From Profit to Purpose:
Firms as Private Providers of Public Goods

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

Preliminary – Please do not circulate

This paper studies how firms allocate their Corporate Social Responsibility (CSR) spending to shed light on the potential social effects of corporate contributions to public goods. We do so using a novel dataset covering the universe of the CSR expenditures of Indian firms over the period 2015-2019, which includes detailed information on the projects and social causes firms invest in. Using textual analysis methods, we construct an index of the technological proximity of firms' industries to social causes to capture the extent to which firms use their production processes for CSR projects. Preliminary results suggest that firms do spend more on causes they have a comparative advantage in, in line with the theoretical literature on the desirability of CSR (Besley and Ghatak, 2007; Hart and Zingales, 2017). However, firms tend to spend in geographic areas where social returns are relatively low.

Keywords: Private provision of public goods, CSR, textual analysis, India
JEL Codes: D62, D64, L21, H41, M14

*Cramer is at LSE (k.f.cramer@lse.ac.uk), Lucie Gadenne is at Queen Mary University of London and the Institute for Fiscal Studies (l.gadenne@qmul.ac.uk), and Noémie Pinardon-Touati is at Columbia University (np2842@columbia.edu). We thank Marco Guttierez Chavez, Nithin Mannil and Bruno Yzeiri for excellent research assistance.

Friedman famously argued that firms should focus on maximizing profits, not on social welfare. The higher the profits, the more shareholders can choose to spend on social causes (Friedman, 1970). Despite this, societies around the globe increasingly ask firms to act in a socially responsible fashion. Firms have followed the call, partly through Corporate Social Responsibility (CSR) activities (Hart and Zingales, 2017; Allcott et al., 2023; Fioretti, 2022; Starks, 2023). Fortune 500 companies spend around 20 billion USD annually on CSR (Iglesias, 2022). In India, the focus of this paper, CSR expenditures represent 0.1% of GDP, roughly one-tenth of total government health expenditures.

A first-order question is whether CSR can be welfare improving. Much of the theoretical literature on the subject argues that CSR is welfare-increasing only if firms fund public goods they have a *comparative advantage* in producing - relative to the public or non-profit sectors (Besley and Ghatak, 2007; Hart and Zingales, 2022).[1] 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 (such as marketing strategies or manager preferences) that lead them to maximize their private returns, potentially at the expense of social returns. There is, to the best of our knowledge, no systematic evidence on how firms allocate their CSR expenditures that could help understand the welfare effects of firms' social impact activities.

This paper seeks to shed light on the motivation behind firms' CSR activities by studying how Indian firms allocate their CSR expenditures across different social causes. We use a novel dataset on the CSR expenditures of the 6,500 largest Indian firms over the period 2015-2019. This data includes detailed information on the projects to which CSR expenditures are allocated (e.g., women's employment training or primary health care centers). To test whether firms use their comparative advantage, we construct an index of technological proximity between the firms' industry and 16 social causes using Natural Language Processing techniques. We find that this index explains firms' allocation of CSR expenditures across social causes. We then show that this efficiency gain appears to be accompanied by equity losses, as firms allocate their spending to areas where social returns are relatively low.

The Indian context is particularly well-suited to learn about firms' CSR strategies for several reasons. In 2013, India became the first country to make CSR

[1]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.

spending compulsory for large firms.[2] Together with setting a minimal share of profits that must be spent on CSR and defining a list of social causes applicable, the law specifies stringent reporting requirements. We find that the reform was well enforced, with a clear and economically significant increase in CSR expenditures to the mandated limit when the law came into force. This legal requirement provides us with, to the best of our knowledge, the first dataset documenting the CSR activities of the quasi-universe of large firms in a major economy.[3] It helps us solve two key challenges when studying firms' CSR choices empirically. First, there is no selection into CSR activities - our sample consists of all firms falling under the CSR mandate. Our results are thus relevant for all large Indian firms and not just a highly selected sub-sample with arguably very specific motivation. Second, the reporting requirement means that all firms document their CSR projects in detail following the same format. This makes it possible to compare CSR allocations across firms and topics. Finally, India is a context of major intrinsic interest. It is the world's fifth largest economy and fourth biggest equity market, but its government is facing substantial development challenges with a low tax-to-GDP ratio.[4] Understanding whether firms' social activities can be harnessed to address these challenges is therefore crucial.

To test whether firms use their comparative advantage, we use Natural Language Processing. We construct an index of technological proximity between the firms' for-profit activity and 16 social causes. 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 contained in our CSR data, we obtain a large corpus of text describing firms' CSR activities in 16 social topics. We then use word embeddings to obtain a vector representation of the text and measure the proximity between industries and social causes using the cosine similarity between these vectors.

We find that firms do follow their comparative advantage when choosing social causes: 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 to the topic by 10%, and the amount spent by 17%. 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 regardless of their technological proximity with

[2]Several countries have since followed suit, see Lin (2021).

[3]Previous contributions have investigated CSR in the Indian context. 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.

[4]India's tax-to-GDP ratio of roughly 12% is comparable to other South Asian countries but low relative to countries of similar income levels globally (Bachas et al., 2023).

their production function. These results are robust to a wide range of robustness checks and 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 and social causes well.

Our measure of proximity between industries and social topics could be correlated to other factors determining firms' choice of causes. Firms selling health care products could, for example, choose to finance local health centers more than firms selling textiles, because their consumers are more ethically minded, or their workers more intrinsically motivated. They may be owned by investors or managed by directors who get a warm glow from charitable spending in the health sector. We consider several such alternative motives for CSR allocation. First, we consider whether downstream firms, or firms with a higher than average wage bill, are more likely to spend on popular causes - those that Indian citizens report caring about in surveys. Second, we use the fact that many investors and directors are present in several firms and introduce investor and director fixed effects in our specifications. Third, we consider whether firms in more regulated industries, or relying more on public procurement contracts, spend more on popular causes, perhaps to obtain favors from the government. We find that none of these factors affect our main results: the effect of the proximity index on CSR allocation is robust to controlling for all potential alternative motivations. Interestingly however some of these factors do explain firms' CSR choices, in line with the predictions of a mostly theoretical literature: firms in all industries allocate more resources to causes favored by consumers and investors, particularly if they are consumer-facing or owned by institutional investors.

Our results indicate that firms use their comparative advantage when deciding how to allocate their CSR spending. The theoretical literature suggests this is a necessary condition for CSR to be welfare maximizing, but it is by no means a sufficient one. Having firms in charge of some social spending may have implications for equity as well as efficiency: firms could be choosing the causes they spend on efficiently but doing so in areas where social returns are low. We compare the allocation of CSR expenditures across areas to that of government expenditures - using the latter as a benchmark for society's equity preferences. We find that, whilst both government and CSR expenditures increase with local development levels, firm expenditures are three times more elastic with respect to local GDP than government expenditures. Much of the difference can be explained by the fact that firms tend to spend more in their home states. The existence of this home bias (which could be efficient if firms have local information) indicates that CSR spending could be systematically biased towards richer areas, as these tend to be where firms are head-

quartered.

Our first contribution to the literature 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). 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 results regarding the spatial allocation of CSR point however to a potential equity-efficiency tradeoff of delegating some public good provision to the private sector which this literature has so far not considered.[5]

Our second contribution lies in providing empirical evidence that enables us to distinguish among different conceptions of CSR. Benabou and Tirole (2010) outline different views of CSR in the literature, and point out that empirical evidence to tell them apart is lacking. Our result that firms use their comparative advantage is consistent with what they call CSR as 'delegated philanthropy' - firms using their technology to produce public goods better than shareholders could. Our evidence that consumer-facing firms spend more on causes that consumers value also suggests a role for 'strategic CSR' - a term first coined by Baron (2001) to characterize firms taking a socially responsible stance to increase their long-term profits. Finally, we find that firms with institutional (i.e., more powerful) investors respond more to investors' preferences when choosing their social causes, suggesting that at least some Indian firms are acting in line with Hart and Zingales (2017)'s argument that firms should maximize shareholder welfare as well as market values.

Our third contribution is to the smaller, but growing, empirical literature on CSR. An earlier literature has focused on the incidence of CSR, in particular on profits (see Margolis et al., 2007; Hong and Shore, 2023, for reviews). More recent papers seek to measure firms' and investors' social impact (see for example Allcott et al., 2023; Kahn et al., 2023), and probe the motivations behind firms' prosocial stances (Fioretti et al., 2023). We build on these papers by considering what information on CSR projects can teach us about both firms' motivations and the potential effects of corporate social expenditures. Our approach in particular complements that of Fioretti (2022) in that we leverage new data to ask whether firms act in a way that could maximize their social impact. Fioretti (2022) uses detailed data on

[5]This paper speaks more generally to the large literature on the private provision of public goods (see Cornes and Sandler, 1996, for a review). Kotchen (2006) in particular studies a case where firms can produce private and public goods jointly, and points out the case where firms have a comparative advantage in producing the public good as of particular interest.

all the pro-social activities of a single firm; our data is limited to CSR expenditures but our results speak to the universe of big firms in a large developing economy.

The paper is organized as follows. Section 1 describes our context of study and data and provides evidence on the implementation of India's CSR mandate. Section 2 outlines a simple conceptual framework that defines our hypothesis of interest and derives empirical tests of different conceptions of CSR. Section 3 presents our main results regarding the effect of technological proximity on firms' choice of social causes, and section 4 sheds light on potential equity implications of these choices.

I. Institutional Setting and Data

A. CSR mandate

India's companies have a long-standing tradition of contributing to public goods. In 1892, the Tata Group established one of India's first philanthropic trusts. Since then, a large number of firms have followed the example, supporting communities with philanthropic efforts. On 29th of August, 2013, India became the first country to legislate on CSR. Section 135 of the Companies Act 2013 by the Ministry of Corporate Affairs (MCA) stipulates that firms above a certain size must form a CSR committee with at least one independent director, formulate a CSR policy, and spend at least 2% of their average profits over the last three years on CSR. Schedule VII of the Act lists the activities that qualify for CSR expenditures. CSR spent within the firm in pursuance of its normal course of business, e.g., employee welfare, does not qualify as CSR. Like most provisions of the law, Section 135 came into effect on April 1st, 2014. Since the Indian fiscal year runs from March to March, the law is applicable from FY 2014-15 onwards. In our notation, year t is the fiscal year ending in March of year t. For instance, the year 2015 is the fiscal year 2014-15.

The CSR mandate applies to firms with profits above INR 50 million, income above INR 10 billion, or a net worth above INR 5 billion in any of the three immediately preceding financial years. This captures a large share of the Indian economy, corresponding to approximately 60% of the formal sector activity. Importantly, these particular thresholds apply only to Section 135 and are not associated with any other requirements in Indian company law. The CSR spending requirement applies to a comply-or-explain basis; if firms do not spend 2% of their profits on CSR, the CSR committee is responsible for explaining why the firm could not reach the target in a report presented at general shareholder meetings. In alignment with firms complying with the policy, we establish an economically strong effect on CSR spending of the policy.

B. Data

B1. CSR project data

Crucially for this study, the policy mandated CSR reporting to the MCA. Since 2015, each firm liable under the policy needs to submit a report with the amount of CSR spent project by project. The data additionally contains the textual description of each project, the name of the associated social topic, and its location (at the district level). The MCA made the data available on their website, and we web-scraped it. Table 1 displays a typical firm-year-project observation in our data. Utilizing the reporting mandate allows us to create a country-wide universe of CSR activities. To our knowledge, this is the most comprehensive dataset on CSR activities that exists. The data contains 11,486 firms, 34,043 firm-years, and 124,812 firm-year-projects. It spans five years (2015 to 2019), 16 social topics, and 35 states. This dataset is the primary input for our empirical analysis.

B2. Firm accounting data

Firm-level accounting data comes from the Prowess database compiled by the Centre for Monitoring the Indian Economy (CMIE). It includes all publicly traded firms as well as a large number of private firms. The dataset contains information from the income statements and balance sheets of companies comprising more than 70% of the economic activity in the formal sector in India and 75% of all corporate taxes collected by the government. It is thus representative of large and medium-sized Indian firms. We consider the data from 2007 to 2019, which contains 40,967 firms and 271,684 firm-year observations. These firms belong to 76 different 2-digit industries in 35 states. We utilize the Prowess data for two purposes. First, we merge it with the MCA CSR project data to empirically test whether firms use their comparative advantage. This test requires information on the firms' industries, which is not available in the MCA data. Second, we utilize Prowess to test for the relevance of the CSR mandate in Section I.C since it provides pre- and post-policy CSR spending.

B3. Merging

We merge the MCA data with Prowess on two levels: firm-level and firm-year-level. The primary reason why we do not only match on firm-year-level is that the industry classification for a given firm does not change over time, which means that we can recover the industry for a firm in the project data, even if we do not have that given year for a given firm in Prowess. The merge is done based on firm's Corporate

Identification Number (CIN).[6] Table A1 shows the quality of the merge. For the firm-level match, we match 61% (7,063) of firms and 65% of firm-years in the project data. We match better for large firms. As a consequence, we match 93% of the average profit over the past three years and 91% of CSR spending in the project data. For the firm-year-level match, we observe a very similar pattern. We match 55% (6,305) of firms and 54% of firm-years as well as 88% of the average profit of the past three years and 86% of the CSR spending. Finally, we take a series of data-cleaning steps, which are described in Appendix B (not yet provided). To conclude, our merged sample captures a large share of firm profits and CSR spending reported to the MCA.

B4. Descriptive statistics

Table 2 shows descriptive statistics for the merged data. Since our main analysis is on the firm $\times$ topic level, we report summary statistics on the firm level, topic level, and industry topic level. In our sample, we have 6,573 firms. The average firm reports a profit of INR 5 billion as a basis for the CSR policy between 2015 and 2019. Correspondingly, the mandated expenditure is INR 87 million (1.7%). The actual CSR spending during the years is slightly below the prescribed one with INR 78 million (1.6%). These 6,573 firms are in 71 industries. Figure 1 shows the top 20 industries by share of CSR contributed. Since CSR spending is closely tied to profits, these are the largest industries in the country by profit measures. The 6,573 firms spend on 16 social topics. The average firm has five projects on 2.5 unique topics per year. Figure 2 depicts all topics by share of CSR allocated. The two largest categories are education and health, with 32% and 17%, respectively. 8% is spent on infrastructure and environmental sustainability. Categories falling in the range of 1%-6% are vocational skills, technology incubators, livelihood enhancement, sanitation, hunger and malnutrition, safe drinking water, vulnerable populations, emergency relief, sports, and women empowerment. Finally, less than 1% is spent on agroforestry and animal welfare. To also capture zero spending on topics, we fill our data on the firm $\times$ topic level. This results in 105,168 firm-topic observations. Finally, Figure 3 depicts the geographical distribution of CSR spending across the country. We observe a concentration in states in the southwest and a lack of spending in the northeast.

