When Private Firms Provide Public Goods: The Allocation of CSR Spending

Kim Fe Cramer, Lucie Gadenne, and Noémie Pinardon-Touati*

April 2025

Abstract

This paper studies how firms allocate their Corporate Social Responsibility (CSR) expenditures to inform the welfare effects of corporate contributions to public goods. We use a novel dataset covering the quasi-universe of firms' CSR expenditures in India over the period 2015-2019, which includes detailed information on CSR projects. We document key stylized facts on the allocation of CSR spending across social topics (e.g., health, education) and locations. Using natural language processing to measure the technological proximity between firms' production technology and topics, we find that firms spend more on topics in which they have a comparative advantage. This is consistent with an efficient allocation of CSR expenditures across topics and the main rationale for CSR in the literature. Considering allocation across locations, however, we find that firms spend more in areas where social returns are low; CSR spending seems less equitably allocated than government expenditures. Overall, our results suggest that CSR mandates may be an efficient but inequitable way to increase public good provision.

Keywords: Private provision of public goods, Corporate Social Responsibility, textual analysis, India

JEL Codes: D62, D64, L21, H41, M14

---

*Kim Fe Cramer: LSE (k.f.cramer@lse.ac.uk), Lucie Gadenne: Queen Mary University of London, IFS, and CEPR (l.gadenne@qmul.ac.uk), and Noémie Pinardon-Touati: Columbia University (np2842@columbia.edu). We thank Michael Best, Nicolas Bonneton, Michele Fioretti, Maitreesh Ghatak, Moqi Groen-Xu, Jessica Jeffers, Anders Jensen, Cynthia Kinnan, Karthik Muralidharan, Jordan Nickerson, Yanos Zylberberg, as well as seminar audiences and participants at the Zurich Public Finance Conference on Developing Countries, EBRD, City University, Indian School of Business, King's College London, London Junior Finance Workshop, LSE, Oxford University, Paris School of Economics, QMUL, Reichman University, Tinbergen Institute, University of London, WAPFIN, WEFIDEV, WashU Annual Finance Conference, and Young Scholars' Webinar, for helpful comments and suggestions. We also thank Marco Guttierez Chavez, Cécile Delcuvellerie, Aristomenis Chrysafis-Progopoulos, Nithin Mannil, and Bruno Yzeiri for excellent research assistance.

1 Introduction

Firms around the globe spend vast amounts on corporate social responsibility (CSR) activities (Hart and Zingales, 2017; Allcott et al., 2023; Fioretti, 2022; Starks, 2023). In India, the focus of this paper, CSR expenditures represent 0.1% of GDP, similar to the share observed in the US.[1] Governments actively seek to promote the private sector's participation in social causes, whether through tax incentives (in place in most countries, see Pickering et al., 2014) or laws that mandate CSR spending by firms, which are increasingly popular in low-and-middle-income countries.

CSR, defined as the allocation of some profits to social causes, is hard to justify by standard economic principles that suggest contributions to such causes are best done by individuals, not firms (Friedman, 1970). One justification stems from the idea that firms may have a *comparative advantage* in producing public goods relative to the public or non-profit sector (Besley and Ghatak, 2007; Hart and Zingales, 2022). This occurs when the public good is naturally bundled with the production of the private good: a firm producing healthcare products may have a comparative advantage in setting up local health projects, for example. In practice, however, firms could engage in CSR for many reasons that lead them to maximize their private returns, potentially at the expense of social returns. There is however no systematic evidence on how firms allocate their CSR expenditures that could help understand the potential welfare effects of CSR.

This paper seeks to shed light on these effects by studying how Indian firms allocate their CSR expenditures. The Indian context is particularly well suited to this analysis for two main reasons. First, in 2013 India became the first country to mandate that large firms allocate a share of profits to a specified list of social causes. The law imposes a common reporting format for CSR projects, making it possible to consider CSR allocations across firms, social topics, and locations. This enables us to construct the first dataset documenting the CSR activities of the quasi-universe of large firms in any economy: we observe all CSR expenditures of the 6,500 largest Indian firms over the period 2015-2019. The data includes detailed description of the CSR projects (e.g., women's employment training or primary health care centers). Second, India is a large emerging economy facing substantial development challenges with limited tax capacity (Das et al., 2023). Whether the CSR mandate is an effective way to increase revenues allocated to public good provision is thus

[1] In the US, charitable giving by corporations represented 27.36 trillion USD in 2023, just under 0.1% of GDP (The Giving Institute, 2023)

a question of major interest.

We start by documenting key facts about CSR expenditures in India. First, we show that the CSR allocation across social topics (e.g. education, health) is similar to how other public good providers allocate their expenditures. Second, we find that firms specialize in topics, which suggests a potential relationship between firms' technologies and their CSR expenditure choices. Third, CSR expenditures are skewed towards a few states, with one state receiving 30% of the spending. Motivated by these facts, we build a conceptual framework in which firms choose how to allocate their CSR expenditures across project types defined by a topic and a state. Firms differ in the technology they use to provide projects in each type to capture the possibility that firms may have different 'comparative advantages' across types. They also have heterogeneous preferences across types which potentially differ from those of the social planner, allowing for wedges between private and social returns. We contrast the socially and privately optimal allocations to clarify what can be learned by our empirical exercises.

We then consider whether CSR is efficiently allocated across projects by looking at whether firms spend more on CSR projects they have a comparative advantage in. Using Natural Language Processing we construct an index of technological proximity between the firms' for-profit activity and CSR topics. For a description of firms' technologies used in their for-profit activity, we rely on the text contained in the industry classification guidelines. From the textual descriptions of projects in the CSR data, we obtain a large corpus of text describing the activities within each topic. We use word embeddings to obtain a vector representation of both texts and measure the proximity between industries and social topics using the cosine similarity between these vectors.

We find that firms' comparative advantage is correlated with how they allocate CSR across social topics: a one standard deviation increase in the proximity between a firm's industry and a social topic increases the probability that the firm allocates any CSR expenditure to the topic by 9%, and the amount spent by 16%.[2] These results are robust to a wide range of robustness checks and are not driven by a particular topic or industry. Changing the method used to construct the proximity index in particular hardly affects our results, suggesting our baseline method captures the proximity between firms' industries

[2] Our conceptual framework clarifies that what we are interested in is the *correlation* between comparative advantage and CSR allocation, not a causal effect. Whether firms spend more on topics they have a comparative advantage in because of their technology or because they have a high preference for these topics is irrelevant from an allocative efficiency perspective. We however present suggestive evidence that at least part of the correlation is due to firms choosing to leverage their technological proximity to topics.

and social causes well.[3] Our results are thus consistent with the idea that firms use their comparative advantage when deciding how to allocate their CSR spending. This provides support for the key necessary condition established by the theoretical literature for CSR to be welfare-enhancing (Besley and Ghatak, 2007; Hart and Zingales, 2022).

Firms allocating expenditures on public goods may however have implications for equity as well as efficiency. To study the equity characteristics of CSR, we consider the allocation of CSR expenditures across areas. We find that CSR expenditures in a state are positively correlated with that state's level of development. Assuming that public goods have higher social returns in poorer areas, this implies that firms spend in areas where social returns are low, reflecting a wedge between private and social returns. A key mechanism behind this finding is that firms concentrate their CSR spending in the state where they are headquartered, and large firms tend to be headquartered in rich states. However, we find that firms spend more in richer states even when spending in their headquarter state is excluded. Finally, we show that the spatial distribution of CSR spending is regressive not only in an absolute sense but also when compared to the allocation of government expenditures across states.

This paper's first contribution is to the empirical literature on CSR. Most of this literature focuses on the relationship between firms' CSR activities and their financial outcomes: the incidence of CSR on profits and shareholder value, the link between CSR and ownership characteristics, and investors' CSR preferences and their effect on the cost of capital (see Margolis et al., 2007; Gillan et al., 2021; Christensen et al., 2021; Hong and Shore, 2023, for reviews).[4] More recent papers seek to measure firms' and investors' social impact (e.g., Allcott et al., 2023; Kahn et al., 2023; Green and Vallee, 2024), or characterize the diverse stakeholder preferences underpinning firms' prosocial stances (e.g., Flammer and Luo, 2017; Fioretti et al., 2023; Christensen et al., 2023; Colonnelli et al., 2024; Conway and Boxell, 2024). Our paper investigates how information on CSR projects can inform the welfare properties of corporate social expenditures. In this respect, our approach complements that of Fioretti (2022) who uses detailed data on the prosocial actions of one firm to estimate its objective function and show that it spends prosocially beyond profit maxi-

[3]Our baseline specification includes firm- and topic- fixed effects to account for the fact that some social causes (e.g., education) are popular with all firms.

[4]Previous contributions have investigated CSR in the Indian context in particular. In the accounting literature, Manchiraju and Rajgopal (2017); Dharmapala and Khanna (2018); Mukherjee et al. (2018); Bhattacharyya and Rahman (2019) investigate the effect of the CSR mandate on firm value, focusing on listed firms. In the strategy literature, Gatignon and Bode (2023) provide a descriptive analysis of Indian firms' CSR strategies.

mization, thereby increasing welfare. By contrast, we consider all large firms in a context in which the amount of prosocial spending is given and consider whether the *allocation* of CSR spending is consistent with social welfare maximization. Finally, our paper's scope is reminiscent of work on charitable giving by individuals that also describes the universe of giving via administrative data, though this work has so far only documented spending patterns in rich countries (see List, 2011, for a review).[5]

Our second contribution is to test whether firms allocate their CSR spending according to the comparative advantage of their industry. The assumption that this is how firms behave is central to much of the theoretical literature on CSR (see Kitzmueller and Shimshack, 2012, for a review); in most models, it is a necessary condition for CSR to increase welfare - that of shareholders (Hart and Zingales, 2017, 2022) or of society (Besley and Ghatak, 2007; Magill et al., 2015; Broccardo et al., 2022).[6] This paper is, to the best of our knowledge, the first to define and empirically implement a test of this assumption. Seen through the light of this literature, our results imply that firms' CSR activities have the potential to be welfare-improving.

Our third contribution is to the literature on the private provision of public goods, of which CSR expenditures are an example. This literature focuses mostly on private provision via privatization or outsourcing (see Hart et al., 1997; Kotchen, 2006; Behaghel et al., 2014; Mukherjee, 2021; Knutsson and Tyrefors, 2022). It emphasizes tradeoffs between the efficiency gains of relying on firms to provide public goods and potential adverse effects in quality or distributional outcomes. This paper points to a similar efficiency-equity tradeoff for CSR activities: our results on comparative advantage suggest that the firm provision of public goods could be efficient but would lead to less spending in deprived areas than if the government had been in charge of the allocation.

Finally, our results speak to debates regarding how to finance development. A large literature focuses on overcoming tax capacity constraints to raise more resources in low- and middle-income countries (LMICs) (see for example Besley and Persson, 2009; Best

[5]Within this literature Card et al. (2010) find that charitable contributions from individuals increase substantially in areas where a firm's headquarters are located, our results suggest corporate contributions exhibit a similar type of home bias.

[6]Note that this condition is necessary but not sufficient for CSR expenditures to increase social or shareholder welfare - one also needs to assume government under-provision of the public good and, for shareholder welfare, shareholder preferences for being socially responsible. CSR expenditures could also increase shareholder welfare if shareholders look to management to solve their free-riding problem (Morgan and Tumlinson, 2019), common ownership leads shareholders to want to maximize industry profit, not firms' profit, etc. Hart and Zingales (2022) however, argue that such considerations are second-order explanations compared to the comparative advantage motivation.

et al., 2015; Gadenne, 2017; Jensen, 2022; Bergeron et al., 2024). Our results suggest that mandating CSR spending can complement such efforts, and indeed, several LMICs recently implemented CSR mandate laws similar to India (Lin, 2021). The focus of our paper is not to compare the CSR mandate to an increase in taxes on large Indian firms. We show, however, that the mandate was well enforced, with a clear and economically significant increase in CSR expenditures. The public framing of the law as asking firms to contribute to development goals, together with the reporting requirements, may have de facto enforced a transfer of resources from the private sector to public good provision in a context where tax enforcement itself is relatively weak.

The paper is organized as follows. Section 2 describes our context of study and data and provides evidence on the implementation of India's CSR mandate. Section 3 provides key stylized facts regarding the allocation of CSR expenditure in our context and motivates the simple conceptual framework that defines our hypothesis of interest and derives empirical tests in section 4. Section 5 considers the efficiency properties of the allocation of CSR expenditures across topics, whilst section 6 studies the equity characteristics of the allocation across locations.

2 Context and Data

2.1 Corporate Social Responsibility in India

Indian companies have a long-standing tradition of contributing to public goods: as early as 1892, the Tata Group established one of India's first philanthropic trusts. In August 2013, India passed section 135 of the Companies Act into law; it came into effect in April 2014. This law mandates that large firms spend at least 2% of their average profits over the last three years on CSR activities. Large firms are defined as those with profits above INR 50 million, income above INR 10 billion, or a net worth above INR 5 billion in any of the three preceding financial years.[7] These firms represent a large share of the Indian economy, corresponding to approximately 60% of the formal sector activity. The act specifies the activities that qualify as CSR expenditures, clarifies that spending occurring within the 'normal course of business' (e.g., employee welfare) does not qualify, and imposes the formation of a CSR committee with at least one independent director. Importantly for our purposes, it also makes reporting of all CSR activities to the Ministry for Corporate Affairs

[7] These thresholds are not associated with any other requirements in Indian law.

(MCA) compulsory. During our study period (2015-2019), the mandate was enforced on a comply-or-explain basis, and since 2019, fines have been imposed for non-compliance. We return to discussing the effects of the law on CSR expenditures after describing our data.

2.2 Data Sources

Our main data source comes from the compulsory reporting of CSR activities to the MCA. Since fiscal year 2014-2015 (hereafter 2015), all liable firms report on each of their CSR projects. The data is available on the MCA website and contains, for each CSR project, the amount spent on the project, the CSR topic this project belongs to (from a pre-specified list defined in the law), a textual description of the project, and the location (state) in which the project occurs. From the 28 CSR topics specified by the law and available in the CSR data, we group similar topics to obtain the 16 topics considered in our analysis (see Appendix B.1.3 for details). Information on the period during which projects are implemented is available in the data; because projects often span multiple years, we aggregate data across years in what follows. We construct a dummy indicating whether the project was implemented by the firm directly or via a third party (typically an NGO) by flagging observations where project descriptions contain NGO names or tokens indicative of third-party implementation ("in partnership", "donation to", ..., see Appendix B.1.4).

The data contains information on 124,813 projects done by 11,487 firms over the period 2015-2019. To the best of our knowledge, this is the most comprehensive dataset on CSR activities for any country in the world. It is comparable in scope to data on charitable giving compiled from US tax data by the Giving USA Foundation (see List, 2011) but contains more information on project types and, crucially for our purposes, provides the official Corporate Identification Number (CIN) of each firm. In the financial year 2018-2019, the total annual CSR expenditure is 142,315 million INR (2,277 million USD).[8]

We combine CSR data with firm-level accounting data to obtain additional information on firms. We use the Prowess database from the Center for Monitoring the Indian Economy, which includes information from the income statements and balance sheets of all publicly traded firms as well as a large number of private firms. Firms in Prowess represent more than 70% of the economic activity in the formal sector in India and 75% of all corporate taxes collected by the government.[9] From this data, we obtain information on firms' in-

[8]Throughout the paper, we denominate in 2015 INR and apply an exchange rate of 0.016 from INR to USD. Relative to GDP in 2019 expressed in 2015 INR, this spending is 0.084%.

[9]See https://prowessdx.cmie.com

dustries at the 2-digit level, which follows the National Industry Classification (NIC). The NIC is India's official industry classification and its documentation gives us an informative textual industry description that we use when studying the allocation of CSR expenditures across topics. We also obtain information on firms' registration locations from Prowess. Finally, we use information on CSR expenditures reported in balance sheet statements to examine the implementation of the reform, using data from 2007 onward.

We merge the CSR and accounting data at the firm level using firms' CINs available in both datasets. The accounting data does not, by design, include all Indian firms and has better coverage of large firms. Our main analysis sample contains all firms present in both datasets: 6,573 firms in 71 industries. This represents 61% of firms and 91% of CSR expenditures in the CSR data.

Table A.1 shows descriptive statistics of the variables used in our analysis. Appendix Figure A1 plots the distribution of CSR expenditures by industry for the 20 largest industries in our data. The CSR shares follow a distribution similar to that of value-added per industry for India, as expected, given that CSR expenditures are a function of profits. No single industry represents more than 12% of total expenditures.

