The Allocation of Corporate Social Responsibility Spending

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

November 2026

Despite the push for corporate social responsibility (CSR), what firms' social spending actually funds remains largely unobserved. Assembling the first dataset on the quasi-universe of corporate social projects in a major economy, India, we document the allocation of CSR spending. Two patterns bear on whether firms should pursue social objectives. First, firms spend more on causes close to their for-profit technology, a necessary condition for CSR to improve welfare. Second, firms spend where they operate, directing CSR toward richer areas, where social returns are likely lower. ESG scores predict neither pattern, so allocations provide new information.

Keywords: corporate social responsibility, ESG, textual analysis
JEL Codes: D64, G30, G38, O16

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

1 Introduction

Firms and investors devote growing resources to social objectives beyond profits. Globally, nearly all of the 250 largest companies report on sustainability, and 25% of assets under management follow sustainability criteria; in the United States, corporations devote roughly 1% of total profits to charitable contributions.[1] Alongside this trend, an active academic debate asks whether firms should pursue social objectives at all. For decision-makers and academics alike, systematic information to assess these social activities is scarce, limited to firms' aggregate spending figures and third-party ESG scores. Hence, a key question remains open: what do firms actually do when they engage in social activities?

In this paper, we assemble the first comprehensive dataset covering the quasi-universe of corporate social projects in a major economy, and use it to establish how firms allocate social spending across social causes and across space. India's 2013 Companies Act requires large firms to spend 2% of profits on a specified list of social causes and report on all their CSR projects. This enables us to construct a dataset on all CSR expenditures of the 6,500 largest Indian firms over the period 2015–2019, together with a detailed description of each CSR project. Annually, this spending amounts to 0.1% of GDP.[2]

Using this data, we describe the allocation of CSR spending across social causes and locations. Our analysis yields three main findings. First, firms allocate more to causes that are technologically close to their own production process. The theoretical literature identifies firms' technological advantage in producing public goods as the key rationale for CSR, and we provide the first evidence that firms' allocations follow it. Second, firms spend more in richer states, which partly reflects where they operate. Assuming social returns negatively correlate with local development, this implies that CSR flows away from the places where returns are highest. These patterns hold for firms that spend more on CSR than is mandated, suggesting our conclusions do not hinge on CSR being mandated rather than voluntary. Finally, these patterns of allocations, which bear on the welfare effects of CSR, are not captured by existing ESG scores.

We begin with a description of the CSR allocation in aggregate. Considering the allocation across 16 social causes, education and health projects receive the largest shares, followed

[1] KPMG (2022); Global Sustainable Investment Alliance (2023); Giving USA Foundation (2026).
[2] We will make the data publicly available upon publication.

by rural infrastructure and environmental sustainability. The CSR portfolio resembles that of other providers of public goods: the allocation across causes highly correlates with both government spending shares and the distribution of NGO activity. This aggregate conceals specialization at the firm level: one third of firms devote more than 90% of their spending to a single cause, and firms with similar product portfolios choose similar causes. Across locations, spending is concentrated in a few states, so that expenditure per capita is highly unequal across Indian states.

We develop a simple framework that characterizes firms' allocation and its welfare implications and yields testable hypotheses. A firm allocates a fixed CSR budget across project types, each defined by a cause and a location, and its allocation reflects how much it values each type and how productive it is at each. Because CSR projects and for-profit goods draw on the same firm technology, project-level productivity rises with the proximity between the firm's for-profit technology and the project's technological requirements; we call this the firm's technological advantage.[3] Comparing firms' allocation to a social planner's isolates two properties that matter for welfare. The first is whether spending follows technological advantage. This is the key argument in favor of firms pursuing social objectives: firms can provide public goods more efficiently than shareholder donations when their for-profit technology also serves their social projects (Besley and Ghatak, 2007; Hart and Zingales, 2022). The second is whether spending follows social returns. Firms could instead engage in social activities for strategic reasons and allocate to maximize private returns at the expense of social ones (Baron, 2001). Neither property needs to hold in the privately optimal allocation; whether it does depends on how firms' preferences across project types correlate with technological advantage and with social returns. Whether they hold is therefore an empirical question. In addition, the relevant empirical objects are correlations rather than causal effects: welfare depends on which firms fund which projects, not on why they do so.[4]

Measuring technological advantage requires a proxy for how close each firm's production process is to a social cause. We construct one from two bodies of text: the industry classification guidelines, which describe the products and production technologies common to firms in each industry, and the descriptions of the projects that make up each cause in the CSR

[3]For example, a healthcare firm running a screening campaign, an engineering firm building rural roads.

[4]For instance, a firm may favor technologically close causes because it is good at them or because it happens to prefer them, and the welfare benefits are the same either way. We leave the question of firms' motives to data that can pin them down.

data. We represent both as vectors of word embeddings and define technological proximity as the cosine similarity between them, obtaining a distinct value for each of 1,136 industry-by-cause pairs. Our key assumption is that semantic proximity between these texts proxies for technological proximity; we derive three tests of this assumption from our framework and show that they hold in our data.

We find that firms allocate more to causes they have a technological advantage in, on both the extensive and the intensive margin. A one-standard-deviation increase in the proximity between a firm's industry and a cause is associated with a one-percentage-point higher spending share on that cause, 16% relative to the mean. This result is highly robust across specifications. Further partitioning CSR projects into 83 finer causes using their descriptions or aggregating them into 10 coarser causes yields nearly identical magnitudes. Our results are robust to using alternative embedding models, an alternative textual source for industries, and dropping any single cause or industry in turn. Findings are also very similar in the sample of firms that spend more on CSR than the amount required by the law. Seen through the lens of the theoretical CSR literature, this result implies that CSR has the potential to be welfare-improving. In terms of magnitude, we find that technological proximity explains roughly one-fifth to one-fourth as much of the variation in firms' CSR allocation across causes as the government's spending share, taken as a proxy for shared priorities across causes.

We next consider firms' CSR allocation across locations. We find that states with 10% higher GDP per capita receive 19% more CSR spending per capita. A similar correlation - more spending in more developed locations - holds across smaller geographical units and for alternative measures of development. Proxying instead for cause-state-specific needs using development indicators (e.g., literacy for the education cause), we find a similar result. This partially reflects home bias: firms spend close to where they operate, and are concentrated in rich states. Excluding spending in headquarter states, the correlation falls by about half but remains statistically significant, so proximity to a firm's own operations cannot account for all of it.[5] Spending close to home may be efficient: firms may know local needs better, or monitor nearby projects more cheaply. In any case, under the assumption that public goods yield higher social returns in poorer areas, these patterns suggest that CSR spending does not follow social returns.

[5]Excluding not only headquarters but also establishments' states - when they can be observed - yields similar results.

Investors cannot systematically observe firms' CSR allocation across projects; the information they use to allocate capital is typically an ESG score. We test whether ESG scores carry information on allocation. To investigate this, we correlate firms' allocations with two scores: the commercial score from Refinitiv, and a score we construct from the mandatory sustainability disclosures of India's largest listed firms. These cover large firms responsible for around 60% of the CSR spending in our sample. We find that firms with high ESG scores are no more likely to direct spending toward causes close to their technology, and no more likely to spend in poorer states. Scores and spending allocations across causes and space therefore capture different dimensions of firms' social activity.

Overall, firms fund projects close to their own production technology, consistent with the efficiency argument for CSR; at the same time, they fund projects in more developed locations, partly following their own geographic footprint, where social returns are likely lower. These patterns - invisible in spending totals and ESG scores - are relevant to investors evaluating firms' ESG policy, to policymakers designing mandates, and to the academic debate on whether firms should pursue social objectives.

Related literature. Our first contribution is to the empirical literature on CSR. The vast majority of this literature focuses on the relationship between firms' CSR activities and their financial outcomes (see Gillan et al., 2021; Edmans and Kacperczyk, 2022; Hong and Shore, 2023; Starks, 2023, for reviews). More recent contributions focus on measuring firms' social impact or characterize the diverse stakeholder preferences underpinning firms' prosocial stances.[6] In this literature, researchers typically observe firms' social activities through one of three measures: a total spending amount that some large firms disclose, an ESG score, or the grants listed in the tax filings of corporate foundations, which identify a recipient but not an activity (Bertrand et al., 2020). We complement this literature by leveraging data on CSR projects to shed light on the allocation of social spending, and examine how properties of this allocation inform the debate on the value of firms' social activities. Our approach in particular complements that in Fioretti (2022), who uses detailed data on all activities of

[6]See e.g., Flammer and Luo (2017); Lins et al. (2017); Liang and Renneboog (2017); Riedl and Smeets (2017); Bertrand et al. (2020); Krueger et al. (2020); Gibson Brandon et al. (2022); Allcott et al. (2023); Cheng et al. (2023); Christensen et al. (2023); Colonnelli et al. (2023); Fioretti et al. (2023); Hartzmark and Shue (2023); Kahn et al. (2023); Conway and Boxell (2024); Green and Vallee (2024); Ferreira and Nikolowa (2025).

one firm to show that it acts prosocially beyond profit maximization, increasing welfare. In contrast, we consider the quasi-universe of CSR spending in a context in which the amount of prosocial spending is given and, through the lens of our framework, examine whether its allocation is consistent with welfare maximization.[7]

Second, our paper relates to the theoretical literature on whether firms should pursue social objectives. We propose an empirical test for a key assumption in this literature: that firms allocate their social spending according to the technological advantage that stems from their for-profit production technology (see Kitzmueller and Shimshack, 2012, for a review). In most models, it is a necessary condition for CSR spending to increase welfare relative to individual donations—whether it is the welfare of shareholders (Hart and Zingales, 2017, 2022) or of society (Besley and Ghatak, 2007; Magill et al., 2015).[8] This paper is, to the best of our knowledge, the first to test and validate this assumption.

Third, the paper also speaks to a broader literature on how institutional and financial reforms shape firm behavior and economic outcomes in developing countries (see, e.g., Ponticelli and Alencar, 2016; Colonnelli and Prem, 2022; Fonseca and Van Doornik, 2022; Barboni and Agarwal, 2023; Higgins, 2024). Policymakers and investors increasingly look to corporate and ESG-oriented capital to fund development. India's CSR mandate offers a large-scale case of such a channel, and similar mandates have been implemented in several developing countries (Lin, 2021). Our results unveil the allocation of these privately-directed development finance flows.[9]

[7]Previous contributions have examined CSR in the Indian context in particular. In the accounting literature, Manchiraju and Rajgopal (2017); Dharmapala and Khanna (2018); Mukherjee et al. (2018); Bhattacharyya and Rahman (2019) investigate the effect of the CSR mandate on firm value, focusing on listed firms. In the strategy literature, Gatignon and Bode (2023) provide a descriptive analysis of Indian firms' CSR strategies. Chhaochharia et al. (2025) consider the mandate's effect on school enrollment, Rajgopal and Tantri (2023) its impact on firms spending voluntarily on CSR.

[8]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 technological advantage motivation.

[9]In that perspective, our results also relate to work on financing development under tax capacity constraints (e.g., Besley and Persson, 2009; Best et al., 2015; Gadenne, 2017; Jensen, 2022; Bergeron et al., 2024) and the private provision of public goods (Hart et al., 1997; Kotchen, 2006; Behaghel et al., 2014; Mukherjee, 2021; Knutsson and Tyrefors, 2022). The latter emphasizes a tradeoff between the efficiency gains of private

Methodologically, our work relates to the growing literature using semantic distance to capture distance in economically relevant space (Gentzkow et al., 2019; Ash and Hansen, 2023), and in particular Hoberg and Phillips (2010, 2016), who use textual analysis to characterize the product space in which firms compete.[10] We extend this approach to the space between firms and social causes, measuring how close a firm's for-profit technology is to the projects that make up each cause, and derive model-based tests to validate that word embeddings adequately capture technological properties of firms' production processes.

The paper proceeds as follows. Section 2 describes the setting and data and provides evidence on the implementation of India's CSR mandate. Section 3 documents key facts about the allocation of CSR expenditure. Section 4 develops a conceptual framework for that allocation, section 5 examines whether spending follows technological advantage, and section 6 investigates whether it follows social returns. Section 7 asks whether ESG scores reflect these allocation properties. Section 8 concludes.

2 Context and Data

2.1 Corporate Social Responsibility in India

In August 2013, India passed into law section 135 of the Companies Act, which mandates that large firms spend at least 2% of their average profits over the last three years on CSR activities. It came into effect in April 2014. Large firms are defined as those with profits above INR 50 million, income above INR 10 billion, or a net worth above INR 5 billion in any of the three preceding financial years.[11] These firms represent a large share of the Indian economy, corresponding to approximately 60% of formal sector activity. The act specifies the activities that qualify as CSR expenditures, clarifies that spending occurring within the 'normal course of business' (e.g., employee welfare) does not qualify, and imposes the formation of a CSR committee with at least one independent director. Importantly

provision and adverse distributional effects; our findings suggest a similar tradeoff.

[10] Beyond product markets, the literature has used text-based measures to describe many firm characteristics: for instance, financial constraints, risk exposures of firms, the information in firms' own disclosure, and corporate culture (Hoberg and Maksimovic, 2015; Hanley and Hoberg, 2019; Cohen et al., 2020; Li et al., 2021; Sautner et al., 2023).

[11] These thresholds are not associated with any other requirements in Indian law. The act also suggests firms should give preference to spending in their local area.

for our purposes, it also makes reporting of all CSR activities to the Ministry of Corporate Affairs (MCA) compulsory. During our study period (2015–2019), the mandate was enforced on a comply-or-explain basis, and since 2019 fines have been imposed for non-compliance. For more details on the provisions of the mandate, see Appendix D.1. Over the period 2015–2019, the total annual CSR expenditure is 142,669 million INR (2,283 million USD) on average, equivalent to 0.1% of GDP.[12] We return to discussing the effects of the law on CSR expenditures after describing our data.

2.2 Data

CSR data. Our main data source is the compulsory reporting of CSR activities to the MCA. Since the fiscal year 2014–2015 (hereafter 2015), all liable firms report on each of their CSR projects. The data is available on the MCA website and contains, for each CSR project, the amount spent, the CSR topic the project belongs to (from a broad pre-specified list defined in the law), a textual description of the project, the location of the project, and an indicator for whether the project was implemented directly by the firm or indirectly via a third party (typically an NGO). After the cleaning procedures outlined in Appendix C, the CSR data contains information on 124,813 projects conducted by 11,487 firms over the period 2015–2019. From the 28 CSR topics specified by the law and available in the CSR data, we group similar topics to obtain the 16 topics considered in our analysis.[13] Because projects often span multiple years, we aggregate data across years in what follows.

Location is not reported for all projects. We observe the state in which a project occurs for projects totaling two-thirds of CSR spending; among those, we observe the district for 27% of expenditures. Overall, we observe CSR expenditures in all 35 states and in 496 districts (78% of all districts).[14]

[12] This is similar to the share observed in the US, where charitable contributions by corporations represented 43.67 billion USD in 2025, roughly 0.15% of GDP (Giving USA Foundation, 2026). Throughout the paper, we denominate in 2015 INR and apply an INR to USD exchange rate of 0.016.

[13] This re-classification exploits the project descriptions to group together topics that contain few projects and are conceptually very similar, e.g., 'environmental sustainability' and 'conservation of natural resources'. See Appendix C.1.4 for a detailed description of the method.

[14] When we do not observe the state, this is either because the variable is not filled in the original data or because the project is specified as occurring in the whole of India. In our regression analysis in Section 6 below, we exclude four states with a population lower than one million, as well as the small state of Chandigarh, for which government spending data is not available.

To the best of our knowledge, this is the most comprehensive dataset on CSR activities for any country in the world. It is comparable in scope to data on charitable giving by individuals in the US compiled by the Giving USA Foundation (see List, 2011) but contains more information on project types and, crucially for our purposes, provides the official Corporate Identification Number (CIN) of each firm.

Accounting data. We combine CSR data with accounting data to obtain additional information on firms. We use the Prowess database from the Center for Monitoring the Indian Economy, which includes information from the income statements and balance sheets of all publicly traded firms as well as a large number of private firms. It is the standard source for information on Indian companies and is widely used in the corporate finance literature (e.g., Khanna and Palepu, 2000; Bertrand et al., 2002; Gopalan et al., 2007; Vig, 2013).

From this data, we obtain information on firms' industries at the 2-digit level, which follows the National Industry Classification (NIC). We use information on CSR expenditures reported in balance sheet statements to examine the implementation of the reform, using data from 2007 onward. We also obtain information on firms' geographical locations: the data contains the location of firms' headquarters, and we use the CapEx dataset from the Center for Monitoring the Indian Economy to measure the location of large capital investments since 1995 to proxy for the presence of firm establishments (following Naaraayanan and Wolfenzon (2024)). In both cases 'location' refers to a state and, where available, a district (see Appendix C.2.5 for more details). Finally, we obtain firm characteristics for the heterogeneity analysis from the data.