[6] In rare cases, a given firm can have multiple CINs, for instance if it changed its state of registration. We augment the merge by CIN by a matching procedure based on subcomponents of the CIN and fuzzy string matching based on firm's name, address of registration, and contact email address.

C. Relevance of CSR mandate

C1. Raw trends

To investigate the relevance of the CSR mandate, we examine raw CSR trends as well as the causal effect of the reform on firm-level CSR spending. We utilize the Prowess data, which, in contrast to the MCA data, provides pre- and post-policy data on CSR spending. Table A2 describes summary statistics of the Prowess data, Discussion A1 the construction of the CSR spending variable. Following the CSR policy, we define liable firms as firms with 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. Table A4 shows the characteristics of treated and control firms in 2014. Liable firms tend to be larger than non-liable firms. As expected, they have higher profits, income, net worth, and CSR spending.

Figure 4 demonstrates how aggregate CSR developed over time. While there was already CSR spending before the policy, we can see a clear and substantial upward trend subsequent to the policy. The total CSR spending almost tripled between 2014 and 2019, rising from INR 59 billion (USD 0.97 billion) to INR 158 billion (USD 2.27 billion). This corresponds to approximately 0.1% of the Indian GDP and 2% of government development spending in 2019. Next, we investigate raw trends separately for liable and non-liable firms (Figure 5). We observe a substantial increase in CSR activity for liable firms. Between 2014 and 2019, CSR spending for liable firms multiplies by a factor of 2.4. CSR spending for non-liable firms, in contrast, has remained stable over the years. Thus, both the aggregate raw trends and the raw trends split by treatment status provide evidence that the CSR reform stimulated an economically meaningful increase in CSR spending.

C2. Difference-in-difference

To identify the causal effect of the policy, we utilize a difference-in-differences (DiD) design in Equation 1. The outcome is CSR spending in year t of firm f, scaled by average profits in the proceeding three years (t-3, t-2, and t-1). $\text{Treated}_{f,t}$ is equal to one if the firm is liable under the CSR regulation in year t, that is, if either profits, income, or net worth, are above their respective thresholds in any of the three preceding financial years. Post_t is a dummy equal to one every year from 2015 onwards. Standard errors are clustered on the firm level. This specification compares firms above and below the thresholds after and before the implementation

of the reform. The coefficient β_1 identifies the causal effect of the policy.

$$\left(\frac{\text{CSR}_{f,t}}{\overline{\text{Profit}}^{3y}_{f,t}}\right) = \beta_0 + \beta_1 \text{Post}_t \times \text{Treated}_{f,t} + \alpha_f + \alpha_t + \varepsilon_{f,t} \tag{1}$$

Figure 6 presents the yearly coefficients from this regression, with three different fixed effect structures. We show results with year fixed effects only, year-industry-state fixed effects only, and year-industry-state with firm fixed effects. The omitted year is 2013. Table 3 summarizes these results. We observe that post-policy, the share of profit used for CSR increases by 1.1%-1.3% for liable firms relative to non-liable firms. This is a substantial increase. However, it is different from the 2% mandated by the policy. This can have three potential reasons. First, the mandate works on a comply-or-explain basis, which leaves firms with a certain degree of freedom in their compliance decision. Second, among liable firms, some were already spending a positive share of their profits on CSR before the reform, as suggested by Figure 5. Third, the reform may also have affected norms on CSR for non-liable firms, even though the raw trends in Figure 4 suggest this effect is limited. To summarize, both trends in the raw data and causal estimates suggest that the CSR policy induced an economically meaningful increase in CSR for liable firms.

Finally, we explore two types of manipulation (see Discussions A2 and A3). We find that bunching around the profit threshold exists but is minimal. Additionally, results are robust to utilizing treatment definitions based on accounting variables from 2013, further suggesting minimal influence of sorting behavior. Second, we investigate the manipulation of CSR reporting. We do not find any evidence of re-labelling. Firms are not less likely to report spending in categories close to CSR, such as expenses related to the environment, employee welfare, employee training, or social amenities. In 2019, 45% of liable firms spent more than 2% of their average profits over the past year on CSR, 29% spent more than 2.2%, and 23% more than 2.4%. This provides suggestive evidence that some firms indeed not only target the 2% policy threshold but prioritize the impact of their CSR projects (A5).

II. Conceptual Framework

This paper studies whether the corporate provision of public goods is socially desirable. The traditional view among economists is that it is not: the most famous statement of this traditional view is Friedman (1970), who argues that private corporations should focus on maximizing profits, while governments should deal with

public goods and externalities. At the core of this argument lies the idea that CSR can always be replicated by individual donations by shareholders or customers so that there is no reason for public goods contributions to be made within the firm.

The main robust theoretical rationale for CSR occurs when the production of the public good is naturally bundled with the production of the private good (Besley and Ghatak (2007), Morgan and Tumlinson (2019), Hart and Zingales (2017)). For instance, it may be less costly for a lumber producer to implement a reforestation program than it would be for a government or an environmental NGO. In this case, firms have a comparative advantage in the provision of the public good. Through this channel, CSR may lead to efficiency gains. Section III tests for this channel.

The counterargument is that there may be a wedge between the CSR allocation that is privately optimal for the firm, and the socially optimal allocation of CSR funds. For instance, firms may have an incentive to undertake project that have a large advertising value for their customers, but may not yield the highest social welfare gains. Section IV investigates such potential wedges.

III. Do Firms Use Their Comparative Advantage?

A. Empirical strategy

To test whether firms use their comparative advantage, we utilize the merged sample of MCA projects and Prowess firm-level account information on the firm $\times$ topic level described in Table 2. Our main regression of interest is Equation 2. First, we investigate on the extensive margin whether firm f allocates any funds to social topic s. Second, on the intensive margin, we examine the share of total CSR spending that firm f allocates to the social topic s. The main explanatory variable is Proximity, describing the technological proximity of an industry i and a social topic s. We employ firm- and topic-fixed effects. The firm-level fixed effects adjust for the fact that certain firms have production functions that are generally closer related to social topics than other firms and thus have generally higher proximity values. The topic-level fixed effects adjust for the pattern that firms across industries have been choosing some topics, such as education, more than others. Finally, we cluster standard errors on the topic-industry level. If firms use their comparative advantage – if they spend more on social topics close to their own production function – we expect $\beta_1 > 0$.

$$(2) \qquad \left(\frac{\text{Spending}_{f,s}}{\text{Spending}_{f}}\right) = \beta_0 + \beta_1 \text{Proximity}_{i(f),s} + \alpha_f + \alpha_s + \varepsilon_{f,s}$$

B. Construction of the proximity variable

To capture the technological advantage of a firm undertaking a given CSR project, we develop the metric Proximity on the industry $\times$ topic level. This metric measures the textual closeness between the firm's for-profit production function and the production function of projects in a given social topic. To obtain a textual description of a firm's for-profit production function, we use its' industry description from the Handbook of the National Industrial Classification (NIC). For each of the 2-digit industries, we recover all the text that contains details about this industry and its sub-components. For the average industry, we obtain 225.0 tokens. Table B1 provides an example of the recovered text for one industry. To obtain a textual description of the production functions of projects in a given social topic, we utilize the textual descriptions of the CSR projects in the MCA data. The average project has a description that contains 4.5 tokens. Table 1 displays some sample observations from this dataset. We apply some basic cleaning procedures to this textual data, detailed in Appendix B. The key assumption is that activities described by a more similar vocabulary are closer in terms of technology.

We encode the two bodies of textual data using word embeddings. Word embeddings are a key tool in Natural Language Processing, 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: similar words have similar vectors. For instance, the word "tree" and the word "forest" will have similar coordinates. In addition, there is an internally consistent geometry on the vector space that allows relating words to one another. For instance, the sum of the vector "king" and of the vector "female" is approximately the vector "queen". We use word embeddings provided by Word2Vec, which is pre-trained on Google News data. This model provides 300-dimensional embeddings for more than 3 billion words. We apply these embeddings to our two bodies of textual data. We obtain a 300-dimensional vector representation of the text for each CSR project description, denoted $\vec{\mathbf{v}}_{p,s(p)}$ for a CSR project p belonging to social topic s, and for each industry description, denoted $\vec{\mathbf{v}}_i$ for each industry i. We leave more details on word embeddings and on the implementation to Appendix B.

In the final step, we measure the proximity between each CSR project and the 71 industries. The proximity can be measured by the cosine similarity between their embedding vectors. The cosine similarity of two vectors is the cosine of the angle between the vectors; that is, it is the dot product of the vectors divided by the product of their lengths. The cosine similarity belongs to the interval $[-1,1]$. More similar vectors have a higher cosine similarity. In particular, two proportional vectors have a cosine similarity of 1, two orthogonal vectors have a similarity of 0,

and two opposite vectors have a similarity of -1. We leverage this insight to construct proximity, the measure of closeness between each CSR project p belonging to social topic s and the 71 industries i (Equation 3):

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

We then define the proximity variable at the industry $\times$ topic level by averaging over projects p belonging to the same social topic s. Denoting $\mathscr{P}_s$ the set of projects belonging to social topic s, we obtain:

$$\text{Proximity}_{s,i} = \frac{1}{\text{card}(\mathscr{P}_s)} \sum_{p \in \mathscr{P}_s} \text{Proximity}_{p,s(p),i} \tag{4}$$

C. Descriptive statistics of proximity variable

Table 2 describes the properties of the proximity variable. On the industry-topic level, the variable has a mean and median of 0.27 and a standard deviation of 0.06. It ranges from 0.12 to 0.58. Figure 7 shows the distribution of the variable. It also indicates an example. The medical and botanical industry has a proximity of 0.19 with vocational skills (around one standard deviation below the mean), 0.26 with hunger and malnutrition (approximately the mean), and 0.33 with health (around one standard deviation above the mean). We will keep this example in mind when interpreting the economic significance of our regressions on the comparative advantage hypothesis. Finally, Figure 8 shows a heat map, showing in which range the proximity variable falls for each combination of the top 20 industries and all topics. Figure 9 shows the same heat map but describes the residual of proximity after taking industry and topic fixed effects. We observe intuitive patterns in both figures. For instance, combinations with a proximity value of more than one standard deviation above the mean include food and beverages $\times$ hunger and malnutrition, chemicals $\times$ sanitation, and civil engineering $\times$ infrastructure. We conclude that the proximity construction successfully captures the closeness of the industry to a given social topic.

D. Main Results

We find that firms do follow their comparative advantage when choosing social causes: 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 to the topic by 10%, and the amount spent by 17%. Table 4 describes the results. As proximity increases by one, any CSR spending increases by 0.35. A one standard

deviation increase in proximity thus increases the likelihood of any CSR spending on a topic by 0.35 $\times$ 0.06 = 0.021. Relative to the mean of any CSR spending, this is 0.021/0.22 = 10%. Turning toward the share of spending, we see that as proximity increases by one, the share of CSR spending on a topic increases by 0.17. A one standard deviation increase in proximity thus increases the share of spending on a given topic by 0.17 $\times$ 0.06 = 0.0102. Expressed relative to the mean of the CSR spending share, this is 0.0102/0.06 = 17%. This means that the medical and botanical industry has approximately a 10% higher likelihood to spend - and spends 17% more - on health than on hunger and malnutrition. This is the first empirical evidence supporting theoretical models that make a case for firms' efficiency in providing public goods.

E. Robustness

Figure 10 summarizes a variety of first-order robustness checks. We test for robustness along different fixed effects structures, levels of clustering, construction of the proximity variable, and level of the data. Across different specifications, the coefficient remains statistically significant at the 5% level and stable, ranging from 0.09 to 0.22 for the CSR share. The only exception is when we omit topic fixed effects; those are necessary for our specification to hold. Additionally, our results are robust to dropping specific topics (Figure A6) or industries (Figure A7), suggesting that the results are not driven by a narrow subset of observations. Coefficients for the CSR share range between 0.11 and 0.19, all significant at the 1% level. We conclude that there is robust empirical evidence for the comparative advantage hypothesis.

Beyond testing for robustness along different econometric specifications, we investigate whether our estimate is biased by alternative drivers of CSR allocation. Companies might choose their topic allocation to cater to consumers, employees, investors, or the government. If these alternative explanations are driving allocation shares and are correlated with the technological proximity index in our regression Equation 2, this could bias our coefficient of interest β_1. To investigate whether our estimate is biased, we require control variables that vary on the firm-topic or industry-topic level to avoid absorption by our fixed effects. To construct respective control variables, we rely on survey data on topic preferences and interact it with firm- or industry-specific measures of dependency on a certain party. For instance, we interact consumer survey data on the importance of different topics with a measure of the downstreamness of an industry, proxying for the dependency of a firm on these consumers. This variable is high for firms in the medical industry for the topic of health if consumers have strong preferences for health and the industry is

downstream.

We utilize survey data from the World Value Survey in India to proxy for the preferences of consumers, employees, and the government (as an aggregator of these preferences). Additionally, we rely on a PwC survey in the U.S. to proxy for preferences of investors. To measure a firm's dependency on consumers, we utilize firm-level data on marketing expenses over total expenses and industry-specific measures of downstreamness. To investigate the reliance on employees, we measure the firm-specific average wage and the total wage bill over income. To understand a firm's dependency on investors, we examine the firm-level share of institutional owners and promoters (in the Indian context, those measure large shareholders with significant control over operations). Finally, to capture government influence, we examine at the industry level whether the industry is strongly regulated and whether procurement contracts play a role. Based on our interacted firm-topic or industry-topic variables, we construct dummies that are equal to one if a firm-topic observation has an interacted variable that is above the median.

We conduct four empirical tests to investigate whether alternative drivers of CSR allocation bias our result. First, we run our baseline Equation 2, but include the constructed dummies that capture alternative explanations as control (Equation A1). Table A7 shows that the coefficient on proximity always remains statistically significant at the 1% level. The magnitude of the coefficient ranges from 0.13 to 0.24. Note that the magnitude does not vary due to the inclusion of the control variable but due to different samples. Our proxies of alternative explanations are not measurable for all firms, industries, or topics. We demonstrate in the same table that when we run the baseline regression without control in the same sample, we recover a highly similar coefficient to the one with the controls. It is not the case that our proxies are simply not capturing alternative explanations for CSR allocation well. In a regression without firm and topic fixed effects, we find that many of them positively predict topic allocation (Table A8 and A9). This evidence suggests that firms do take into account the preferences of consumers, employees, and investors. Crucially, however, this does not bias our result.

