone carving at night, or being able to stay a little longer to sell vegetables at the market. Similarly, control families quoted clothing repairs or small household retail as income-generating activities that they believe could be enabled by having access to solar lamps.

Aside from effects on the extensive margin, the lamp can also affect how existing activities are operated. In the bottom panel of Table 4.17 we present results on whether the lamp influences the timing, location and manager of the activity, estimated on the subsample of households that engage in at least one non-farm activity.

We find evidence that the lamp increases the likelihood that the activity is carried out primarily during the hours of darkness by around 7 percentage points (compared to a mean of 9 percentage points in the sample). This result confirms the previous indications of 'task-shifting' of productive activities into the evening, particularly for men, seen in Section 4.5.2.

[21] In Table 4.17 we still identify a statistically significant impact on the number of non-farm activities when controlling for the imputed baseline incidence of non-farm activities as per Equation 4.2. However, the result disappears if we control for the imputed baseline number of non-farm activities.

Table 4.18: Incomes

Outcome	CX	CTL	GEO	IV
Farm income	-137.44*	-137.26*	-153.70	-297.85*
	(72.51)	(73.63)	(97.34)	(160.10)
Non-farm income	293.08*	191.11*	405.28*	635.14*
	(115.60)	(95.02)	(207.74)	(253.47)
Total income	155.64	53.85	251.58	337.29
	(141.08)	(124.24)	(222.84)	(304.89)
Equivalised total income	6.98	-0.22	27.11	15.12
	(14.11)	(12.75)	(19.63)	(30.50)

*Standard errors in parentheses. * p<0.1, ** p<0.01, *** p<0.001.*

On the other hand, access to light does not appear to influence the location of the activity. Indeed, contrary to commonly held expectations around the increase in home-based production, the coefficient on whether the activity is carried out at home is negative (thought not significantly different from zero).

Treatment is also not associated with any change in the gender of the person involved in the productive activity. The coefficients on the involvement of the woman and/or the man are jumpy across specifications and all statistically insignificant. This indicates that the treatment is gender-neutral, meaning it does not alter the existing gender balance in the involvement in non-farm work. As we saw in Table 4.3, men are twice as likely than women to be managing the non-farm activity.

Incomes

As a final link in the possible causal chain triggered by access to light, we present some summary results on treatment effects on income levels, as seen in Table 4.18.

We find strong evidence of an economically large shift in incomes from farm to non-farm activities. We identify a statistically significant intention-to-treat effect on the level of income from non-farm sources of around 290 Kenyan Shillings (USD $2.90) per week, or equivalent to 20% of the mean household income in our sample. Similarly, we find a statistically significant intention-to-treat effect whereby incomes from farm activities fall by 140 Kenyan Shillings per week (USD $1.40; 9% of mean household income) Importantly, introducing controls, notably the imputed incidence of non-farm income-generating activities at baseline, has no effect on the estimate on farm income while the estimate for non-farm incomes is attenuated but still sizeable.[22] This indicates the effects are driven by changes in the intensive margin. While the effect on farm income becomes marginally insignificant, the inclusion of geographical fixed effects does not dramatically alter the story.

At the same time, we do not detect any change in total income. The treatment effect

[22] Both results are also robust to the inclusion of the imputed number of non-farm activities at baseline.

Table 4.19: Savings

Outcome	CX	CTL	GEO	IV
Weeky expenditure on lighting fuel (Kshs)	-21.87**	-24.35***	-6.10	-47.48**
	(7.07)	(6.94)	(8.16)	(15.08)
Do you feel you are able to set a side part of your income as savings?	-2.14	-4.56	-2.02	-4.65
	(3.48)	(3.39)	(6.01)	(7.59)
Savings set in a savings institution?	-1.49	-3.11	-3.23	-3.26
	(3.53)	(3.45)	(6.42)	(7.69)

*Standard errors in parentheses. * p<0.1, ** p<0.01, *** p<0.001.*

on total household income and total equivalised household income are generally positive, but not statistically significant.[23] This raises the question as to why families might be induced toward this shift if it leaves them no better off. One explanation could relate to the diversification of risks across income streams. Alternatively, it could be that the process of transformation in economic activities is still ongoing and is not fully developed at only 7 months from treatment.

4.5.4 Savings

There is plenty of evidence showing that solar lighting products generate net savings on alternative fuel expenditures. If re-invested, these savings could constitute an additional, or even the primary, mechanism through which access to light can lead to changes in household economic circumstances. In this section, we present a small set of results attempting to gauge the strength of this mechanism in the context of our project. Results are displayed in Table 4.19.