We merge the CSR and Prowess accounting data at the firm level using firms' CINs. The Prowess accounting data does not, by design, include all Indian firms and has better coverage of large firms. We match 61% of firms and 91% of CSR expenditures in the CSR data, as well as 99% of the post-2015 CSR expenditures reported in the accounting data. After the merge and a second set of cleaning steps outlined in Appendix C, our main analysis sample consists of 86,334 projects by 6,573 firms distributed across 71 2-digit industries.

Table A.1 shows descriptive statistics of the variables used in our analysis. Firms are large, and the distribution has a long right tail. We systematically present results with and without weighting by firm size (proxied by total CSR expenditures) in what follows. Figure A.1 plots the distribution of CSR expenditures by industry for the 20 largest industries in

our data. The CSR shares follow a distribution similar to that of value-added per industry for India, as expected given that CSR expenditures are a function of profits.

Textual data. Our analysis exploits the description of projects in the CSR data. To use this text, we perform a number of data cleaning steps described in Appendix C.1.2. In particular, we filter out uninformative tokens (words) and observations. After cleaning, the average project observation contains 4.3 informative tokens (standard deviation is 4.0) and the average CSR topic contains 23,509 tokens (standard deviation is 30,763).[15] Figures C.2–C.4 provide a visualization of this data by showing word clouds for the project descriptions by topic.

We encode the textual data using word embeddings. Word embeddings are a natural language processing method in which individual words are represented as real-valued vectors in a high-dimensional space. These vectors are meant to capture the meaning of words so that similar words have similar vectors. In addition, an internally consistent geometry on the vector space allows words to be related. We use the word embeddings provided by the pre-trained Word2Vec model released by Google. The model contains 300-dimensional vectors for 3 million words. We obtain a vector representation of the text describing each CSR project p, denoted $\mathbf{v}_p$. The details of the implementation of Word2Vec are in Appendix C.1.3. We use Word2Vec embeddings as our baseline due to their methodological transparency: representing documents as the mean of token-level embeddings enables intuitive and interpretable exploration of semantic patterns. We present results using OpenAI's NLP model as a robustness check.

In section 5, we additionally exploit textual data characterizing the firms' industries. For each 2-digit industry, the Handbook of the National Industrial Classification provides a description of the products and production technologies common to firms in the industry. After cleaning, this text yields an average of 250 informative tokens per industry (standard deviation is 225). Table C.4 shows an example of text for one industry. Using word embeddings, we obtain a vector representation of the text describing each industry i, denoted $\mathbf{v}_i$. In robustness tests, we exploit SEC 10-K filings for US firms to produce an alternative corpus describing industries (see Appendix C.2 for details). The magnitude of embedding vectors does not encode meaningful information, so we normalize all embedding vectors to

[15] Before filtering out uninformative tokens, the average is 7.4 tokens per observation.

have norm 1.

Notation. Throughout the text, we use the notation $\|\mathbf{x}\|$ to denote the Euclidean norm of vector $\mathbf{x}$ and $\cos(\mathbf{x}, \mathbf{y})$ for the cosine similarity between vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}.$$

ESG scores. For a subset of firms, we observe two ESG scores. The first is the Refinitiv ESG score, a widely used commercial score that summarizes a firm's environmental, social, and governance practices into a single number; it covers the larger listed firms in our sample, 449 in total (for construction details, see London Stock Exchange Group (2024)). The second is a measure we construct ourselves from firms' Business Responsibility and Sustainability Reports (BRSR). The BRSR is a standardized sustainability disclosure that the Securities and Exchange Board of India (SEBI), the securities market regulator, requires of the largest listed companies (Securities and Exchange Board of India, 2021). We collect the reports that firms file with the National Stock Exchange, India's largest stock exchange by equity trading volume.

To construct a BRSR score, we concentrate on the indicators that the circular (Securities and Exchange Board of India, 2023) defines as BRSR Core, which groups nine ESG attributes and specifies the indicators a firm must report under each. Appendix C.2.6 describes details on the construction, including the indicators that enter the score. We sign each indicator so that a higher value is better, convert it to a percentile rank among that year's filers, average the ranks within each attribute, and weight the attributes equally.[16] The resulting score runs from 0 to 100 and covers 722 firms. We also describe robustness to the construction of this score.

We obtain a firm-level average of the ESG scores. The CSR data ends in 2019. ESG scores are infrequently filled in these years in Refinitiv (spanning 2011-2023) and are not available in BRSR (spanning 2023-2026). We therefore average each score over all years in which a firm is scored, capturing a firm's sustained ESG profile as observed after the CSR data period.

[16] This approach largely follows the scoring approach of London Stock Exchange Group (2024); details are provided in Appendix C.2.6.

Firms with ESG scores are relatively few but large: they account for 59% (Refinitiv) and 61% (BRSR) of the CSR spending in our sample. The two scores are complementary: the commercial Refinitiv score is a signal investors frequently use, while the disclosure-based score covers more firms and is less influenced by voluntary reporting. Notably, the two agree only weakly: among the 418 firms both cover, their correlation is 0.28, in line with a correlation range of 0.38 to 0.71 among six commercial ESG raters documented by Berg et al. (2022).

2.3 Implementation of the CSR Mandate

This section briefly describes evidence on the implementation of the CSR mandate. In Figure 1(a), we plot the evolution of total CSR expenditures in India over time, as reported in the accounting data. We see a large increase from 2015 onward: aggregate CSR spending roughly tripled since the mandate was implemented. Figure 1(b) plots the evolution of CSR spending as a share of profits separately among liable firms (defined as firms whose income, profits, or net worth are above the thresholds defined in the law) and all other firms in the accounting data. The aggregate increase in CSR spending comes from liable firms. Appendix D investigates the evolution of CSR expenditures in liable and non-liable firms over the period more formally by conducting a difference-in-differences exercise (see in particular Figure D.1). Results suggest the mandate led to an increase in CSR spending as a share of profits of 1 percentage point in the first year of its implementation, and up to nearly 1.5 points at the end of the period. The average profit share of CSR among liable firms is 2.3% at the end of the period, suggesting the mandate was well respected overall (see also Chhaochharia et al., 2025).[17]

Figure 1(b) also shows that many firms spend on CSR beyond the scope of the mandate: 13% of firms in our main sample spend more than 1% of profits on CSR in 2014 (prior to the mandate), 18% spend more than 2.5% after 2015. Together, these two sets of firms represent

[17] Appendix D also considers whether firms reclassified pre-existing expenditures as CSR after 2015 to reach the mandated threshold. After the mandate, we see that expenditures on donations to charities decrease while CSR expenditures (a distinct budget line in Prowess) increase. Most charitable donations enter the scope of the CSR mandate, so this suggests that firms understand the mandate and are respecting it by classifying these expenditures as CSR. We see small or no effects on other expenditures (e.g., employee welfare) that do not belong to the scope of the mandate and may have been opportunistically re-classified as CSR. We also test whether the increase in CSR expenditures crowded out government expenditures at the state level, but are underpowered to conclude with certainty on that question (see Table D.4).

24% of our sample and constitute what we call the 'voluntary CSR' sample. Their behavior may indicate intrinsic preferences for CSR activities. In what follows, we systematically consider whether these firms allocate their expenditures differently. This allows us to investigate whether our conclusions hinge on the spending being mandated.

3 Key Facts About CSR

What firms actually do when they engage in social activities has largely remained a black box: spending levels and ESG scores are observable, but the underlying activities are not. This section uses our data to open this black box. We document four key facts on the allocation of CSR spending by firms in India, which show that firms' choices are systematic along two dimensions—which social topics firms fund, and where. These facts motivate the two questions the rest of the paper takes up: whether the allocation across topics follows firms' technological advantage, and whether the allocation across locations follows social returns.

We obtain these facts on our main analysis sample and systematically reproduce them on i) the sample of all firms in the CSR data (before the merge with the accounting data) and ii) the voluntary CSR sample; we find very similar patterns (see Appendix B for more details).[18]

Fact 1: CSR spending is concentrated in health and education. Table 1 shows the allocation of CSR spending across topics. This table also clarifies the meaning of the topics by listing the three most common project types within each topic. For each topic, the most common project types are identified by partitioning projects into types by estimating a k-means clustering algorithm on project embeddings (see also Figures C.2–C.4 for word clouds for each topic).[19] We see that firms finance a wide range of projects and that projects typically do not involve the type of products that firms sell.

The largest social topic in terms of spending is education (32% of the total). Common education projects involve school construction or renovation and the promotion of education

[18] Facts 1 and 4 on the allocation across topics and states have overlaps with official CSR reports (Ministry of Corporate Affairs, 2021) and the descriptive analysis in Gatignon and Bode (2023).

[19] See the full list of clusters and implementation details in Appendix C.3.

for differently-abled children. The second largest topic is health (17% of spending), with projects focused on preventive healthcare, medical treatment, medical equipment, or mobile health camps. Rural infrastructure and environmental sustainability follow, with 8% of spending each. Rural infrastructure involves mostly small-scale infrastructure in rural areas (e.g., rural roads, street lights); for environmental sustainability, conservation and tree plantation are the most frequent types of projects. The other social topics all receive equal to or less than 6% of spending. In section 5, we ask whether this allocation follows firms' technological advantage.

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

The allocation of CSR spending across topics is significantly correlated with that of government spending and NGO activity. In both cases, the pairwise correlation is around 0.8 and statistically significant at the 1% level. Notably, the three types of public goods providers allocate almost precisely the same share to education. Firms differ from the government and NGOs in that they allocate less to vulnerable populations and more to industry and technology, vocational skills, and water and sanitation projects. Overall, Figure 2 suggests that different public good providers agree to a large extent on the relative valuation of public goods across topics. We return to this when we compare the role of technological advantage in explaining the allocation of CSR across topics to that of preferences for topics common to firms and the government.

Fact 3: Firms specialize in topics, and similar firms specialize similarly. Firms' CSR spending is highly concentrated across topics. The firm-level distribution of spending shares across topics has an average Herfindahl-Hirschman Index equal to 0.63, and 34% of firms allocate more than 90% of their spending to only one topic. This is not only the result of indivisibilities: in the sample of firms reporting multiple projects, 20% of firms allocate more than 90% of their spending to only one topic.

Moreover, firms that sell similar products tend to allocate their CSR expenditures across topics similarly. We regress the cosine similarity between firms' CSR shares across topics on the cosine similarity between firms' sales shares across products and find a statistically significant correlation (see Table A.2). These specialization patterns suggest a link between firms' for-profit production processes and their choice of CSR spending. This is a key building block of our conceptual framework in section 4.

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

4 Conceptual Framework

The facts above raise two questions. The first is positive: what drives firms' choices of topics and locations? The second is normative: are these the choices in line with those a social planner would make, and what does this imply for the debate on whether firms should pursue social objectives? This section develops a simple framework that addresses both.

In this framework, a firm allocates CSR expenditures across project types by considering how much it values each type and how productive it is at each. Firm-project specific productivity increases in the proximity between the firm's for-profit technology and the project's technological requirements. Turning to the normative implications, comparing firms' allocation across projects to that of a social planner highlights two properties that matter for welfare: whether spending follows firms' technological advantage stemming from their for-profit technology, as the key argument in favor of CSR presumes (Besley and Ghatak, 2007; Hart and Zingales, 2022), and whether it follows social returns. We take these to the data in sections 5 and 6. Model derivations are in Appendix E.

4.1 Model

Set-up. Production occurs across projects, indexed by $p \in \mathcal{P}$. $\mathcal{P}$ can be partitioned into $\mathcal{P}^{CSR}$ for CSR projects and $\mathcal{P}^{pri}$ for for-profit goods sold in competitive markets. Our object of interest is how firms f allocate CSR spending, constrained to be higher than an exogenous amount E, across CSR project types $p \in \mathcal{P}^{CSR}$. In our application, we think of CSR project types as being characterized by a social topic and a location.

Projects are produced by combining tasks $\tau \in \mathcal{T}$. The relative importance of tasks varies across projects as characterized by the vector $\boldsymbol{\phi}_\mathbf{p} = [\phi_{p\tau}]_{\tau\in\mathcal{T}}$ such that $\forall p,\ \sum_\tau \phi_{p\tau} = 1$. Firms are endowed with a task-specific productivity vector $\mathbf{z_f} = [z_{f\tau}]_{\tau\in\mathcal{T}}$ used to produce both for-profit and CSR projects.[20] Firm f's output in project p is:

$$y_{fp} = \left[\prod_{\tau\in\mathcal{T}} (\exp(z_{f\tau})\ell_{fp\tau})^{\phi_{p\tau}}\right]^\rho$$

where $\rho < 1$. $\ell_{fp\tau}$ is the labor assigned by firm f to task τ and project type p, and we denote $\ell_{fp} = \sum_\tau \ell_{fp\tau}$ total labor allocated to project p. The wage w is taken to be exogenous.

$\boldsymbol{\phi}_\mathbf{p}$ captures the technological requirements of project p. We interpret tasks in a broad sense as any type of work, material, or immaterial input that is required to complete a project. $\mathbf{z_f}$ captures firm productivity at each of the tasks, and can include physical capital, organizational capital, any type of specialized knowledge, know-how, or information. We assume for convenience that $||\boldsymbol{\phi}_\mathbf{p}|| = \phi, \forall p$ and $||\mathbf{z_f}|| = 1, \forall f$, and relax these assumptions in Appendix E.

The firm allocates labor across projects to maximize:

$$\max_{\{\ell_{fp}\}} \sum_{p\in\mathcal{P}^{pri}} (\zeta_{fp} y_{fp} - w\ell_{fp}) + \sum_{p\in\mathcal{P}^{CSR}} (\zeta_{fp} y_{fp} - w\ell_{fp}) \tag{1}$$

When p is a for-profit good, ζ_{fp} is the market price of good p. When p is a CSR project ζ_{fp} captures how much firm f values project type p. These preferences could reflect both private benefits (strategic considerations, e.g., consumer or investor pressure) and/or warm-

[20] Our modeling choices borrow heavily from the task-based production framework (e.g., Acemoglu and Restrepo, 2018), but the key insights are applicable to a general multi-dimensional sorting framework (e.g., Lindenlaub, 2017).

glow utility. The mandate imposes $\sum_{p \in \mathcal{P}^{CSR}} w\ell_{fp} \geq E$; firms whose valuation of CSR is high enough spend more than the mandated amount.

Characterization of firms' optimal CSR allocation. Conditional on allocating labor ℓ_{fp} to producing project p, the firm allocates labor $\ell_{fp\tau}$ across tasks in a way that maximizes project-level profits. The optimal labor allocation satisfies: $\frac{\ell_{fp\tau}}{\ell_{fp}} = \phi_{p\tau}$. This yields an expression for project output as a function of firm-project productivity α_{fp} and project-level labor ℓ_{fp} :

$$y_{fp} = \exp(\alpha_{fp})\ell_{fp}^{\rho}, \text{ with } \alpha_{fp} = \rho\phi \cos(\boldsymbol{\phi}_{\mathbf{p}}, \mathbf{z}_{\mathbf{f}}) + \kappa_p. \tag{2}$$

$\cos(\boldsymbol{\phi}_{\mathbf{p}}, \mathbf{z}_{\mathbf{f}})$ is the cosine similarity between the vector of returns to tasks for that project type $\boldsymbol{\phi}_{\mathbf{p}}$ and the firm's vector of task-specific technologies $\mathbf{z}_{\mathbf{f}}$. κ_p is a project-level constant. The parameter α_{fp} captures firm f's productivity in providing type p. We label this parameter *technological advantage*. Equation (2) shows that firms are more productive in projects that put a large weight on their most productive tasks: technological advantage increases with the proximity between the firm's own technology and the project's technological requirements.

Second, each firm maximizes its utility in (1) subject to $\sum_{p \in \mathcal{P}^{CSR}} w\ell_{fp} \geq E$. This yields an expression for the share of CSR expenditures allocated to project p, as in:

$$s_{fp} = \frac{\left[\zeta_{fp} \exp(\alpha_{fp})\right]^{1/(1-\rho)}}{\sum_{q \in \mathcal{P}^{CSR}} \left[\zeta_{fq} \exp(\alpha_{fq})\right]^{1/(1-\rho)}} \tag{3}$$

The share firm f allocates to project type p is increasing in its (relative) technological advantage in this type α_{fp} and in the preference parameter ζ_{fp}. Equation (3) holds whether or not the mandate binds: shares do not depend on the total amount spent, so a firm spending above E allocates in the same proportions.