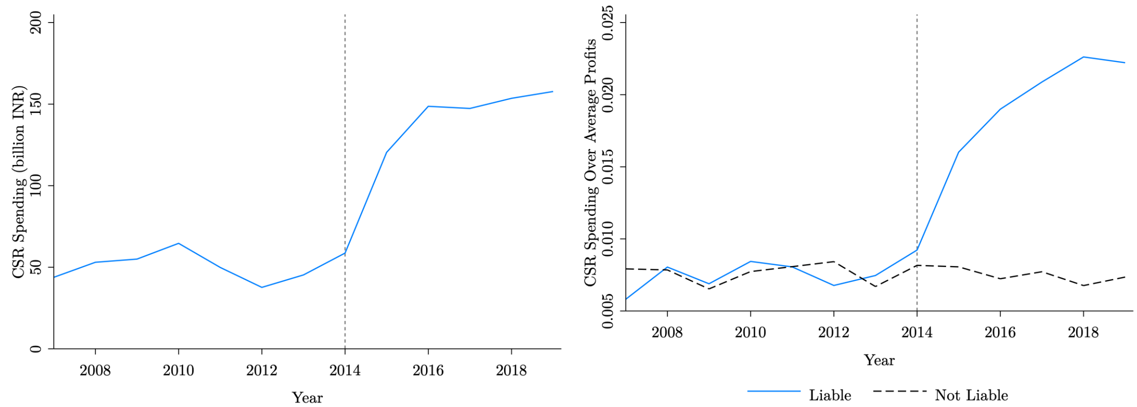
2.3 Implementation of the CSR Mandate

This section briefly describes evidence regarding how well the CSR mandate was implemented. In Figure 1(a), we plot the evolution of total CSR expenditures in India over time, as reported in the accounting data. We see a large increase from 2015 (fiscal year 2014-2015) onward, the year the mandate came into force: aggregate CSR spending roughly tripled since the mandate was implemented.

Figure 1(b) plots the evolution of CSR spending as a share of profits separately among liable firms (defined as firms whose income, profits, or net worth are above the thresholds defined in the law) and all other firms in the accounting data. We see two important takeaways. First, all the aggregate increase in CSR spending seems to come from liable firms - the CSR share of profits remains roughly stable over time among non-liable firms. Appendix C investigates the evolution of CSR expenditures in liable and non-liable firms over the period more formally by conducting a difference-in-differences exercise (see in particular Figure C1). Results suggest the mandate led to an increase in CSR spending as a share of profits of 1 percentage point in the first year of its implementation, and up to nearly 1.5 points at the end of the period. The average profit share of CSR among liable

Figure 1: CSR Spending Over Time

(a) CSR Spending in Aggregate (b) CSR Spending by Firms' Liability Status



Notes: This figure presents CSR spending over time. Data is from Prowess (2007-2019). In Figure 1(a), CSR spending is aggregated over all firms and denominated in 2015 billion INR. Figure 1(b) depicts the CSR spending of a given firm in a given year over average profits in the past three years. The blue line (solid) depicts the mean over firms that are liable under the policy and the black line (dashed) depicts the mean over firms that are not liable.

firms is 2.3% at the end of the period, suggesting the mandate was well respected overall.

Second, Figure 1(b) also shows that some firms are voluntarily spending on CSR prior to the mandate. Among firms in our main sample, 13.6% of firms spend more than 1% of their profits in CSR already in 2014 (see Table A.1). Similarly, some firms are voluntarily spending more than 2% of their profits on CSR during our period of interest: 19.5% of firms in our sample spend more than 2.5% on CSR on average over the period 2015-2019. In what follows, we consider whether our results differ when we consider only firms that voluntarily spend more than what is required by the law, as their spending patterns suggest they may have intrinsic preferences for CSR activities. Does this lead them to allocate their CSR expenditures differently from those that only spend because of the mandate?

3 Key Facts About CSR in India

This section documents four key facts on the allocation of CSR spending by firms in India. This is the first descriptive evidence of the allocation of the quasi-universe of CSR spending in a large country, which is of intrinsic interest given the large amounts spent on CSR. In addition, these facts motivate our analysis of the efficiency-equity trade-off associated with firms deciding on the allocation of public goods.

Fact 1: CSR spending is concentrated on health and education. Table 1 shows the allocation of CSR spending across social topics and the share of each topic in total CSR expenditures. This table also clarifies the meaning of the social topics by listing examples of the most common projects within each category, and Appendix Figures B3-B5 show word clouds for each social topic.[10] We see that firms finance a very wide range of projects.

The largest social topic in terms of spending is education (32% of the total). Common education projects involve school construction or renovation and the promotion of education for differently-abled children or children from underprivileged backgrounds. The second largest topic is health (17% of spending), with projects focused on preventive healthcare, mobile health, or the organization of medical/checkup camps. Infrastructure and environmental sustainability follow, with around 8% of spending each. Infrastructure involves mostly small-scale infrastructure in rural areas (e.g., rural roads, street lights), while for environmental sustainability, tree plantation is the most frequent type of project. The other social topics all receive less than 6% of spending. It is worth noting that, sometimes, several social topics contribute to the same broader cause. For instance, both vocational skills and livelihood enhancement (which together account for 11% of spending) promote income generation. Section 5 below considers the determinants of firms' allocation of CSR across topics.

Fact 2: Firms' allocation across social topics correlates with the allocation of other public good providers. Figure 2 compares the allocation of CSR spending across social topics to the allocation of spending by other key public goods providers: the government and NGOs. To make this comparison feasible, we aggregate several topics together. We leave the details of the mapping between the CSR topics and the spending categories for other agents, as well as the respective data sources, to Appendix B.2.

The allocation of CSR spending across social topics is significantly correlated with that of government spending and NGO activity. In both cases, the pairwise correlation is around 0.8 and statistically significant at the 1% level. Notably, the three types of public goods providers allocate almost precisely the same share to education. Firms differ from the government and NGOs in that they allocate markedly less to vulnerable populations. Meanwhile, firms allocate more of their spending to industry and technology, vocational skills, as well as water and sanitation projects. Overall, Figure 2 suggests that different

[10]Most frequent project types in Table 1 are identified from counts of tokens and bigrams (sequences of two consecutive tokens) for each social topic.

Table 1: Spending Share and Most Frequent Project Types by Social Topic

Social Topic	CSR Share	Most Common Project Types
Education	32%	School construction, promoting education for differently-abled and for underprivileged
Health	17%	Preventive healthcare, medical/checkup camps, mobile health
Infrastructure	8%	Rural road, street lights, village infrastructure (community centers, walls...)
Environmental sustainability	8%	Tree plantation, protection/conservation projects, solar energy
Vocational skills	6%	Vocational training, skill acquisition
Technology incubators	5%	Computer lab, mobile science, scientific research
Livelihood enhancement	5%	Opportunities for differently-abled or underprivileged, support to income generation
Sanitation	5%	Toilet construction, hygiene awareness
Hunger and malnutrition	4%	Food distribution, midday meal scheme
Safe drinking water	2%	Water tank, reverse osmosis plant, water purification
Vulnerable populations	2%	Support to old age, veterans, hostel for widows, orphans
Emergency relief	2%	Disaster relief, Prime Minister Relief Fund
Sports	2%	Rural sport, sport equipment
Women empowerment	1%	Gender equality
Agroforestry	1%	Farmer training programs, organic farming, soil conservation
Animal welfare	0%	Animal protection, cow sheds

Notes: This table displays the share of total CSR spending and the most frequent project types by social topic. The project types are drawn from tabulations of the most common tokens and bigrams by social topic. The full wordclouds by social topic can be found in Appendix Figures B3-B5.

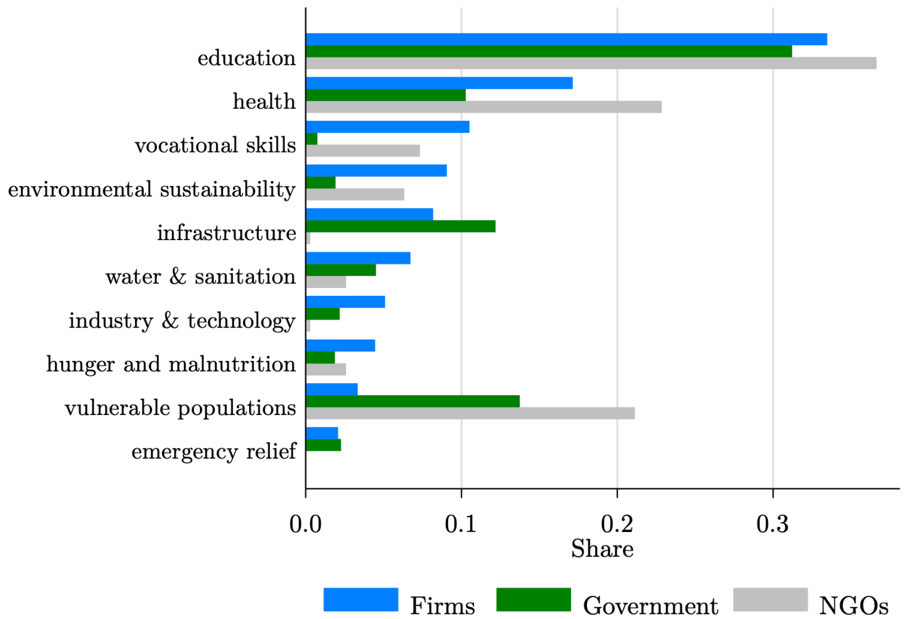
public good providers agree to a large extent on the relative valuation of public goods across social topics.

Fact 3: Firms specialize in social topics. Firms' CSR spending is highly concentrated across social topics. Aggregating all CSR spending by the firm over our sample period, the distribution of spending shares across topics has an average Herfindahl-Hirschman Index equal to 0.63, and 35% of firms allocate more than 90% of their spending to only one topic.

This is not only the result of indivisibilities: even subsetting on the firm×year observations where firms report multiple projects, 20% of firms allocate more than 90% of their spending to only one topic. This specialization suggests a potential link between firms' technology and their choice of CSR spending. This is a key building block of our conceptual framework in Section 4 and we consider whether firms efficiently specialize across social topics in Section 5.

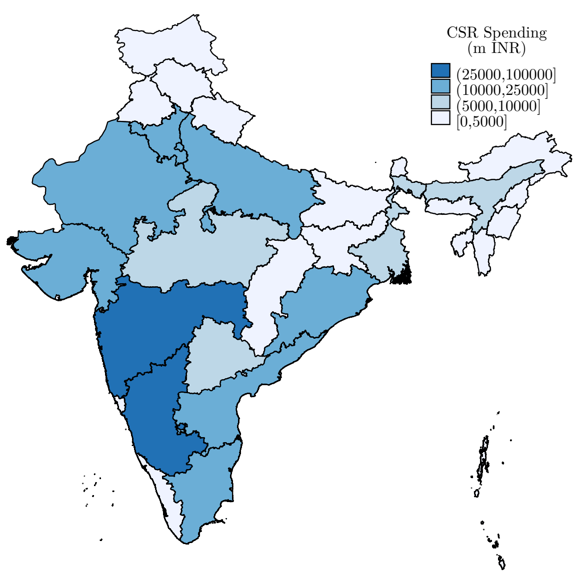
Fact 4: CSR spending is highly concentrated geographically. Figure 3 shows the distribution of CSR spending across states. Almost 30% of CSR spending funds projects in the state of Maharashtra. Six states (Maharashtra, Karnataka, Gujarat, Tamil Nadu, Andhra Pradesh, and Delhi) receive 66% of the spending. This does not simply reflect the distribution of the population: Maharashtra concentrates only 9% of the population (and these six states 34%). This concentration of CSR spending in a few states thus leads to large discrepancies in CSR spending per capita. In Section 6, we explore both the determinants and the implications of the geographical allocation of CSR expenditures.

Figure 2: Allocation Across Social Topics by Public Goods Provider



Notes: This figure displays the share of each social topic in total CSR spending, total government spending, and number of NGOs. See Appendix B.2 for the mapping of CSR social topics to government spending and NGO data. NGO data does not include emergency relief.

Figure 3: Allocation of CSR Spending Across States



Notes: This figure depicts CSR spending aggregated by state, denominated in 2015 million INR.

4 Conceptual Framework

This section provides a simple conceptual framework that compares the private allocation of CSR expenditures by firms to the socially optimal allocation to guide our empirical analysis. We are interested in the allocation across project types, which we define in our empirical analysis as either topics or locations. Firms differ in the type-specific production function they use to produce projects from CSR expenditures. These differences in 'public good technology' enable us to capture the idea in the literature that firms may have a comparative advantage in producing public goods because of the technology they use in their profit-maximizing activities. Firms also differ in their preferences across project types. Social welfare is increasing in the amount of public good provided by each project, with different social returns across project types.

Set-Up. Our object of interest is how firms f allocate an exogenous CSR amount E across a set of project types p. We denote s_{fp} the share that firm f allocates to type p. Shares become projects via production functions that vary at the firm and type level: the amount of project type p produced by firm f is given by:

$$g_{fp} = \alpha_{fp}(s_{fp}E)^{\rho} \tag{1}$$

where $\rho < 1$. The parameter α_{fp} thus captures how efficient firm f is at producing project p. We label α_{fp} f's comparative advantage in providing type p.[11]

Firms obtain utility U from their vector of project types g_{fp}, defined in the following way:

$$U_f = \sum_p \beta_{fp}\alpha_{fp}(s_{fp}E)^{\rho} \tag{2}$$

where the β_{fp} capture firm preferences, so how much firm f values projects of type p.

Social welfare is a function of the projects funded by all firms and is defined as follows:

$$W = \sum_p \gamma_p \sum_f \alpha_{fp}(s_{fp}E)^{\rho} \tag{3}$$

where the γ_p terms capture the social welfare returns to project type p.

[11] Note that here α_{fp} is strictly speaking an absolute and not comparative advantage but we use this terminology in line with the literature, e.g., Besley and Ghatak (2007).

Socially Optimal Allocation. Maximizing social welfare in expression (3) subject to $\sum_p s_{fp} = 1, \forall f$ yields:

$$s^*_{fp} = \frac{(\gamma_p \alpha_{fp})^{1/(1-\rho)}}{\sum_i (\gamma_i \alpha_{fi})^{1/(1-\rho)}} \tag{4}$$

The socially optimal amount firm f allocates to a project type p is increasing in its comparative advantage in this type, α_{fp}, and in the social returns parameter γ_p. We define a CSR allocation as *allocatively efficient* if firms with a higher comparative advantage on a project type spend more on that type. The socially optimal allocation satisfies allocative efficiency.

Privately Optimal Allocation. Each firm maximizes its utility in expression (2) subject to $\sum_p s_{fp} = 1$. This yields:

$$s^P_{fp} = \frac{(\beta_{fp} \alpha_{fp})^{1/(1-\rho)}}{\sum_i (\beta_{fi} \alpha_{fi})^{1/(1-\rho)}} \tag{5}$$

When firms internalize social welfare ($\beta_{fp} = \gamma_p, \forall f, \forall p$) or have no preferences across project types ($\beta_{fp} = \beta_f, \forall f, \forall p$), allocative efficiency holds, and firms spend more on project types they have a comparative advantage in. However, when firms' preferences across types are different from those of the social planner, allocative efficiency may not hold. In particular, if the correlation between firms' preferences across types and their comparative advantage across types is negative and large, allocative efficiency will not hold.

Hypotheses Taken to the Data. In what follows, we start by considering how CSR expenditures are allocated across one dimension of project type, that of social topic indexed by d (section 5). We do not impose any shape on the distribution of social returns across topics and test whether the private allocation is allocatively efficient by considering whether $\frac{\partial s^P_{fd}}{\partial \alpha_{fd}} > 0$. Our object of interest is thus the correlation between firms' comparative advantage across topics (proxied using a method described below) and the share they spend on topics: allocative efficiency holds if this correlation is positive, even if firms' comparative advantage and preferences are positively correlated. We allow for any pattern of aggregate firm preferences towards topics by including topic fixed effects. We consider a potential determinant of these aggregate firm preferences, and particularly whether they are correlated with the government's preferences across topics, as a second step.

We then consider in section 6 how CSR spending is allocated across the other dimension of project type, location (states), indexed by s. There, we use a location's economic development to proxy for the social returns of spending in the location, and test whether $\frac{\partial s_{fs}^{P}}{\partial \gamma_s} > 0$.

5 Allocative Efficiency: Do Firms use Their Comparative Advantage?

This section investigates the CSR allocation across social topics. Motivated by our definition of allocative efficiency above, we consider whether firms spend more on topics for which their production technology gives them a comparative advantage. We start by explaining how we construct a proxy for comparative advantage using the textual proximity between firms' industries and topics' project descriptions. We then explain our empirical strategy and present our results.

5.1 Construction of a Proxy for Comparative Advantage

Testing for allocative efficiency requires knowing which types of CSR projects are naturally bundled with the firm's for-profit production process, giving them a technological advantage in implementing these CSR projects. This is difficult since there exist no data on firm$\times$project-specific CSR productivities. We circumvent this challenge by exploiting the idea that if projects in a given CSR topic require a technology that is close to the firm's for-profit technology, this technological proximity will be reflected in a semantic proximity between descriptions of the CSR social topic and descriptions of the firm's production function.[12] For instance, consider whether pharmaceutical firms are more efficient at undertaking CSR projects in health than financial firms. Our premise is that this would be reflected in a higher semantic proximity between the description of a pharmaceutical firm's production function and the description of health CSR projects than the semantic proximity between the description of a financial firm's production function and the description of these projects.