In line with previous work, we identify a statistically significant intention-to-treat effect causing a fall in expenditure on alternative lighting fuel. However, at 20 Kenyan Shillings per week (USD $0.20), or a mere 1.4% of average weekly incomes in our sample, it is considerably smaller than identified in other studies (Grimm et al., 2014; Hassan & Lucchino, 2014; IDinsight, 2015).[24] In line with this, we find no evidence of treatment effects on self-reported savings behaviour. We conclude that any effect of access to light on savings appears to be small, if present at all.

[23] All the results presented in this section omit the top 1% of incomes. As these were several dozens of times the average in the sample, they raised the concern that they could distort the results. Indeed, including these, we find much larger treatment effects on non-farm incomes and statistically insignificant reduction in farm incomes, thereby leading to large statistically significant impacts on overall income. The omission of outliers therefore reflects a cautious approach.

[24] The treatment effect is also not robust to the inclusion of geographical fixed effects.

4.6 Conclusion

Economic development and structural change are long processes involving the evolution and transformation of all aspects of the economy and society. Existing work shows that large scale electrification has an important part to play in this transformation. However, considering a quarter of humanity lives in areas where off-grid solutions such as solar lamps or home systems are the only options allowing some immediate form of energy access, there is very little research into whether, albeit commensurate to their much smaller size, these can trigger similar mechanisms of socio-economic transformation.

To our knowledge, this paper is one of very few recent works tackling this issue, and the only one that finds evidence that small-scale lighting solutions can help stimulate the very first steps in the direction of economic transformation. By exploiting experimental variation in the ownership of solar lamps, we identify treatment effects leading to a shift in household livelihoods from agricultural to non-farm economic activities. We find robust evidence indicating that household income streams change in an economically substantial way in line with this shift. These results are complemented by treatment effects causing a reduction in cash crop cultivation. At the same time, we find tentative evidence indicating a substantial treatment effect on the incidence and number of non-farm income-generating activities the household is involved in.

The emerging picture is backed up by detailed evidence on household time use. We find robust evidence that treated households reduce the time they dedicate to agricultural work, and some indication of an increase in time spent in non-farm contexts (specifically on social engagements and non-farm work). Beyond aggregate time use, detailed time diaries reveal strong evidence of 'task-shifting', particularly between morning and the late afternoon and evenings. This confirms statistically relevant increases in non-farm work and social engagements for men in the evenings, and for women in the late afternoons. Women increase time spent on informal work in the evenings.

Our results suggest that the shift in the household economic activities emerges primarily because of a changed use of time, rather than through saving and investment. Indeed, we find, at best, economically minor effects on savings on alternative lighting fuels expenditure, and no change in savings behaviour. Noting the observed changes in income streams, and in the *timing* of activities more than in the *total time* dedicated to each, we speculate that access to light allows a productivity-increasing reallocation of activities across the times of day.

It is often argued that the flexibility to reallocate activities over an extended range of hours of the day is particularly relevant to the increased economic participation and empowerment of women. This paper speaks to directly to this topic and delivers some sobering results. In line with expectations, we do find evidence that access to light allows women extend their day into the evenings, but this additional time is primarily dedicated to house chores. Importantly, we find a corresponding reduction in the incidence of house chores amongst men, who are therefore the ultimate beneficiaries of this additional time. Like men, women in treated households increase their engagement in non-farm work and social activities in the late afternoon. However, we find no evidence that access to light influences the gender distribution among those responsible for non-farm income-generating activities.

This paper therefore contributes novel evidence indicating that a cheap and renewable source of energy used exclusively for *lightling* can indeed help reap at least a modest fraction the benefits of full scale electrification. In particular, it can favour household diversification away from farm livelihoods toward non-farm micro-entrepreourial endevours by allowing an improved re-allocation of activities over the course of the day. Contrary to the common narrative about the time constraints of women, however, the effect of access to light does not appear to flow in any larger part to women. If anything, we find that effects of the intervention we study are gender-neutral, in the sense that they do not alter, but rather emerge within, the prevailing balance of power between genders. As perhaps would be the case in most of the world, this means that women, albeit benefitting, may not be reaping their full fair share of these benefits.