Combined with equation (2), firms thus allocate more to projects whose technological requirements are close to their own technology. This logic applies to both CSR projects and for-profit goods. It is in line with resource-based theories of the firm proposing that firms diversify into products that use common capabilities, supported by empirical evidence (Maksimovic and Phillips, 2002; Boehm et al., 2022; Koh and Raval, 2025). This framework implies that firms' projects, both private and public, are determined by a common firm-level productivity vector. A direct implication is that firms making similar for-profit product

choices have a similar vector $\mathbf{z_f}$, and hence make similar CSR allocation choices. This is in line with the evidence in Table A.2 discussed in section 3.

Note that equation (3) holds irrespective of the source of project-specific productivities α_{fp}. The predictions from (3) thus remain unchanged if project-specific productivity comes from, for instance, firms' locations instead of their technologies, making them more productive in projects located close to their private activities.

Implications across industries. In anticipation of our empirical exercise, it is useful to define a notion for industries. We define an industry as a set of private goods $\mathcal{P}_i \subset \mathcal{P}^{pri}$ centered around a technological vector $\boldsymbol{\phi_i}$: $\forall p \in \mathcal{P}_i, \boldsymbol{\phi_p} = \boldsymbol{\phi_i} + \boldsymbol{\epsilon_p}$, with $\boldsymbol{\epsilon_p}$ mean-zero and i.i.d. In addition, we assume that firms belonging to industry i have a productivity vector centered around $\boldsymbol{\phi_i}$: $\mathbf{z_f} = \boldsymbol{\phi_i} + \boldsymbol{\epsilon_f}$. This ensures that firms in industry i have $p \in \mathcal{P}_i$ as their main product, in line with how firms are classified across industries in standard datasets. For a firm f in industry i, we can then write:

$$\alpha_{fp} = \rho\phi \cos(\boldsymbol{\phi_p}, \boldsymbol{\phi_i}) + \kappa_p \tag{4}$$

Firms are more productive in, and hence will all else equal allocate a larger expenditure share to, CSR projects that have technological requirements similar to those of their for-profit industry.

Would a social planner allocate CSR differently? Let social welfare be an increasing function of the projects funded by all firms, defined as:

$$W = \sum_{p \in \mathcal{P}^{CSR}} \mu_p \sum_f y_{fp} \tag{5}$$

where the μ_p terms capture the social returns to project type p.

Maximizing social welfare, taking each firm's total CSR expenditure as given and choosing only its allocation across types, yields:

$$s^*_{fp} = \frac{\left[\mu_p \exp(\alpha_{fp})\right]^{1/(1-\rho)}}{\sum_{q \in \mathcal{P}^{CSR}} \left[\mu_q \exp(\alpha_{fq})\right]^{1/(1-\rho)}} \tag{6}$$

The socially optimal share firm f allocates to a project type p is increasing in its (relative) technological advantage in this type, α_{fp}, and in the social returns parameter μ_p.

Comparing equations (3) and (6), the privately and socially optimal allocations differ only to the extent that firms' preferences across project types ζ_{fp} differ from social returns μ_p. In addition, welfare in (5) depends on firms' preferences only through the shares they choose. That is, welfare only depends on the realized allocation of firm CSR, and not on the reasons why the firm chose this specific allocation.

4.2 Hypotheses Taken to the Data

Does CSR follow firms' technological advantage? We say that the CSR allocation follows firms' technological advantage if, conditional on a project type p, firms with a higher technological advantage allocate a larger share of their spending to that type: $\left(\frac{ds_{fp}}{d\alpha_{fp}}\right)_p > 0$. This property qualitatively characterizes one dimension of the socially optimal allocation: abstracting from differences in social returns across project types, firms that are relatively more productive at producing a given public good should allocate more resources to it. It is also the empirical counterpart of the argument in Besley and Ghatak (2007) and Hart and Zingales (2022). In these models, a necessary condition for CSR to be welfare-improving, in particular relative to shareholders' donations, is that firms leverage their for-profit technology. This condition is relevant only if firms in practice direct their spending to the project types in which that technology gives them an advantage.[21]

From equation (3), when firms internalize social welfare ($\zeta_{fp} = \mu_p, \forall f, \forall p$) or have no preferences across project types ($\zeta_{fp} = \zeta_f, \forall p$), firms do spend more on project types in which they have a technological advantage. However, when firms' preferences across project types differ substantially, this property may not hold.

In section 5, we test for this property by focusing on one dimension of project type along which there is likely substantial variation in technological advantage across firms: social topics, indexed by d. Formally, we test whether the correlation between firms' topic-level technological advantage and spending share is positive. Following (4), we measure techno-

[21] Besley and Ghatak (2007) and Hart and Zingales (2022) consider if CSR can be welfare-improving relative to other contribution mechanisms, while we consider, for a given amount of CSR expenditure, the welfare resulting from an allocation. The two questions are linked: the better CSR is allocated, the stronger the case for it relative to alternative uses of the same funds.

logical advantage as the proximity between the technology of a firm's for-profit industry and the technological requirements of CSR projects in a topic, $\cos(\boldsymbol{\phi_d}, \boldsymbol{\phi_i})$, as detailed below.[22]

Does CSR follow social returns? Does CSR spending increase in social returns μ_p across project types $\frac{ds_{fp}}{d\mu_p} > 0$? The socially optimal allocation satisfies this property unless technological advantage is strongly negatively correlated with social returns. The privately optimal allocation need not satisfy it, since social returns do not enter equation (3). In particular, if firms' preferences across project types diverge from social returns, $\frac{ds_{fp}}{d\mu_p} > 0$ has no reason to hold.

In section 6, we ask whether CSR spending increases with social returns. To do so, we consider how CSR spending is allocated across the other dimension of project type, locations, indexed by s. To proxy for the social returns to spending, we use a location's economic development and assume that social returns are higher in locations with lower economic development. Under this assumption, we then test whether $\frac{ds_{fs}}{d\mu_s} > 0$ by looking at the correlation between economic development and CSR expenditures across locations.

5 Does Spending Follow Technological Advantage?

Motivated by our conceptual framework, we consider whether firms spend more on topics for which their for-profit production processes give them a technological advantage. We begin by explaining how we construct a proxy for technological advantage and validating that it captures technological properties of firms' production processes. We then outline our empirical strategy, present our results, and examine mechanisms and heterogeneity across firms.

5.1 Construction of a Proxy for Technological Advantage

Testing whether spending follows technological advantage requires a measure of firm×project-specific CSR productivities. We exploit the insight that firms will be more productive in

[22] A key assumption of our empirical exercise—that follows from our model assumptions here—is that firms' for-profit industry choice reflects their productivity vectors. Sufficient conditions are standard profit-maximizing and price-taking behavior (see proof in Appendix E). This allows for distortions as long as the distribution of these distortions is not such that it is more profitable for firms to specialize in products where they have low productivity.

CSR projects that are technologically close to their for-profit production process. This insight has been repeatedly discussed in the literature under the term 'bundling' (see e.g. Besley and Ghatak, 2007) and is formalized by our micro-foundation for firm×project-specific CSR productivity in section 4.1. We operationalize it by building a measure of technological proximity for each pair of industry i and social topic d, meant to proxy for the model object $\cos(\boldsymbol{\phi_d}, \boldsymbol{\phi_i})$.

To proxy for $\cos(\boldsymbol{\phi_d}, \boldsymbol{\phi_i})$, we make the assumption that if projects in a given CSR topic require a technology that is close to the firm's for-profit technology, this *technological* proximity will be reflected in a *semantic* proximity between descriptions of the CSR topic and descriptions of the firm's production function. For instance, consider whether pharmaceutical firms are more efficient at undertaking CSR projects in health than financial firms. Our premise is that this would be reflected in a higher semantic proximity between the description of a pharmaceutical firm's production function and the description of health CSR projects than the semantic proximity between the description of a financial firm's production function and the description of these projects.

To construct this measure, we exploit embeddings of textual data describing CSR topics and industries, as described in section 2.2. For each industry i, we use the description in the NIC handbook to obtain the embedding $\mathbf{v}_i$. For each topic d, we define $\mathbf{v}_d$ as the average of the project embeddings $\mathbf{v}_p$ for projects p belonging to topic d. We proxy for $\cos(\boldsymbol{\phi_d}, \boldsymbol{\phi_i})$, the technological proximity between industry i and topic d, as:

$$\text{Proximity}_{id} = \cos(\mathbf{v}_d, \mathbf{v}_i) \tag{7}$$

The key assumption underlying our measurement exercise is that the semantic proximity between texts describing industry i and topic d appropriately proxies for the proximity of the technological requirements of industry i and topic d. Formally, we assume $\cos(\mathbf{v}_d, \mathbf{v}_i) \approx \cos(\boldsymbol{\phi_d}, \boldsymbol{\phi_i})$.[23] After describing our measure, we propose three tests to validate this assumption below.

Higher values indicate higher similarity, but the variable has no cardinal interpretation.

[23] Let $\mathcal{T}$, $\mathcal{W}$ be the dimensions of the task and the embedding space, respectively. A sufficient condition is that $\mathbf{v} = \mathbf{A}\boldsymbol{\phi}$ where $\mathbf{A} \in \mathbb{R}^{\mathcal{W}\times\mathcal{T}}$ is a linear isometry. This requires $\mathcal{T} \leq \mathcal{W}$, and ensures $\cos(\mathbf{v}_d, \mathbf{v}_i) = \cos(\boldsymbol{\phi_d}, \boldsymbol{\phi_i})$. Alternatively, when $\mathcal{T} \geq \mathcal{W}$, $\mathbf{A}$ must be a random projection and we apply the Johnson–Lindenstrauss lemma to obtain $\cos(\mathbf{v}_d, \mathbf{v}_i) \approx \cos(\boldsymbol{\phi_d}, \boldsymbol{\phi_i})$.

We standardize it to have a mean of zero and a standard deviation of one to ease the interpretation of our results. In robustness checks, we use alternative measures for $\mathbf{v}_d$ and $\mathbf{v}_i$. We obtain a distinct technological proximity value for each of the 1,136 (71×16) industry×topic combinations in our data. We also consider the robustness of our results to measuring proximity at a more granular level. To do so, we partition projects into 83 project types using a k-means clustering algorithm on the project embeddings (see Appendix C.3 for details), which yields 5,893 (71×83) distinct technological proximity values.

Description of Proximity$_{id}$. Figure 4 depicts the distribution of the proximity variable across topics and industries in a heatmap for all 16 topics and the 16 largest industries in our data. Deeper blue colors indicate higher proximity, and light gray indicates lower proximity. We see that the distribution of the variable is reasonably intuitive: for example, the health topic has a particularly high proximity with the pharmaceutical industry, and the civil engineering industry has a high proximity with the sanitation, safe drinking water, rural infrastructure, and environmental topics – all topics that require some degree of engineering. We also see that some topics/industries have consistently low or high proximity with many industries/topics (see, for example, the 'safe drinking water' topic). This may reflect true technological patterns, or be due to less desirable characteristics of our textual corpus, such as the recurrence of some non-technical terms in project descriptions. Our regression results below control for topic and firm fixed effects throughout to allow for this possibility, and we consider the robustness of our results to the exclusion of each topic or industry in turn.[24] Figure A.3 plots the distribution of the variable at the firm×topic level, with some examples. One standard deviation in proximity corresponds roughly to the difference in proximity for the pharmaceutical industry between the topics 'hunger and malnutrition' (medium proximity) and 'health' (high proximity).

In Tables C.5–C.8, we investigate which tokens lead a topic×industry pair to have a high or low similarity. The patterns are very intuitive. For instance, for the CSR topic 'rural infrastructure', the closest industry is 'civil engineering'. Looking at the sets of tokens closest to the centroids of the embeddings vector $\mathbf{v}_d$ and $\mathbf{v}_i$ for this pair, we see that both sets have significant overlap, with tokens related to construction driving their high cosine

[24] Figure A.2 plots the heatmap of the distribution of the residuals of the proximity variable after removing firm and topic fixed effects.

similarity.

Validation. We propose three tests of our assumption that $\cos(\mathbf{v}_d, \mathbf{v}_i) \approx \cos(\boldsymbol{\phi}_\mathbf{d}, \boldsymbol{\phi}_\mathbf{i})$. We provide the intuition for these tests below and leave formal derivations to Appendix E.

Test 1. Industry predicts semantic proximity across firms. From our definition of industries, it follows that the technology vectors of firms in the same industry are more similar than those of firms in different industries. If word embeddings appropriately represent technology vectors, then this property should hold for embeddings. We implement this test using the firm-level business descriptions contained in 10-K filings for US firms (there is no equivalent firm-level data for India). Table A.3 shows that firms in the same 2-digit industry have a higher cosine similarity than firms in different industries by 0.118 (compared to an average across-industry value of 0.679). This number rises to 0.146 and 0.155 when we compare firms in the same 3-digit and 4-digit industries, respectively.

Test 2. Semantic proximity across industries predicts firms' sales shares across industries. Our model predicts that a firm in industry i will have a higher sales share in products belonging to industry $i' \neq i$ if industry i has high technological proximity with i' (high $\cos(\boldsymbol{\phi}_\mathbf{i}, \boldsymbol{\phi}_{\mathbf{i'}})$). If the industry embedding vectors $\mathbf{v}_i$ appropriately represent technological requirements, this property should hold for semantic proximity between i and i' (high $\cos(\mathbf{v}_\mathbf{i}, \mathbf{v}_{\mathbf{i'}})$). Table A.4 shows that this is indeed the case.

Test 3. Semantic proximity across industries predicts input-output proximity. Finally, if industry embedding vectors capture technological requirements, we should see that industries with similar embeddings also use similar inputs, sell similar products, or develop supplier-customer relationships. Using the input-output matrix for India, Table A.5 shows that industry pairs (i, i') with similar products or inputs, or strong supplier-customer links, indeed have high textual proximity $\cos(\mathbf{v}_\mathbf{i}, \mathbf{v}_{\mathbf{i'}})$.

The approach of using textual data and semantic proximity to characterize firm proximity in economically relevant dimensions was pioneered by Hoberg and Phillips (2016). They interpret semantic similarity as product portfolio similarity, and hence proximity in the competitive space. While test 1 is in line with their interpretation of semantic proximity, tests 2 and 3 suggest that semantic similarity also captures technological proximity well.[25]

[25]In particular, test 1 would work if the text describing firms' activity was simply listing the products

5.2 Empirical Strategy

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

$$y_{fd} = \beta \text{Proximity}_{i(f)d} + \gamma_f + \gamma_d + \varepsilon_{fd} \tag{8}$$

where y_{fd} is an increasing function of CSR expenditures at the firm f and topic d level, $\text{Proximity}_{i(f)d}$ is our proxy for technological advantage defined at firm's industry $i(f)$ and topic level, defined above, and standard errors are clustered at the topic×industry level. Motivated by our conceptual framework above, we include firm fixed effects γ_f to ensure β captures a correlation with the firm's relative technical advantage on a topic, and topic fixed effects γ_d to capture preferences across topics that are common to all firms (we investigate what could be driving such common preferences below).

This specification tests whether spending follows technological advantage, as defined above: $\beta > 0$ indicates that firms spend more on topics in which they have a technological advantage. Note that the property does not require that firms do so because they choose to leverage this advantage: it still holds if the allocation is due to firms having, for example, high preferences for topics they have an advantage in (if the α_{fd} and ζ_{fd} terms in the conceptual framework above are positively correlated). Our object of interest is thus the correlation between CSR expenditures and technological advantage.

Our first outcome variable is the share of CSR expenditures a firm spends on the topic. We then consider the extensive margin decision by using an indicator for whether the firm spends any amount on the topic, and finally the intensive margin decision using the share of CSR expenditures spent on the topic, conditional on this share being positive. Our baseline specification gives equal weight to all firms but we also present results obtained by weighing each firm by its total CSR expenditures to consider how technological proximity relates to the aggregate CSR allocation.

produced by firms. Tests 2 and 3 suggest that the embeddings capture information about production processes that allow them to have predictive power for patterns of firm scope across industries and input-output relatedness.

5.3 Results

Table 2 presents the results of estimating specification (8).[26] The correlation between the proxy for technological advantage and how much firms spend on a topic is positive regardless of the outcome variable used. In Panel A, column 1, we see that a one-standard-deviation higher technological proximity between a firm's industry and a topic is associated with a one-percentage-point higher share that the firm spends on that topic (16% relative to the mean). It is also associated with a two-percentage-point higher probability that the firm spends on the topic (9% relative to the mean) and, conditional on spending a positive amount, with a two-percentage-point higher share (8%). The estimates are larger in Panel B, where we weigh each firm by its total CSR expenditures (the effect on the unconditional spending share is 29% of the mean), suggesting a stronger correlation among larger firms.