[12]The approach consisting in using semantic proximity to measure technological proximity has been used by Hoberg and Phillips (2016) to measure the technological distance between all pairs of listed firms in the United States.

We operationalize this insight by defining a measure of technological proximity for each pair of industry i and social topic d. We exploit two corpora of text. For each 2-digit industry i, the Handbook of the National Industrial Classification provides a description of the products and production technologies common to firms in the industry. After cleaning, this text yields an average of 250 informative tokens per industry (standard deviation is 225). The second corpus consists of the description of all CSR projects within a topic d in the CSR data. After cleaning, the average topic contains 23,311 tokens (standard deviation is 30,661). Table B3 shows an example of the full text for one industry i, Figures B3-B5 show word clouds for the project descriptions associated with each social topic d, and Appendix B.3 provides more details on the textual data pre-processing.

We encode and compare these two corpora using word embeddings. Word embeddings are a Natural Language Processing method in which individual words are represented as real-valued vectors in a high-dimensional space. These vectors are meant to capture the meaning of words so that similar words have similar vectors. In addition, an internally consistent geometry on the vector space allows words to be related.[13] We use the word embeddings provided by the pre-trained Word2Vec model released by Google.[14] The model contains 300-dimensional vectors for 3 million words.

We obtain a vector representation of the text describing each industry i and each CSR project p, denoted $\vec{\mathbf{v}}_i$ and $\vec{\mathbf{v}}_p$, respectively (see implementation details in Appendix B.1.2). We measure the textual proximity between each project d and each industry i using the cosine similarity between their vectors. We then define the textual proximity between each topic d and each industry i by averaging across projects p belonging to topic d. Denoting $\mathscr{P}_d$ the set of N_d projects in topic d,

$$\text{Proximity}_{d,i} = \frac{1}{N_d} \sum_{p \in \mathscr{P}_d} \frac{\vec{\mathbf{v}}_p \cdot \vec{\mathbf{v}}_i}{||\vec{\mathbf{v}}_p|| \, ||\vec{\mathbf{v}}_i||} \tag{6}$$

This variable is our proxy for the comparative advantage that firms in an industry have to implement projects in a topic. Higher values indicate higher similarity, but the variable has no cardinal interpretation. We standardize it to have a mean of zero and a standard deviation of one to ease the interpretation of our results.

Figure 4 depicts the distribution of the proximity variable across topics and industries in

[13] Quantifying semantic similarity using word embeddings is superior to using word counts that miss distinct words with similar meaning.

[14] The model can be downloaded at https://code.google.com/archive/p/word2vec/.

a heatmap for all topics and the 16 largest industries in our data. Deeper blue colors indicate higher proximity, and light grey indicates lowest proximity. We see that the distribution of the variable is reasonably intuitive: for example, the health topic has a particularly high proximity with the medical industry, and the civil engineering industry has a high proximity with the sanitation, safe drinking water, infrastructure, and environmental topics - all topics which require some degree of engineering. We also see that some topics/industries have consistently low or high proximity with most industries/topics (see, for example, the 'sport' and 'safe drinking water' topics or the information provision industry). This may reflect true technological patterns, or be due to less desirable characteristics of our textual corpus, such as the recurrence of some non-technical terms in project descriptions. Our regression results below control for topic and firm fixed effects throughout to allow for this possibility, and we consider the robustness of our results to the exclusion of each topic or industry in turn.[15] Appendix Figure A3 plots the distribution of the variable at the firm×topic level, with some examples. One standard deviation in proximity corresponds roughly to the difference in proximity between the topics 'hunger' and 'health' for the medical and pharmaceutical industry.

We test our assumption that technological proximity correlates with semantic proximity by analyzing semantic proximity between each pair of industries (i, i'). Looking at industry pairs has the advantage that we can construct benchmark measures of technological proximity based on features of the input-output matrix. In Appendix Table A.2, we show that industry pairs that sell similar products, use similar inputs, or have strong supplier-customer ties have a high textual proximity. This suggests that our method appropriately captures similarities in production technologies.

5.2 Empirical Strategy

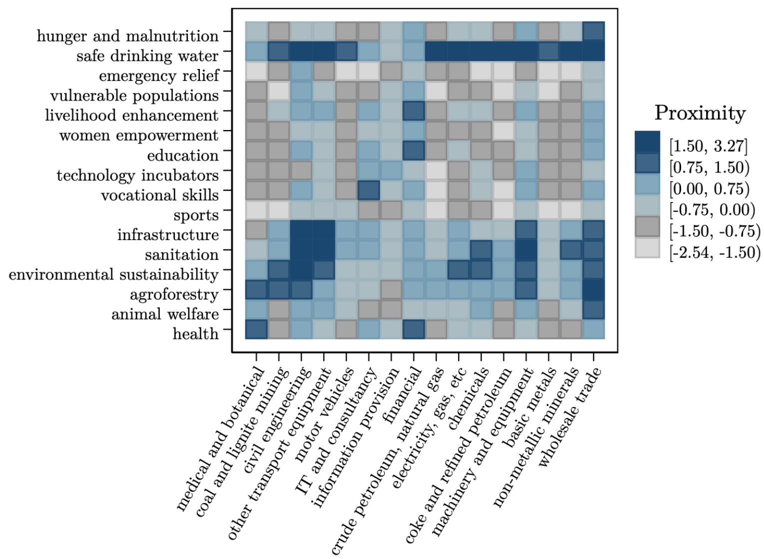
We consider whether firms f's comparative advantage is correlated with how they allocate CSR expenditures across topics d using the following specification:

$$y_{fd} = \beta \text{Proximity}_{i(f)d} + \gamma_f + \gamma_d + \varepsilon_{fd} \tag{7}$$

where y_{fd} is an increasing function of CSR expenditures at the firm f and topic d level, $\text{Proximity}_{i(f)d}$ is our proxy for comparative advantage defined above using the firm's

[15] Appendix Figure A2 plots the heatmap of the distribution of the residuals of the proximity variable after removing firm and topic fixed effects.

Figure 4: Proximity Across Topics and Largest Industries



Notes: The figures depict all topics and the largest 16 industries by total CSR spending. The unit of observation is at the firm-topic level. Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1.

industry $i(f)$, γ_f and γ_d are firm and topic fixed effects, and standard errors are clustered at the topic$\times$industry level.

This specification tests for allocative efficiency, as defined above: $\beta > 0$ would indicate that firms spend more on topics in which they have a comparative advantage in. Note that our definition of allocative efficiency does not require that firms spend more on topics on which they have a comparative advantage *because* they choose to leverage this comparative advantage: allocative efficiency still holds if the allocation is due to firms having, for example, high preferences for topics they have a comparative advantage in (if the α_{fd} and β_{fd} terms in the conceptual framework above are positively correlated). Our object of interest is thus the correlation between CSR expenditures and comparative advantage. Whether our results are driven by firms choosing to leverage their comparative advantage or because they happen to prefer topics in which they have a comparative advantage is nonetheless of intrinsic interest. We return to this in section 5.4 below.

Our first outcome variable is the share of CSR expenditures a firm spends on the topic. We then consider the extensive margin decision by using an indicator for whether the firm spends any amount on the topic, and finally the intensive margin decision using the share of CSR expenditures spent on the topic, conditional on this share being positive. Our preferred specification uses levels of the outcome variables because CSR expenditures are null for many firms×topics. We also consider results using logs of outcome variables when running the specification at the industry×topic level. Our baseline specification gives equal weight to all firms but we also present results obtained by weighing each firm by its total CSR expenditures to consider how textual proximity affects the aggregate CSR allocation. We include firm and topic fixed effects to allow for the fact that some industries and topics have particularly high proximity to all industries or topics, as seen in Figure 4. Topic fixed effects also capture preferences for topics that are shared by all firms. We discuss what could be driving such preferences below.

5.3 Results

Table 2 presents our results: the correlation between our proxy for comparative advantage and how much firms spend on a topic is positive regardless of the outcome variable used. In Panel A we see that a one standard deviation in technological proximity between a firm's industry and a topic increases the share that the firm spends on that topic by one percentage point, a 16% increase relative to the mean. It increases the probability that the firm spends on the topic by two percentage points (9% relative to the mean) and the share spent, conditional on spending a positive amount, by two percentage points (8%). The effects of proximity on CSR expenditure outcomes are larger in Panel B, where we weigh each firm by its total CSR expenditures (the effect on the unconditional spending share is 29%), suggesting larger effects for larger firms.[16]

Figure A4 presents a series of robustness tests. We see that our results are robust to choices made when defining our proximity variable: we obtain similar results when aggregating across projects within an industry differently or using OpenAI's NLP model to construct the proximity variable instead of Word2Vec. Results remain statistically significant when we cluster standard errors at the industry or topic level. Figure A5 shows that results are stable when we exclude each topic, or each of the 20 largest industries, in turn.[17] Finally, Table A.3 presents results obtained by aggregating our data at the industry × topic

[16]The effect on the spending probability is smaller relative to the mean (11%) when we weigh by total CSR spending because larger firms (that by definition spend more on CSR) spend on a larger number of topics.

[17]The only exception is when we exclude the health topic. The coefficient drops slightly but remains statistically significant and indistinguishable from the coefficient obtained using our main specification.

Table 2: Effect of Technological Proximity on CSR Expenditure Shares

	CSR Share Unconditional$_{f,d}$ (1)	Any CSR Spending$_{f,d}$ (2)	CSR Share Conditional$_{f,d}$ (3)
	Panel A: Not Weighted		
Proximity$_{i(f),d}$	0.010***	0.021***	0.021***
	(0.003)	(0.003)	(0.006)
Avg dep var	0.062	0.223	0.280
Firm FE	✓	✓	✓
Topic FE	✓	✓	✓
R-squared	0.24	0.33	0.36
Observations	105,168	105,168	21,684
	Panel B: Weighted by Total CSR Spending		
Proximity$_{i(f),d}$	0.018***	0.046***	0.025***
	(0.005)	(0.008)	(0.009)
Avg dep var	0.062	0.415	0.151
Firm FE	✓	✓	✓
Topic FE	✓	✓	✓
R-squared	0.27	0.37	0.33
Observations	105,168	105,168	21,684

Notes: This table describes the effect of proximity on CSR spending, derived from Equation 7. The unit of observation is at the firm-topic level. The dependent variables are an indicator for any CSR spending by firm (f) in a given topic (d), the share of CSR spending of a firm (f) over topics (d), and the share of CSR spending conditional on any spending. Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. In Panel A, observations are unweighted. In Panel B, observations are weighted by the total CSR spending of each firm, winsorized at the 1st and 99th percentile. Standard errors are clustered at the industry-topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

level, in levels and in logs.[18] Results are similar to those obtained in Panel B of Table 2 regardless of the specification used.

Our results suggest that firms allocate more CSR expenditures to topics in which they have a comparative advantage. How much of the overall allocation across topics does this explain? Given the similarities between the allocation of funds across topics by firms and the government described above, we compare how much technological proximity and government preferences explain aggregate allocation. We regress CSR expenditure share at the firm$\times$topic level on both the proximity variable and the government's expenditure share on the topic in Table A.5.[19] We find that technological proximity has an explanatory power

[18]These results are obtained on a sample excluding the 153 industry$\times$ topic pairs for which there is no CSR spending.

[19]We slightly re-define topics so that government and CSR expenditures are comparable, in line with the

approximately four times lower than the government's spending share: a one standard deviation increase in the government's spending share increases the CSR share by 0.07-0.09, while a one standard deviation increase in technological proximity increases the CSR share by 0.02. Comparative advantage thus plays a role in explaining the overall allocation, but that role is smaller than that played by social preferences across topics, proxied for by the government's allocation.

5.4 Mechanisms and Heterogeneity

Seen through the lens of our conceptual framework, these results suggest the allocation of CSR expenditures across topics is allocatively efficient. But do firms spend more on topics they have a comparative advantage in because of this comparative advantage or because they have an intrinsic preference for spending on those topics?

We fundamentally cannot disentangle firms' preferences from their comparative advantage across topics, but we use information on the mode of implementation of projects to provide suggestive evidence on the question. Intuitively, firms that outsource their project implementation to NGOs are not using their production function to implement the project themselves, so they are not leveraging their comparative advantage for this project. If preferences are the only determinant of firms' allocation across topics, the decision to implement the project themselves or outsource it should be orthogonal to their comparative advantage across topics. If, however, they seek to leverage their comparative advantage for at least some projects, we should see that they are less likely to outsource projects in topics they have more of an advantage in. In Column 1 of Table 3, we see that firms are indeed slightly less likely to outsource projects on topics for which they have a higher comparative advantage: a one standard deviation increase in proximity decreases the probability of outsourcing by roughly 5%. When splitting topics into directly and indirectly implemented projects in Columns 2 and 3, we also see a stronger correlation with proximity for directly implemented projects. Whilst these results must be treated with caution (project implementation mode is itself endogenous to topic proximity), this evidence suggests that at least part of the effect of proximity on allocation across topics is due to firms choosing to leverage their comparative advantage, and not purely due to preferences.

We present additional heterogeneity results in Figure A6. We start by considering whether firms that voluntarily spend on CSR before the mandate or spend more than 2.5%

approach in section 3 above.

Table 3: Direct and Indirect Implementation

		CSR Share Unconditional$_{f,d}$	
	Indirect Implementation$_{p,f,d}$ (1)	Direct Projects (2)	Indirect Projects (3)
Proximity$_{i(f),d}$	-0.008**	0.009***	0.002**
	(0.003)	(0.002)	(0.001)
Avg dep var	0.148	0.054	0.009
Firm FE	✓	✓	✓
Topic FE	✓	✓	✓
R-squared	0.36	0.20	0.10
Observations	74,735	96,992	96,992

Notes: This table describes the effect of proximity based on implementation mode, which is direct or indirect. Column 1 is derived from Equation 8. The unit of observation is on the firm-project level. The dependent variable is an indicator that is one if the project is implemented indirectly. Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. Column 2 and 3 are derived from Equation 9. The unit of observation is on the firm-topic-implementation-level; the data is filled by topic and implementation mode. The dependent variable is the share of CSR spending of a firm (f) over topics (d). In Column 2, the sample is limited to directly implemented projects, and in Column 3, the sample is limited to indirectly implemented projects. In Columns 2 and 3, observations are weighted by the inverse of the number of projects by topic to give equal weight to all topics as in the main proximity regressions. Standard errors are clustered at the topic-industry level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

$$\text{Indirect Implementation}_{p,f,d} = \beta_0 + \text{Proximity}_{i(f),d} + \alpha_f + \alpha_d + \varepsilon_{p,f,d} \tag{8}$$

$$\left(\frac{\text{Spending}_{f,d}}{\text{Spending}_d}\right) = \beta_0 + \beta_1 \text{Proximity}_{i(f),d} + \alpha_f + \alpha_d + \varepsilon_{f,d} \tag{9}$$

of their profits on CSR after the mandate behave differently. These firms likely have more intrinsic motivation for CSR than those that were forced to spend on CSR because of the mandate. Results suggest the correlation with technological proximity is smaller for these firms, perhaps because their strong preferences for some topics (not positively correlated with their comparative advantage) are what leads them to spend on CSR voluntarily. The differences across subsamples are not however statistically significant. We then consider whether firms that could be particularly beholden to some types of stakeholders behave differently, perhaps in response to stakeholder pressure. We find no evidence that firms with different ownership structures (publicly listed firms or firms with one dominant stakeholder), firms in which employees may have more bargaining power (proxied by the labor share, the average wage, or training expenses), or firms that rely more on their reputation with final consumers (proxied by downstreamness or advertising expenses) behave differently. We also see no difference among firms for whom having a good relationship with the government could be more important (either because they operate in heavily regulated industries or because they compete with government-owned firms).

6 Is CSR Allocated Equitably?

In this section we turn to the allocation of CSR across locations (states) to consider how equitable the allocation is. A natural proxy for the potential equity returns to spending on public goods in any particular state is that state's level of economic development. In what follows, we think of an allocation as more equitable the more the correlation between state-level expenditure shares and GDP per capita is negative. As shown above, however, CSR expenditures are concentrated in a few states, with 30% going to just Maharashtra, the richest state (in total GDP) in our data.[20] To consider more generally how equitable the CSR allocation is, we run the following specification at the firm f and state s level:

$$y_{fs} = \beta \text{GSP}_s + \gamma_f + \varepsilon_{f,s} \tag{10}$$

where GSP_s is the state's gross product per capita in logs, γ_f are firm fixed effects and we control throughout for state population.

[20]Maharashtra is also the 6th richest state in our data in per capita terms, out of 29.