4.7 Annex

(a) Non-farm work

(b) House chores

(c) In transit

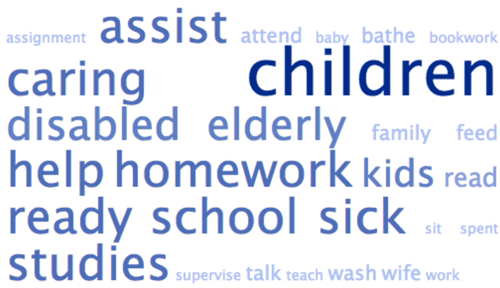
(d) Livestock

Figure 4.8: Activity word clouds - 1

(a) Social engagements

(b) Agriculture

(c) Personal care



(d) Family care

(e) Sleep and rest

(f) Shopping and sourcing goods

Figure 4.9: Activity word clouds - 2

Bibliography

Aklin, M., Bayer, P., Harish, S., & Urpelainen, J. (2015). Does Basic Energy Access Generate Socio-Economic Benefits? A Field Experiment with Distributed Solar Power in India.

Alter, K. (2007). Social enterprise typology.

Ambler, K., Aycinena, D., & Yang, D. (2015). Channeling Remittances to Education: A Field Experiment among Migrants from El Salvador. *American Economic Journal: Applied Economics*, 7(2), 207–32.

Angrist, J. & Lavy, V. (2009). The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial. *American Economic Review*, 99(4), 1384–1414.

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

Bagnoli, M. & Watts, S. G. (2003). Selling to Socially Responsible Consumers: Competition and The Private Provision of Public Goods. *Journal of Economics & Management Strategy*, 12(3), 419–445.

Baird, S., Bohren, J. A., McIntosh, C., & Ozler, B. (2014). *Designing Experiments to Measure Spillover Effects*. SSRN Scholarly Paper ID 2505070, Social Science Research Network, Rochester, NY.

Baron, D. P. (2001). Private Politics, Corporate Social Responsibility, and Integrated Strategy. *Journal of Economics & Management Strategy*, 10(1), 7–45.

Bell, R. M. & McCaffrey, D. F. (2003). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*.

Ben-Ner, A., Karaca-Mandic, P., & Ren, T. (2012). Ownership and Quality in Markets with Asymmetric Information: Evidence from Nursing Homes. *The B.E. Journal of Economic Analysis & Policy*, 12(1).

Ben-Ner, A. & Van Hoomissen, T. (1991). Nonprofit Organizations in the Mixed Economy. *Annals of Public and Cooperative Economics*, 62(4), 519–550.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.

Bester, C. A., Conley, T. G., & Hansen, C. B. (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, 165(2), 137–151.

Bilodeau, M. & Slivinski, A. (1998). Rational Nonprofit Entrepreneurship. *Journal of Economics & Management Strategy*, 7(4), 551–571.

BIS (2010). *BIS Small Business Survey 2010: a user guide for analysts*. Technical report.

Bonke, J. (2005). Paid Work and Unpaid Work: Diary Information Versus Questionnaire Information. *Social Indicators Research*, 70(3), 349–368.

Brewer, M., Crossley, T. F., & Joyce, R. (2013). *Inference with Difference-in-Differences Revisited*. IZA Discussion Paper 7742, Institute for the Study of Labor (IZA).

Brickley, J. A., Lu, S. F., & Wedig, G. J. (2011). *Ownership, Branding and Multiple Audience Concerns: Evidence from Nursing Homes*. SSRN Scholarly Paper ID 1918460, Social Science Research Network, Rochester, NY.

Brown, E. & Slivinski, A. (2006). Nonprofits and the market. In *The Nonprofit Sector: A Research Handbook*. New Haven & London: Yale University Press.

Bruhn, M. & McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4), 200–232.

Budlender, D. (2007). A Critical Review of Selected Time Use Surveys. *UNRISD Publications*.

Bugni, F., Canay, I., & Shaikh, A. (2015). *Inference under covariate-adaptive randomization*. CeMMAP working paper CWP45/15, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Burde, D. & Linden, L. L. (2013). Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools. *American Economic Journal: Applied Economics*, 5(3), 27–40.

Bnabou, R. & Tirole, J. (2010). Individual and Corporate Social Responsibility. *Economica*, 77(305), 1–19.

Calveras, A. & Ganuza, J. J. (2014). *Building a reputation as a socially responsible firm*. Economics Working Paper, Department of Economics and Business, Universitat Pompeu Fabra.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3), 414–427.

Cameron, A. C. & Miller, D. L. (2013). *A Practitioner's Guide to Cluster-Robust Inference*.

Chou, S.-Y. (2002). Asymmetric information, ownership and quality of care: an empirical analysis of nursing homes. *Journal of Health Economics*, 21(2), 293–311.

Crifo, P. & Forget, V. D. (2014). The Economics of Corporate Social Responsibility: A Firm-Level Perspective Survey. *Journal of Economic Surveys*, (pp. n/a–n/a).

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 5(2), 29–57.

Dinkelman, T. (2011). The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review*, 101(7), 3078–3108.