Our baseline test exploits technological proximity across 16 social topics, defined using the topic labels provided in the MCA data. In Table A.6 Panel A, we construct technological proximity at a more granular level using a partition of projects into 83 project types obtained from a k-means clustering algorithm applied to project embeddings (see Appendix C.3 for details).[27] The results are very similar: a one-standard-deviation higher technological proximity between a firm's industry and a project type is associated with a share allocated to that type that is higher by 14% relative to the mean. In Panel B, we instead aggregate similar topics into the 10 coarser topics of Figure 2 to construct technological proximity. The results are again highly similar, with a corresponding estimate of 17% relative to the mean.

Figure A.5 presents a series of robustness checks. We first consider results when changing choices made to obtain word embeddings: using OpenAI's NLP model to construct the proximity variable instead of Word2Vec or using text from 10-K filings in the US instead of from the NIC handbook, following Hoberg and Phillips (2016). We then change the method we use to aggregate information from projects at the topic level. Our baseline averages the cosine similarity at the project level across topics. We consider using the median similarity instead, or placing more weight on projects whose descriptions have more informative content (weighing each project by the number of tokens or aggregating at the topic level before computing the cosine similarity). We also cluster standard errors at the topic or industry

[26] Figure A.4 depicts the relationship graphically.

[27] The main limitation of this approach is that assigning projects to types requires sufficiently informative project descriptions, which reduces the sample by 24%.

level. Estimates are remarkably similar across these specifications. Figure A.6 shows that results are also stable when we exclude each topic, or each of the 20 largest industries, in turn.[28] Finally, Table A.7 presents results obtained by aggregating our data at the industry and topic level, in levels and in logs. Results are similar to those obtained in Panel B of Table 2 regardless of the specification used.

Our results thus suggest that firms allocate more CSR expenditures to topics in which they have a technological advantage. Seen through the lens of the theoretical literature, a necessary condition for CSR to be welfare-improving is therefore satisfied. To assess its empirical relevance, we ask how much of the overall allocation across topics technological proximity can explain. Given the similarities between the allocation of funds across topics by firms and the government described above, we compare how much technological proximity and government preferences explain the aggregate allocation. In Table A.8 we estimate a version of equation (8) in which we replace topic fixed effects, which capture common preferences for topics across all firms, with the government's expenditure share on each topic. This is equivalent to assuming that firms' preferences for topics are identical to the government's and using government expenditure shares to proxy for the latter.[29] This specification explains nearly as much of the variation in the data as our baseline specification, suggesting firms' preferences across topics are indeed similar to those of the government. Comparing across coefficients, we find that the explanatory role of technological proximity is one-fifth to one-fourth that of the government's spending share. Shared priorities across topics thus play a bigger role in explaining firms' CSR allocation than technological advantage.

Do firms spend more on topics that are technologically close because of preferences or because of technology? We fundamentally cannot disentangle unobserved firm preferences from their technological advantage across topics, and our interpretation of the results above holds regardless of what drives the correlation. We nevertheless provide two pieces of suggestive evidence on this question.

We first use information on the mode of implementation of projects. Firms that outsource

[28] The only exception is when we exclude the health topic. The coefficient drops slightly, suggesting technological advantage is particularly important for this topic, but remains statistically significant and indistinguishable from the coefficient obtained using our baseline specification.

[29] To do this, we re-classify topics so that government and CSR expenditures are comparable, in line with the approach in section 3 above. This is the same classification into 10 topics as that in the robustness checks.

CSR projects to third parties likely make less direct use of their own technology than those implementing projects themselves. If preferences are the only determinant of firms' allocation across topics, the decision to outsource might be orthogonal to technological advantage across topics. If, however, firms choose at least some projects because they want to leverage their for-profit production technologies, we should see that they are less likely to outsource projects in topics they have an advantage in. In Table A.9, we see that firms are indeed less likely to outsource projects on topics for which they have a technological advantage. While consistent with the idea that firms take into account their technological advantages across topics when deciding how to allocate CSR, this evidence is subject to two caveats. First, firms with a technological advantage in a topic may also be particularly good at choosing better third-party implementers for that topic. Second, firms may obtain larger private benefits from implementing a project themselves, such as visibility among customers, and these benefits may be larger for topics close to their production function.[30]

Second, we consider whether the correlation between firms' technological advantage and their CSR allocation is affected by their physical capital, measured as fixed assets over total assets. In our framework, physical capital is part of what makes a firm productive at a task, so firms with more physical capital should have more technology to carry over to CSR projects. If preferences alone drove the allocation, there would be no reason for the correlation with technological proximity to depend on this share. The correlation is indeed almost twice as large for firms with an above-median share of fixed assets, although the difference is not statistically significant (Figure A.7).

Stakeholder influence. In Figure A.7 we ask whether firms that are particularly beholden to different types of stakeholders allocate differently, perhaps in response to stakeholder influence. We find no evidence that firms with different ownership structures (publicly listed firms or firms with controlling shareholders), firms in which employees may have more bargaining power (proxied by the average wage, labor share, or training expenses), or firms that rely more on their reputation with final consumers (proxied by advertising expenses or downstreamness) allocate differently. The correlation with technological proximity is smaller, but not significantly so, among firms that may value a good relationship with the government

[30]The definition of third-party implementation is imprecise in the reporting requirements, giving rise to measurement concerns for this variable (see Appendix C.1.5).

(because they operate in heavily regulated industries or compete with government-owned firms), perhaps because of strategic considerations.

Heterogeneity by other firm characteristics. Figure A.7 also examines firms' financial characteristics. The correlation does not vary with proxies for financial constraints (leverage, cash holdings over assets, dividend payments, or interest coverage), suggesting that the tracking of technological advantage is not driven by firms that lack the resources to develop capabilities outside their existing technology. Consistent with this, an alternative proxy for financial constraints—whether the firm belongs to a business group, and thus has access to internal capital markets—also shows no differential correlation. Finally, Figure A.8 shows estimates by decile of the firm-size distribution. They are somewhat smaller for the first three deciles, in line with the weighted–unweighted comparison in Table 2, though the differences across deciles are not statistically significant. The tracking of technological advantage is thus a broad feature of the data rather than a behavior of particular types of firms.

External validity: is 'voluntary' CSR allocated similarly? To consider whether our findings could extend to the CSR choices made by firms whose total CSR expenditures are not constrained by a mandate, we consider whether the correlation between technological proximity and spending shares is different among firms that spend more than the amount prescribed by the law - the 'voluntary CSR' sample defined above (Table B.2). The unweighted coefficients (Panel A) are highly similar in this sample, and all are statistically significant at 1%. The weighted results are slightly smaller (e.g., 0.012 versus 0.018 for the unconditional CSR share). This suggests that technological advantage may matter slightly less for large firms with a strong preference for CSR compared to other large firms, perhaps because they also have strong intrinsic preferences for some topics. Overall, our results suggest that voluntary CSR spending is allocated similarly to mandated CSR.

6 Does Spending Follow Social Returns?

Empirical strategy. We now consider the allocation of CSR expenditures across locations by running the following specification at the firm f and state s level:

$$y_{fs} = \beta \log \text{GDP}_s + \gamma_f + \varepsilon_{fs} \tag{9}$$

where $\log \text{GDP}_s$ is the state's log gross product per capita, γ_f are firm fixed effects, and we control throughout for state population.

Results. Results are presented in Table 3 (columns 1, 3, and 5). We see that CSR expenditure is positively correlated with state GDP per capita at both the extensive and intensive margin. Figure 5(a) plots the resulting state-level relationship between CSR spending and state GDP (both per capita). Each blue point represents one state, and we show the linear fit of the corresponding regression using states' population as weight. States with 10% higher GDP per capita receive 19% more CSR spending per capita. We observe very similar results when we consider the sample of all firms in the CSR data or the sample of firms that voluntarily spend more than the mandate requires (Figure B.2 and Table B.3). We also obtain similar results when we proxy for state development using a multidimensional poverty index instead of GDP per capita (Figure A.9). In sum, CSR funds flow towards the most developed Indian states. Under the assumption that CSR expenditures yield higher social returns in less developed areas, these results indicate that firms' CSR allocation choices do not follow social returns.

However, development levels vary substantially within states in India, and firms could be allocating their CSR expenditures to the poorest areas within states. We cannot completely rule out this possibility, but two pieces of evidence suggest this is not the case. First, we find similar patterns when restricting the analysis to CSR expenditures allocated to the 'rural infrastructure' topic, which by definition are allocated to poorer regions within each state (see Figure A.11). Second, in Figure A.12, we replicate the analysis in Figure 5 at the district level. We see again a very similar pattern, with more CSR expenditures in richer districts, regardless of the proxy for district development used.

Finally, local development may be only a weak proxy for needs and therefore social returns

to expenditures on specific CSR causes. In Figure A.13, we instead consider the correlation between CSR expenditures at the state×topic level and a proxy for state-level needs on that CSR topic. For four topics, we can define a plausible proxy for need (e.g., for education, we consider the state-level literacy rate). We find again a similar pattern: CSR expenditures tend to flow to areas with relatively low need, even when considering topic-specific expenditures.

Mechanisms and heterogeneity. Why are CSR expenditures concentrated in rich states? One reason could be that firms concentrate their spending where they are headquartered, and corporate headquarters are concentrated in rich states. To test this, columns 2, 4, and 6 in Table 3 control for an indicator for whether the firm is headquartered in the state. The coefficients for the indicator are large, reflecting the fact that around 60% of CSR spending occurs in firms' headquarter states. The coefficient for state GDP per capita indeed falls and becomes null for the intensive margin. Using past capital spending to proxy for firms' establishment locations, Table A.10 includes in addition an indicator for whether the firm has an establishment in the state. We see that this is also a strong predictor of CSR spending in the state.[31] CSR thus exhibits a home bias familiar from other allocation decisions: just as investors overweight nearby assets and individual donors give around corporate headquarters (Coval and Moskowitz, 1999; Card et al., 2010), firms direct their social spending toward the places they operate in.

The concentration of CSR spending where firms are located could be efficient: firms may possess information or technologies that are particularly suited to finding out, or meeting, the needs of areas in which they operate. Note however that when we exclude expenditures in headquarter states in Figure 5(b), the aggregate slope falls but remains statistically significant and large (this is also true when we exclude expenditures in states in which firms have establishments, see Figure A.10). This suggests that efficiency considerations linked to firms' locations alone are unlikely to fully explain why firms spend more in richer states.[32] Regardless of the mechanism underlying the geographical allocation of CSR, the evidence suggests that CSR funds do not follow social returns, at least when comparing medium-to-

[31] 84% of CSR spending occurs either in firms' headquarter states or in states in which they have establishments.

[32] It also indicates that the tilt toward richer states is not merely a by-product of the mandate's local-area provision, which is poorly enforced.

large (states or districts) geographical units.

We finally ask how allocation across locations varies with firm characteristics. The correlation between CSR spending and state development varies little with firms' stakeholder or financial characteristics: the tilt toward richer states is a broad feature of the data rather than a behavior of particular types of firms (Figure A.14). Similarly, there is little evidence of firm size mattering for the correlation between CSR spending and state development (Figure A.16(a)). The home bias, in contrast, is weaker in three groups of firms (Figures A.15 and A.16(b)). Business-group-affiliated firms concentrate less spending in their headquarter states, perhaps because groups can channel CSR through affiliated firms operating in other states. High-wage firms also spend less locally, consistent with local spending partly serving firms' standing in local labor markets, which matters less for firms recruiting in national markets. Larger firms, which operate in more states, also show a weaker home bias, concentrated in the top two size deciles. These patterns are consistent with the home bias reflecting the costs and benefits of spending close to a firm's own operations, rather than a uniform constraint.

Comparison to alternative uses of CSR funds. Does the CSR allocation follow social returns more or less closely than alternative uses of the same funds? One comparison point is the allocation of government expenditures per capita, assuming the government could have increased taxes instead of imposing a CSR mandate. In Figure 5(c), we see that state-level government expenditures per capita (restricted to the topics covered by the CSR data, green dots) increase slightly with state development, but the slope is only one-fifth of the slope for CSR expenditures (around one-third if we exclude spending in headquarter states).[33] The allocation of CSR across space thus tilts toward richer states more strongly than the allocation of government expenditures.

Another potential benchmark is that CSR expenditures could have been redistributed to shareholders instead. We cannot locate shareholders, but they are likely much richer than the average Indian citizen and located in richer states. A counterfactual allocation of CSR expenditures to profits could thus have led to an allocation tilted even more toward richer areas.

[33]The small positive slope reflects the fact that richer states both get less inter-governmental transfers from the central government and collect more tax revenues; the latter effect dominates.

7 Do ESG Scores Reflect Firms' CSR Allocations?

Investors and researchers rarely observe what firms fund. What they observe instead is typically an ESG score, and investors increasingly condition capital on such scores. We therefore ask whether ESG scores carry information about the two allocation properties we study. If they do, an ESG score is a usable summary of allocation; if they do not, scores and allocation capture different dimensions of firms' social activity.

We do this by constructing firm-level summary measures of how firms allocate their CSR expenditures. How much a firm's allocation follows its technological advantage is captured by the within-firm relationship between topic spending shares and technological proximity, holding constant how much all firms spend on each topic (TechProx_f). The extent to which it allocates expenditures towards richer states is captured by the within-firm relationship between state spending shares and log state GDP per capita (GDP tilt_f). Finally, Home_f captures how much more a firm spends in its headquarter state than the state's population would imply. Each measure is the firm-specific slope in one of the specifications estimated above, so that the coefficients in Tables 2 and 3 are weighted averages of them.[34] We then regress the firm-specific allocation measures on the firm's ESG score:

$$m_f = \delta \, \text{ESG}_f + \mu_{i(f)} + \varepsilon_f \tag{10}$$

where m_f is one of the three allocation measures, $\mu_{i(f)}$ are 2-digit NIC industry fixed effects, and some specifications additionally control for size.[35] Both m_f and ESG_f are standardized within each estimation sample, so that δ is measured in standard deviations of the allocation measure per standard deviation of the ESG score. The unit of observation is the firm; standard errors are heteroskedasticity-robust. We use the ESG scores and subsamples of firms for which they are available, as described in section 2.2; Table A.11 shows that our key results on allocation are similar on these subsamples.

Table 4 reports the results. Considering columns 1 and 2, we see that ESG scores are

[34] Specifically, for each specification, we partial out fixed effects and controls and take the within-firm slope of the residualized outcome on the residualized regressor (using the same observation weights), so that the pooled coefficient is the denominator-weighted mean of the firm-level slopes. For TechProx and GDP tilt, this recovers the coefficients in Tables 2 and 3 exactly; for Home it recovers the coefficient from a specification with the headquarter indicator alone.

[35] We use log assets as the size control but results are similar when using a firm's CSR spending instead.

uncorrelated with whether a firm's spending follows its technological advantage: the coefficients are of varying sign, small in magnitude, and far from statistically significant. This is also true when examining the correlation with local development (columns 3 and 4). When considering the correlation with home bias (columns 5 and 6), we see a negative and statistically significant correlation for the Refinitiv score that disappears once we control for firm size. Figure A.17 depicts these relationships graphically; we see that the lack of correlation is not an artifact of our specification choices. Tables A.12 and A.13 show that these results are also unaffected by changing our data construction choices, and Appendix C.2.6 discusses both tables in detail.

Overall, we see essentially no correlation between either ESG score and our measures of the properties of firms' CSR allocations, which bear on the welfare effects of CSR. This suggests that ESG scores and data on firms' CSR allocation in practice capture different dimensions of firms' social activity.

8 Conclusion

What do firms actually do when they engage in social activities? For decision-makers and academics alike, assessing firms' social activities requires knowing what firms do when they pursue them, yet what firms fund is rarely observed systematically: available information is limited to firms' aggregate spending figures and third-party ESG scores. In this paper, we assemble the first comprehensive dataset covering the quasi-universe of corporate social projects in a major economy, India, and use it to describe how firms allocate social spending across social causes and across space.

In aggregate, the CSR portfolio resembles that of other providers of public goods: education and health receive the largest shares, and the allocation across causes is highly correlated with both government spending shares and the distribution of NGO activity. This aggregate conceals specialization at the firm level: one third of firms devote more than 90% of their spending to a single cause, and firms with similar product portfolios choose similar causes. Across locations, spending is concentrated in a few states.