Table 4: Effect of State-Level Characteristics and Firm Location on CSR Spending

	CSR Share Unconditional$_{f,s}$		Any CSR Spending$_{f,s}$		CSR Share Conditional$_{f,s}$	
	(1)	(2)	(3)	(4)	(4)	(6)
Log(GDP per 1m People)$_s$	0.100***	0.014***	0.161***	0.059***	0.141***	-0.002
	(0.031)	(0.005)	(0.041)	(0.008)	(0.020)	(0.010)
1(Firm Location State)$_{f,s}$		0.605***		0.721***		0.335***
		(0.028)		(0.024)		(0.024)
Avg dep var	0.030	0.030	0.068	0.068	0.444	0.444
Firm FE	✓	✓	✓	✓	✓	✓
R-squared	0.09	0.51	0.19	0.44	0.42	0.56
Observations	191,143	191,143	191,143	191,143	9,463	9,463

Notes: This table describes the effect of state-level characteristics and firm location on CSR spending. Data is from MCA, Prowess, and RBI. The unit of observation is at the firm-state level. The dependent variables are an indicator for any CSR spending by firm (f) in a given state (s), the share of CSR spending of a firm (f) over states (s), and the share of CSR spending conditional on any spending. The independent variables are the log of state-level GDP per 1 million people and an indicator that equals one if the firm is located in the state as per government records. We control for the log of population in millions. We exclude six states with a population that is lower than one million or missing, as well as the small state of Chandigarh, which does not have government spending data. Observations are weighted by the 2011 population. Standard errors reported in parentheses are clustered by state. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Results are presented in Table 4, Columns 1, 3, and 5. We see that, regardless of the CSR outcome variable we consider, CSR expenditures are positively correlated with state GDP per capita. To consider how this affects the allocation of aggregate CSR spending, Figure 5(a) plots CSR spending as a function of state GDP (both per capita, blue dots) as well as the linear fit of a regression of CSR spending on GDP per capita using states' population as weight. We see that a 10% increase in state GDP per capita increases CSR spending in that state by 18%.

The allocation of CSR across space is thus inequitable insofar as CSR flows more to richer states. Is this more or less inequitable than alternative uses of CSR funds? One simple comparison point is the allocation of government expenditures per capita, as the government could have chosen to tax firms more instead of imposing a CSR mandate. Figure 5(a) also plots state-level government expenditures (restricted to the topics covered by the CSR data) as a function of state GDP (both per capita, green dots). We see that government expenditure is also slightly increasing in state GDP per capita, reflecting the fact that richer states also collect more tax revenues, but the slope with state economic development is much lower - about one fifth of the slope for CSR expenditures. The allocation of CSR across space is thus much more inequitable than the allocation of government expenditures.

This allocation could potentially be explained by a form of firm ‘comparative advantage’, this time across locations. While large firms in India typically operate at the national level, they nevertheless all have roots in one particular location, in which they have their headquarters.[21] This could give them particularly good information on the needs of these locations, making spending in their headquarter locations more efficient than spending elsewhere. Headquarters tend to be in rich states, so this could lead to a positive correlation with state GDP per capita. To consider this, Columns 2, 4, and 6 of Table 4 include as an additional regressor an indicator for whether the firm’s headquarter is in that state. The coefficients for the indicator are very large, reflecting the fact that around 60% of CSR spending occurs in firms’ headquarter states.[22] As expected, the coefficient for state GDP

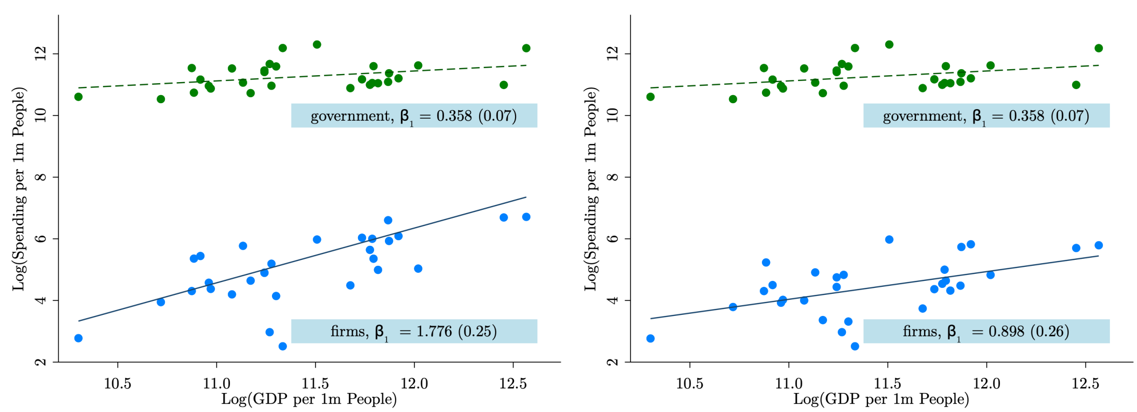
[21] We use registration state to proxy for headquarter presence.

[22] This can be seen visually in Figure A7 which show maps of CSR “creation” and CSR spending by state.

Figure 5: Effect of State-Level GDP on CSR and Government Spending

(a) CSR Spending

(b) CSR Spending Outside Headquarter States



Notes: This figure presents the effect of state-level GDP on firm (CSR) and government spending, derived from Equation 10. The scattered dots indicate state-level observations, blue for firm spending and green for government spending. The lines indicate fitted linear approximations, blue for firm spending (solid) and green for government spending (dashed). Data is from MCA, Prowess, and RBI. The unit of observation is at the state level. The dependent variables are the log of spending (in millions, denominated by 2015 INR) per one million people by firms (CSR) and the government, aggregated from 2015 to 2019. The independent variable is the log of state-level GDP (in millions, denominated by 2015 INR) per one million people in 2013. Population numbers are from the 2011 Population Census. We exclude six states with a population that is lower than one million or missing, as well as the small state of Chandigarh, which does not have government spending data. Observations are weighted by the 2011 population.

per capita falls and even becomes null for the intensive margin of CSR spending (Column 6). In Figure 5(b), we plot CSR spending per state excluding spending in headquarter states and we see that the slope falls by roughly 50%. It remains statistically significant, however, and more than twice as large as that for government expenditures. This suggests efficiency considerations linked to firms' own locations alone cannot explain why firms spend more in richer states.

These is one important caveat to our finding that CSR expenditures are inequitably allocated. Had CSR expenditures not been spent on CSR or taxed, they may have been redistributed to shareholders. We cannot locate shareholders, but they are likely much richer than the average Indian citizen and located in richer states. A counterfactual allocation of CSR expenditures to profits or wages could thus have led to an even more inequitable allocation of these funds across locations.

Finally, in Figure A8, we consider whether CSR expenditures are allocated to locations and sectors that may have higher needs by plotting the correlation between CSR expenditures at the topic and state level and a proxy for the states 'needs' on that topic. For four social topics, we have found a state-level development indicator that is a plausible proxy for need (e.g., for education, we consider the literacy rate). Higher values mean better development outcomes. We see a positive correlation for each of these topics. This confirms that CSR expenditures tend to flow to areas where social returns are relatively low, even when considering topic-specific expenditures.

7 Conclusion

In this paper, we use a novel dataset on the quasi-universe of the CSR expenditures of Indian firms to shed light on the potential welfare effects of CSR. We reach two main conclusions. First, we find evidence consistent with the idea that firms spend more on CSR projects they have a comparative advantage in, i.e., projects they may be particularly good at providing because of the technology they use in their private good production processes. We do so by constructing a proxy for the technological proximity between firms' industries and a CSR topic (e.g., health, education), using the textual proximity between the descriptions of industries and topics. Seen through the lens of the theoretical literature on CSR, this suggests CSR can efficiently contribute to public good provision. However, we find that differences in proximity across industries and topics explain a relatively small share of the aggregate allocation of CSR, suggesting other considerations also loom large in determining firms' CSR strategies. Second, we find that firms spend substantially more on CSR in richer states, in part because they spend more in states where their headquarters are located. Put together, our results suggest that mandating CSR may be an efficient way to increase expenditures on public good provision in our context, but this will come at an equity cost.

References

Allcott, Hunt, Giovanni Montanari, Bora Ozaltun, and Brandon Tan, "An Economic View of Corporate Social Impact," NBER Working Papers 31803, National Bureau of Economic Research, Inc October 2023.

Antràs, Pol, Davin Chor, Thibault Fally, and Russell Hillberry, "Measuring the upstreamness of production and trade flows," *American Economic Review*, 2012, *102* (3), 412–416.

Awasthi, Kshitij, Sai Yayavaram, Rejie George, and Trilochan Sastry, "Classification for regulated industries: A new index," *IIMB Management Review*, 2019, *31* (3), 309–315.

Behaghel, Luc, Bruno Crépon, and Marc Gurgand, "Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment," *American Economic Journal: Applied Economics*, October 2014, *6* (4), 142–74.

Bergeron, Augustin, Gabriel Tourek, and Jonathan L Weigel, "The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the drc," *Econometrica*, 2024, *92* (4), 1163–1193.

Besley, Timothy and Maitreesh Ghatak, "Retailing public goods: The economics of corporate social responsibility," *Journal of public Economics*, 2007, *91* (9), 1645–1663.

_ **and Torsten Persson**, "The Origins of State Capacity: Property Rights, Taxation, and Politics," *The American Economic Review*, 2009, *Vol. 99(4)*, 1218–1244.

Best, Michael, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan," *Journal of Political Economy*, 2015, *123* (6).

Bhattacharyya, Asit and Md Lutfur Rahman, "Mandatory CSR expenditure and firm performance," *Journal of Contemporary Accounting & Economics*, 2019, *15* (3), 100163.

Broccardo, Eleonora, Oliver Hart, and Luigi Zingales, "Exit versus Voice," *Journal of Political Economy*, 2022, *130* (12), 3101–3145.

Card, David, Kevin F. Hallock, and Enrico Moretti, "The geography of giving: The effect of corporate headquarters on local charities," *Journal of Public Economics*, 2010, *94* (3), 222–234.

Christensen, H, Emmanuel T De George, Anthony Joffre, and Daniele Macciocchi, "Consumer responses to the revelation of corporate social irresponsibility," 2023.

Christensen, Hans B, Luzi Hail, and Christian Leuz, "Mandatory CSR and sustainability reporting: Economic analysis and literature review," *Review of accounting studies*, 2021, *26* (3), 1176–1248.

Colonnelli, Emanuele, Niels Joachim Gormsen, and Tim McQuade, "Selfish corporations," *Review of Economic Studies*, 2024, *91* (3), 1498–1536.

Conway, Jacob and Levi Boxell, "Consuming values," *Available at SSRN 4855718*, 2024.

Das, Satadru, Lucie Gadenne, Tushar Nandi, and Ross Warwick, "Does going cashless make you tax-rich? Evidence from India's demonetization experiment," *Journal of Public Economics*, 2023, *224*, 104907.

Dharmapala, Dhammika and Vikramaditya Khanna, "The impact of mandated corporate social responsibility: Evidence from India's Companies Act of 2013," *International Review of law and Economics*, 2018, *56*, 92–104.

Fioretti, Michele, "Caring or Pretending to Care? Social Impact, Firms' Objectives, and Welfare," *Journal of Political Economy*, 2022, *130* (11), 2898–2942.

_, Victor Saint-Jean, and Simon C. Smith, "The Shared Cost of Pursuing Shareholder Value," 2023.

Flammer, Caroline and Jiao Luo, "Corporate social responsibility as an employee governance tool: Evidence from a quasi-experiment," *Strategic Management Journal*, 2017, *38* (2), 163–183.

Friedman, Milton, "The social responsibility of business is to increase its profits," 1970.

Gadenne, Lucie, "Tax Me, but Spend Wisely? Sources of Public Finance and Government Accountability," *American Economic Journal: Applied Economics*, January 2017, *9* (1), 274–314.

Gatignon, Aline and Christiane Bode, "When few give to many and many give to few: Corporate social responsibility strategies under India's legal mandate," *Strategic Management Journal*, 2023, *44* (9), 2099–2127.

Gillan, Stuart L, Andrew Koch, and Laura T Starks, "Firms and social responsibility: A review of ESG and CSR research in corporate finance," *Journal of Corporate Finance*, 2021, *66*, 101889.

Green, Daniel and Boris Vallee, "Measurement and effects of bank exit policies," 2024.

Hart, Oliver and Luigi Zingales, "Companies should maximize shareholder welfare not market value," *ECGI-Finance Working Paper*, 2017, (521).

_ , Andrei Shleifer, and Robert W. Vishny, "The Proper Scope of Government: Theory and an Application to Prisons*," *The Quarterly Journal of Economics*, 11 1997, *112* (4), 1127–1161.

Hart, Oliver D. and Luigi Zingales, "The New Corporate Governance," NBER Working Papers 29975, National Bureau of Economic Research, Inc April 2022.

Hoberg, Gerard and Gordon Phillips, "Text-based network industries and endogenous product differentiation," *Journal of political economy*, 2016, *124* (5), 1423–1465.

Hong, Harrison and Edward Shore, "Corporate Social Responsibility," *Annual Review of Financial Economics*, 11 2023, *15*, 327–350.

Jensen, Anders, "Employment Structure and the Rise of the Modern Tax System," *American Economic Review*, January 2022, *112* (1), 213–234.

Kahn, Matthew E, John Matsusaka, and Chong Shu, "Divestment and Engagement: The Effect of Green Investors on Corporate Carbon Emissions," Working Paper 31791, National Bureau of Economic Research October 2023.

Kitzmueller, Markus and Jay Shimshack, "Economic Perspectives on Corporate Social Responsibility," *Journal of Economic Literature*, March 2012, *50* (1), 51–84.

Knutsson, Daniel and Björn Tyrefors, "The Quality and Efficiency of Public and Private Firms: Evidence from Ambulance Services*," *The Quarterly Journal of Economics*, 02 2022, *137* (4), 2213–2262.

Kotchen, Matthew J., "Green Markets and Private Provision of Public Goods.," *Journal of Political Economy*, 2006, *114* (4), 816 – 834.

Lin, Li-Wei, "Mandatory Corporate Social Responsibility Legislation around the World: Emergent Varieties and National Experiences," *University of Pennsylvania Journal of Business Law*, 2021.

List, John A., "The Market for Charitable Giving," *Journal of Economic Perspectives*, June 2011, *25* (2), 157–80.

Magill, Michael, Martine Quinzii, and Jean-Charles Rochet, "A Theory of the Stakeholder Corporation," *Econometrica*, September 2015, *83* (5), 1685–1725.

Manchiraju, Hariom and Shivaram Rajgopal, "Does corporate social responsibility (CSR) create shareholder value? Evidence from the Indian Companies Act 2013," *Journal of Accounting Research*, 2017, *55* (5), 1257–1300.

Margolis, Joshua, Hillary Elfenbein, and James Walsh, "Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance," 01 2007.

Mikolov, Tomas, "Google code archive - Word2Vec," Jul 2013.

Morgan, John and Justin Tumlinson, "Corporate provision of public goods," *Management Science*, 2019, *65* (10), 4489–4504.

Mukherjee, Abhishek, Ron Bird, and Geeta Duppati, "Mandatory Corporate Social Responsibility: The Indian experience," *Journal of Contemporary Accounting & Economics*, 2018, *14* (3), 254–265.

Mukherjee, Anita, "Impacts of Private Prison Contracting on Inmate Time Served and Recidivism," *American Economic Journal: Economic Policy*, May 2021, *13* (2), 408–38.

Pickering, Adam, Elaine Quick, and Toni Ann Kruse, "Rules to Give By: A Global Philanthropy Legal Environment Index," 2014.

Rios, Anthony and Brandon Lwowski, "An empirical study of the downstream reliability of pre-trained word embeddings," in "Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)" 2020.

Starks, Laura, "Presidential Address: Sustainable Finance and ESG Issues—Value versus Values," *The Journal of Finance*, 2023, *78* (4), 1837–1872.

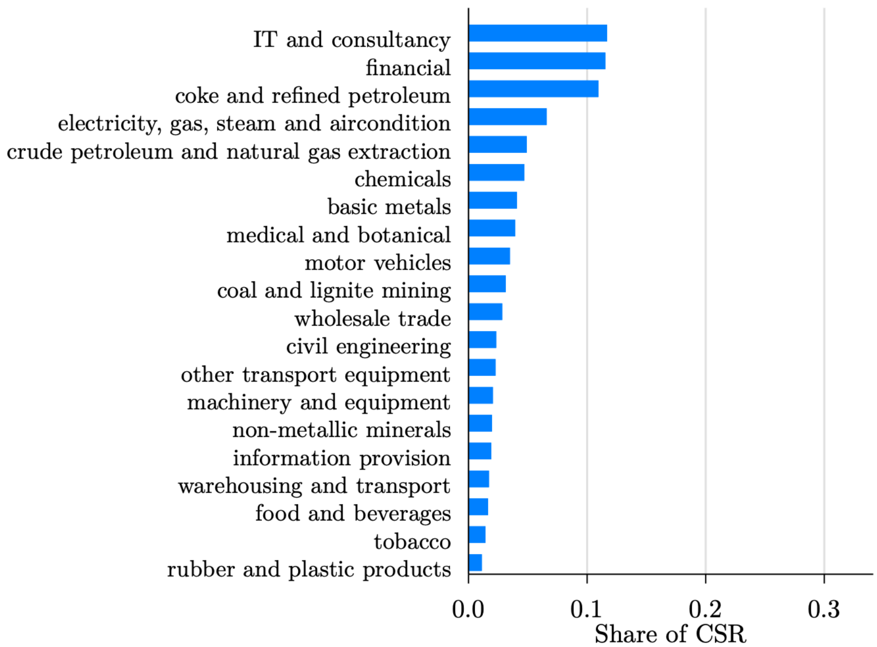
The Giving Institute, *Giving USA: The Annual Report on Philanthropy in America 2023*, Chicago, IL: Giving USA Foundation, 2023. Comprehensive report on charitable giving trends and data in the United States.