In a second test, we want to measure the differential impact of proximity for firms with low and high dependency on consumers, employees, investors, and the government. Thus, we interact the proximity variable with a dummy that is one if a firm- and industry-specific dependency measure is above the median (Equation A4). Table A10 shows that there is little evidence of a differential reaction in CSR allocation. For instance, firms that have a downstream index below the median have a 0.23 increase in CSR share due to proximity, while those above the median show a 0.24 increase in CSR share.

In a third test, we control for director and investor fixed effects. Utilizing data

from Prowess on directors and investors, we exploit the fact that multiple firms share identical directors and investors. If these parties do not substantially influence CSR allocation, we expect fixed effects to make little difference. Indeed, we recover similar coefficients in Table A11 for the proximity with and without director or investor fixed effects.

Finally, we test whether proximity has stronger predictive power for projects that are directly implemented by the firm, as opposed to donations to a third-party organization. We expect this to be the case if the proximity truly captures the technological advantage of firms in undertaking certain CSR projects. By contrast, if the proximity metric just captures alternative explanations of investing in certain CSR projects, there should be no link between our proximity metric and the mode of implementation. For this test, we split our main sample into projects that are implemented directly by the firms as opposed to indirectly through third parties. To measure whether a project is indirectly implemented, we utilize the textual description of the project. According to this classification, 15% of projects are classified as indirectly implemented, as described in Appendix B. First, we observe that the proximity variable negatively predicts third-party implementation (see Table A12). This is in alignment with our predictions; if the firm has a higher comparative advantage, it should be more likely to directly implement the project. Second, the coefficient and r-square in our baseline regression Equation 2 is substantially smaller in the sub-sample of indirectly implemented projects than in the sub-sample of directly implemented projects (see Table A13). Note that these tests should be interpreted with some caveats in mind. We can only identify third-party implementation for projects that have a sufficient textual description. Additionally, in the second test, whether a project is directly or indirectly implemented is itself an outcome of our proximity variable. To summarize, collectively, these tests suggest that alternative explanations of CSR allocation do not bias our main coefficient of interest.

IV. Is There an Efficiency-Equity Trade-off?

After demonstrating that firms use their comparative advantage, so there is potential for an efficiency gain, we turn towards equity. Since firms maximize private returns, not social returns, they might not sufficiently take into account equity considerations. To test whether there is potential for an efficiency-equity trade-off, we focus on one specific dimension of equity: the geographical distribution of resources. This has the advantage that we have a clear prediction of where CSR spending should flow: under the assumption that the social returns of projects are larger in more disadvantaged areas, CSR spending should flow to states with lower GDP or other development indicators. To test whether firms spend more in states

that are better off, we run regression Equation 5. We regress the log of state-level CSR spending (in million rupees, aggregated over 2015 to 2019) on the log of state-level GDP (in million rupees, in 2011), scaling both variables by population. Here k denotes state. To obtain CSR spending location, we utilize the MCA project-level information. The state location is available for 84% of projects. GDP data is obtained from the RBI. If firms spend more in richer states, we expect $\beta_1 > 0$. Beyond that, we are interested in comparing the geographical distribution of CSR spending to that of government spending. For this purpose, we repeat 5 with the log of state-level government spending (in million rupees, aggregated over 2015 to 2019) as the dependent variable. We restrict our attention to government spending on the social causes covered by CSR. We collect state-level government spending from the RBI.

$$\log(\text{Spending per 1m people})_k = \beta_0 + \beta_1 \log(\text{GDP per 1m people})_k + \varepsilon_k \tag{5}$$

In a second step, we repeat the exercise on the state-topic level (Equation 6). The dependent variables are the logs of state-level CSR and government spending on a given social topic. The independent variable stacks topic-specific development indicators. For instance, for education we utilize literacy rates. A higher development indicator value means that the state is better off on this dimension. The development indicators are standardized to have a mean of zero and a standard deviation of one. Here s denotes social topic.

$$\log(\text{Spending per 1m people})_{k,s} = \beta_0 + \beta_1 \mathrm{D}_{k,s} + \alpha_s + \varepsilon_{k,s} \tag{6}$$

Table 5 shows the results. Column 1 and 2 refer to Equation 5, Column 3 and 4 to Equation 6. Columns 1 and 3 report firms' CSR spending, while Columns 2 and 4 refer to government spending. We find that if GDP per one million people increases by 10%, CSR spending increases by 13.5%. In contrast, government spending shows a weaker response with 4.1%. This suggests that firms' expenditures are three times more elastic with respect to local GDP than government expenditures. We reject equality of the coefficients for firm and government spending at the 1% level. Similarly, firms' spending is five times more elastic to local development indicators than government spending. Here, we reject equality of the coefficients for firm and government spending at the 10% level. Figure 11 and Figure 12 show the results graphically.

Why is CSR more likely to be distributed in states that are better off compared to government spending? We show that one likely explanation is that firms are spending more in their own state. To obtain firms' location, we utilize the address registered with the Registrar of Companies reported by Prowess. Figure 13 shows

CSR resources by firm location and by spending location. We observe a strong overlap. Indeed, 60% of CSR resources are spent in the state where the firm is registered. Figure 14 shows that if we eliminate same-state spending, increasing GDP per one million people by 10% increases CSR spending only by 10.3%, compared to 13.5% in the full sample. To summarize, that firms spend in their own state seems to be an important factor for the efficiency-equity trade-off. Note that firms might spend more in their own states for various reasons. For instance, they might face lower implementation costs or are more aware of local needs; both could increase efficiency of spending, but come at the cost of equity.

V. 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 welfare effects of CSR. A robust finding of the theoretical literature on CSR is that CSR can be welfare-improving only if firms have a comparative advantage in the provision of some public goods. We test this hypothesis by constructing an index of technological proximity between a firm's for-profit activity and various social causes, and showing that technological proximity does predict the allocation of firm's CSR spending across social causes. We then show that this efficiency gain appears to be accompanied by equity losses, as firms allocate their spending to areas where social returns are relatively low. These results highlight an important trade-off associated to firms acting as providers of public goods.

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.

Bachas, Pierre, Lucie Gadenne, Davi Bhering, and Radhika Kaul, "Responses of Firms to Taxation and the Link to Informality: Evidence from India's GST," *Hidden Potential: Rethinking Informality in South Asia*, 2023.

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

Benabou, Roland and Jean Tirole, "Individual and Corporate Social Responsibility," *Economica*, 2010, *77* (305), 1–19.

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

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.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri, " Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records *," *The Quarterly Journal of Economics*, 05 2011, *126* (2), 749–804.

Cornes, R. and T. Sandler, *The Theory of Externalities, Public Goods and Club Goods*, Cambridge University Press, 1996.

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.

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

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.

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

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

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

Iglesias, Orial, "The End Of CSR (As We Know It) And The Rise Of Businesses With A Conscience," *Forbes*, 2022.

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.

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.

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.

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.

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

Tables

Table 1: Project-Level CSR Data Example

Firm	Year	Topic	State	District	Spent	Description
Lohiya Edible Oils	2015	hunger malnutrition	Telangana	Hyderabad	1,400,000	food subsidies
The Andhra Sugars	2017	environmental sustainability	Gujarat	West Godavari	10,000,000	promotion of green energy
Welspun India	2019	vocational skills	Gujarat	Kutch	7,500,000	Career guidance program at Ratnal village

Table 2: Summary Statistics Merged Project and Firm Accounting Data

	Mean	SD	Median	Min	Max
Firm-level					
Avg profit past 3y (m INR)	5,055	46,978	427	-432,203	1,864,825
Income (m INR)	50,023	424,301	6,397	0	23,389,356
Net worth (m INR)	30,566	254,386	3,225	-180,965	13,795,679
Mandated CSR (m INR)	87.02	701.84	9.24	0.00	32,455.87
CSR spending (m INR)	78.21	707.16	8.43	0.00	34,242.65
CSR/avg profit past 3y (%)	0.05	0.83	0.01	0.00	46.82
Unique projects per year (nr)	4.76	8.18	2.60	1.00	271.34
Unique topics per year (nr)	2.53	1.84	1.94	1.00	13.95
Topic-level					
Any CSR spending on topic (yes/no)	0.22	0.18	0.18	0.04	0.75
CSR spending (m INR)	4.11	5.25	2.94	0.26	20.94
CSR spending, non-zero (m INR)	14.97	7.74	15.57	2.94	28.07
CSR share (%)	0.06	0.09	0.04	0.01	0.38
CSR share, non-zero (%)	0.21	0.10	0.18	0.12	0.51
Industry-topic-level					
Proximity	0.27	0.06	0.27	0.12	0.58
Unique observations					
Unique firm-topics (nr)	105,168				
Unique industry-topics (nr)	1,136				
Unique firms (nr)	6,573				
Unique industries (nr)	71				
Unique topics (nr)	16				

Notes: This table shows the merged MCA and Prowess data (2015 to 2019). Firm-level and topic-level variables are summed over all years, except unique projects per year and unique topics per year. Those are depicted as the average over the years since there is no unique project-level identifier. All monetary variables are presented in real terms, denominated in 2015 rupees. These variables are not winsorized. CSR share (%) is the main outcome of interest in Equation 2. Proximity is the textual measure of closeness between an industry and a topic defined in Section III.B.

Table 3: Effect of the Policy on CSR Spending (Diff-in-Diff)

	$CSR_{f,t}/\overline{Profit}^{3y}_{f,t}$		
	(1)	(2)	(3)
Treated × Post	0.013***	0.012***	0.011***
	(0.001)	(0.001)	(0.001)
Constant	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)
Year FE	Yes		
Year × Ind. × State FE		Yes	Yes
Firm FE			Yes
Group FE	Yes	Yes	Yes
R-squared	0.01	0.06	0.37
Observations	233,523	226,335	222,660

Notes: This table shows the causal effect of the policy on CSR spending over average profits in the past three years. Data is from Prowess (2007-2019), on the firm-year level. Equation 1 describes the regression. All monetary variables are presented in real terms, denominated in 2015 rupees. Variables are winsorized at the 99th percentile. Profits are additionally winsorized at the 1st percentile. Standard errors are clustered at the firm level. ∗∗∗, ∗∗ and ∗ indicate significance at the 1%, 5% and 10% levels.

Table 4: Comparative Advantage

	Any CSR Spending	CSR Spending Share
	(1)	(2)
Proximity	0.35***	0.17***
	(0.06)	(0.05)
Firm FE	Yes	Yes
Topic FE	Yes	Yes
R-squared	0.33	0.24
Unique firms	6,573	6,573
Unique topics	16	16
Observations	105,168	105,168

Notes: This table shows a test for whether firms use their comparative advantage when engaging in CSR activities. Data is from MCA and Prowess (2015-2019), on the firm-topic level. Equation 2 describes the regression. Spending share is the share of CSR that a firm allocates to a given topic. Proximity is the textual measure of closeness between an industry and a topic defined in Section III.B. Standard errors are clustered at the industry-topic level. ∗∗∗, ∗∗ and ∗ indicate significance at the 1%, 5% and 10% levels.

Table 5: Equity - Distribution Across States

	Log(Spending per 1m people)			
	Firms	Government	Firms	Government
	(1)	(2)	(3)	(4)
Log(GDP per 1m people)	1.35***	0.41**		
	(0.29)	(0.16)		
Development indicators (standardized)			0.22**	0.03
			(0.09)	(0.06)
$\beta(\text{firms}) = \beta(\text{gov})$	0.00	.	0.07	.
Topic FE	No	No	Yes	Yes
R-squared	0.44	0.19	0.44	0.72
Unique states	30	30	30	30
Unique topics	.	.	8	8
Observations	30	30	222	222

Notes: This table shows how firms and governments allocate their spending geographically across states. Data is from MCA, Prowess, and RBI, on the state level. Equations 5 (columns 1 and 2) and 6 (columns 3 and 4) describe the regressions. The outcome variable is the log of spending (real million rupees) per one million people by firms (CSR) or the government, aggregated from 2015 to 2019. The independent variables are the log of GDP (real million rupees) per one million people (2011), or standardized topic-specific development indicators (e.g., literacy rate, various years). Variables are not winsorized and standard errors are not clustered. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Figures

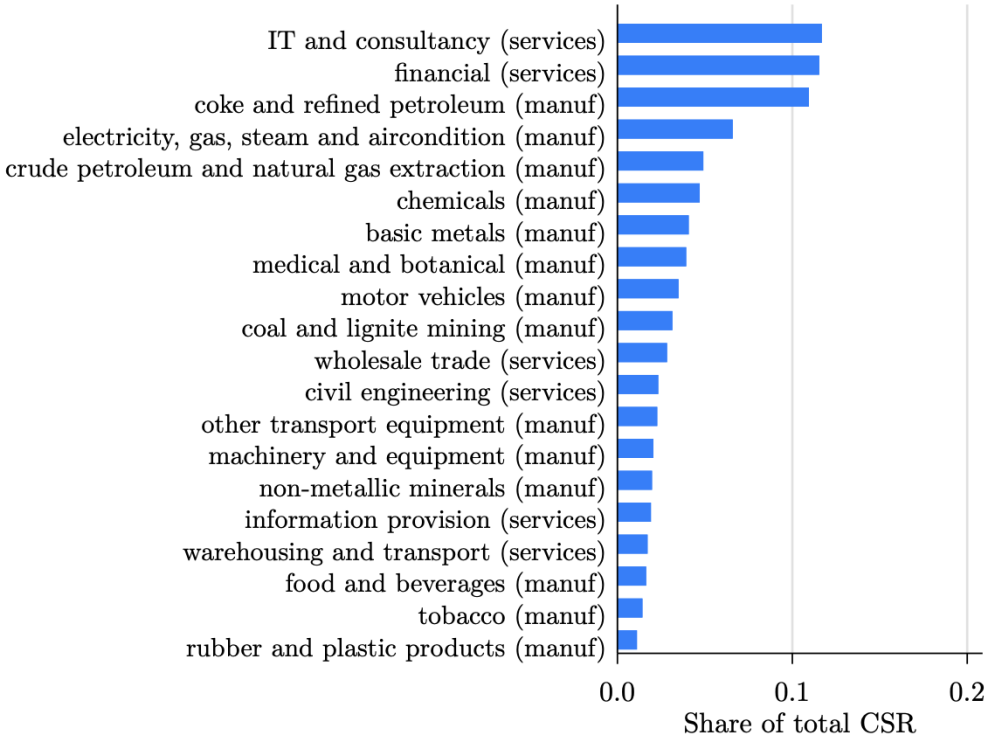


Figure 1. CSR Spending Share by Top 20 Industries. Data is from MCA and Prowess (2015-2019). Share of total CSR is aggregated CSR spending by industry over total CSR spending.