Two results inform the discussion of whether firms should pursue social objectives. First, CSR spending follows firms' technological advantage: the higher the technological proximity between a firm's industry and a cause, the larger the share of spending it receives. The

theoretical literature identifies this advantage as the key rationale for CSR, and these are the first estimates showing that firms' allocation is consistent with it. Second, CSR spending does not follow social returns: firms spend close to where they operate, and firms are concentrated in the most developed states. Under the assumption that public goods yield higher social returns in less developed areas, CSR is not directed toward the locations where it is needed most. These results hold for firms that spend more on CSR than is mandated, suggesting our conclusions do not hinge on CSR being mandated rather than voluntary.

Finally, these patterns of allocation, which bear on the welfare effects of CSR, are not captured by existing ESG scores: firms with high scores are no more likely to direct spending toward causes close to their technology, and no more likely to spend in poorer states. Scores and spending allocations therefore capture different dimensions of firms' social activity. The allocations we establish, invisible in spending totals and ESG scores, are relevant to investors evaluating firms' ESG policy, to policymakers designing mandates, and to the academic debate on whether firms should pursue social objectives.

References

Acemoglu, Daron and Pascual Restrepo, "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, June 2018, *108* (6), 1488–1542.

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.

Alok, Shashwat and Meghana Ayyagari, "Politics, State Ownership, and Corporate Investments," *The Review of Financial Studies*, 2020, *33* (7), 3031–3087.

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

Ash, Elliott and Stephen Hansen, "Text Algorithms in Economics," *Annual Review of Economics*, 2023, *15* (Volume 15, 2023), 659–688.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad, "Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform," *The World Bank Economic Review*, 2021, *35* (4).

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

Barboni, Giorgia and Parul Agarwal, "How do flexible microfinance contracts improve repayment rates and business outcomes? Experimental evidence from India," 2023. working paper, available at SSRN 4358795.

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

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

Berg, Florian, Julian F Kölbel, and Roberto Rigobon, "Aggregate confusion: The divergence of ESG ratings," *Review of finance*, 2022, *26* (6), 1315–1344.

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

Bertrand, Marianne, Matilde Bombardini, Raymond Fisman, and Francesco Trebbi, "Tax-Exempt Lobbying: Corporate Philanthropy as a Tool for Political Influence," *American Economic Review*, July 2020, *110* (7), 2065–2102.

_ , Paras Mehta, and Sendhil Mullainathan, "Ferreting out tunneling: An application to Indian business groups," *The quarterly journal of economics*, 2002, *117* (1), 121–148.

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

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

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

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

Boehm, Johannes, Swati Dhingra, and John Morrow, "The comparative advantage of firms," *Journal of Political Economy*, 2022, *130* (12), 3025–3100.

Brandon, Rajna Gibson, Simon Glossner, Philipp Krueger, Pedro Matos, and Tom Steffen, "Do Responsible Investors Invest Responsibly?*," *Review of Finance*, 09 2022, *26* (6), 1389–1432.

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

Cheng, Ing-Haw, Harrison Hong, and Kelly Shue, "Do managers do good with other people's money?," *The Review of Corporate Finance Studies*, 2023, *12* (3), 443–487.

Chhaochharia, Vidhi, Rik Sen, and Jing Xu, "Corporate Benevolence and Societal Impact: Evidence from India's CSR Reform," 2025. Mimeo, University of Georgia.

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

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, "Lazy prices," *The Journal of Finance*, 2020, *75* (3), 1371–1415.

Colonnelli, Emanuele and Mounu Prem, "Corruption and firms," *Review of Economic Studies*, 2022, *89* (2), 695–732.

_ , Timothy McQuade, Gabriel Ramos, Thomas Rauter, and Olivia Xiong, "Polarizing Corporations: Does Talent Flow to Good Firms?," Working Paper 31913, National Bureau of Economic Research November 2023.

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

Coval, Joshua D and Tobias J Moskowitz, "Home bias at home: Local equity preference in domestic portfolios," *The Journal of Finance*, 1999, *54* (6), 2045–2073.

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.

Edmans, Alex and Marcin Kacperczyk, "Sustainable finance," *Review of Finance*, 2022, *26* (6), 1309–1313.

Ferreira, Daniel and Radoslawa Nikolowa, "Polarization, purpose and profit," *Journal of Financial Economics*, 2025, *172*, 104147.

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

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

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

Fonseca, Julia and Bernardus Van Doornik, "Financial Development and Labor Markets: Evidence from Brazil," *Journal of Financial Economics*, 2022, *143* (1), 550–568.

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

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

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, September 2019, *57* (3), 535–74.

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

Giving USA Foundation, *Giving USA 2026: The Annual Report on Philanthropy for the Year 2025*, Chicago, IL: Giving USA Foundation, 2026. Researched and written by the Indiana University Lilly Family School of Philanthropy.

Global Sustainable Investment Alliance, "Global Sustainable Investment Review 2022," Technical Report, Global Sustainable Investment Alliance November 2023. Accessed: 2026-10-20.

Gopalan, Radhakrishnan, Vikram Nanda, and Amit Seru, "Affiliated Firms and Financial Support: Evidence from Indian Business Groups," *Journal of Financial Economics*, 2007, *86* (3), 759–795.

Government of India, "State-Wise Number and Percentage of Population Below Poverty Line in India: 2011-12," 2014. Planning Commission, Government of India, https://www.data.gov.in/resource/state-wise-number-and-percentage-population-below-poverty-line-india-2011-12-based.

_ , "State/UT-wise Prevalence of Underweight, Stunting, Wasting and Anaemia Among Children Under 5 Years of Age as per National Family Health Survey-4, 2015-16," 2020. Ministry of Women and Child Development, Government of India, https://ap.data.gov.in/resource/stateut-wise-prevalence-underweight-stunting-wasting-and-anaemia-among-children-under-5.

_ , "State-Wise Infant Mortality Rate, 2013," 2021. Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, Government of India, https://www.rbi.org.in/scripts/PublicationsView.aspx?id=20670.

_ , "State-Wise Literacy Rate, 2011," 2021. Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, Government of India, https://www.rbi.org.in/scripts/PublicationsView.aspx?id=20665.

_ , "State/UT-wise Details of Headcount Ratio, Intensity and Multi-Dimensional Poverty Index (MPI), 2015-16," 2024. https://www.data.gov.in/resource/stateut-wise-details-headcount-ratio-intensity-and-multi-dimensional-poverty-index-mpi.

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

Hanley, Kathleen Weiss and Gerard Hoberg, "Dynamic interpretation of emerging risks in the financial sector," *The Review of Financial Studies*, 2019, *32* (12), 4543–4603.

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

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

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

Hartzmark, Samuel and Kelly Shue, "Counterproductive Impact Investing: The Impact Elasticity of Brown and Green Firms," *SSRN Electronic Journal*, 01 2023.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "A Bright Idea for Measuring Economic Growth," *American Economic Review*, 2011.

Higgins, Sean, "Financial Technology Adoption: Network Externalities of Cashless Payments in Mexico," *American Economic Review*, 2024, *114* (11), 3469–3512.

Hoberg, Gerard and Gordon Phillips, "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis," *The Review of Financial Studies*, 08 2010, *23* (10), 3773–3811.

_ **and** _ , "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, 2016, *124* (5), 1423–1465.

_ **and Vojislav Maksimovic**, "Redefining financial constraints: A text-based analysis," *The Review of Financial Studies*, 2015, *28* (5), 1312–1352.

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

ICRISAT, "ICRISAT-District Level Data," 2020. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), https://data.icrisat.org/dld/ .

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

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

Khanna, Tarun and Krishna Palepu, "Is Group Affiliation Profitable in Emerging Markets? An Analysis of Diversified Indian Business Groups," *Journal of Finance*, 2000, *55* (2), 867–891.

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

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

Koh, Paul and Devesh Raval, "Economies of Scope from Shared Inputs," 2025.

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

KPMG, "Big Shifts, Small Steps: Survey of Sustainability Reporting 2022," Technical Report, KPMG International October 2022. Accessed: 2026-10-20.

Krueger, Philipp, Zacharias Sautner, and Laura T. Starks, "The Importance of Climate Risks for Institutional Investors," *The Review of Financial Studies*, 2020, *33* (3), 1067–1111.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, "Measuring corporate culture using machine learning," *The Review of Financial Studies*, 2021, *34* (7), 3265–3315.

Liang, Hao and Luc Renneboog, "On the Foundations of Corporate Social Responsibility," *The Journal of Finance*, 2017, *72* (2), 853–910.

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

Lindenlaub, Ilse, "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 2017, *84* (2 (299)), 718–789.

Lins, Karl V., Henri Servaes, and Ane Tamayo, "Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis," *The Journal of Finance*, 2017, *72* (4), 1785–1824.

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

Lloyd, Stuart, "Least squares quantization in PCM," *IEEE transactions on information theory*, 1982, *28* (2), 129–137.

London Stock Exchange Group, "Environmental, Social and Governance Scores from LSEG," 2024. https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf.

MacQueen, James, "Some methods for classification and analysis of multivariate observations," in "Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics," Vol. 5 University of California press 1967, pp. 281–298.

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

Maksimovic, Vojislav and Gordon Phillips, "Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence," *The Journal of Finance*, 2002, *57* (2), 721–767.

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.

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

Ministry of Corporate Affairs, "Compendium on Corporate Social Responsibility in India," Technical Report, Ministry of Corporate Affairs, Government of India 2021.

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

Mosaik, "Global High Resolution Estimates of the United Nations Human Development Index," 2023. https://www.mosaiks.org/data-sets .

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

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

Naaraayanan, S Lakshmi and Daniel Wolfenzon, "Business group spillovers," *The Review of Financial Studies*, 2024, *37* (1), 231–264.

Ponticelli, Jacopo and Leonardo S. Alencar, "Court enforcement, bank loans, and firm investment: Evidence from a bankruptcy reform in Brazil," *Quarterly Journal of Economics*, 2016, *131* (3), 1365–1413.

Rajgopal, Shivaram and Prasanna Tantri, "Does a government mandate crowd out voluntary corporate social responsibility? Evidence from India," *Journal of Accounting Research*, 2023, *61* (1), 415–447.

Riedl, Arno and Paul Smeets, "Why Do Investors Hold Socially Responsible Mutual Funds?," *The Journal of Finance*, 2017, *72* (6), 2505–2550.

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

Rousseeuw, Peter J, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, 1987, *20*, 53–65.

Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang, "Firm-level climate change exposure," *The Journal of Finance*, 2023, *78* (3), 1449–1498.

Securities and Exchange Board of India, "Business Responsibility and Sustainability Reporting by Listed Entities," 2021. https://www.sebi.gov.in/legal/circulars/may-2021/business-responsibility-and-sustainability-reporting-by-listed-entities_50096.html.

_ , "Annexure I: Format of BRSR Core," 2023. https://www.sebi.gov.in/sebi_data/commondocs/jul-2023/Annexure_I-Format-of-BRSR-Core_p.pdf.

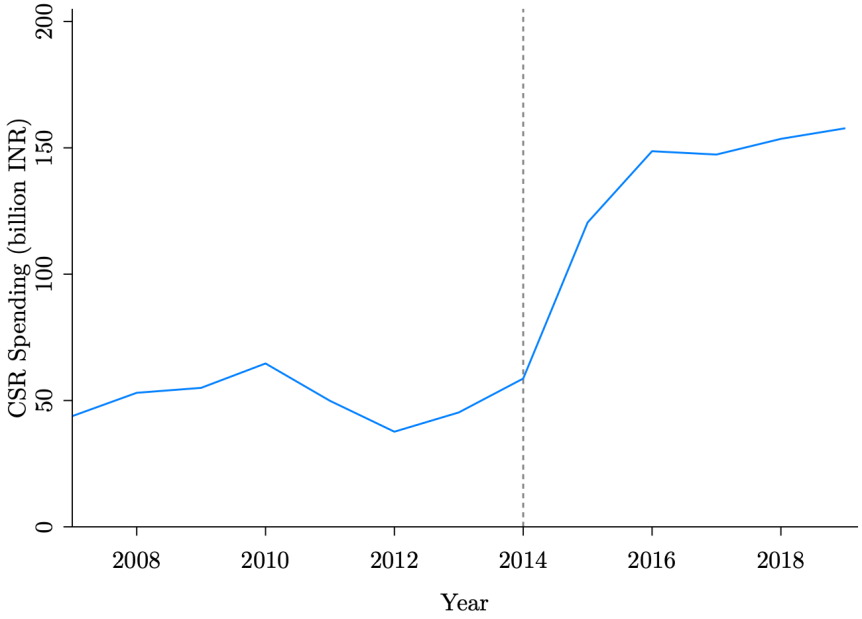
SHRUG, "The Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG), Version 2.1," 2024. https://www.devdatalab.org/shrug.

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

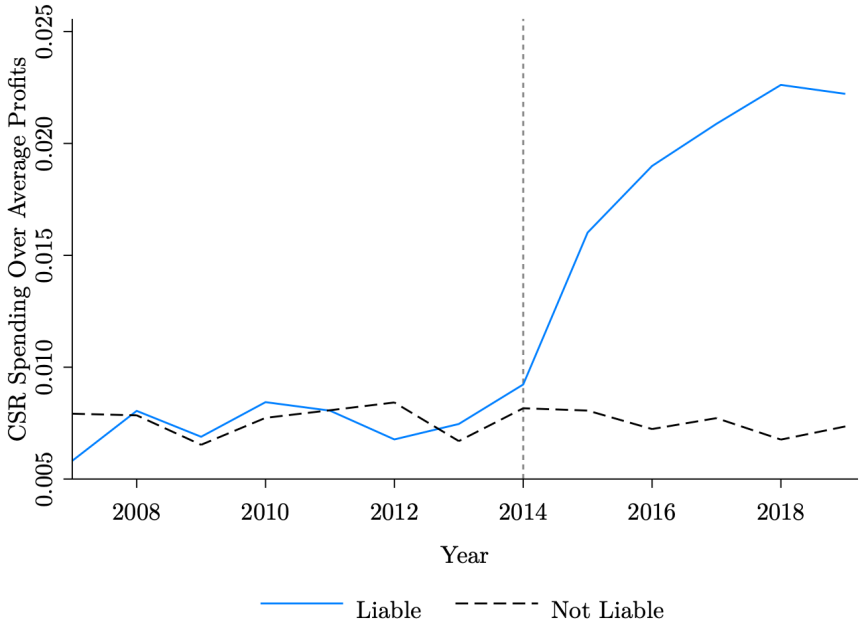
Vig, Vikrant, "Access to collateral and corporate debt structure: Evidence from a natural experiment," *Journal of Finance*, 2013, *68* (3), 881–928.