Appendices

(for Online Publication Only)

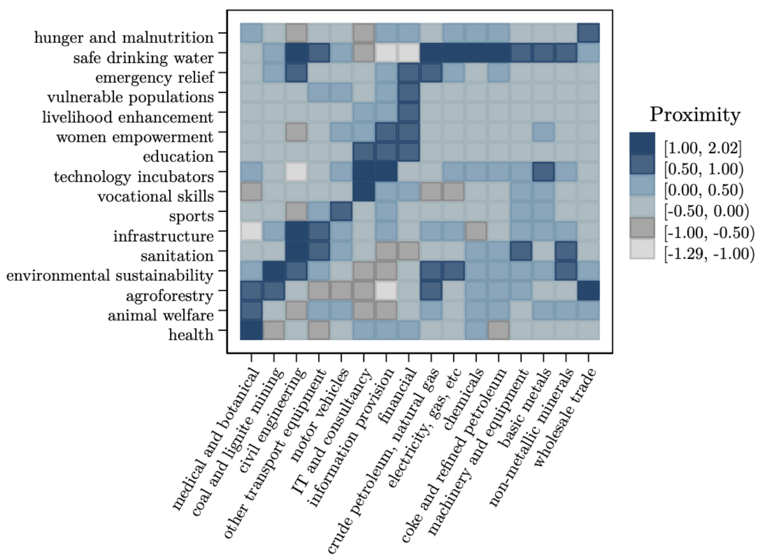
A Additional Tables and Figures

Figure A1: CSR Spending Share by Industry



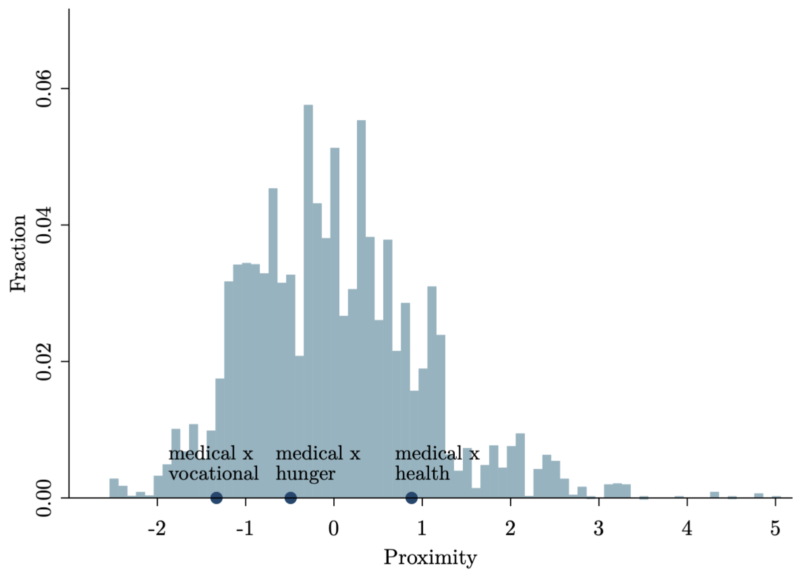
Notes: This figure depicts the share of each industry in total CSR expenditures for the 20 largest industries in the data.

Figure A2: Proximity Across Topics and Largest Industries With Fixed Effects



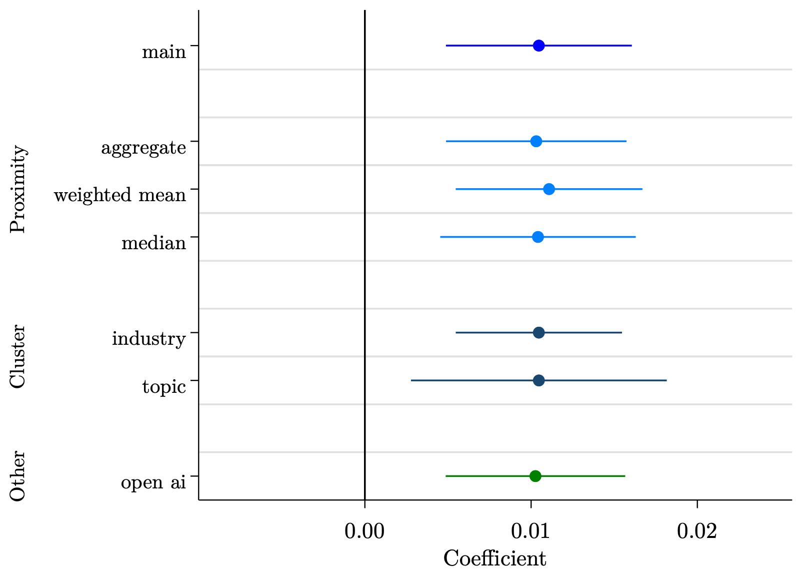
Notes: This figure depicts all topics and the largest 16 industries by total CSR spending. The unit of observation is on the firm-topic level. Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. We residualise the proximity measure on firm and topic fixed effects.

Figure A3: Distribution of Proximity



Notes: This figure depicts the distribution of the proximity variable. Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. The unit of observation is on the industry-topic level. The figure shows three examples, corresponding to approximately one standard deviation below the mean (medical and botanical × vocational skills), the mean (medical and botanical × hunger and malnutrition), and one standard deviation above the mean (medical and botanical × health).

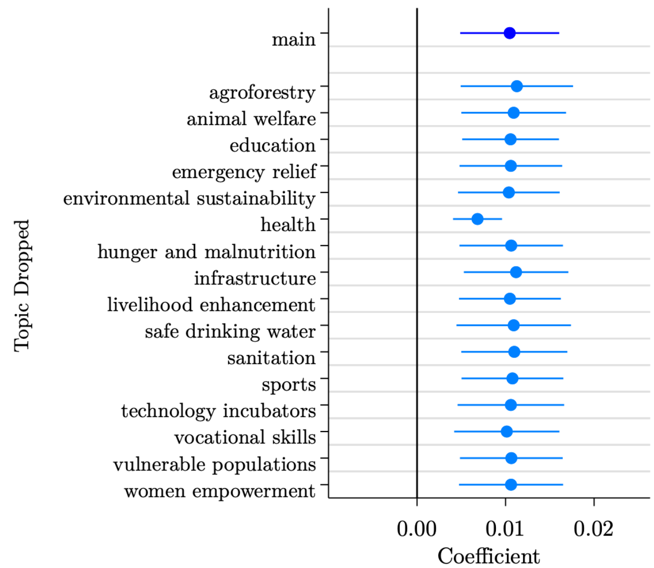
Figure A4: Effect of Proximity on CSR Spending, Robustness



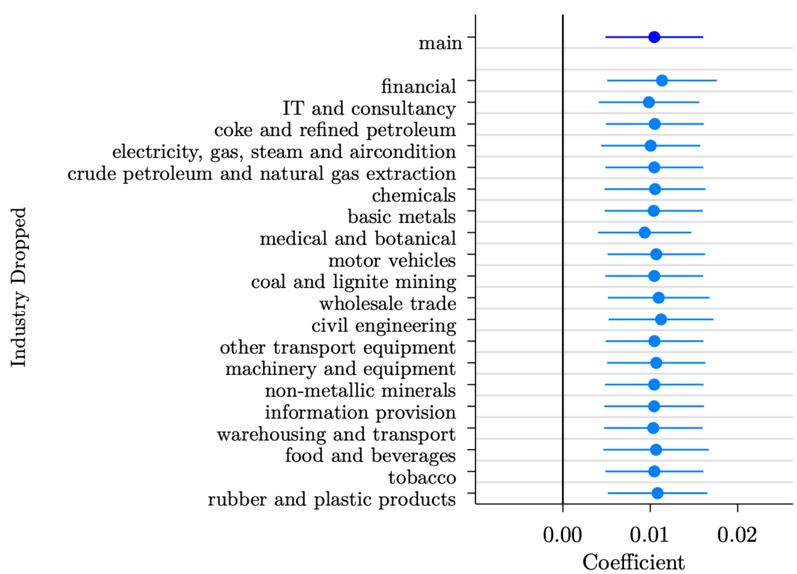
Notes: This figure describes robustness for the effect of proximity on CSR spending, derived from Equation 7. The dependent variable is the share of CSR spending that firm (f) spends on topic (d). Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. Row 1 describes the main specification. Rows 2 to 4 describe different versions of the proximity variable; in the main specification, we calculate the distance between each industry and project, and then take the mean across projects. Row 2 first aggregates the text across all projects and then calculates the distance to industries. Row 3 weights the distance between industry and projects by the number of informative tokens. Row 4 takes the median instead of the mean across projects. Row 5 clusters standard errors on the industry level and Row 6 on the topic level. Row 7 utilizes the proximity variable as in our main specification, but uses the Open AI language model to create it. The figure shows 95% confidence intervals.

Figure A5: Effect of Proximity on CSR Spending, Robustness

(a) Dropping Each Topic in Turn

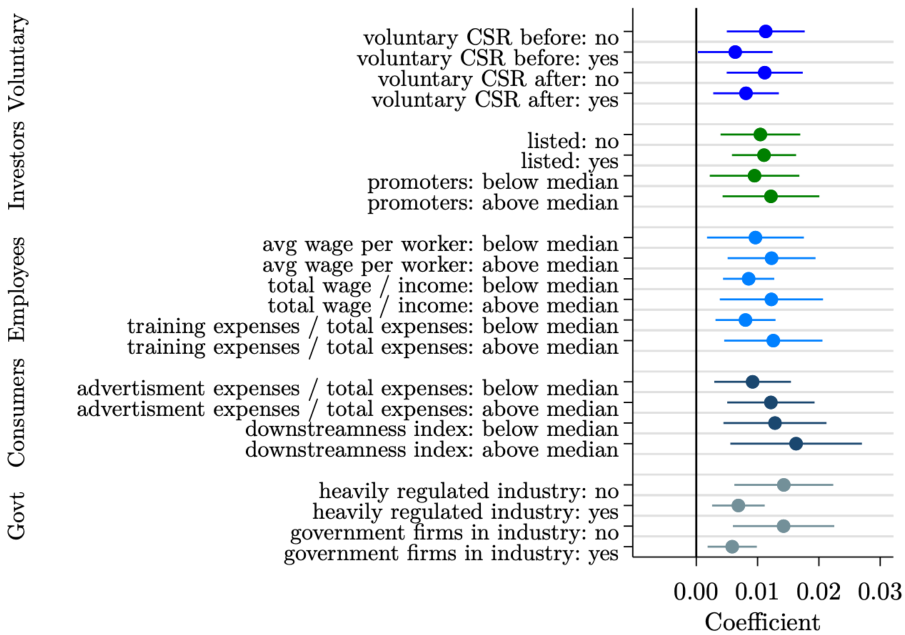


(b) Dropping Each of the Top 20 Industries in Turn



Notes: This figure describes robustness for the effect of proximity on CSR spending, dropping individual topics or industries, derived from Equation 7. The dependent variable is the share of CSR spending that firm (f) spends on topic (d). Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. Figure A5(a) describes robustness to dropping individual topics. Figure A5(b) describes robustness to dropping individual industries. The figure shows 95% confidence intervals.

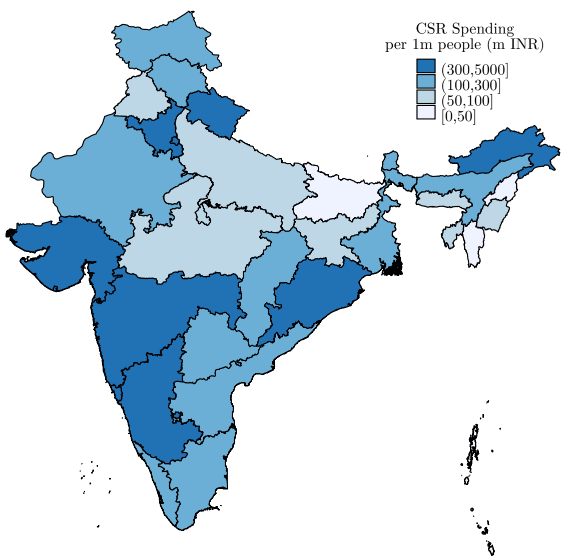
Figure A6: Effect of Proximity on CSR Spending, Heterogeneity



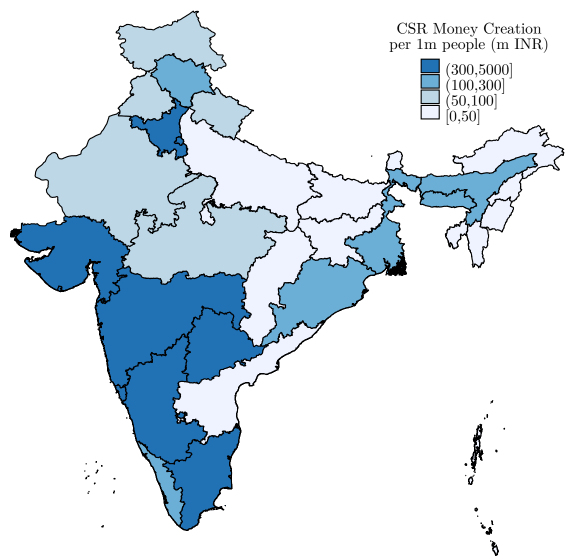
Notes: This figure describes heterogeneity for the effect of proximity on CSR spending, derived from Equation 7. The dependent variable is the share of CSR spending that firm (f) spends on topic (d). Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. In the first group, the sample is split by firms that spend voluntarily and those who do not. Firms are defined as spending voluntarily before the policy if they spend more than 1% of their profits in 2014 on CSR. Firms are defined as spending voluntarily after the policy if they spend more than 2.5% of their profits on average in the years 2015 to 2019. In the second group, the sample is split by how exposed firms are to investors, measured by whether the firms are listed on stock exchanges and whether the equity share of promoters is below or above median. Promoters in the Indian context are investors who own a significant stake in the company and play a key role in its management and decision-making. In the third group, the sample is split by how exposed firms are to employees, measured by the average wage per worker, the total wage bill over total income, and employee training expenses over total expenses. In the fourth group, the sample is split by how exposed firms are to consumers, as measured by advertisement expenses over total expenses, and a downstreamness index of the industry (obtained from Antràs et al. (2012)). In the fifth group, the sample is split by how exposed firms are to the government, as measured by whether the industry is heavily regulated (obtained from Awasthi et al. (2019)) and whether government firms are present in the Prowess sample for that industry. If not otherwise specified, data is obtained from the 2013 Prowess accounting data. The figure shows 95% confidence intervals.

Figure A7: Geographical Distribution of CSR Spending and Creation by State

(a) CSR Spending by State

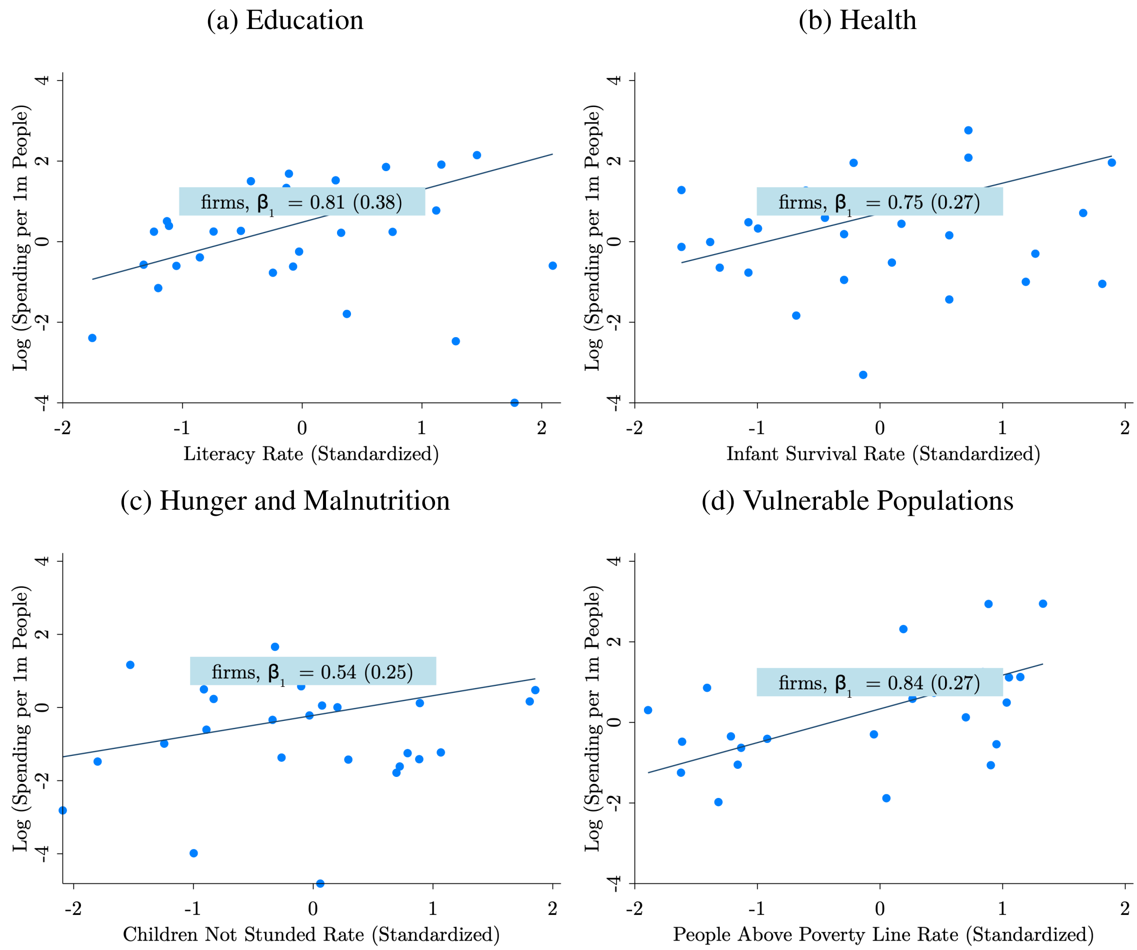


(b) CSR Creation by State



Notes: This figure depicts CSR geographical patterns, aggregated by states. Figure A7(a) depicts CSR spending by state, and Figure A7(b) depicts CSR creation by state, as defined by the firm's official address as per government records. CSR spending and creation is per million people, denominated in 2015 million INR.