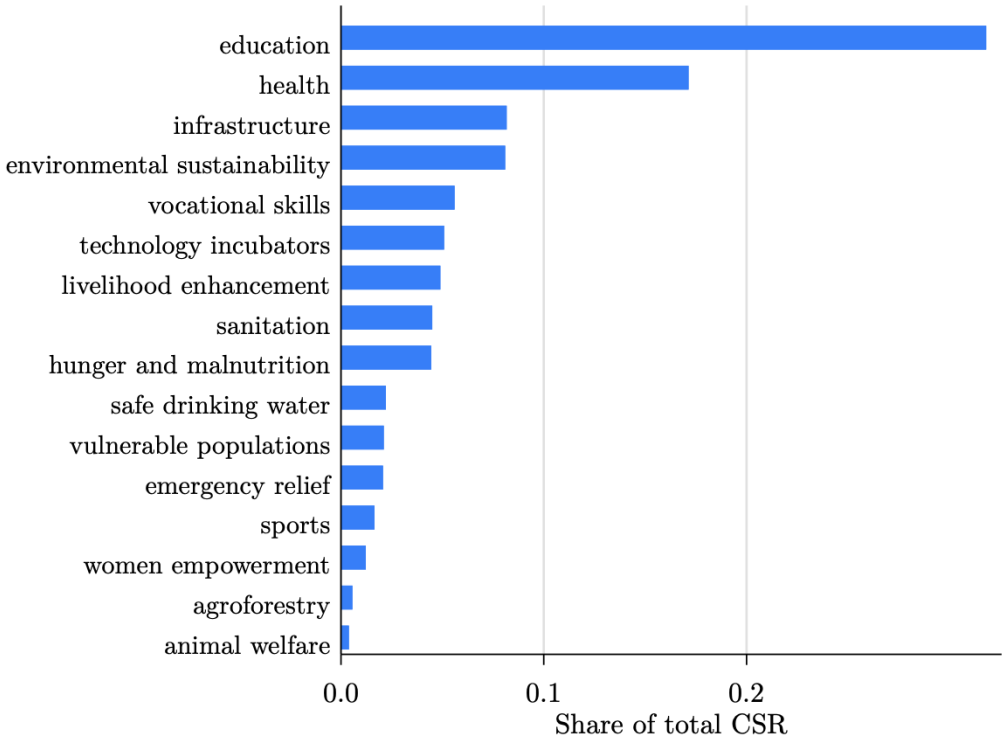


Figure 2. CSR Spending Share by Topics. Data is from MCA and Prowess (2015-2019). Share of total CSR is aggregated CSR spending by topic over total CSR spending.

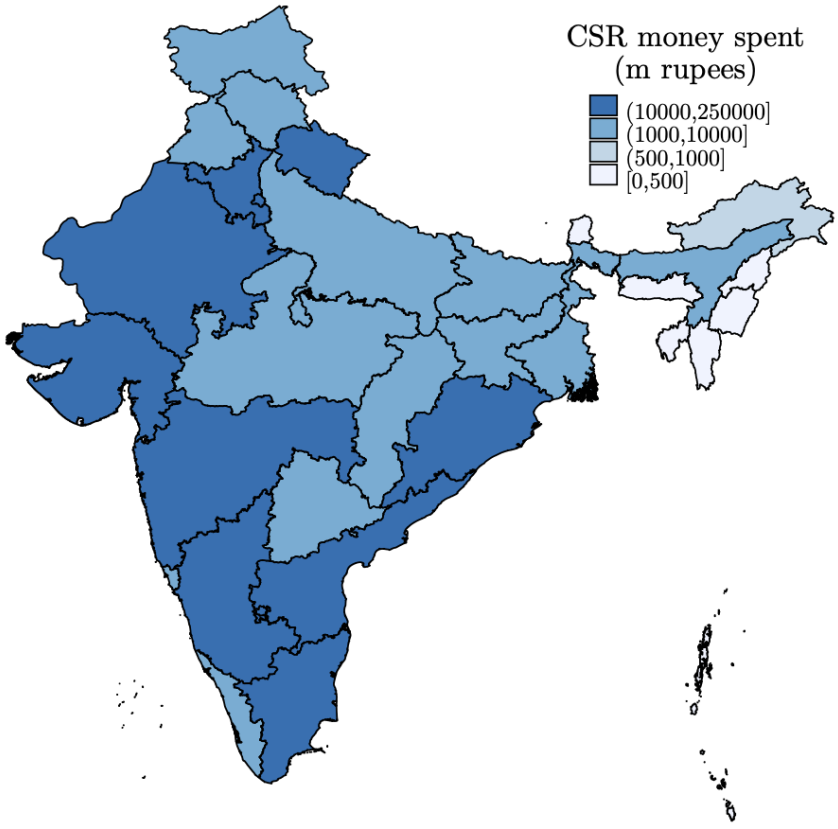


Figure 3. Geographical Distribution of CSR Spending. Data is from MCA and Prowess (2015-2019), aggregated by states. CSR money spent is in millions, presented in real terms, denominated in 2015 rupees.

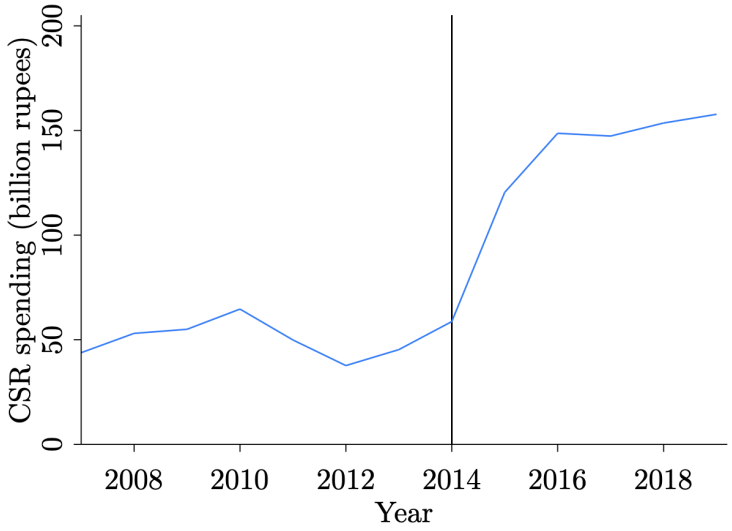


Figure 4. Total CSR Spending Tripled Between 2014 to 2019. Data is from Prowess (2007-2019), aggregated by years. CSR spending in billions, presented in real terms, denominated in 2015 rupees.

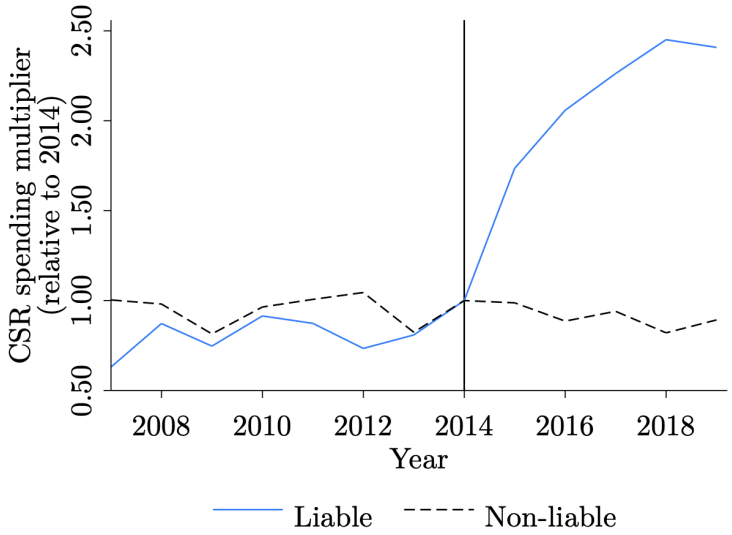


Figure 5. Raw Trends Indicate Substantial Increase in CSR Spending for Liable Firms. Data is from Prowess (2007-2019), aggregated by years and liability status. The blue line (solid) indicates liable firms and the black line (dashed) represents non-liable firms. Underlying CSR spending data presented in real terms, denominated in 2015 rupees.

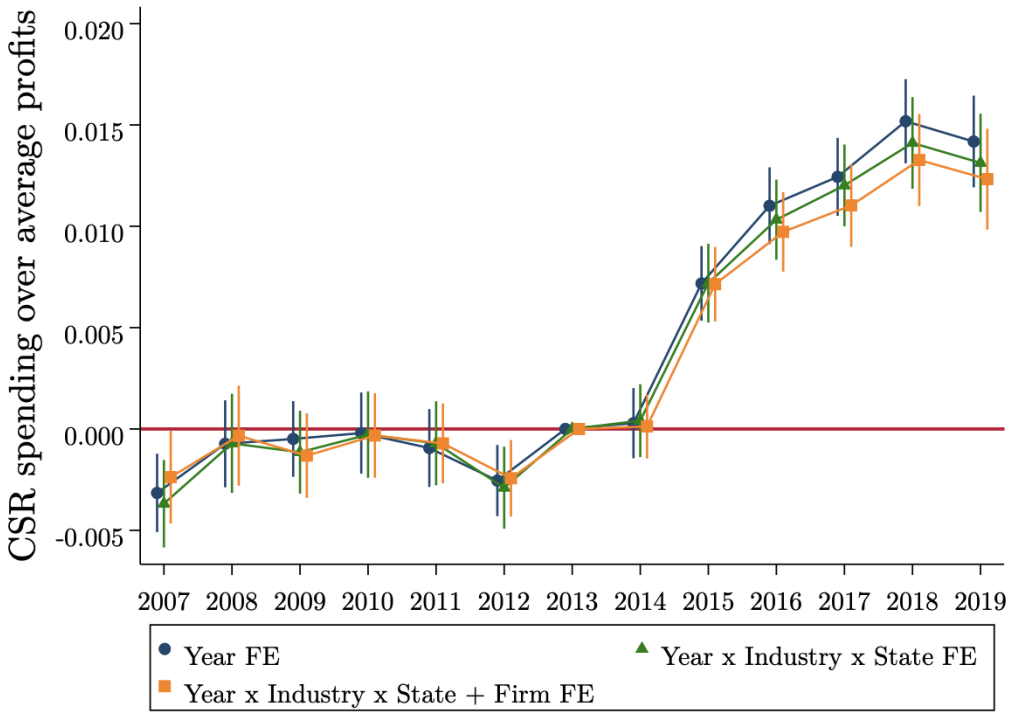


Figure 6. Difference-in-Difference Shows Increase in CSR Spending for Liable Firms. Data is from Prowess (2007-2019), on the firm-year level. Equation 1 describes the regression, with different fixed effect structures. The main outcome of interest is $\text{CSR}_{f,t}/\overline{\text{Profit}}^{3y}_{f,t}$, the CSR spending over average profits in the preceding three years. CSR spending (million rupees) and average profit over the past three years (million rupees) are expressed in real terms, denominated in 2015 rupees. Variables are winsorized at the 99th percentile. The profit measure is additionally winsorized at the 1st percentile. Treatment status is one if the firm's profits, income, or net worth cross the respective thresholds defined by the policy in any of the three preceding financial years. Standard errors are clustered at the firm level. The figure shows 95% confidence intervals.

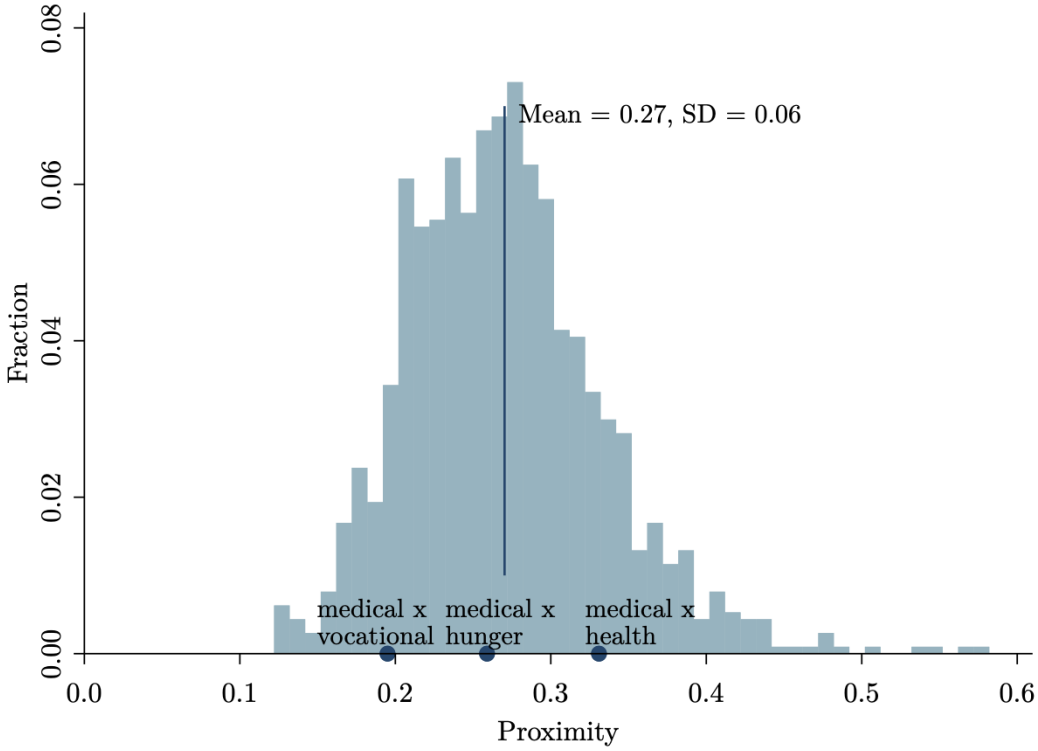


Figure 7. Distribution of Technological Proximity. Data is from MCA and Prowess (2015-2019), on the industry-topic level. Proximity is the textual measure of closeness between an industry and a topic defined in Section III.B. The histogram shows three examples, corresponding to approximately one standard deviation below the mean (medical and botanical $\times$ vocational skills), the mean (medical and botanical $\times$ hunger and malnutrition), and one standard deviation above the mean (medical and botanical $\times$ health).