Figure 1: CSR Spending Over Time

(a) Aggregate CSR Spending

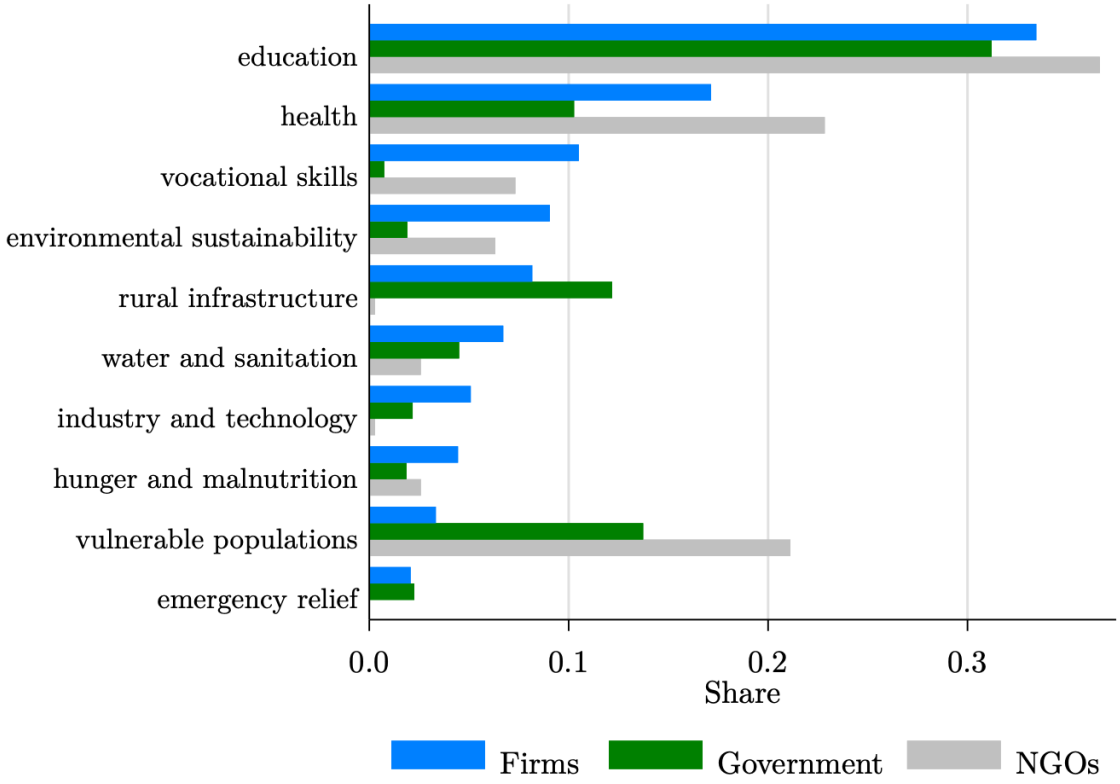


(b) CSR Spending by Firms' Liability Status



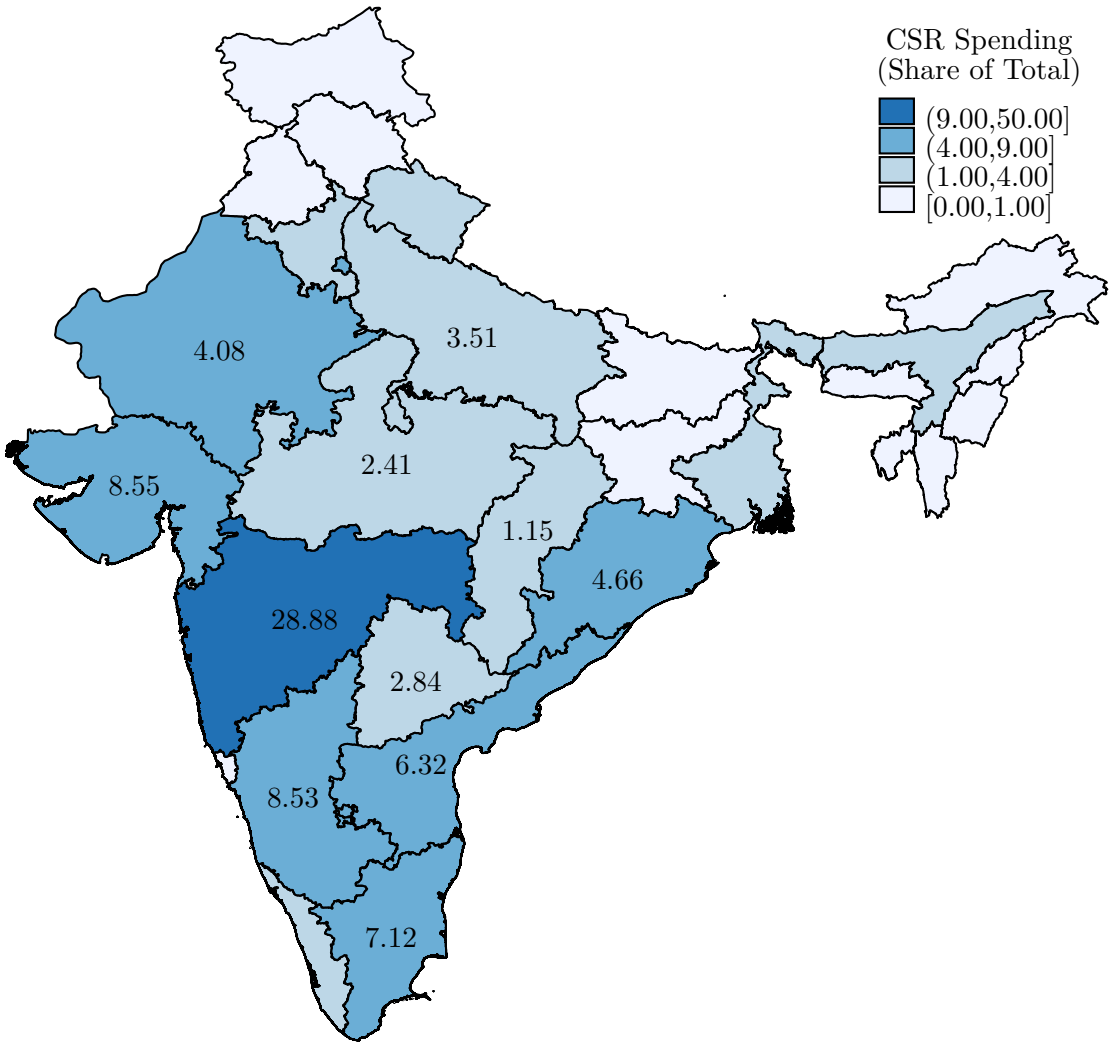
Notes: This figure depicts CSR spending over time. In Figure 1(a), CSR spending is aggregated over all firms and denominated in 2015 billion INR. Figure 1(b) depicts the CSR spending of a given firm in a given year over average profits in the past three years. The blue line (solid) depicts the mean over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

Figure 2: Allocation Across Topics by Public Goods Providers



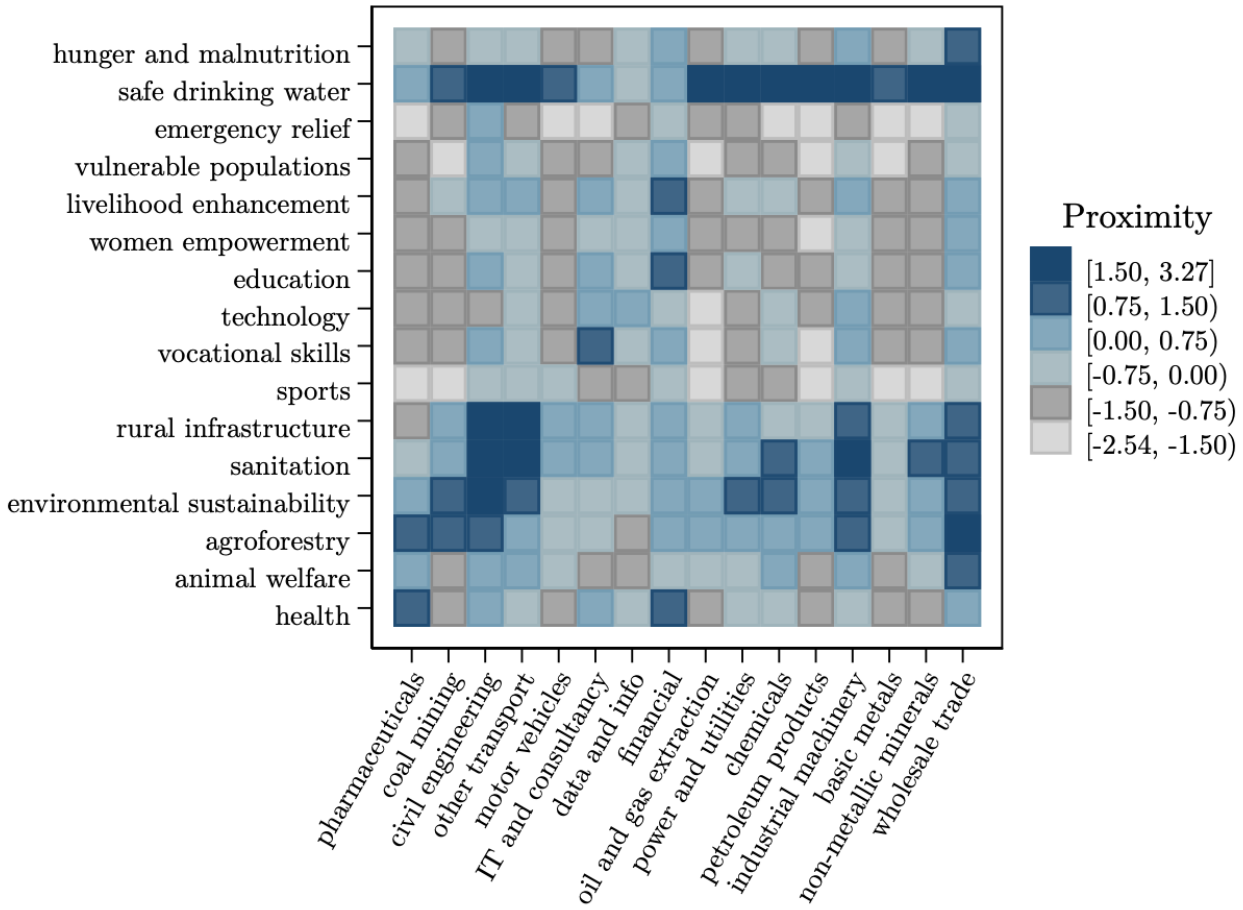
Notes: This figure depicts the share of each topic in terms of total CSR spending, total government spending, and number of NGOs. See Appendix C.2 for the mapping of CSR topics to government spending and the number of NGOs. NGO data does not include emergency relief.

Figure 3: Allocation of CSR Spending Across States



Notes: This figure depicts CSR spending shares by state.

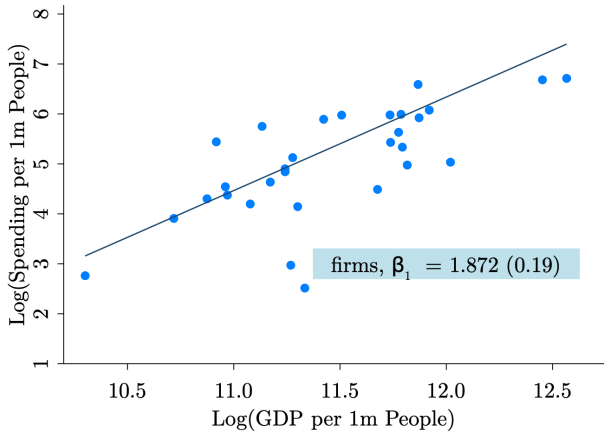
Figure 4: Proximity Across Topics and Largest Industries



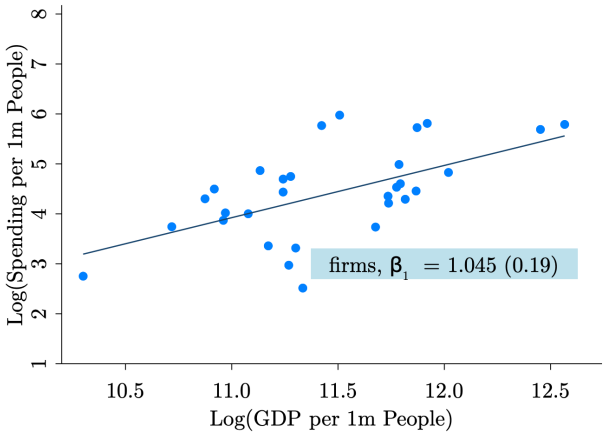
Notes: This figure depicts all 16 topics and the largest 16 industries by total CSR spending. The unit of observation is at the industry×topic level. $\text{Proximity}_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1.

Figure 5: CSR Spending, Government Spending, and State-Level GDP

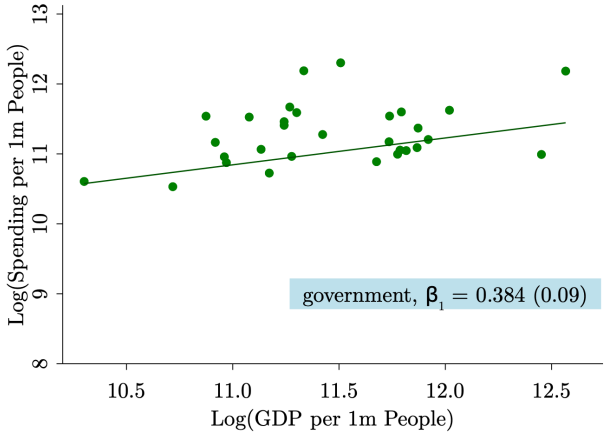
(a) CSR Spending



(b) CSR Spending Outside HQ States



(c) Government Spending



Notes: This figure depicts the relationship between state-level GDP and spending per capita. The scattered dots indicate state-level observations. The lines indicate fitted linear approximations. The unit of observation is at the state level. The dependent variable is the log of spending (in millions, denominated in 2015 INR) per one million people, aggregated from 2015 to 2019. Figure 5(a) depicts CSR spending by firms, including spending in their headquarter state. Figure 5(b) depicts CSR spending by firms, excluding spending in their headquarter state. Figure 5(c) depicts government spending on topics covered by the CSR data. The independent variable is the log of state-level GDP (in millions, denominated in 2015 INR) per one million people in 2013. Observations are weighted by the 2011 population.

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

Topic	Share	Three Most Frequent Project Types
Education	32%	special education, scholarships & support for meritorious students, school construction & infrastructure
Health	17%	preventive health, medical treatment and patient care, medical infrastructure and equipment
Rural infrastructure	8%	construction (roads, walls, street light), construction (community centers), rural transformation
Environmental sustainability	8%	ecology/conservation projects, tree plantation, green energy
Vocational skills	6%	vocational training, skill acquisition, staff training
Technology	5%	education technology, mobile science labs & support to labs, computers & other equipment donations
Livelihood enhancement	5%	sustainable livelihood and education, disability inclusion, economic development
Sanitation	5%	sanitation infrastructure, toilet construction, cleanliness campaigns
Hunger and malnutrition	4%	eradicating hunger, midday meal scheme, food distribution
Safe drinking water	2%	water supply infrastructure (tanks, pumps, wells...), water safety programs, water purification
Vulnerable populations	2%	hostels for old age, widows & orphans, social welfare programs, veterans support
Emergency relief	2%	flood relief, disaster relief, contribution to prime minister relief fund
Sports	2%	competition organization, support to olympic/nationally-recognized sports, support to sport clubs
Women empowerment	1%	education and young women empowerment, gender equality, disadvantaged women and girls
Agroforestry	1%	sustainable agriculture, farmer training, agronomy
Animal welfare	0%	animal care, animal shelters, cow sheds

Notes: This table displays the share of total CSR spending and the most frequent project types by topic. The project types are the three largest clusters obtained by k-means clustering within each topic, as described in Appendix C.3.

Table 2: CSR Spending and Proximity

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

Notes: This table describes the relationship between CSR spending and proximity, derived from equation (8). The unit of observation is at the firm×topic level. The dependent variables are the share of CSR spending of a firm (f) over topics (d), an indicator for any CSR spending by firm (f) in a given topic (d), and the share of CSR spending conditional on any spending. Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. In Panel A, observations are unweighted. In Panel B, observations are weighted by the total CSR spending of each firm, winsorized at the 1st and 99th percentile. Standard errors are clustered at the industry×topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table 3: CSR Spending, State-Level Characteristics, and Firm Headquarters

	CSR Share Unconditional$_{fs}$		Any CSR Spending$_{fs}$		CSR Share Conditional$_{fs}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Log(GDP per 1m People)$_s$	0.096***	0.014***	0.156***	0.058***	0.145***	-0.001
	(0.031)	(0.004)	(0.040)	(0.008)	(0.022)	(0.011)
1(Firm Headquarter State)$_{fs}$		0.601***		0.722***		0.333***
		(0.026)		(0.022)		(0.023)
Avg dep var	0.029	0.029	0.067	0.067	0.437	0.437
Firm FE	✓	✓	✓	✓	✓	✓
R-squared	0.08	0.52	0.18	0.44	0.42	0.57
Observations	196,415	196,415	196,415	196,415	9,736	9,736

Notes: This table describes the relationship between state-level characteristics as well as firm headquarters and CSR spending, derived from equation (9). The unit of observation is at the firm×state level. The dependent variables are the share of CSR spending of a firm (f) over states (s), an indicator for any CSR spending by firm (f) in a given state (s), and the share of CSR spending conditional on any spending. The denominator in the shares is the firm's total CSR spending, including projects whose state cannot be identified. The independent variables are the log of state-level GDP per 1 million people and an indicator that equals one if the firm is headquartered in the state as per government records. We control for the log of population in millions. Observations are weighted by the 2011 population. Standard errors are clustered at the state-level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table 4: ESG Scores and Allocation

	TechProx $(\text{std})_f$		GDP tilt $(\text{std})_f$		Home $(\text{std})_f$	
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Refinitiv			
ESG score $(\text{std})_f$	0.046	0.026	-0.016	0.041	-0.167***	-0.085
	(0.052)	(0.056)	(0.049)	(0.053)	(0.055)	(0.055)
Log(assets)_f		0.041		-0.114***		-0.166***
		(0.040)		(0.035)		(0.036)
Industry FE	✓	✓	✓	✓	✓	✓
Size control		✓		✓		✓
R-squared	0.12	0.13	0.18	0.21	0.15	0.20
Observations	449	449	449	449	449	449
			Panel B: BRSR			
ESG score $(\text{std})_f$	-0.022	-0.025	-0.027	-0.010	-0.050	-0.021
	(0.043)	(0.043)	(0.043)	(0.042)	(0.044)	(0.043)
Log(assets)_f		0.019		-0.111***		-0.187***
		(0.023)		(0.024)		(0.026)
Industry FE	✓	✓	✓	✓	✓	✓
Size control		✓		✓		✓
R-squared	0.07	0.07	0.14	0.17	0.08	0.15
Observations	722	722	722	722	722	722