Figure A8: Correlation Development Indicators and CSR Topic Spending



Notes: This figure presents the effect of state-level development indicators on firm (CSR) spending on a topic. The scattered dots indicate state-level observations. The lines indicate fitted linear approximations. Data is from MCA, Prowess, and RBI. The unit of observation is at the state level. The dependent variable is the log of CSR spending (in millions, denominated by 2015 INR) per one million people in a given topic, aggregated from 2015 to 2019. In Figure A8(a), the independent variable is the literacy rate in 2011 obtained from the Reserve Bank of India. In Figure A8(b), the independent variable is the survival rate for infants in 2013 obtained from the Reserve Bank of India. In Figure A8(c), the independent variable is the rate of children not stunted in 2016 obtained from the Reserve Bank of India. In Figure A8(d), the independent variable is the rate of people living above the poverty line in 2010 obtained from the Government of India Planning Commission, 2013. Population numbers are from the 2011 Population Census. We exclude six states that have a population that is lower than one million or missing, as well as the small state of Chandigarh, which does not have government spending data. Observations are weighted by the 2011 population.

Table A.1: Summary Statistics

	Mean	SD	Median
Firm-level			
Income (m INR)	15,137	96,366	3,008
Voluntary CSR before policy (yes/no)	0.136	0.342	0.000
Voluntary CSR after policy (yes/no)	0.195	0.397	0.000
Firm-topic-level			
CSR share unconditional (%)	0.062	0.189	0.000
Any CSR spending (yes/no)	0.223	0.416	0.000
CSR share conditional (%)	0.280	0.314	0.142
Firm-state-level			
CSR share unconditional (%)	0.030	0.153	0.000
Any CSR spending (yes/no)	0.068	0.252	0.000
CSR share conditional (%)	0.444	0.402	0.305
Observations			
Unique firms (nr)	105,168		
Unique firm-topics (nr)	6,573		
Unique firm-states (nr)	191,143		

Notes: This table describes the merged MCA and Prowess data (2015 to 2019). Income is calculated as an annual average over the time period, in real terms, denominated in 2015 INR. Voluntary CSR before the policy is an indicator that equals one if the firm spends more than 1% of its profits on CSR in 2014. Voluntary CSR after the policy is an indicator that equals one if the firm spends more than 2.5% of its profits on CSR annually on average in the years 2015 to 2019. CSR share unconditional is the share of CSR expenditure a firm spends on a topic or in a state. Any CSR spending is an indicator that equals one if the firm spends on a topic or state. CSR share conditional is the CSR share conditional on any spending on a topic or state. CSR spending is aggregated over the time period. Variables are not winsorized.

Table A.2: Proximity Across Industries

	Industry Proximity$_{i,i'}$		
	(1)	(2)	(3)
Benchmark$_{i,i'}$	0.45***	0.20***	0.23***
	(0.02)	(0.01)	(0.01)
Benchmark	Leontief	Input	Output
R-squared	0.23	0.13	0.18
Observations	2,025	2,025	2,025

Notes: This table presents correlations between textual proximity between pairs of industries, and other benchmarks that capture proximity across industries. For each pair of industry i,i', Industry proximity$_{ii'}$ is the proximity between industry i and i' as defined by the cosine similarity of their respective embedding vectors. Leontief$_{ii'}$ is the ii' entry of the Input-Output matrix for the Indian economy. Input$_{ii'}$ is the cosine similarity of the input shares of industries i and i'. Output$_{ii'}$ is the cosine similarity of the output shares of industries i and i'.

Table A.3: Effect of Proximity on CSR Spending, Industry Level

	CSR Share Unconditional$_{f,d}$	Log(CSR Share Unconditional)$_{f,d}$
	(1)	(2)
Proximity$_{i(f),d}$	0.020***	0.340***
	(0.006)	(0.063)
Avg dep var	0.072	
Industry FE	✓	✓
Topic FE	✓	✓
R-squared	0.66	0.53
Observations	983	983

Notes: This table describes the effect of proximity on CSR spending, derived from Equation 7, on the industry level. The unit of observation is at the industry-topic level. The dependent variable is the share of CSR spending of a firm (f) over topics (d). Proximity$_{i(f),d}$ is the textual measure of closeness between an industry and a topic defined in Section 5.1. In Column 1, the outcome is in levels. In Column 2, the outcome is log transformed. Standard errors are clustered at the industry-topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.4: Correlation CSR Proximity and Product Proximity

	CSR Proximity$_{f,f'}$	
	(1)	(2)
Product Proximity$_{f,f'}$	0.11***	0.11***
	(0.01)	(0.01)
Firm Clustering		✓
R-squared	0.00	0.00
Observations	316,969	316,969

Notes: This table shows the correlation of the CSR proximity and the product proximity across all pairs of firms. Product proximity$_{ff'}$ is the cosine similarity of the vectors of product shares of firms f and f'. Let $\theta_{fj} = (\text{Sales}_{fj}/\text{Sales}_f)$ be the share of product j in the sales of firm f. Then,

$$\text{Product proximity}_{ff'} = \frac{\sum_j \theta_{fj}\theta_{f'j}}{\sum_j \theta_{fj}^2 \sum_j \theta_{f'j}^2} \tag{A.1}$$

CSR proximity$_{ff'}$ is the cosine similarity of the vectors of CSR shares of firms f and f'. Let $\psi_{fd} = (\text{CSR}_{fd}/\text{CSR}_f)$ be the share of topic d in the CSR of firm f. Then,

$$\text{CSR proximity}_{ff'} = \frac{\sum_d \psi_{fd}\psi_{f'd}}{\sum_d \psi_{fd}^2 \sum_t \psi_{f'd}^2} \tag{A.2}$$

Table A.5: Comparison Explanatory Power of Proximity and Government Spending

	CSR Share Unconditional$_{f,d}$			
Proximity$_{i(f),d}$	0.017***	0.018*		
	(0.005)	(0.010)		
Government Share$_d$			0.093***	0.071***
			(0.004)	(0.006)
Weight	None	CSR spending	None	CSR spending
Var. explained	0.167	0.181	0.928	0.712
Firm FE	✓	✓	✓	✓
Topic FE	✓	✓		
R-squared	0.21	0.25	0.16	0.16
Observations	65,730	65,730	65,730	65,730

Notes: This table compares the explanatory power of technological proximity and government spending shares on the allocation of CSR. Social topics are aggregated at a level consistent between CSR topics and government expenditures, as indicated in Table B2. Variation explained is the standard deviation of the explanatory variable multiplied by the estimated coefficient, divided by the mean of the outcome variable. Standard errors are clustered at the industry-topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

B Data Appendix

B.1 CSR Data

B.1.1 Cleaning of Project Descriptions Data

We execute the following steps to clean the CSR project descriptions:

1. Convert the text to lowercase, removing special characters and numbers
2. Tokenize the text (splitting strings into tokens), lemmatize, and stem the tokens
3. Translate Hindi tokens to English
4. Remove uninformative tokens:
   - Create a list of common uninformative token sequences: it includes common token sequences found in the project descriptions but unrelated to CSR projects, e.g., "CSR overheads", "project not found", "administration expenditures", "detailed in report", etc.
   - Remove tokens flagged as uninformative
5. Flag uninformative observations, defined as satisfying at least one of the following criteria:
   (a) No project description
   (b) Project description equal (or highly similar) to the title of the social topic or groupings of social topics on the CSR portal: this avoids using project descriptions that are just a repetition of the title of the social topic
   (c) "Word salads": project description lists the titles of many different social topics.
   (d) Project description where more than 60% of original bigrams correspond to uninformative token sequences defined above
   (e) Project description Word2Vec embedding is empty (e.g., if the only remaining token following the cleaning is a proper noun)

The raw data contains 238,602 projects. Removing the observations flagged in step 5, our dataset contains 107,996 projects.

B.1.2 Vectorization of the Textual Data

We employ the methodology introduced by Mikolov (2013) to transform all textual tokens into numerical representations suitable for analysis.

This method involves encoding words as numerical vectors, known as word embeddings, which capture semantic meaning based on the context in which they appear. There are two primary ways to generate these embeddings: (1) training them on a custom corpus containing domain-specific documents or (2) utilizing pre-trained embeddings derived from a large, general-purpose corpus that includes the words of interest. We opt for the latter approach because (1) we lacked access to a sufficiently large and diverse collection of documents to train reliable embeddings, and (2) existing research supports the effectiveness of pre-trained word embeddings (Rios and Lwowski, 2020). Specifically, we use the 300-dimensional embeddings from the Word2vec model of Mikolov (2013), which were trained on Google News data encompassing approximately 3 billion words.

After obtaining word embeddings for each token, we aggregate them into a single vector representation for each project description using a weighted average:

$$\vec{\mathbf{v}}_p = \frac{1}{N_p} \sum_{j=1}^{N_p} w_{j,p} * \vec{\mathbf{v}}_{j,p} \tag{B.1}$$

where $\vec{\mathbf{v}}_p$ is the embedding vector associated to project p, $\vec{\mathbf{v}}_{j,p}$ is the embedding vector associated to token j, $w_{j,p}$ is the weight of word j, and N_p is the number of words in project p.

The weights $w_{j,p}$ are determined using the Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF assigns importance to each word based on how frequently it appears in a specific document (Term Frequency) while reducing the weight of commonly occurring words across all documents (Inverse Document Frequency). TF-IDF is widely used in natural language processing and information retrieval. To avoid TF-IDF systematically down-weighting terms frequently appearing in the largest social topics (e.g., "school" in the social topic education), we construct the TF-IDF weights in a corpus that has an equal number of projects for each social topic. To construct this corpus, we use all observations in the topic with the largest number of observations ($N_{max} = 39,213$) and sample N_{max} observations with replacement in all the social topics with a number of observations smaller than N_{max}.

B.1.3 Classification Across Social Topics

The initial data contains 28 different social topics. This initial classification has three issues: (i) some social topics have only a handful of observations so that the project descriptions would be insufficient to reliably estimate proximity with the firms' industries; (ii) some social topics have a large overlap in terms of the vocabulary they use; (iii) some observations are blatantly misclassified (e.g., a project with description "school construction" being classified as sanitation). We proceed as follows.

Step 1: Manually assign the 28 original social topics into 19 aggregated topics. We reclassify all topics with less than 2,000 observations unless there is no sufficiently close topic. This mapping is detailed in Table B1. Figure B1 shows the number of observations in each initial topic, with colors indicating topics grouped together. Figure B2 reports the average pairwise similarity between the initial social topics.

Step 2: Automatic correction for misclassifications. The intuition for the procedure is that we detect an observation as being misclassified if it is significantly more similar to the average description in other social topics than in its own social topic.

1. Compute $\vec{\mathbf{v}}_d$ be the average embedding of projects in social topic d:

$$\vec{\mathbf{v}}_d = \frac{\sum_{p=1}^{N_d} \vec{\mathbf{v}}_{p,d(p)}}{N_d} \tag{B.2}$$

2. For each project p in social topic d, compute:
   - The similarity between p and its own topic $d(p)$: $\text{OwnProximity}_p = \cos(\vec{\mathbf{v}}_p, \vec{\mathbf{v}}_{d(p)})$
   - The largest similarity between p and any topic d': $\text{Max1Proximity}_p = \max_{d'} \cos(\vec{\mathbf{v}}_p, \vec{\mathbf{v}}_{d'})$
   - The topic d' with the maximum similarity: $\text{IsMax}_p = \arg\max_{d'} \cos(\vec{\mathbf{v}}_p, \vec{\mathbf{v}}_{d'})$
   - The second-largest similarity between p and any topic d': $\text{Max2Proximity}_p = \max_{d' \neq \text{IsMax}_p} \cos(\vec{\mathbf{v}}_p, \vec{\mathbf{v}}_{d'})$
3. If $\text{Max1Proximity}_p > \lambda^{miscl} \times \text{OwnProximity}_p$, we say observation p is misclassified
   - If $\text{Max1Proximity}_p > \lambda^{recl} \times \text{Max2Proximity}_p$, we reclassify observation p to topic IsMax_p
   - Otherwise, we say observation p cannot be classified
4. Discard observations that cannot be classified
5. Repeat until the social topic assigned in iteration n is the same as the social topic assigned in iteration $n+1$.

In our implementation, we use $\lambda^{miscl} = \lambda^{recl} = 1.2$.

We impose a number of additional rules manually:

1. Observations in agro-forestry contain projects related to farming as well as projects related to environmental sustainability. Because of the small number of observations in agro-forestry, the algorithm tends to reclassify environmental projects in agro-forestry. We manually assign the projects containing the tokens ["tree plantation", "protection flora fauna", "maintenance flora fauna", "biodiversity protection", "environmental sustainability"] to Environmental Sustainability.
2. Rural development contains a mix of infrastructure projects and of projects corresponding to the other social topics but implemented in rural areas. We reclassify observations in this social topic using $\lambda^{miscl} = \lambda^{recl} = 1.05$ so that observations related to other topics are reclassified, and observations remaining in Rural Development mostly consist of infrastructure projects.
3. Slum area development is small (304 observations) and consists of highly heterogeneous projects. We force the reclassification of projects in this social topic into the closest social topic.
4. Other central government funds consist of heterogeneous projects, so that it does not make sense to define the proximity between firm technologies and this social topic. We force the reclassification of projects in this social topic into the closest social topic.

Step 3: Assigning social topics to observations with missing value. Having obtained a clean definition of each social topic, we attempt to classify observations with a missing social topic. Using a methodology similar to the one described above, we assign an observation p with missing social topic to topic d if p has significantly higher proximity to d than all other topics d'.

Starting from 107,996 observations, we obtain 92,979 observations with a valid social topic assignment.

B.1.4 Definition of Third-Party Implementation

To define third-party implementation, we use two strategies:

- Check for tokens and bigrams indicative of third-party implementation: ["gift", "offering", "grant", "endowment", "input", "participation", "support", "contribution", "donation", "partnership", "ngo", "benefaction", "collaboration", "alliance", "association", "joint venture", "cooperation", "affiliation", "organization", "agency", "implemented by", "bureau", "department", "authority", "office", "establishment", "advisory", "counsel", "guidance", "expert advice", "consulting", "foundation", "in collaboration", "professional services", "fund", "executed by", "carried out by", "enacted by", "put into effect by", "non-profit organization", "operated through", "voluntary organization", "nonprofit group", "together", "jointly", "in partnership", "working together", "donations to"]

- Check for mention of NGOs using a list of the 23 NGOs most frequently present in our dataset

B.1.5 Description of the Final Dataset

Figures B3 to B5 show word clouds of the project descriptions by social topic.

B.2 Other Data Sources

B.2.1 Government Expenditures

We obtain government expenditures data from the Reserve Bank of India for 2010-2021, retrieved on 08th December 2022.[23] We use state level expenditures, this includes expenditures from central government transfers to state and covers most expenditures in India that correspond to the social topics in the CSR data except some food subsidies implemented directly by the central government. We use the 'revised' expenditure variables and consider all regular and capital expenditures labeled 'development expenditures'. We exclude from our analysis expenditures that have no equivalent in the CSR data: energy and transport, tax collection expenditures, interest payments, 'organs of state' (this includes police and judiciary), grants to lower levels of governments (these represent 1% of state expenditures), and other (which includes things like tourism expenditures).

The mapping between government expenditures and CSR social topics is described in Table B2.