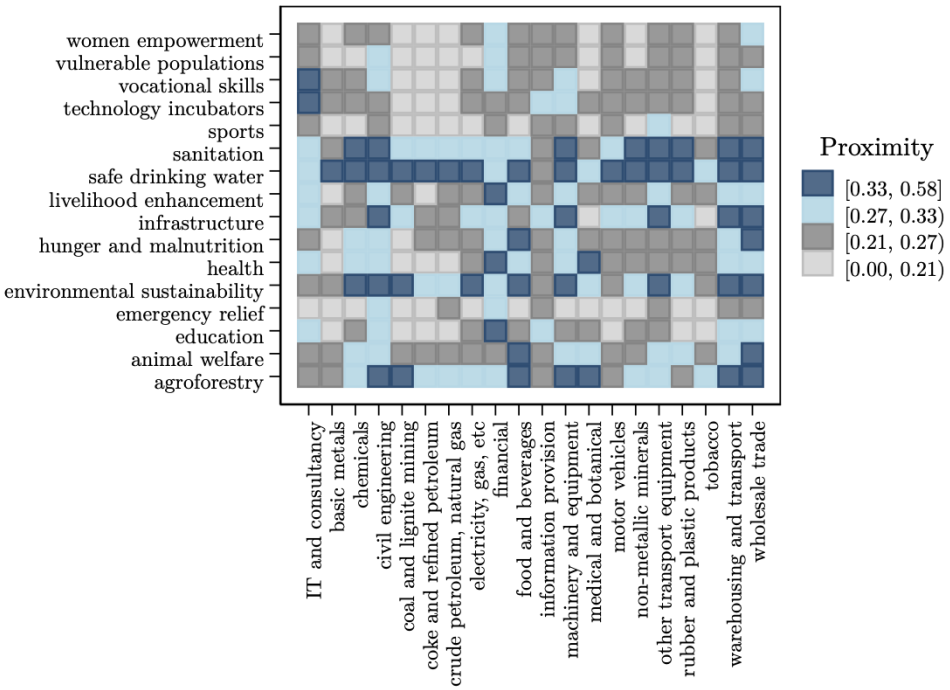


Figure 8. Distribution of Technological Proximity. Data is from MCA and Prowess (2015-2019), on the industry-topic level. Proximity is the textual measure of closeness between an industry and a topic defined in Section III.B. The heat plot shows the proximity for all topics and the top 20 industries by CSR spending share.

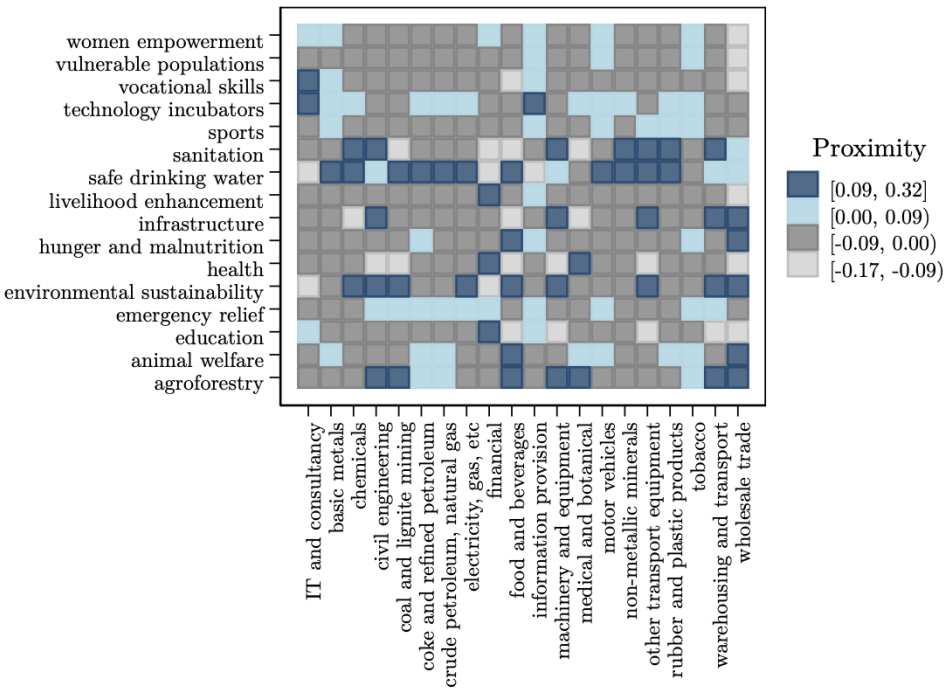


Figure 9. Distribution of Technological Proximity with Fixed Effects. Data is from MCA and Prowess (2015-2019), on the industry-topic level. Proximity is the textual measure of closeness between an industry and a topic defined in Section III.B. The heat plot shows the residuals of proximity after industry and topic fixed effects for all topics and the top 20 industries by CSR spending share.

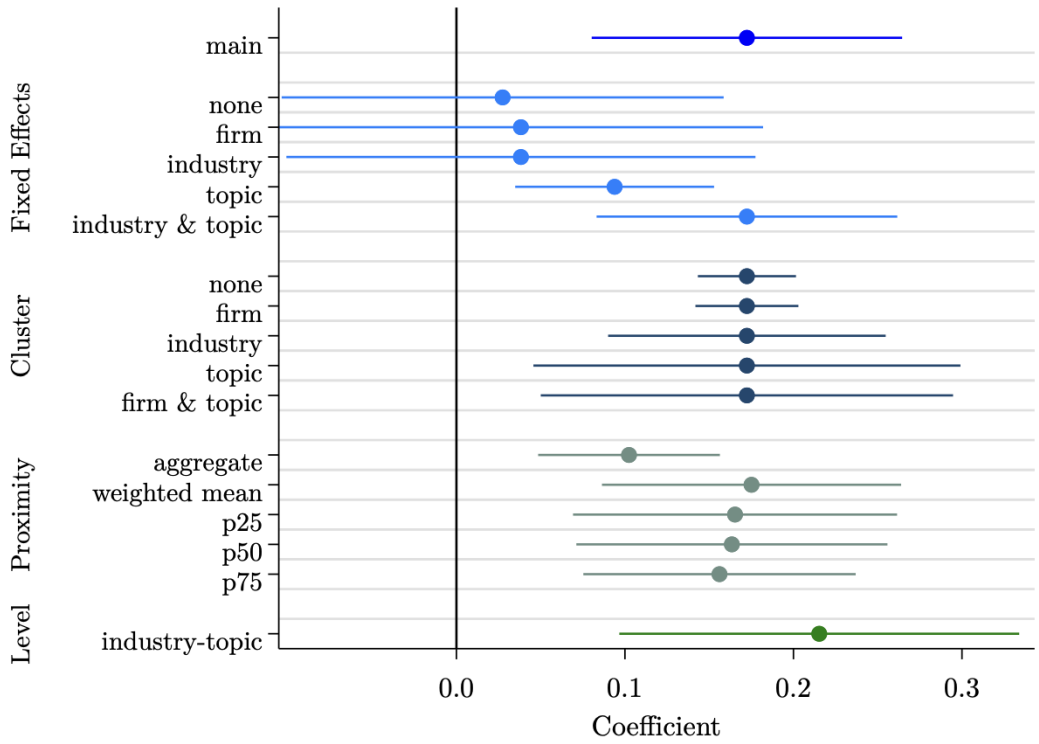


Figure 10. Comparative Advantage Robustness Tests Summary. Data is from MCA and Prowess (2015-2019). Equation 2 describes the regression. The main outcome is spending share, the share of CSR that a firm allocates to a given topic. The main independent variable is proximity. Rows 2 to 6 describe different fixed effect structures; in the main regression, we employ firm and topic fixed effects. Row 7 to 11 show different clustering of standard errors; in the main regression, we cluster standard errors at the topic $\times$ industry level. Row 12 to 16 show different versions of the proximity variable; in the main regression, we take the mean across projects. Row 17 repeats our main regression, collapsing the data on the industry-topic level. The figure shows 95% confidence intervals.

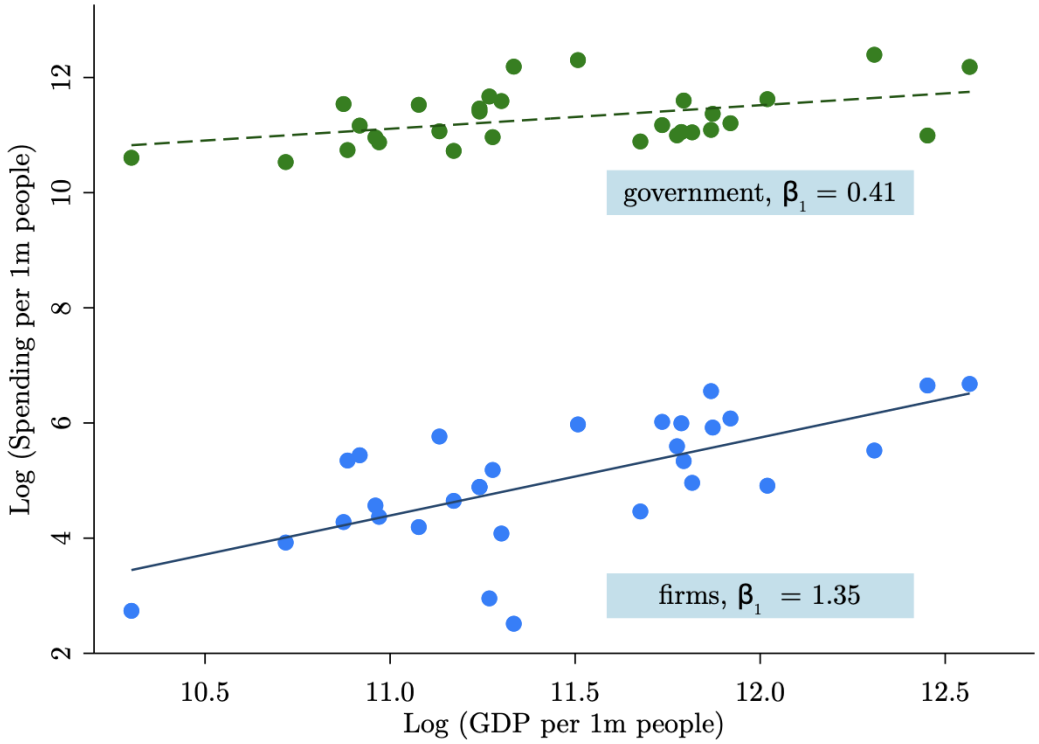


Figure 11. CSR Spending is More Sensitive to GDP than Government Spending. Data is from MCA, Prowess, and RBI, on the state level. Equation 5 describes the regression. The main outcome variable is the log of spending per one million people, either from firms or the government (2015-2019). State-level GDP is measured in 2011. The underlying spending and GDP data is in millions, in real terms, denominated in 2015 rupees. The scattered dots indicate state-level observations, blue for CSR spending and green for government spending. The lines indicate fitted linear approximations, blue for CSR spending (solid) and green for government spending (dashed).

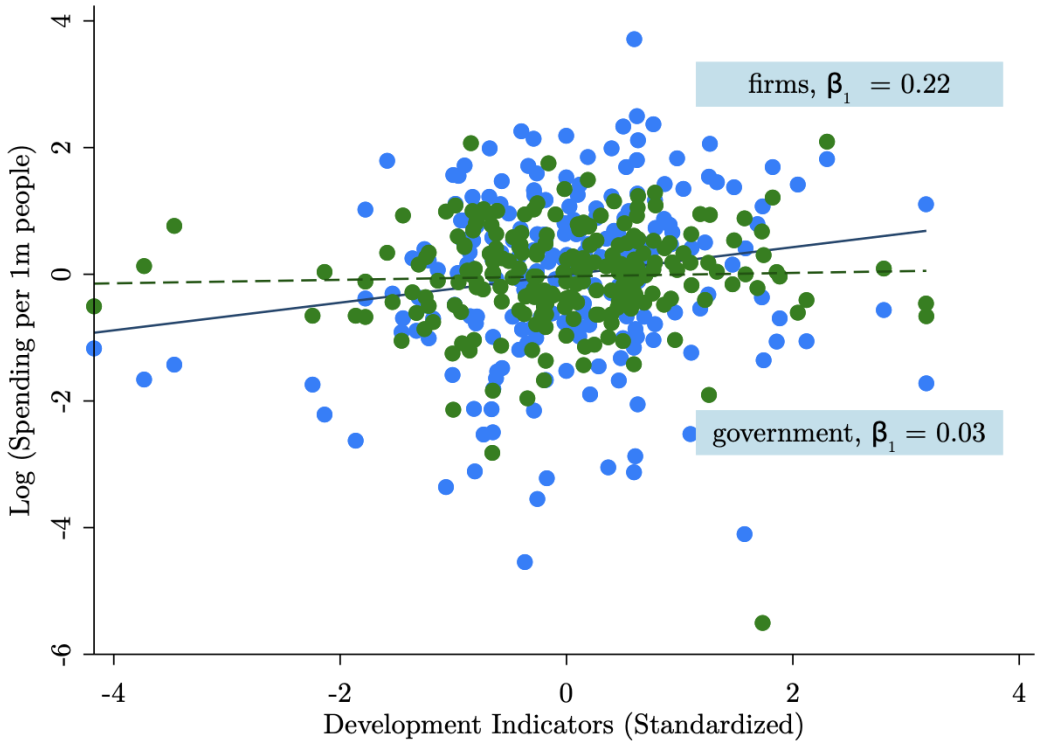
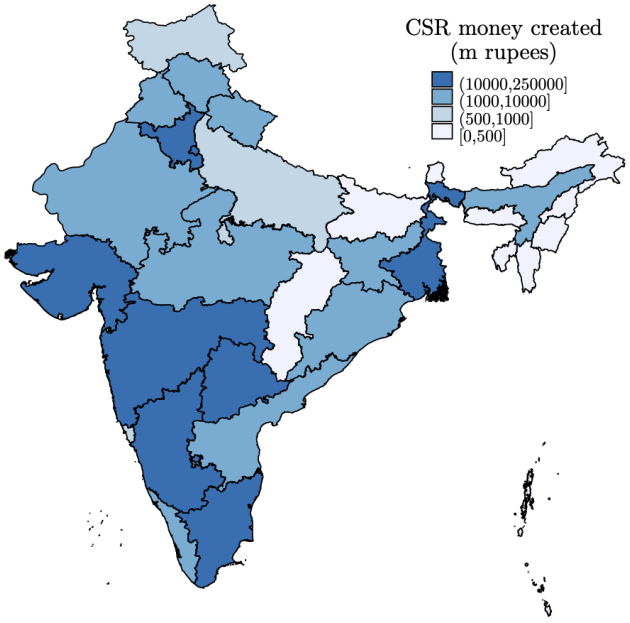
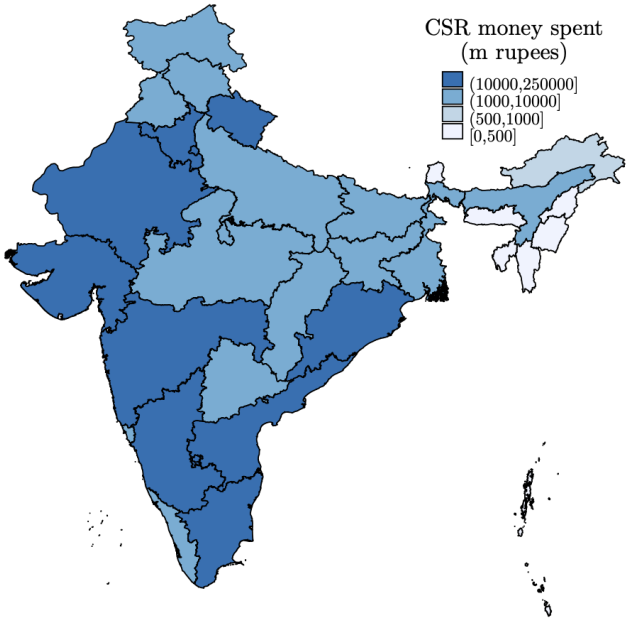


Figure 12. CSR Spending is More Sensitive to Development Indicators than Government Spending. Data is from MCA, Prowess, and RBI, on the state-topic level. Equation 5 describes the regression. The main outcome variable is the log of spending per one million people, either from firms or the government. The underlying spending data is in millions, in real terms, denominated in 2015 rupees. Spending is demeaned by topic fixed effects. The main independent variable is a stacked variable containing development indicators for different topics measured in various years depending on data availability, transformed such that a higher indicator always means a better status. Development indicators are standardized with a mean of zero and a standard deviation of one. The scattered dots indicate state-level observations, blue for CSR spending and green for government spending. The lines indicate fitted linear approximations, blue for CSR spending (solid) and green for government spending (dashed).