Donald, S. G. & Lang, K. (2007). Inference with Difference-in-Differences and Other Panel Data. *The Review of Economics and Statistics*, 89(2), 221–233.

DTI (2002). *Social Enterprise: A Strategy for success*. London: HM Treasury.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *The American Economic Review*, 91(4), 795–813.

DWP (2016). Chapter 2: Work Programme Claimant Groups.

Economides, N. (1998). *Trademarks*. SSRN Scholarly Paper ID 61148, Social Science Research Network, Rochester, NY.

Gasser, T. & Mller, H.-G. (1979). Kernel estimation of regression functions. In T. Gasser & M. Rosenblatt (Eds.), *Smoothing Techniques for Curve Estimation*, number 757 in Lecture Notes in Mathematics (pp. 23–68). Springer Berlin Heidelberg. DOI: 10.1007/BFb0098489.

Gertler, P. J., Patrinos, H. A., & Rubio-Codina, M. (2012). Empowering parents to improve education: Evidence from rural Mexico. *Journal of Development Economics*, 99(1), 68–79.

Glaeser, E. L. & Shleifer, A. (2001). Not-for-profit entrepreneurs. *Journal of Public Economics*, 81(1), 99–115.

Glewwe, P. (2002). Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. *Journal of Economic Literature*, 40(2), 436–482.

Grabowski, D. C. & Hirth, R. A. (2003). Competitive spillovers across non-profit and for-profit nursing homes. *Journal of Health Economics*, 22(1), 1–22.

Grabowski, D. C. & Stevenson, D. G. (2008). Ownership Conversions and Nursing Home Performance. *Health Services Research*, 43(4), 1184–1203.

Grimm, M., Munyehirwe, A., Peters, J., & Sievert, M. (2014). *A First Step up the Energy Ladder? Low Cost Solar Kits and Household's Welfare in Rural Rwanda*. IZA Discussion Paper 8594, Institute for the Study of Labor (IZA).

Hansmann, H. (1980). The Role of Nonprofit Enterprise. *The Yale Law Journal*, 89(5), 835.

Hansmann, H. (2008). *Firm Ownership and Organizational Form*. SSRN Scholarly Paper ID 2101327, Social Science Research Network, Rochester, NY.

Harsdorff, M. & Bamanyaki, P. (2009). *Impact Assessment of the Solar Electrification of Micro Enterprises, Households and the Development of the Rural Solar Market*. Technical report.

Hassan, F. & Lucchino, P. (2014). Powering Education. *ENEL Foundation Working Paper*, 17.

Hassan, F. & Lucchino, P. (2016). Powering Education 2.

IDinsight (2015). *d.light Solar Home System Impact Evaluation*. Technical report.

IEA (2013). *World Energy Report*. Technical report.

IEA (2014). *World Energy Report*. Technical report.

Innes, R. & Sam, A. (2008). Voluntary Pollution Reductions and the Enforcement of Environmental Law: An Empirical Study of the 33/50 Program. *Journal of Law and Economics*, 51(2), 271–296.

Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling. *Quarterly Journal of Economics*, 125(2), 515–548.

Kan, M. Y. (2007). Measuring Housework Participation: The Gap between Stylised Questionnaire Estimates and Diary-based Estimates. *Social Indicators Research*, 86(3), 381–400.

Khandker, S. R., Barnes, D. F., & Samad, H. A. (2012). The Welfare Impacts of Rural Electrification in Bangladesh. *The Energy Journal*, Volume 33(Number 1).

Khandker, S. R., Barnes, D. F., & Samad, H. A. (2013). Welfare Impacts of Rural Electrification: A Panel Data Analysis from Vietnam. *Economic Development and Cultural Change*, 61(3), 659–692.

Kitzmueller, M. & Shimshack, J. (2012). Economic Perspectives on Corporate Social Responsibility. *Journal of Economic Literature*, 50(1), 51–84.

Landes, W. M. & Posner, R. A. (1987). Trademark Law: An Economic Perspective. *Journal of Law and Economics*, 30(2), 265–309.

Leete, L. (2001). Whither the Nonprofit Wage Differential? Estimates from the 1990 Census. *Journal of Labor Economics*, 19(1), 136–170.

Liang, K.-Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22.

Liebermann, Y. & Flint-Goor, A. (1996). Message strategy by product-class type: A matching model. *International Journal of Research in Marketing*, 13(3), 237–249.

Lipscomb, M., Mobarak, A. M., & Barham, T. (2013). Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. *American Economic Journal: Applied Economics*, 5(2), 200–231.

Lowe, M. (2014). Night Lights and ArcGIS: A Brief Guide.