[23] https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=State%20Finances%20:%20A%20Study%20of%20Budgets

B.2.2 NGO Expenditures

To compare CSR spending with NGO activity by social topic, we exploit a report prepared by the state of Haryana that collects data on the 150 highest-capability NGOs operating in the state and classifies them by sustainable development goal (which we map to our social topics).[24] The mapping between sustainable development goals and CSR social topics is described in Table B2.

B.3 Construction of the Proximity Variable

The *technological proximity index* is a measure of semantic similarity between the social topics and industries. Two textual components were needed in order to calculate it: (i) the description of the social topics (obtained from the steps described above) and (ii) the industry descriptions.

B.3.1 Textual Data on Industry Descriptions

The industry descriptions come from the National Industry Classification report from 2008. We extracted the descriptions from its "Detailed Structure" and "Explanatory Notes" sections. We manually clean the text for typos and mentions of products/services that should be excluded from each industry. Table B3 shows an example of the full text for one industry i. We obtain the word embeddings associated with the description of each industry i. We clean the text as described in steps 1-4 of section B.1.1. We obtain word embeddings using the same Word2vec model, again applying TF-IDF. For each industry i, we obtain an embeddings vector $\vec{\mathbf{v}}_i$.

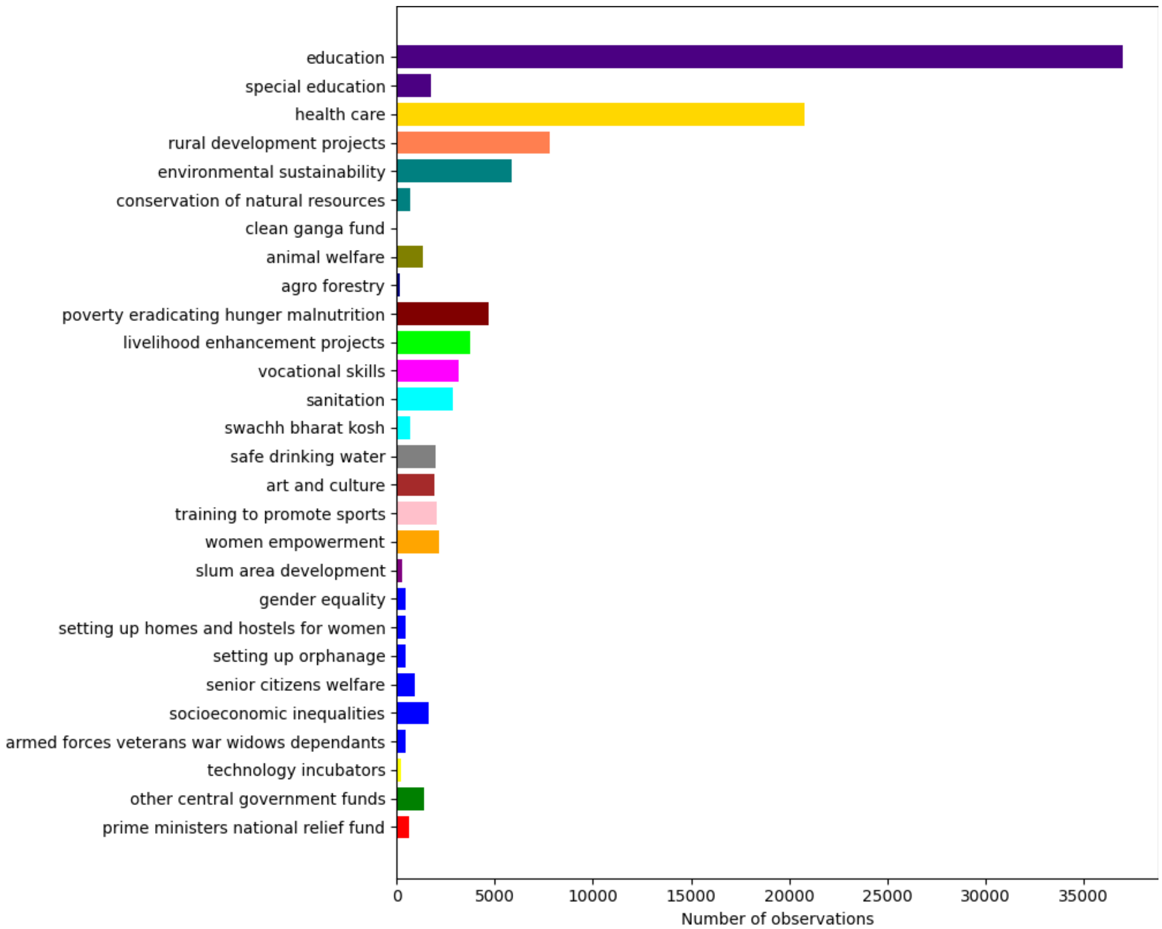
B.3.2 Construction of the Proximity Variable

For each project p, we estimate the cosine similarity of the embedding of project p with the embedding of all industries i:

$$\text{Proximity}_{p,i} = \cos(\vec{\mathbf{v}}_p, \vec{\mathbf{v}}_i) \tag{B.3}$$

[24] Report link.

Figure B1: Number of Observations by Initial Social Topic



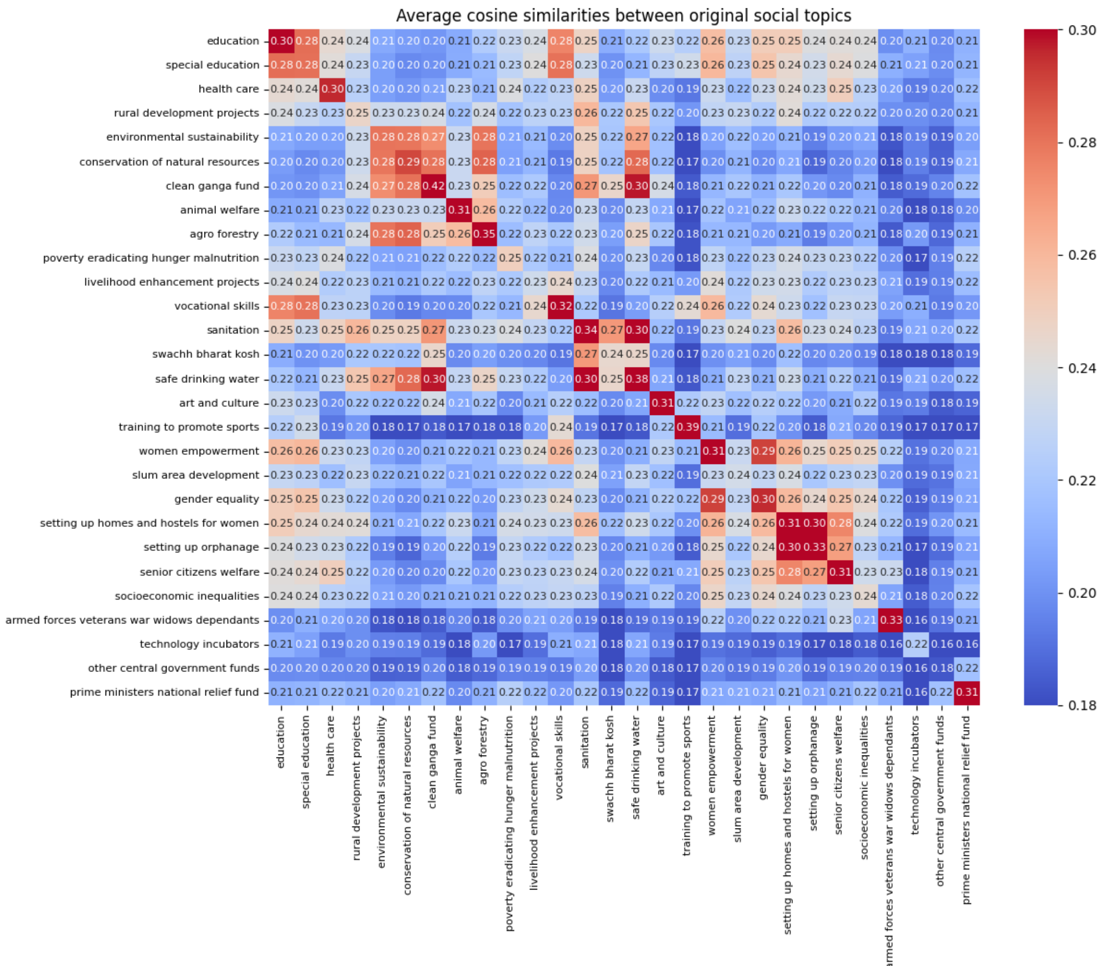
Notes: This figure reports the number of observations by initial social topic. The bar colors reflect the groupings of social topics.

We then obtain the proximity between social topic d and industry i by averaging this metric over projects:

$$\text{Proximity}_{d,i} = \frac{1}{N_d} \sum_{p \in \mathscr{P}_d} \text{Proximity}_{p,i} \tag{B.4}$$

We also construct $\text{Proximity}_{d,i}$ by taking the median or the mean across projects weighted by token count. The correlation coefficient between these different variants exceeds 0.98.

Figure B2: Average Cosine Similarity Between Initial Social Topics

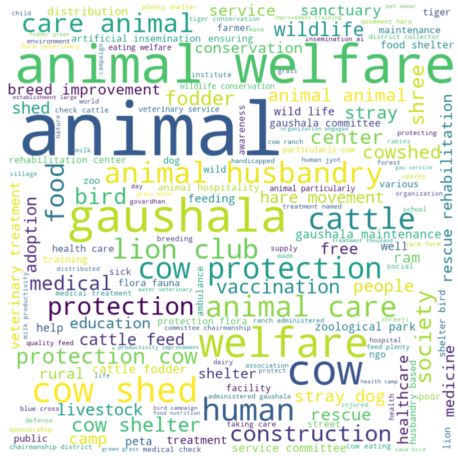


Notes: This figure reports the average cosine similarity of projects for each pair of social topics. The diagonal elements report the average cosine similarity of projects within a social topic.

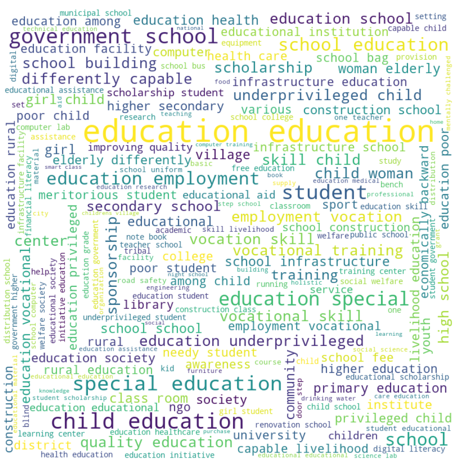
Figure B3: Project Descriptions by Social Topics: Word Clouds

(a) Agro-forestry

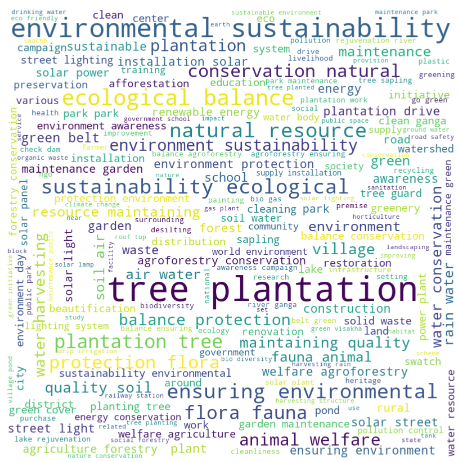
(b) Animal Welfare



(c) Education



(d) Environmental Sustainability



(e) Health Care

(f) Livelihood Enhancement

Notes: This figure shows word clouds of CSR project descriptions by social topic.

Figure B4: Project Descriptions by Social Topics: Word Clouds

(a) Hunger and Malnutrition
(b) Emergency Relief
(c) Infrastructure
(d) Safe Drinking Water
(e) Sanitation
(f) Technology

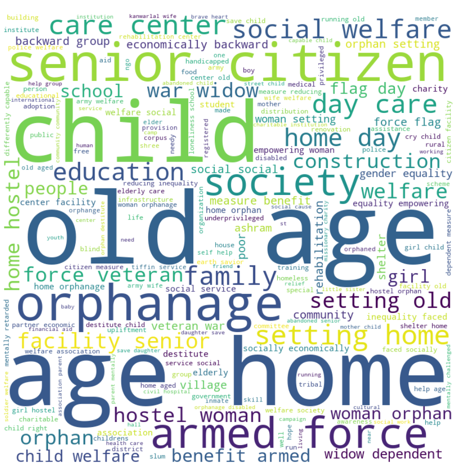
Notes: This figure shows word clouds of CSR project descriptions by social topic.

Figure B5: Project Descriptions by Social Topics: Word Clouds

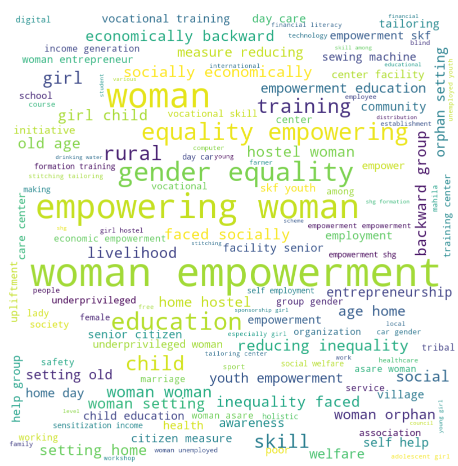
(a) Sports

(b) Vocational Skills

(c) Vulnerable Populations



(d) Women Empowerment



Notes: This figure shows word clouds of CSR project descriptions by social topic.

Table B1: Definition of Social Topics

Initial	Final
Education	Education
Special education	Education
Health care	Health care
Rural development projects	Rural development projects
Environmental sustainability	Environmental sustainability
Conservation of natural resources	Environmental sustainability
Clean Ganga Fund	Environmental sustainability
Animal welfare	Animal welfare
Agro forestry	Agro forestry
Poverty eradicating hunger malnutrition	Poverty eradicating hunger malnutrition
Livelihood enhancement projects	Livelihood enhancement projects
Vocational skills	Vocational skills
Sanitation	Sanitation
Swachh bharat kosh	Sanitation
Safe drinking water	Safe drinking water
Art and culture	Art and culture
Training to promote sports	Training to promote sports
Women empowerment	Women empowerment
Slum area development	Slum area development
Gender equality	Welfare
Setting up homes and hostels for women	Welfare
Armed forces veterans war widows dependents	Welfare
Senior citizens welfare	Welfare
Setting up orphanage	Welfare
Socioeconomic inequalities	Welfare
Technology incubators	Technology incubators
Other central government funds	Other central government funds
Prime Ministers National Relief Fund	Prime Ministers National Relief Fund

Notes: This table reports the mapping between initial social topics and the aggregated topics.

Table B2: Mapping of CSR Topics to Government Spending and SDGs

Category	CSR	Government	Social Development Goals
Education	Education + sports	Education, sports, art & culture	Quality education
Vulnerable populations	Vulnerable populations + women empowerment	Social security & welfare	No poverty + gender equality + reduced inequality
Environmental sustainability	Environmental sustainability + agroforestry + animal welfare	Soil & water conservation + forestry & wild life + irrigation	Responsible consumption & production + climate action + life below water + life on land + (1/3) affordable & clean energy
Health	Health	Medical & public health + family welfare	Good health & well-being
Water & sanitation	Safe water + sanitation	Water supply & sanitation	Clean water & sanitation
Industry & technology	Technology incubators	Science, technology & environment	Industry, innovation & infrastructure + (1/3) affordable & clean energy
Infrastructure	Infrastructure	Rural development	(1/3) Affordable & clean energy
Vocational skills	Vocational skills + livelihood enhancement	Labor & labor welfare	Decent work & economic growth
Hunger & malnutrition	Hunger & malnutrition	Nutrition	Zero hunger
Emergency relief	Emergency relief	Relief on account of natural calamities	NA

Notes: This table reports the mapping of CSR topics to government spending and social development goals.

Table B3: Example of Text Recovered from NIC Handbook (Division 16)

Panel A: Text from Explanatory Notes
16 Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials. This division includes the manufacture of wood products, such as lumber, plywood, veneers, wood containers, wood flooring, wood trusses, and prefabricated wood buildings. The production processes include sawing, planning, shaping, laminating, and assembling of wood products starting from logs that are cut into bolts, or lumber that may then be cut further, or shaped by lathes or other shaping tools. The lumber or other transformed wood shapes may also be subsequently planed or smoothed, and assembled into finished products, such as wood containers. With the exception of sawmilling, this division is subdivided mainly based on the specific products manufactured.
161 Sawmilling and planning of wood.
162 Manufacture of products of wood, cork, straw and plaiting materials. This group includes the manufacture of products of wood, cork, straw or plaiting materials, including basic shapes as well as assembled products.