(a) CSR Firm Location



(b) CSR Project Location

Figure 13. CSR Firm and Project Location Show Strong Overlap. Data is from MCA, Prowess, and RBI, on the state level. We depict CSR expenses by state of origin (state of registration of the firm) and state of spending. The variables are in millions, in real terms, denominated in 2015 rupees.

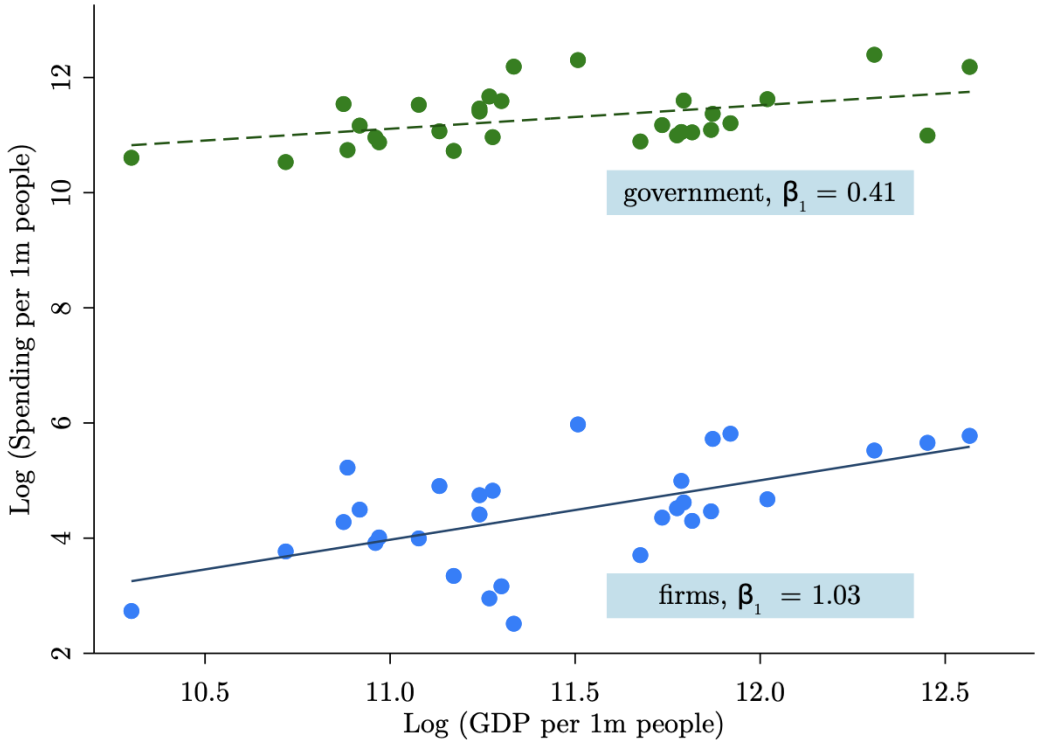


Figure 14. CSR Spending is Less Sensitive to GDP if Same State is Eliminated. Data is from MCA, Prowess, and RBI, on the state level. Equation 5 describes the regression. The main outcome variable is the log of spending per one million people, either from firms or the government (2015-2019). State-level GDP is measured in 2011. The underlying spending and GDP data is in millions, in real terms, denominated in 2015 rupees. The scattered dots indicate state-level observations, blue for CSR spending and green for government spending. The lines indicate fitted linear approximations, blue for CSR spending (solid) and green for government spending (dashed). In contrast to Figure 11, we eliminate CSR that is spent in the same state as the firm is registered.

Appendix A

Tables

Table A1: Merge of CSR Project and Firm Accounting Data

	MCA only	Prowess match	Matched (%)
Firm match			
Firms (nr)	4,423	7,063	61
Firm-years (nr)	11,925	22,118	65
Avg profit past 3 years (m INR)	2,210,746	27,892,015	93
CSR spending (m INR)	50,759	522,823	91
Firm-year match			
Firms (nr)	5,181	6,305	55
Firm-years (nr)	15,686	18,357	54
Avg profit past 3 years (m INR)	3,593,131	26,509,630	88
CSR spending (m INR)	80,442	493,141	86

Notes: This table shows the total over observations conditional on an unsuccessful merge ("MCA only") and a successful merge ("Prowess match"). Years 2015 to 2019. All monetary variables are presented in real terms, denominated in 2015 rupees. These variables are winsorized at the 98th percentile. Profit measures are additionally winsorized at the 2nd percentile.

Table A2: Summary Statistics Firm-Year-Level Financial Accounts Data

	Mean	SD	Median	Min	Max
Firm-year-level					
Profit (m INR)	329	5,912	4	-258,049	441,335
Avg profit past 3 years (m INR)	284	4,719	4	-107,115	358,093
Income (m INR)	4,787	54,382	273	0	5,620,830
Net worth (m INR)	2,589	28,886	131	-273,693	3,553,998
Treatment status (1=liable)	0.28	0.45	0.00	0.00	1.00
CSR spending (m INR)	4.59	98.58	0.00	0.00	15,660.00
CSR / avg profit past 3 years (%)	32.35	11,047.19	0.00	0.00	5,033,165.00
Any CSR committee (1=yes)	0.13	0.34	0.00	0.00	1.00
CSR directors (nr)	0.45	1.20	0.00	0.00	14.00
Unique observations					
Unique firm-years (nr)	245,464				
Unique firms (nr)	39,053				
Unique years (nr)	13				
Unique 2-digit industries (nr)	77				
Unique states (nr)	35				
Unique districts (nr)	423				

Notes: Data is from Prowess (2007-2019). All monetary variables are presented in real terms, denominated in 2015 rupees. These variables are not winsorized. CSR over average profits in the past three years (%) is the main outcome of interest in Equation 1.

Table A3: Firm-Level Accounting Data - Two Similar CSR Measures

	Mean	SD	Median	Min	Max
Firm-year-level					
CSR spending (m INR) - constructed	7.91	131.97	0.00	0.00	15,642.70
CSR spending (m INR) - explicit	6.30	93.77	0.00	0.00	7,447.12

Notes: Data is from Prowess (2007-2019). The analysis is limited to the years 2016 to 2019 to capture the years in which CSR was explicitly collected. CSR constructed is the sum of social and community expenses as well as donations, which are collected before and after the policy. CSR explicit is directly collected in financial statements in response to the policy. All monetary variables are presented in real terms, denominated in 2015 rupees. These variables are not winsorized. For more details see Discussion A1.

Table A4: Treated and Control Firms Before the Policy (2014)

	All (1)	Control (2)	Treated (3)	Difference (4)
Profit (m INR)	287	-114	1,225	0.00
	(5,409)	(1,761)	(9,444)	
Avg profit past 3 years (m INR)	262	-120	1,155	0.00
	(4,564)	(1,674)	(7,867)	
Income (m INR)	5,033	1,359	13,629	0.00
	(61,210)	(14,126)	(109,270)	
Net worth (m INR)	2,522	337	7,633	0.00
	(26,368)	(7,575)	(46,373)	
CSR spending (m INR)	2.77	0.29	8.57	0.00
	(56.13)	(13.74)	(100.15)	
CSR / avg profit past 3 years (%)	0.05	0.01	0.12	0.33
	(4.95)	(0.22)	(8.84)	
Observations	20,793	14,566	6,227	20,793

Notes: Data is from Prowess (2007-2019), on the firm-year level. The last column shows the p-value of a t-test on the difference between control and treatment in 2014. All monetary variables are presented in real terms, denominated in 2015 rupees. These variables are winsorized at the 99th percentile. Profit and net worth measures are additionally winsorized at the 1st percentile.

Table A5: Results Robust to Treatment Based on 2013 Accounting

	$CSR_{f,t}/\overline{Profit}^{3y}_{f,t}$					
	DID	DID	DID	IV-DID	IV-DID	IV-DID
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	0.013***	0.012***	0.011***	0.012***	0.012***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Year FE	Yes			Yes		
Year × Ind. × State FE		Yes	Yes		Yes	Yes
Firm FE			Yes			Yes
Group FE	Yes	Yes	Yes	Yes	Yes	Yes
First-Stage F Stat				26,200	19,151	17,582
R-squared	0.01	0.06	0.37	0.00	0.00	0.00
Observations	233,523	226,335	222,660	202,154	195,259	193,994

Notes: This table shows the causal effect of the policy on CSR spending over average profits in the past three years. Columns 4 to 6 instrument the true treatment status with a treatment status calculated from 2013 accounting variables. Data is from Prowess (2007-2019), on the firm-year level. Equation 1 describes the regression. All monetary variables are presented in real terms, denominated in 2015 rupees. Variables are winsorized at the 99th percentile. Profits are additionally winsorized at the 1st percentile. Standard errors are clustered at the firm level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A6: No Evidence of Relabelling

	Percent of Total Expenses (x100)			
	Environment	Employee Welfare	Employee Training	Social Amenities
	(1)	(2)	(3)	(4)
Treated × Post	-0.03	0.40	-0.04	-0.00
	(0.02)	(0.43)	(0.05)	(0.00)
Year × Ind. × State FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Group FE	Yes	Yes	Yes	Yes
R-squared	0.28	0.31	0.20	0.15
Observations	233,207	233,207	233,207	233,207

Notes: This table shows the causal effect of the policy on other expense categories to test for manipulation via relabelling. Data is from Prowess (2007-2019), on the firm-year level. Equation 1 describes the regression. All monetary variables are presented in real terms, denominated in 2015 rupees. Variables are winsorized at the 99th percentile. Standard errors are clustered at the firm level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A7: Robustness to Alternative Explanations

	CSR Spending Share							
	Consumers		Employees		Investors		Government	
	Marketing	Downtream	Avg wage	Wage bill/income	Institutional	Promoters	Regulated	Procurement
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proximity	0.13***	0.16***	0.16***	0.13***	0.23***	0.24***	0.13***	0.13***
	(0.04)	(0.04)	(0.05)	(0.03)	(0.06)	(0.06)	(0.03)	(0.03)
Coefficient β_2	-0.01	0.00	0.01**	0.00	0.01	0.00	0.00	-0.01
	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Topic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.32	0.33	0.35	0.32	0.22	0.22	0.32	0.32
Unique firms	5,561	4,266	1,982	5,848	1,507	1,507	5,905	5,905
Unique topics	10	10	10	10	9	9	10	10
Observations	55,610	42,660	19,820	58,480	13,563	13,563	59,050	59,050
Baseline coef	0.13	0.16	0.16	0.13	0.24	0.24	0.13	0.13

Notes: This table tests the robustness of proximity to alternative drivers of CSR spending allocation. Data is from MCA and Prowess (2015-2019), on the firm-topic level. Equation A1 describes the regression. The only difference to Equation 2 is a dummy control variable. This dummy control variable is one if the interaction between a topic preference survey measure and a firm- or industry-specific dependency measure of a firm on a party (e.g., consumers) is above the median. Different columns present different dependency measures. Survey preferences for consumers, employees, and the government are based on the World Value Survey, and for investors on a PwC report. Sample sizes vary because the control measure is not always available for all topics, firms, or industries. The last line shows the coefficient of the baseline regression in the respective sample. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

$$(\frac{\text{Spending}_{f,s}}{\text{Spending}_f}) = \beta_0 + \beta_1 \text{Proximity}_{i(f),s} + \beta_2 1(\text{Preference}_s \times \text{Dependency}_{f/i} > \text{Median}) + \alpha_f + \alpha_s + \varepsilon_{f,s} \quad \text{(A1)}$$

Table A8: Some Alternative Explanations Predict Allocation Without Fixed Effects

	CSR Spending Share							
	Consumers		Employees		Investors		Government	
	Marketing	Downtream	Avg wage	Wage bill/income	Institutional	Promoters	Regulated	Procurement
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient β_1	0.000	0.032**	0.011**	0.009	0.017***	0.039***	0.033	0.006
	(0.004)	(0.014)	(0.005)	(0.005)	(0.003)	(0.007)	(0.024)	(0.023)
Firm FE	No	No	No	No	No	No	No	No
Topic FE	No	No	No	No	No	No	No	No
R-squared	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00
Unique firms	6,573	6,573	6,573	6,573	6,573	6,573	6,573	6,573
Unique topics	16	16	16	16	16	16	16	16
Observations	55,610	42,660	19,820	58,480	13,563	13,563	59,050	59,050

Notes: This table tests for the predictive power of alternative drivers of CSR spending allocation without fixed effects. Data is from MCA and Prowess (2015-2019), on the firm-topic level. Equation A2 describes the regression. The dummy variable is one if the interaction between a topic preference survey measure and a firm- or industry-specific dependency measure of a firm on a party (e.g., consumers) is above the median. Different columns present different dependency measures. Survey preferences for consumers, employees, and the government are based on the World Value Survey, and for investors on a PwC report. Sample sizes vary because the measure is not always available for all topics, firms, or industries. Standard errors are clustered at the industry-topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

$$\left(\frac{\text{Spending}_{f,s}}{\text{Spending}_{f}}\right) = \beta_0 + \beta_1 1(\text{Preference}_s \times \text{Dependency}_{f/i} > \text{Median}) + \varepsilon_{f,s} \tag{A2}$$

Table A9: Survey Measures Predict Allocation Without Fixed Effects

	CSR Spending Share	
	World Value Survey (1)	PwC Survey (2)
Coefficient β_1	0.061***	0.052***
	(0.013)	(0.008)
Firm FE	No	No
Topic FE	No	No
R-squared	0.02	0.03
Unique firms	6,573	6,573
Unique topics	10	9
Observations	65,730	59,157