Notes: This table relates firm-level ESG scores to how a firm allocates its CSR spending. The unit of observation is the firm. Each column estimates

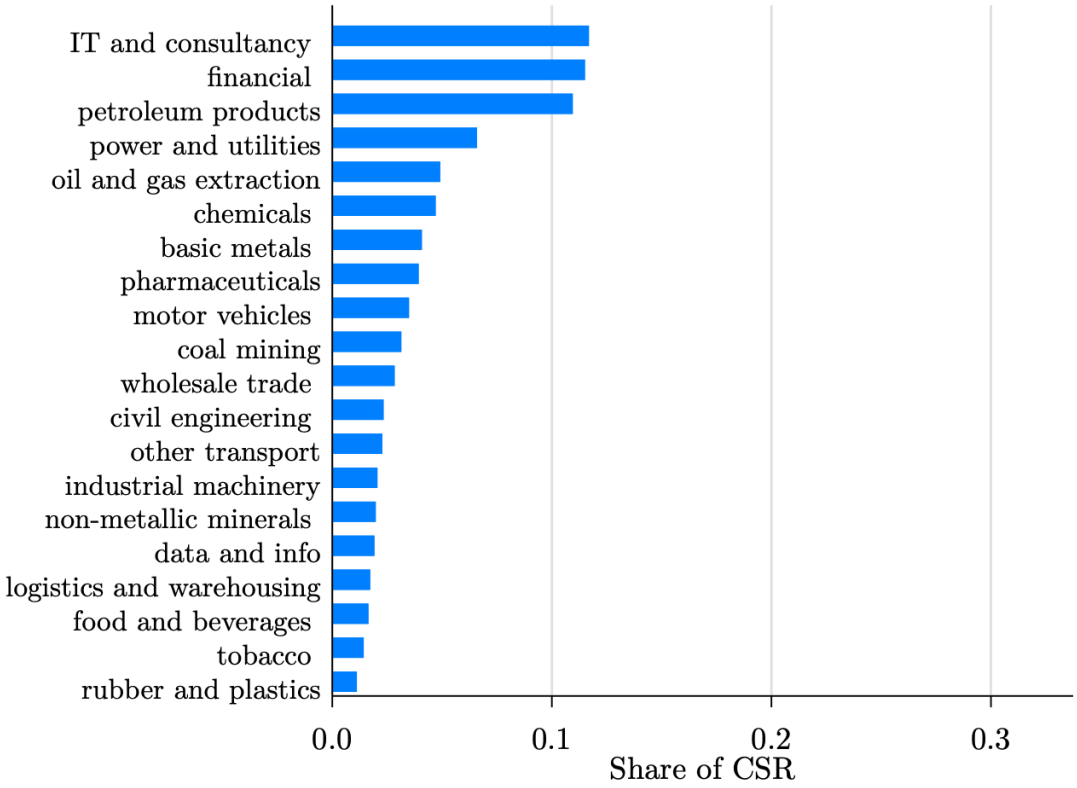
$$m_f = \delta\, \text{ESG}_f + \mu_{i(f)} + \varepsilon_f,$$

where m_f is one of three firm-level allocation measures defined below, $\mu_{i(f)}$ are 2-digit NIC industry fixed effects, a log-assets control enters in the even-numbered columns only, and the main coefficient of interest is δ. Each measure m_f takes a pooled specification estimated elsewhere in the paper and obtains firm-specific slopes. *TechProx* comes from equation (8), which relates CSR shares to technological proximity. *GDP tilt* comes from equation (9), which relates CSR shares to log state GDP per capita. *Home* comes from that same specification with an indicator for the firm's headquarter state replacing state GDP. ESG_f is, in both panels, averaged across the years in which a firm is observed: in Panel A the Refinitiv ESG score, and in Panel B a score of the environmental, social and governance pillars built from firms' BRSR filings with the National Stock Exchange. Both sides of the regression are standardized: ESG_f and the allocation measure m_f, so that δ is measured in standard deviations of the allocation measure per standard deviation of the ESG score. The firm-specific slopes are obtained by partialling the source specification's fixed effects and controls out of both the outcome and the regressor, and then computing, for each firm separately, the ratio of the weighted cross-product of the two residualized variables to the weighted sum of squares of the residualized regressor, using that specification's observation weights. This is the within-firm OLS slope of the residualized outcome on the residualized regressor, and the pooled coefficient is recovered exactly as the denominator-weighted mean of these firm-level slopes. Standard errors are heteroskedasticity-robust. Log(assets) is obtained from the 2013 accounting data. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Internet Appendices

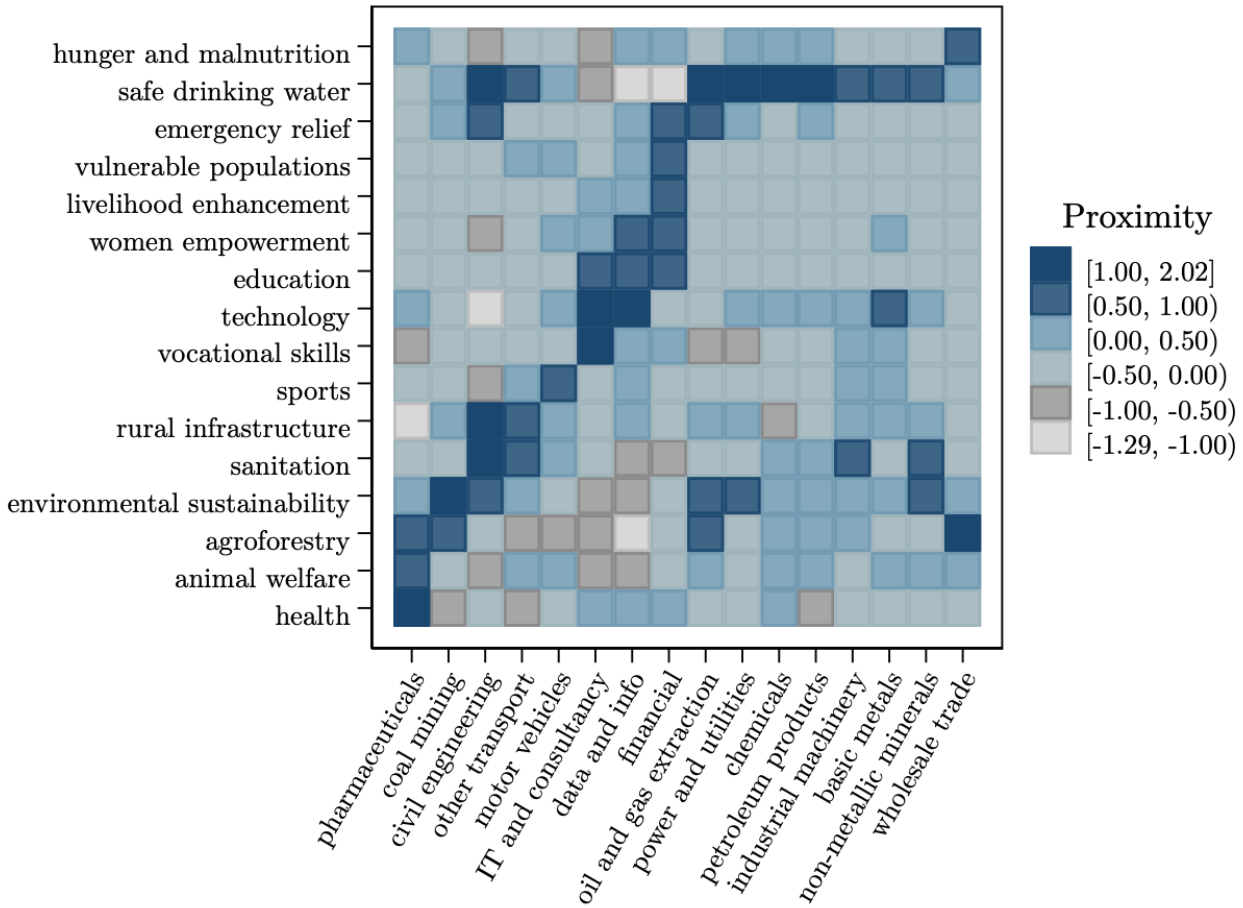
A Additional Tables and Figures

Figure A.1: CSR Spending Share by Industry



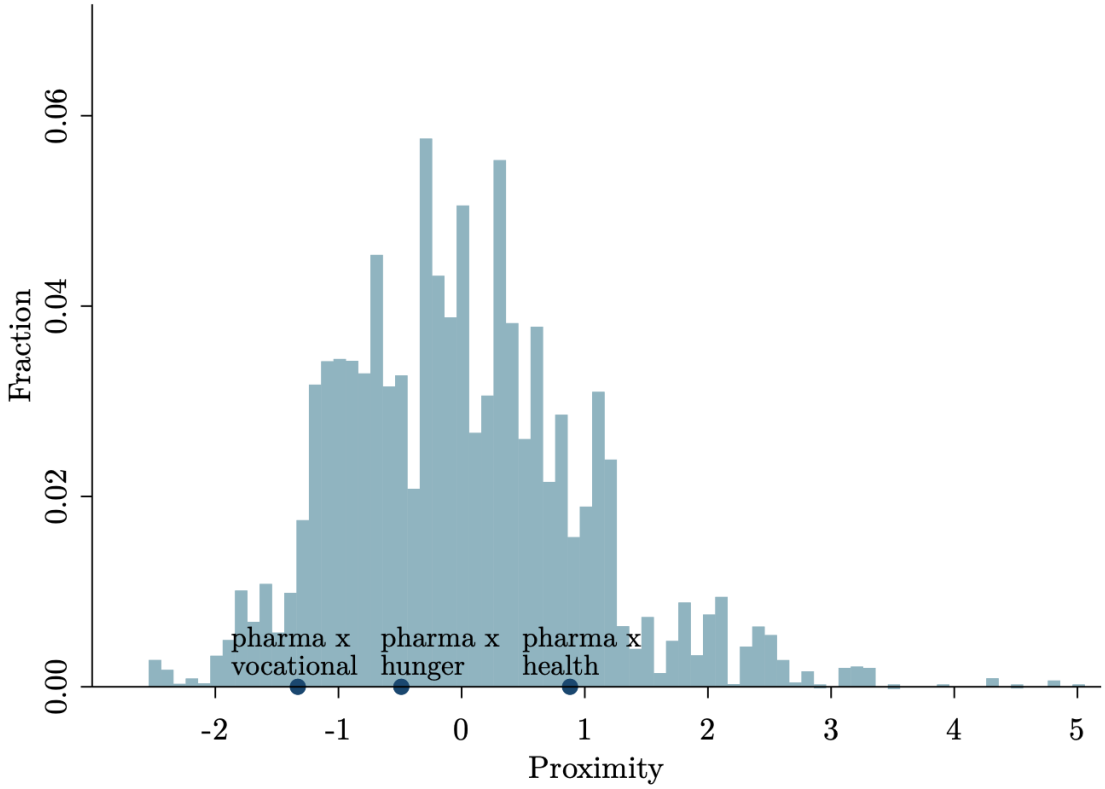
Notes: This figure depicts the share of each industry in total CSR expenditures for the 20 largest industries in the data.

Figure A.2: Proximity Across Topics and Largest Industries With Fixed Effects



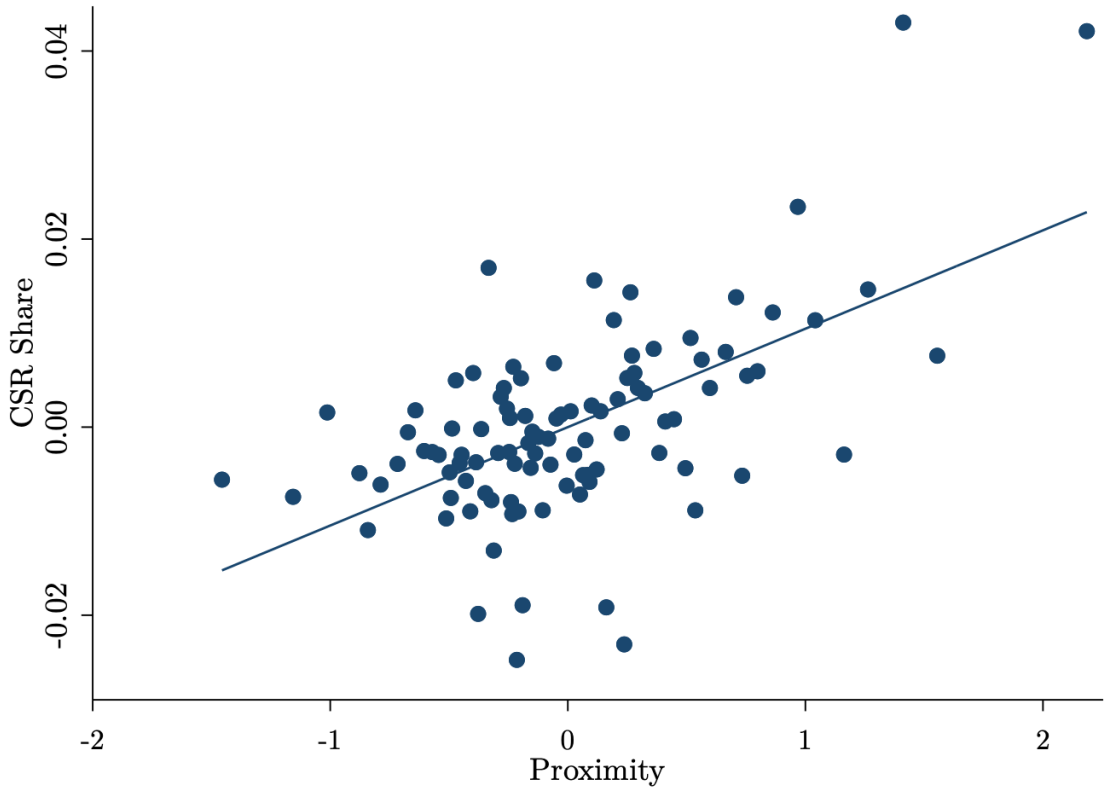
Notes: This figure depicts the proximity variable for all 16 topics and the largest 16 industries by total CSR spending. The unit of observation is at the industry×topic level. $\text{Proximity}_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. The proximity measure is residualized on firm and topic fixed effects.

Figure A.3: Distribution of Proximity



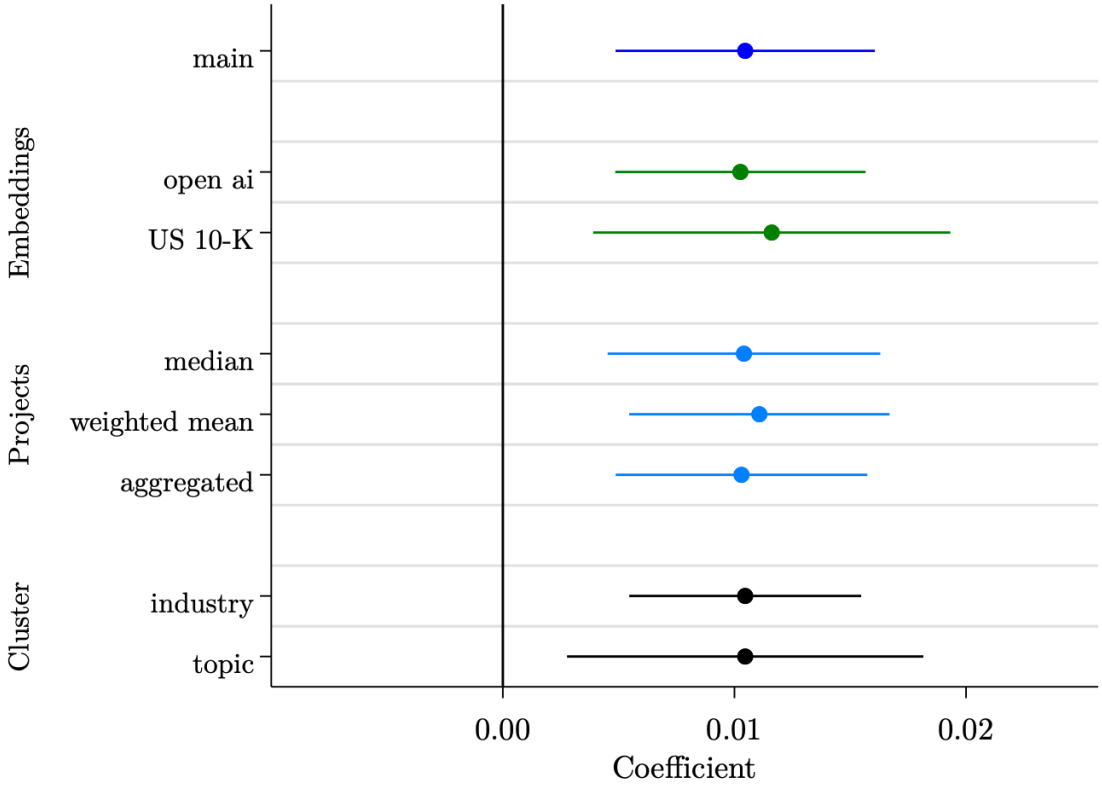
Notes: This figure depicts the distribution of the proximity variable. Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. The unit of observation is at the firm×topic level. The figure shows three examples, corresponding to approximately one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

Figure A.4: CSR Spending and Proximity



Notes: This figure describes the relationship between CSR spending share and proximity. The unit of observation is at the firm×topic level, and the 105,168 observations are binned into 100 equal-sized bins. The variable on the y-axis is the unconditional CSR spending share for a firm (f) over topics (d). Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. The variables are residualized on firm and topic fixed effects.