Panel B: Text from Further Industry Breakdown
Saw milling and planing of wood
  Saw milling and planing of wood
    Sawing and planing of wood
    Manufacture of unassembled wooden flooring including parquet flooring
    Manufacture of wooden railway sleepers
    Activities related to saw milling and planing of wood n.e.c.
Manufacture of products of wood, cork, straw and plaiting materials
  Manufacture of veneer sheets; manufacture of plywood, laminboard, particle board and other panels and board
    Manufacture of ply wood and veneer sheets
    Manufacture of particle board and fibreboard including densified wood
    Manufacture of flush doors and other boards or panels
    Manufacture of other plywood products n.e.c.
  Manufacture of builders' carpentry and joinery
    Manufacture of structural wooden goods [intended to be used primarily in the construction industry such as beams, rafters, roof struts, glue-laminated and metal connected, prefabricated wooden roof trusses, doors, windows, shutters and their frames, whether or not containing metal fittings, stairs, railings, wooden beadings and mouldings, shingles and shakes etc.]
    Manufacture of prefabricated buildings, or elements thereof, predominantly of wood
    Manufacture of builders' carpentry and joinery n.e.c.
  Manufacture of wooden containers
    Manufacture of wooden boxes, barrels, vats, tubs, packing cases etc.
    Manufacture of plywood chests
    Manufacture of market basketry, grain storage bins and similar products made of bamboo or reed
    Manufacture of other wooden containers and products entirely or mainly of cane, rattan, bamboo, willow, fibre, leaves and grass n.e.c.
  Manufacture of other products of wood; manufacture of articles of cork, straw and plaiting materials
    Manufacture of wooden industrial goods
    Manufacture of cork and cork products
    Manufacture of wooden agricultural implements
    Manufacture of various articles made of bamboo, cane and grass
    Manufacture of broomsticks
    Manufacture of articles made of palm leaf, dhak leaf, screw-pine leaf and khajoor leaf; articles of vegetables fibre etc,.
    Manufacture of products of pith and shalapith
    Manufacture of other wood products n.e.c.

C Implementation of the CSR Mandate

This appendix provides more detail on the implementation of the CSR mandate. To estimate the effect of the CSR mandate on CSR expenditures, we compare the evolution of CSR expenditures, as reported in the Prowess data, among liable and non-liable firms before and after the mandate's implementation. We define firms as liable under the Act if they have profits above INR 50 million, income above INR 10 billion, or net worth above INR 5 billion in any of the three preceding financial years, as observed in the Prowess Data. All other firms present in this data constitute the non-liable group. We estimate the following difference-in-differences specification:

$$\left(\frac{\text{CSR}_{f,y}}{\overline{\text{Profit}}^{3y}_{f,y}}\right) = \beta \text{Post}_y \times \text{Treated}_{f,y} + \gamma_y + \gamma_f + \gamma_g + \varepsilon_{f,y} \tag{C.1}$$

where f indexes the firm and y the year, the outcome variable is CSR spending scaled by average profits in the proceeding three years (y-3, y-2, and y-1). $\text{Treated}_{f,y}$ is equal to one if the firm is liable under the CSR regulation in year y, Post_y is a dummy equal to one every year from 2015 onwards, γ_y are year fixed effects and γ_f are firm fixed effects. γ_g are group fixed effects, which indicate the liability status of a given firm in a given year. Standard errors are clustered at the firm level.

We define CSR spending in Prowess as the sum of two variables: social and community expenses and donations. Social and community expenses are expenses incurred by firms for the benefit of society in general. Donations include donations for social causes, religious purposes, or political parties. Both social and community expenses, as well as donations, are reported in the schedules or notes to financial statements of the annual reports under the break-up of expenses or under welfare expenses. In 2015, in alignment with the introduction of the policy, Prowess began to collect explicit CSR data. Since this variable was not available before, we do not utilize it to estimate the policy impact, which requires pre- and post-policy data. As expected, this explicitly collected CSR data closely maps our constructed version. After the policy, in the years in which CSR was explicitly collected, the average annual CSR spending that we construct is 7.14 million INR, and the one explicitly collected 5.51 million INR.

Table C1, Panel A, describes the results. In Column 1, we observe that the share of CSR spending over average profits increases by 1.1% for liable firms relative to non-liable firms after the mandate is implemented. This effect remains stable if we replace the year fixed effects with year $\times$ industry $\times$ state fixed effects in Column 2. Note that the effect is not 2% because non-liable firms also spend on CSR. The CSR spending of non-liable firms before and after the policy is 0.8% on average, while that of liable firms rises from 0.7% to 2.0% on average (see also Figure 1(b) for raw trends).

The key identification assumption is parallel trends. This ensures that pre-existing trends between liable and non-liable firms do not influence the estimate. While this assumption is untestable, Figure C1 documents parallel pre-trends in an event study analysis.

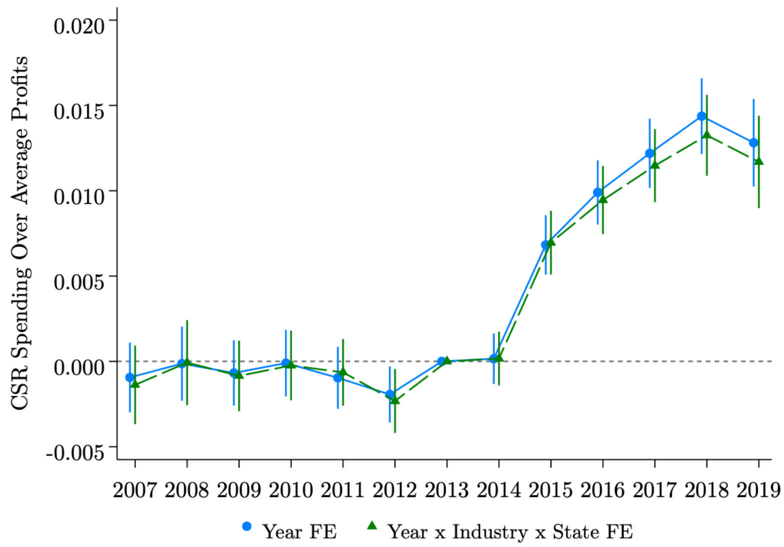
We next discuss manipulation by firms under the policy. Note that in this study, we focus on the allocation of CSR spending across topics and locations. Our design does not rely on firms not manipulating their level of spending. Nonetheless, we briefly discuss the manipulation of the level of spending here. To address manipulation, the government is relying on provisions such as mandatory disclosures, board and CSR committee accountability with an independent director, and audit of accounts for monitoring. In 2019, the government also introduced fines. Firms might also face substantial reputation concerns if they manipulate.

Firms have two possible options to manipulate their level of spending. First, firms might manipulate their accounting variables to change their treatment status. We initially investigate which threshold is most binding: income, net worth, or profit (Figure C2(a) to C2(c)). We observe that for income and net worth, only 6% of firm-years have values higher than the respective threshold. In contrast, 32% of firm-years have values higher than the respective threshold for profits. This suggests that profit is the binding threshold for the majority of firms. Figure C2(d) depicts the distribution of profits before the policy, between 2007 and 2014. Figure C2(e) depicts it after the policy, between 2015 and 2019. Visual inspection shows only minor bunching below the size threshold. Additionally, Figure C2(f) plots the actual and counterfactual profit distribution post-policy, which are visually similar. This evidence suggests that only a small minority of firms manipulated their liability status. To address this dimension of manipulation, we further test a version of the difference-in-difference specification in Equation C.1 in Columns 3 and 4 of Table C1, but instrument the treatment status Treated$_{f,y}$ with a pre-policy variable Treated$_f$. The latter is an indicator equal to one if the firm is liable under the CSR regulation in the year 2014, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years (2011-2013). Results are quantitatively similar under this specification.

Second, firms might manipulate by wrongly relabeling some of their expenditures as

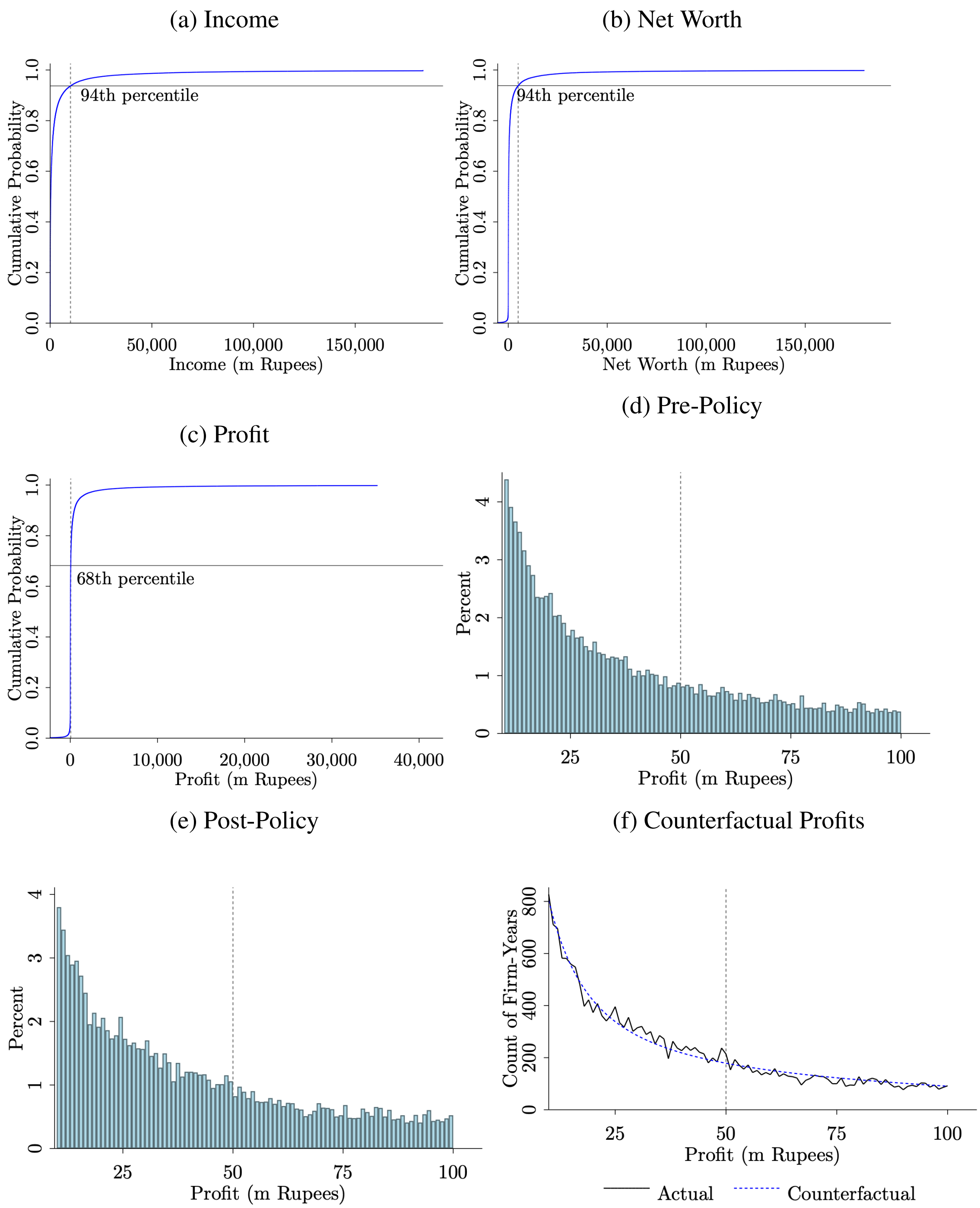
CSR to increase their total CSR spending. The following expenditure categories are excluded from CSR spending: a) activities undertaken in pursuance of the normal course of business of the company, b) contributions to any political party, c) activities benefiting employees, d) activities for deriving marketing benefits for products or services. To test whether firms are relabeling, we run Equation C.1, using as dependent variables expenditures reported in Prowess that firms could plausibly relabel as CSR. Table C2 reports the results. The first column is our CSR variable, the sum of social expenses and donations; the second and third columns present results for each of these categories in turn: we see some substitution away from donations, which may not all have been expenditures that would have counted as CSR in the 2013 law. We see little effect on expenditures on the (work) environment, employee welfare or training, social amenities, or advertisement. There appears to be a decrease in marketing expenses, which could indicate relabeling, but might also be consistent with firms spending less on this dimension because overall the policy has a negative effect on their firm outcomes. Overall, relabeling is very difficult to test and, while illegal, is more likely to affect the spending level than the liability status change.

Figure C1: Effect of the Policy on CSR Spending



Notes: This figure describes the effect of the policy on CSR spending, derived from Equation C.1. Data is from Prowess (2007-2019). The unit of observation is at the firm-year level. The dependent variable is the CSR spending of a given firm (f) in a given year (y) over average profits in the past three years. The independent variables are the interactions of Treated$_{f,y}$ with year indicators. Treated$_{f,y}$ is an indicator equal to one if the firm is liable under the CSR regulation in year y, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years. The green dashed line replaces the year fixed effects in the regression with year $\times$ industry $\times$ state fixed effects. All monetary variables are presented in real terms, denominated in 2015 INR. Variables are winsorized at the 99th percentile. Profits are additionally winsorized at the 1st percentile. Standard errors are clustered at the firm level. The figure shows 95% confidence intervals.

Figure C2: Manipulation of Liability Status



Notes: This figure tests for manipulation of the liability status. Data is from Prowess (2007-2019), on the firm-year level. Figures C2(a) to C2(c) show the cumulative probability for the three size thresholds. Figure C2(d) shows the profit distribution between 2007 and 2014. Figure C2(e) shows the profit distribution between 2015 and 2019. Figure C2(f) plots the actual and counterfactual profit distribution post-policy.

Table C1: Effect of the Policy on CSR Spending

	$CSR_{f,y}/\overline{\text{Profit}}^{3y}_{f,y}$			
	DID (1)	DID (2)	DID-IV (3)	DID-IV (4)
$\text{Treated}_{f,y} \times \text{Post}_y$	0.011*** (0.001)	0.011*** (0.001)		
$\text{Treated}_{f} \times \text{Post}_y$			0.012*** (0.001)	0.012*** (0.001)
Firm FE	✓	✓	✓	✓
Group FE	✓	✓	✓	✓
Year FE	✓		✓	
Year × Industry × State FE		✓		✓
F Statistic			231	174
R-squared	0.31	0.34	0.00	0.00
Observations	197,729	197,729	197,729	197,729

Notes: This table describes the effect of the policy on CSR spending, derived from Equation C.1. Data is from Prowess (2007-2019). The unit of observation is at the firm-year level. In Columns 1 and 2, the dependent variable is the CSR spending of a given firm (f) in a given year (y) over average profits in the past three years. The independent variable is the interaction of $\text{Treated}_{f,y}$ with Post_y. $\text{Treated}_{f,y}$ is an indicator equal to one if the firm is liable under the CSR regulation in year y, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years. Post_y is a dummy equal to one every year from 2015 onwards. In Columns 3 and 4, we instrument the time-varying liability variable $\text{Treated}_{f,y}$ with Treated_f, which is an indicator equal to one if the firm is liable under the CSR regulation in the year 2014, that is, if either profits, income, or net worth, are above their respective thresholds in any of the three preceding financial years (2011-2013). All monetary variables are presented in real terms, denominated in 2015 INR. Variables are winsorized at the 99th percentile. Profits are additionally winsorized at the 1st percentile. Standard errors are clustered at the firm level.

Table C2: Manipulation by Relabeling

	CSR (Social + Donations) (1)	Social (2)	Donations (3)	Environment (4)	Employee Welfare (5)
$\text{Treated}_{f,y} \times \text{Post}_y$	0.44**	0.49***	-0.05**	-0.03	0.51
	(0.19)	(0.18)	(0.03)	(0.03)	(0.44)
Avg dep var	0.18	0.07	0.11	0.02	0.53
Firm FE	✓	✓	✓	✓	✓
Group FE	✓	✓	✓	✓	✓
Year × Ind. × State FE	✓	✓	✓	✓	✓
R-squared	0.30	0.29	0.35	0.31	0.24
Observations	197,729	197,729	197,729	197,729	197,729
	Employee Training (6)	Social Amenities (7)	Advertising (8)	Marketing (9)	
$\text{Treated}_{f,y} \times \text{Post}_y$	0.01*	-0.00	0.05	-0.15***	
	(0.01)	(0.00)	(0.03)	(0.05)	
Avg dep var	0.03	0.00	0.60	1.14	
Firm FE	✓	✓	✓	✓	
Group FE	✓	✓	✓	✓	
Year × Ind. × State FE	✓	✓	✓	✓	
R-squared	0.39	0.18	0.42	0.55	
Observations	197,729	197,729	197,729	197,729	

Notes: This table tests for manipulation by relabeling regular business expenses, derived from Equation C.1. Data is from Prowess (2007-2019). The unit of observation is at the firm-year level. In Column 1, the dependent variable is CSR spending, defined as the sum of social expenses and donations. Columns 2 and 3 split the components of this CSR variable. Columns 4 to 9 have as dependent variables expenses that the firm could have possibly relabeled. The independent variable is the interaction of $\text{Treated}_{f,y}$ with Post_y. $\text{Treated}_{f,y}$ is an indicator equal to one if the firm is liable under the CSR regulation in year y, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years. Post_y is a dummy equal to one every year from 2015 onwards. All monetary variables are presented in real terms, denominated in 2015 INR. Variables are winsorized at the 99th percentile. Profits are additionally winsorized at the 1st percentile. Standard errors are clustered at the firm level.