Notes: This table tests for the predictive power of topic-level survey measures, studying CSR spending allocation without fixed effects. Data is from MCA and Prowess (2015-2019), on the firm-topic level. Equation A3 describes the regression. The dummy variable is one if a topic preference survey measure is above the median. Survey preferences for consumers, employees, and the government are based on the World Value Survey, and for investors on a PwC report. Sample sizes vary because the measure is not always available for all topics, firms, or industries. Standard errors are clustered at the industry-topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

$$(\frac{\text{Spending}_{f,s}}{\text{Spending}_f}) = \beta_0 + \beta_1 1(\text{Preference}_s > \text{Median}) + \varepsilon_{f,s} \tag{A3}$$

Table A10: Robustness to Alternative Explanations

	CSR Spending Share							
	Consumers		Employees		Investors		Government	
	Marketing	Downstream	Avg wage	Wage bill/income	Institutional	Promoters	Regulated	Procurement
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proximity	0.17***	0.23***	0.20***	0.18***	0.16***	0.20***	0.19***	0.14***
	(0.06)	(0.06)	(0.06)	(0.03)	(0.05)	(0.05)	(0.06)	(0.04)
Coefficient β_2	0.00	0.01	0.01	-0.01	0.06	-0.00	-0.02	0.03
	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.02)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Topic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.24	0.25	0.26	0.24	0.30	0.30	0.24	0.24
Unique firms	5,561	4,266	1,982	5,848	1,507	1,507	5,905	5,905
Unique topics	16	16	16	16	16	16	16	16
Observations	88,976	68,256	31,712	93,568	24,112	24,112	94,480	94,480
Baseline coef	0.18	0.23	0.20	0.18	0.20	0.20	0.18	0.18

Notes: This table tests the robustness of proximity to alternative drivers of CSR spending allocation. Data is from MCA and Prowess (2015-2019), on the firm-topic level. Equation A4 describes the regression. The only difference to Equation 2 is that the proximity variable is interacted with a dummy. This dummy variable is one if a firm- or industry-specific dependency measure of a firm on a party (e.g., consumers) is above the median. Different columns present different dependency measures. Sample sizes vary because the measure is not always available for all topics, firms, or industries. The last line shows the coefficient of the baseline regression in the respective sample.

$$(\text{A4}) \qquad \left(\frac{\text{Spending}_{f,s}}{\text{Spending}_{f}}\right) = \beta_0 + \beta_1 \text{Proximity}_{i(f),s} + \beta_2 \text{Proximity}_{i(f),s} \times 1(\text{Dependency}_{f/i} > \text{Median}) + \alpha_f + \alpha_s + \varepsilon_{f,s}$$

Table A11: Director and Investor Fixed Effects

	CSR Spending Share			
	Directors Fixed effects (1)	Directors No fixed effects (2)	Investors Fixed effects (3)	Investors No fixed effects (4)
Proximity	0.17***	0.19***	0.18***	0.22***
	(0.05)	(0.06)	(0.04)	(0.05)
Topic FE	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	Yes
Director FE	Yes	No	No	No
Investor FE	No	No	Yes	No
R-squared	0.23	0.23	0.31	0.31
Unique firms	5,894	5,894	1,534	1,534
Unique topics	16	16	16	16
Observations	892,352	892,352	203,472	203,472

Notes: This table tests the robustness of proximity to alternative drivers of CSR spending allocation. Data is from MCA and Prowess (2015-2019), on the firm-topic level. The only difference to Equation 2 is the fixed effects structure. In Columns 1 and 3, we include fixed effects for directors and investors, respectively. Columns 2 and 4 repeat our baseline regression on the same sample, which is smaller than our baseline sample because not all firms have directors and investors recorded in Prowess. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A12: Proximity Predicts Indirect Implementation Negatively

	Indirect Implementation (1)
Proximity	-0.14**
	(0.06)
Firm FE	Yes
Topic FE	Yes
R-squared	0.36
Unique firms	5,361
Unique topics	16
Observations	74,742

Notes: This table shows that proximity predicts indirect implementation negatively. Data is from MCA and Prowess (2015-2019). Equation A5 describes the regression. The data is on the project level, weighted by the inverse of the number of projects by topic to give equal weight to all topics as in our main regressions. The main outcome is an indicator that is one if a project was indirectly implemented, through a third party. Standard errors are clustered at the topic-industry level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

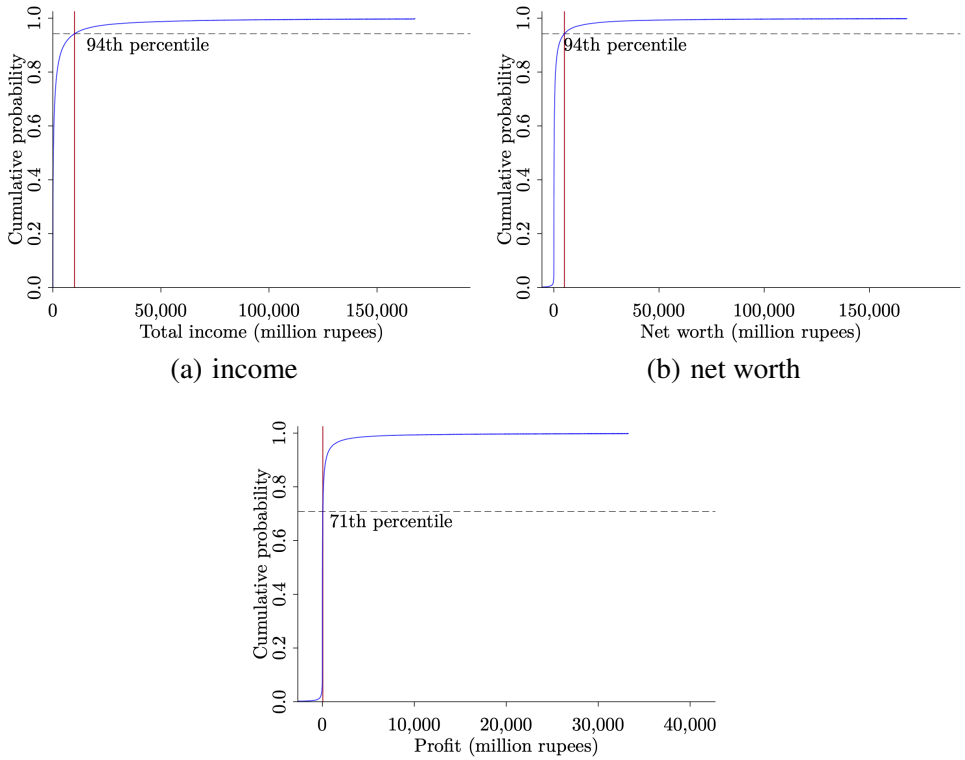
$$\text{Indirect Implementation}_{p,f,s} = \beta_0 + \text{Proximity}_{i(f),s} + \alpha_f + \alpha_s + \varepsilon_{p,f,s} \tag{A5}$$

Table A13: Stronger Effects for Directly Implemented Projects

	CSR Spending Share	
	Direct Projects (1)	Indirect Projects (2)
Proximity	0.15***	0.03**
	(0.04)	(0.01)
Firm FE	Yes	Yes
Topic FE	Yes	Yes
R-squared	0.20	0.10
Unique firms	6,062	6,062
Unique topics	16	16
Observations	96,992	96,992

Notes: This table tests the robustness of proximity to alternative drivers of CSR spending allocation. Data is from MCA and Prowess (2015-2019), on the firm-topic level. The only difference to Equation 2 is the underlying data sample. We split our main sample into two samples: projects directly and indirectly implemented. We hypothesize that we find a stronger relationship for directly implemented projects. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

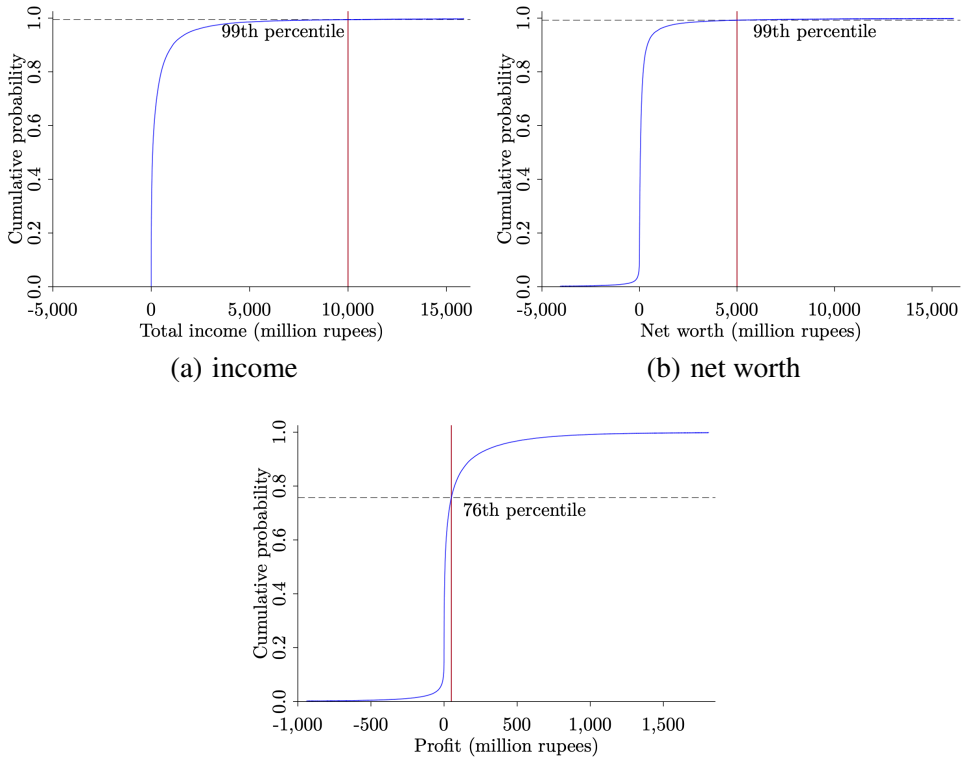
Figures



(a) income (b) net worth

(c) profit

Figure A1. Thresholds of total income, net worth, and profits. Data is from Prowess (2007-2019), on the firm-year level. The red lines indicate the policy threshold, and the blue lines indicate the unconditional cumulative distribution. For more details see Discussion A2.

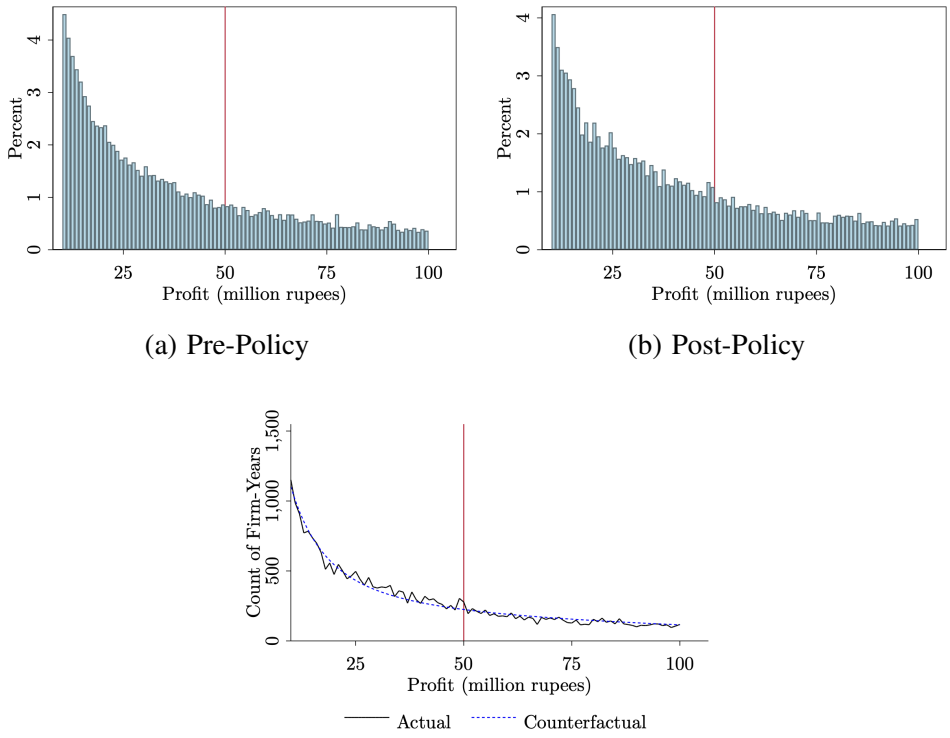


(a) income

(b) net worth

(c) profit

Figure A2. Thresholds of total income, net worth, and profits, conditional. Data is from Prowess (2007-2019), on the firm-year level. The red lines indicate the policy threshold, and the blue lines indicate the cumulative distribution, conditional on the other two variables being below their respective thresholds. This indicates that profits are the binding measure for the policy. For more details see Discussion A2.



(a) Pre-Policy

(b) Post-Policy

(c) profit

Figure A3. Distribution of profits. Data is from Prowess (2007-2019), on the firm-year level. Pre-policy refers to years between 2007 and 2014 and post-policy to years between 2015 and 2019. The red lines indicate the policy threshold. The last figure plots the actual and counterfactual profit distribution. This indicates that profits are the binding measure for the policy. For more details see Discussion A2.