Lucas, A. M. & Mbiti, I. M. (2012). Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya. *American Economic Journal: Applied Economics*, 4(4), 226–53.

Lucchino, P. (2017). Social enterprise: is there anything in this fuzzy concept?

Lyon, T. P. & Maxwell, J. W. (2011). Greenwash: Corporate Environmental Disclosure under Threat of Audit. *Journal of Economics & Management Strategy*, 20(1), 3–41.

Mackinnon, J. G. & Webb, M. D. (2016). Wild Bootstrap Inference for Wildly Different Cluster Sizes. *Journal of Applied Econometrics*, (pp. n/a–n/a).

Malani, A., Philipson, T., & David, G. (2003). *Theories of Firm Behavior in the Nonprofit Sector. A Synthesis and Empirical Evaluation.* NBER Chapters, National Bureau of Economic Research, Inc.

Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2009). *Does it Pay to Be Good...And Does it Matter? A Meta-Analysis of the Relationship between Corporate Social and Financial Performance*. SSRN Scholarly Paper ID 1866371, Social Science Research Network, Rochester, NY.

McIntosh, C., Alegra, T., Ordez, G., & Zenteno, R. (2014). *Infrastructure upgrading and budgeting spillovers: Mexico's Habitat experiment*. Technical Report 036, UC Berkeley, Center for Effective Global Action.

McWilliams, A. & Siegel, D. (2001). Corporate Social Responsibility: A Theory of the Firm Perspective. *The Academy of Management Review*, 26(1), 117–127.

Miguel, E. & Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1), 159–217.

Muralidharan, K. & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 119(1), 39–77.

Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability and Its Application*, 9.

Nelson, P. (1974). Advertising as Information. *Journal of Political Economy*, 82(4), 729–754.

O'Neill, C., Harrington, C., Kitchener, M., & Saliba, D. (2003). Quality of Care in Nursing Homes: An Analysis of Relationships among Profit, Quality, and Ownership. *Medical Care*, 41(12), 1318–1330.

Paull, G. (2002). Biases in the reporting of labour market dynamics.

Peattie, K. & Morley, A. S. (2008). Social Enterprises: Diversity and Dynamics, Contexts and Contributions.

Png, I. P. L. & Reitman, D. (1995). Why Are Some Products Branded and Others Not? *Journal of Law and Economics*, 38(1), 207–224. ArticleType: research-article / Full publication date: Apr., 1995 / Copyright 1995 The University of Chicago.

Preston, A. E. (1989). The Nonprofit Worker in a For-Profit World. *Journal of Labor Economics*, 7(4), 438–463.

Ramello, G. B. (2006). What's in a Sign ? Trademark Law and Economic Theory. *Journal of Economic Surveys*, 20(4), 547–565.

Rosenbaum, P. R. (2002). Covariance Adjustment in Randomized Experiments and Observational Studies. *Statist. Sci.*, (3), 286–327.

Rud, J. P. (2012). Electricity provision and industrial development: Evidence from India. *Journal of Development Economics*, 97(2), 352–367.

Rugaimukamu, D. (2005). Report from Pilot Survey Study - Time Use Study and Advanced Census Analysis in Tanzania.

Schlesinger, M. & Gray, B. H. (2006). Nonprofit organizations and health care: Some paradoxes of persistent scrutiny. In *The Nonprofit Sector: A Research Handbook*. New Haven & London: Yale University Press.

Schultz, T. P. (2004). School subsidies for the poor: evaluating the Mexican Progresa poverty program. *New Research on Education in Developing Economies*, 74(1), 199–250.

Siegel, D. S. & Vitaliano, D. F. (2007). An Empirical Analysis of the Strategic Use of Corporate Social Responsibility. *Journal of Economics & Management Strategy*, 16(3), 773–792.

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 111–147.

The Wye Group (2007). *Rural Households Livelihood and Well-Being - Statistics on Rural Development and Agriculture Household Income*. Technical report, United Nations.

United Nations Statistics Division (2005). *Guide to Producing Statistics on Time Use: Measuring Paid and Unpaid Work*. Technical Report 93.

van de Walle, D., Ravallion, M., Mendiratta, V., & Koolwal, G. (2013). *Long-term impacts of household electrification in rural India*. Policy Research Working Paper Series 6527, The World Bank.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhy: The Indian Journal of Statistics; Series A*, 26.

Weisbrod, B. (1975). Toward a theory of the voluntary nonprofit sector in a three-sector economy. In E. S. Phelps (Ed.), *Altruism, Morality, and Economic Theory*. New York: Russell Sage Foundation.

World Bank (2012). Gender differences in employment and why they matter. In *World Development Report 2012: Gender Equality and Development*.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298), 348–368.