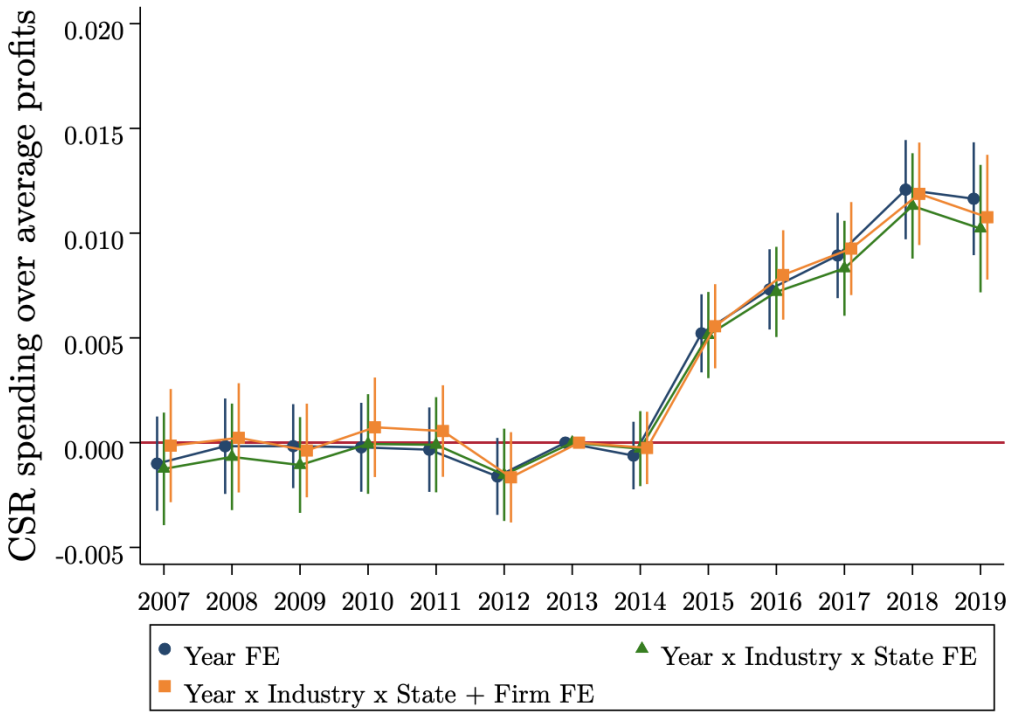


Figure A4. Difference-in-Difference based on 2013 accounting variables. Data is from Prowess (2007-2019), on the firm-year level. Equation 1 describes the regression, with different fixed effect structures. In contrast to Figure 6, treatment is based on 2013 accounting variables. The main outcome of interest is $\text{CSR}_{f,t}/\overline{\text{Profit}}^{3y}_{f,t}$, the CSR spending over average profits in the preceding three years. CSR spending (million rupees) and average profit over the past three years (million rupees) are expressed in real terms, denominated in 2015 rupees. Variables are winsorized at the 99th percentile. The profit measure is additionally winsorized at the 1st percentile. Treatment status is one if the firm's profits, income, or net worth cross the respective thresholds defined by the policy in any of the three preceding financial years. Standard errors are clustered at the firm level. The figure shows 95% confidence intervals.

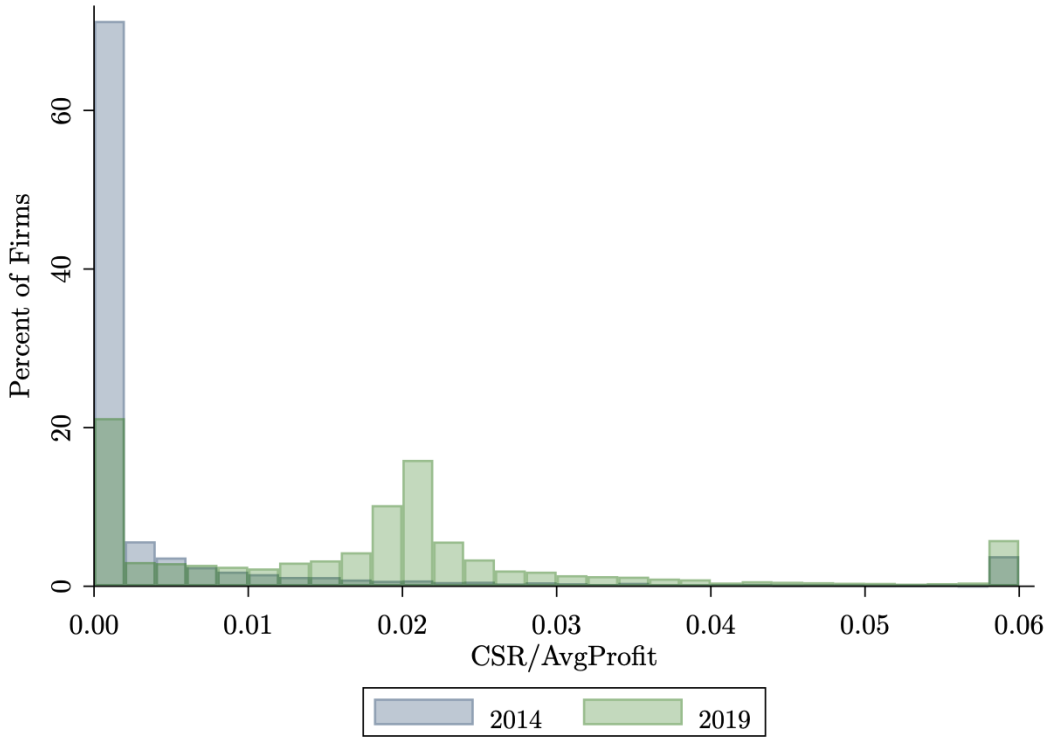


Figure A5. Histogram of CSR spending over average profits for liable firms. Data is from Prowess (2014 and 2019). The main outcome of interest is $\text{CSR}_{f,t}/\overline{\text{Profit}}^{3y}_{f,t}$ for a given year, the CSR spending over average profits in the preceding three years. CSR spending (million rupees) and average profit over the past three years (million rupees) are expressed in real terms, denominated in 2015 rupees. Variables are winsorized at the 99th percentile. The profit measure is additionally winsorized at the 1st percentile. Treatment status is one if the firm's profits, income, or net worth cross the respective thresholds defined by the policy in any of the three preceding financial years.

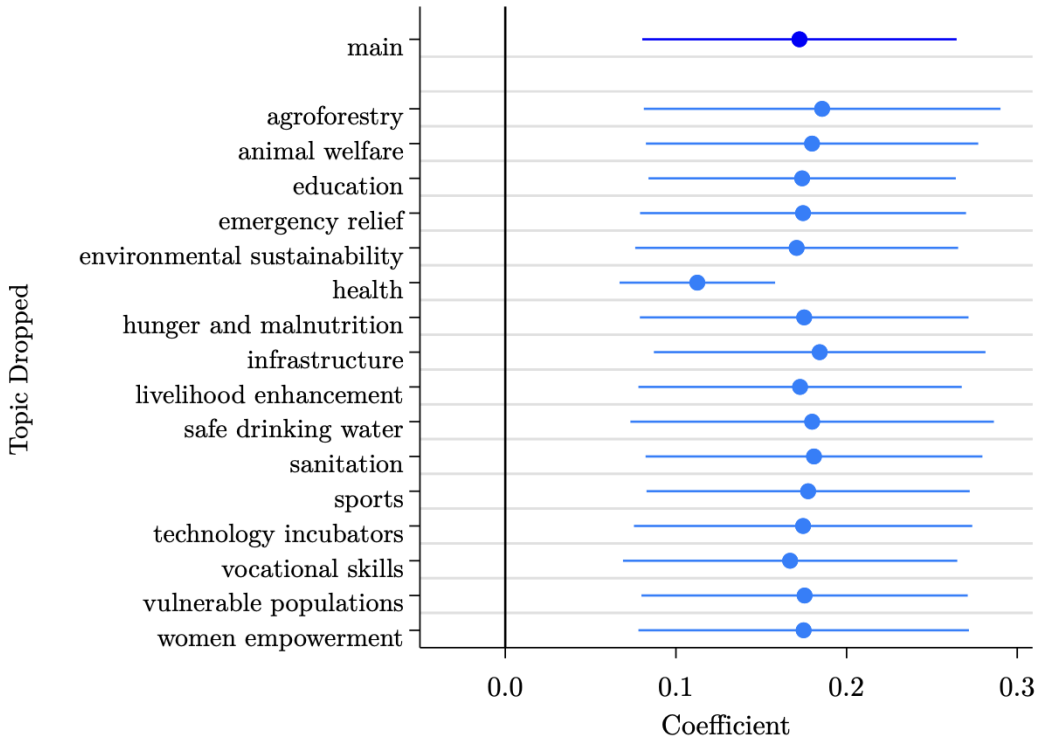


Figure A6. Comparative advantage test robust to dropping topics. Data is from MCA and Prowess (2015-2019). Equation 2 describes the regression. The main outcome is spending share, the share of CSR that a firm allocates to a given topic. The main independent variable is proximity. Each row drops one specific topic from the sample. The figure shows 95% confidence intervals.

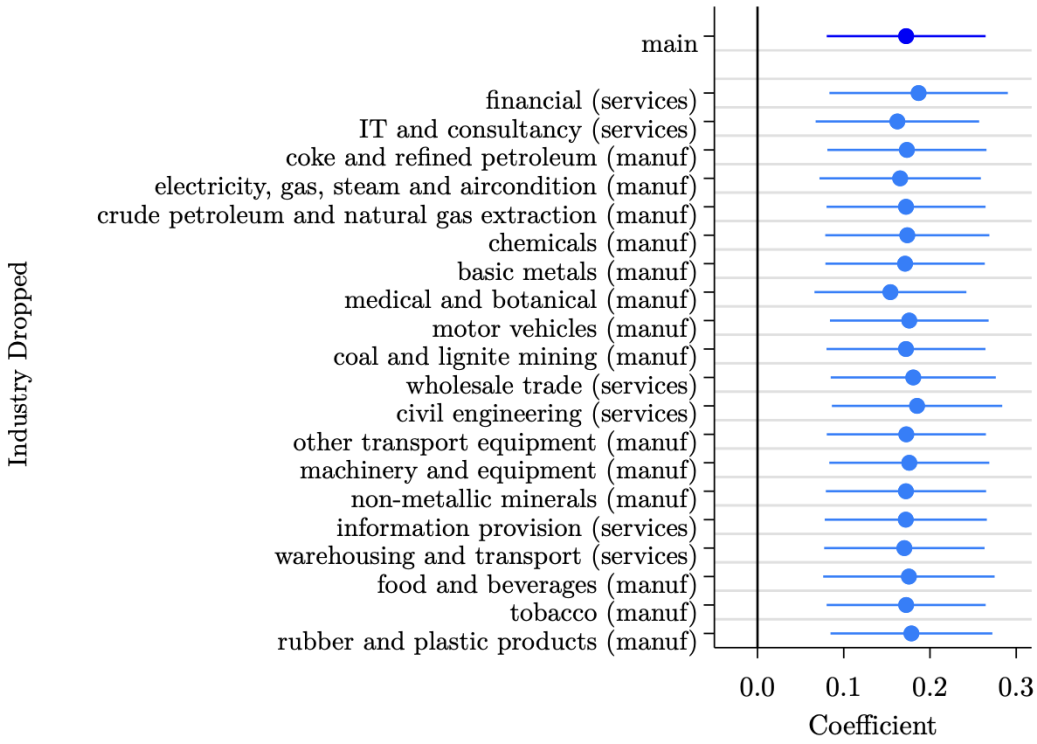


Figure A7. Comparative advantage test robust to dropping industries. Data is from MCA and Prowess (2015-2019). Equation 2 describes the regression. The main outcome is spending share, the share of CSR that a firm allocates to a given topic. The main independent variable is proximity. Each row drops one specific top-20 industry from the sample. The figure shows 95% confidence intervals.

Discussions

Discussion A1: CSR spending variables in Prowess. Our Prowess CSR spending in Table A2 is the sum of two variables reported by Prowess: social and community expenses as well as donations. Social and community expenses are expenses incurred by firms for the benefit of society in general. Donations are reported separately in Prowess. They 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. Due to the close relationship of these two variables, we define CSR as the sum of the two. 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 for estimating the policy impact, which requires pre- and post-policy data. As expected, this explicitly collected CSR data closely maps our constructed version, as can be observed in Table A3; the means are INR 3.42 million and INR 3.76 million, respectively.

Discussion A2: Manipulation of treatment status. We next explore the degree of manipulation by firms under the policy. First, we discuss that firms may try to manipulate their financial accounts so as not to appear liable to the legislator. This could lead to a bias in our causal estimate since firms that manipulate their earnings to fall just below the threshold are also those with the highest dis-utility for CSR spending. In other words, the positive effect on CSR spending could arise from a tendency of firms just below the threshold to spend less. To investigate whether firms manipulate their financial accounts, we initially examine which of the three thresholds is binding. Liable firms 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 shown in Figure A1, only 6% of firm-year observations lie above the income threshold. Similarly, only 6% of firm-year observations are above the net worth threshold. In contrast, 29% of firm-year observations cross the profit threshold. We, therefore, expect that profits are the main determinant of a firm's liability status. To further investigate this, we look at the distributions of each variable, conditional on the two other variables being below their respective thresholds, plotted in Figure A2. The probability that a firm in a given year becomes liable because income or net worth cross their thresholds while the two other variables remain below is less than 1%. On the other hand, given that income and net worth are below their thresholds, the probability that profits cross the threshold is 24%. Therefore, the key sorting variable is profits.

We first provide visual evidence of the profit distribution. We plot the distribution before and after 2014 in Figure A3. The distributions are limited to the range of INR 10 million to INR 100 million, relatively close to the threshold of INR 50 million. Indeed, we observe some evidence of bunching post-policy. However, visually, the distortion appears to be small, suggesting that only a few firms manipulated their accounts. To quantify the magnitude of the distortion, we characterize the abnormal mass in the distribution by estimating the counterfactual distribution, absent the threshold, as in Chetty et al. (2011). We use the range from zero profits to INR 200 million. Figure A3 plots the actual and counterfactual distribution. The difference yields the abnormal mass to the left of the threshold and the missing mass to the right of the cutoff. We observe that the counterfactual distribution is remarkably similar to the actual distribution. The missing mass to the right of the cutoff corresponds to around 94 firm$\times$year observations, which is very small compared to the almost 23,000 firm$\times$year observations located to the right of the cutoff. This suggests that bunching exists but is minimal. Finally, to further alleviate concerns of bias due to bunching, we utilize treatment definitions based on accounting variables from 2013, the last year before the policy was known. We repeat our difference-in-difference analysis with this new treatment, depicted in Figure A4. Results look very similar to Figure 6 with the original treatment definition. Additionally, we compute the Wald-DiD estimator in columns (4) to (6) of Table A5, instrumenting our previous treatment status with treatment status based on 2013 accounting variables. We find that estimates are very similar to before, providing further evidence that the influence of bunching is minimal.

Discussion A3: Manipulation of CSR reporting. Firms have a second potential dimension on which they can manipulate: They could manipulate their CSR spending reports, reallocating expenses on other categories into this category. According to the policy, CSR spent within the firm in pursuance of its normal course of business, e.g., employment welfare, does not qualify as CSR. To investigate whether firms reallocate spending, we repeat our DiD analysis with accounting categories from which firms could have reallocated. The results are depicted in Table A6 show that there is no evidence of manipulation via CSR relabelling. Finally, one might also raise the concern that firms had CSR spending before, but did not report them. With the policy, they might start reporting CSR for the first time. One imperfect test for this is whether firms are more likely to have any CSR reporting after the policy. Indeed, we find this is the case. Note, however, that this positive effect might simply come from firms truly occurring CSR for the first time. This result is thus no evidence that firms engage in this type of manipulation.

Table B1: Example of text recovered from NIC handbook (division 16, Manufacture of wood and products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials)

.

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.