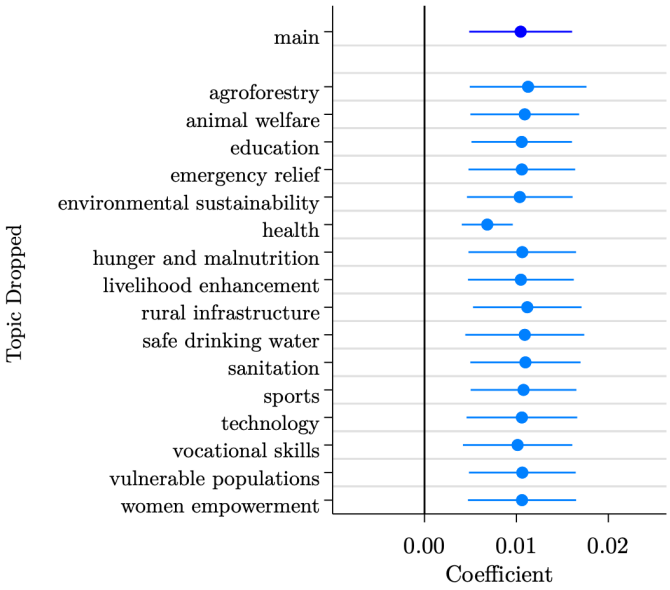
Figure A.5: CSR Spending and Proximity, Robustness I



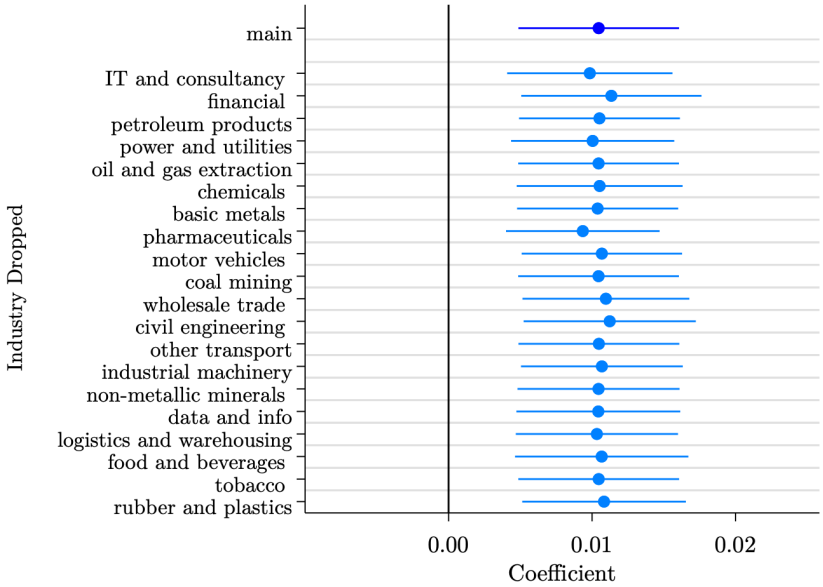
Notes: This figure describes the robustness of the relationship between CSR spending and proximity, derived from equation (8). The unit of observation is at the firm×topic level. The dependent variable is the share of CSR spending that firm (f) spends on topic (d). Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. Row 1 describes the main specification from Table 2, Panel A, column 1. Rows 2 and 3 show results for different word embeddings, using either OpenAI's NLP model or text from 10-K filings in the US. Rows 4 to 6 demonstrate different methods of aggregating information from projects at the topic level. The baseline averages the cosine similarity at the project level across topics. Here we consider using the topic level median instead, weighting each project by the number of tokens, or aggregating at the topic level before constructing the cosine similarity. Rows 7 and 8 cluster standard errors at the industry or topic level. The figure shows 95% confidence intervals.

Figure A.6: CSR Spending and Proximity, Robustness II

(a) Dropping Each Topic in Turn

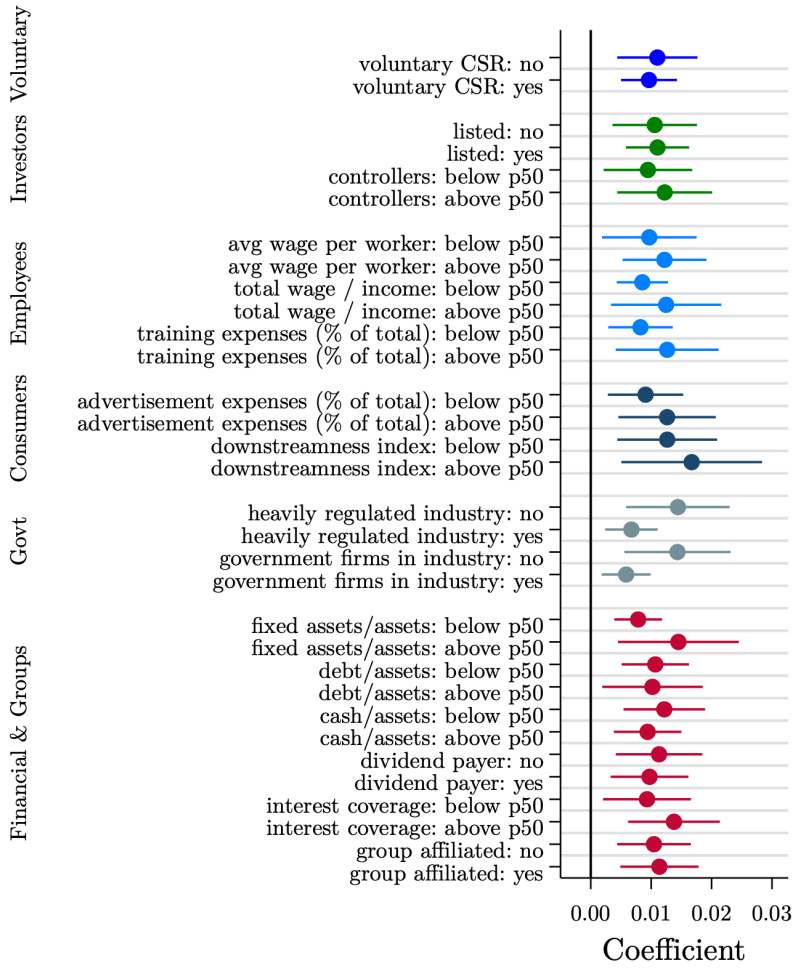


(b) Dropping Each of the Top 20 Industries in Turn



Notes: This figure describes the robustness of the relationship between CSR spending and proximity, derived from equation (8), dropping individual topics or industries. The unit of observation is at the firm×topic level. The dependent variable is the share of CSR spending that firm (f) spends on topic (d). The first row in each figure refers to the main specification from Table 2, Panel A, column 1. Figure A.6(a) describes robustness to dropping individual topics. Figure A.6(b) describes robustness to dropping individual industries. The figure shows 95% confidence intervals.

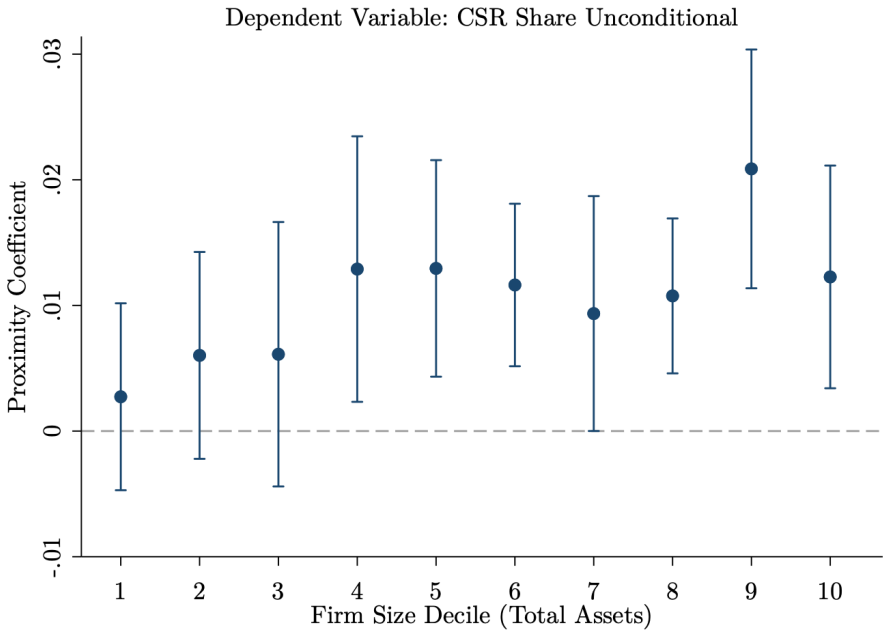
Figure A.7: CSR Spending and Proximity, Heterogeneity



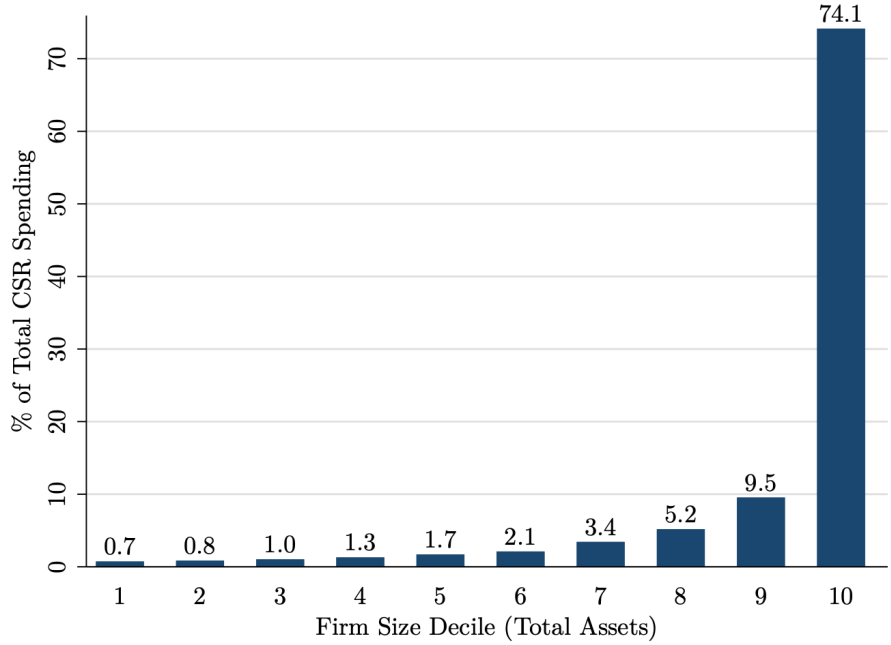
Notes: This figure describes the heterogeneity of the relationship between CSR spending and proximity, derived from equation (8). The dependent variable is the share of CSR spending that firm (f) spends on topic (d). In the first group, the sample is split into firms that spend voluntarily and those that do not. Firms are defined as spending voluntarily if they spend more than 1% of their profits on CSR in 2014 before the policy or if they spend more than 2.5% of their profits on average in the years 2015 to 2019 after the policy. In the second group, the sample is split by how exposed firms are to investors, measured by whether the firms are listed on stock exchanges and whether the equity share of controlling shareholders is below or above median. In the third group, the sample is split by how exposed firms are to employees, measured by the average wage per worker, the total wage bill over total income, and employee training expenses over total expenses. In the fourth group, the sample is split by how exposed firms are to consumers, measured by advertisement expenses over total expenses, and a downstreamness index of the industry (obtained from Antràs et al. (2012)). In the fifth group, the sample is split by how exposed firms are to the government, measured by whether the industry is heavily regulated (based on Awasthi et al. (2019)) and whether government firms are present in the Prowess sample for that industry. In the sixth group, interest coverage is defined as EBITDA over interest expense and business group affiliation follows Bertrand et al. (2002). If not otherwise specified, heterogeneity variables are obtained from the 2013 Prowess accounting data. The figure shows 95% confidence intervals.

Figure A.8: CSR Spending and Proximity by Firm Size

(a) Regression Results



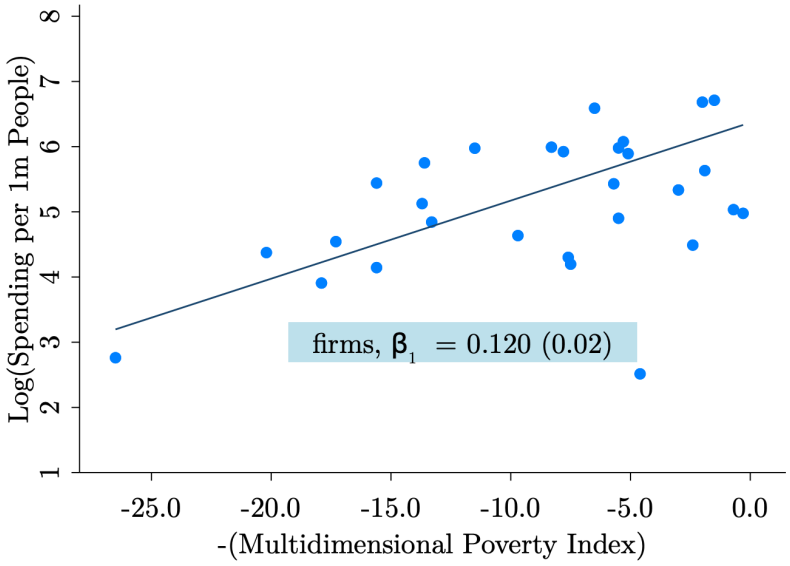
(b) Spending by Decile



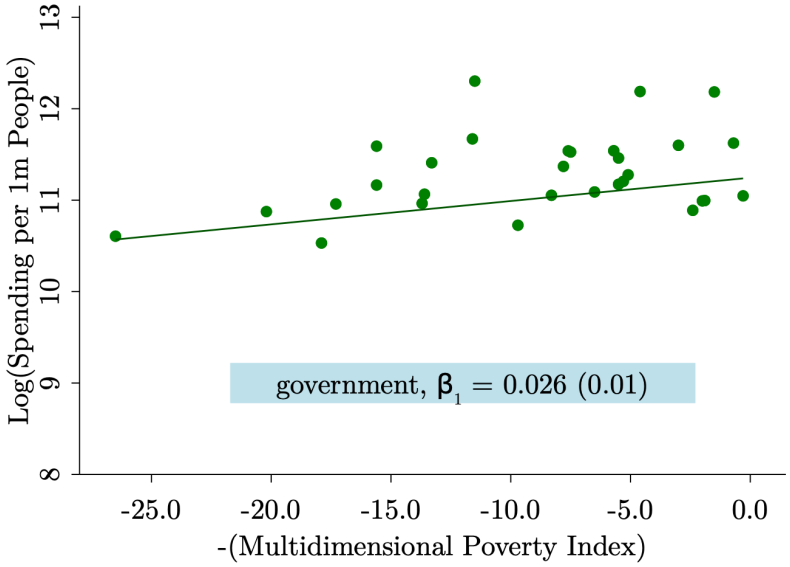
Notes: Figure A.8(a) plots the coefficient on technological proximity obtained by running equation (8) separately for each decile of the firm size distribution, where firm size is measured by total assets in 2013 in the Prowess data. The figure shows 95% confidence intervals. Figure A.8(b) plots the share of each firm size decile in total CSR spending.

Figure A.9: CSR Spending, Government Spending, and Poverty Index

(a) CSR Spending

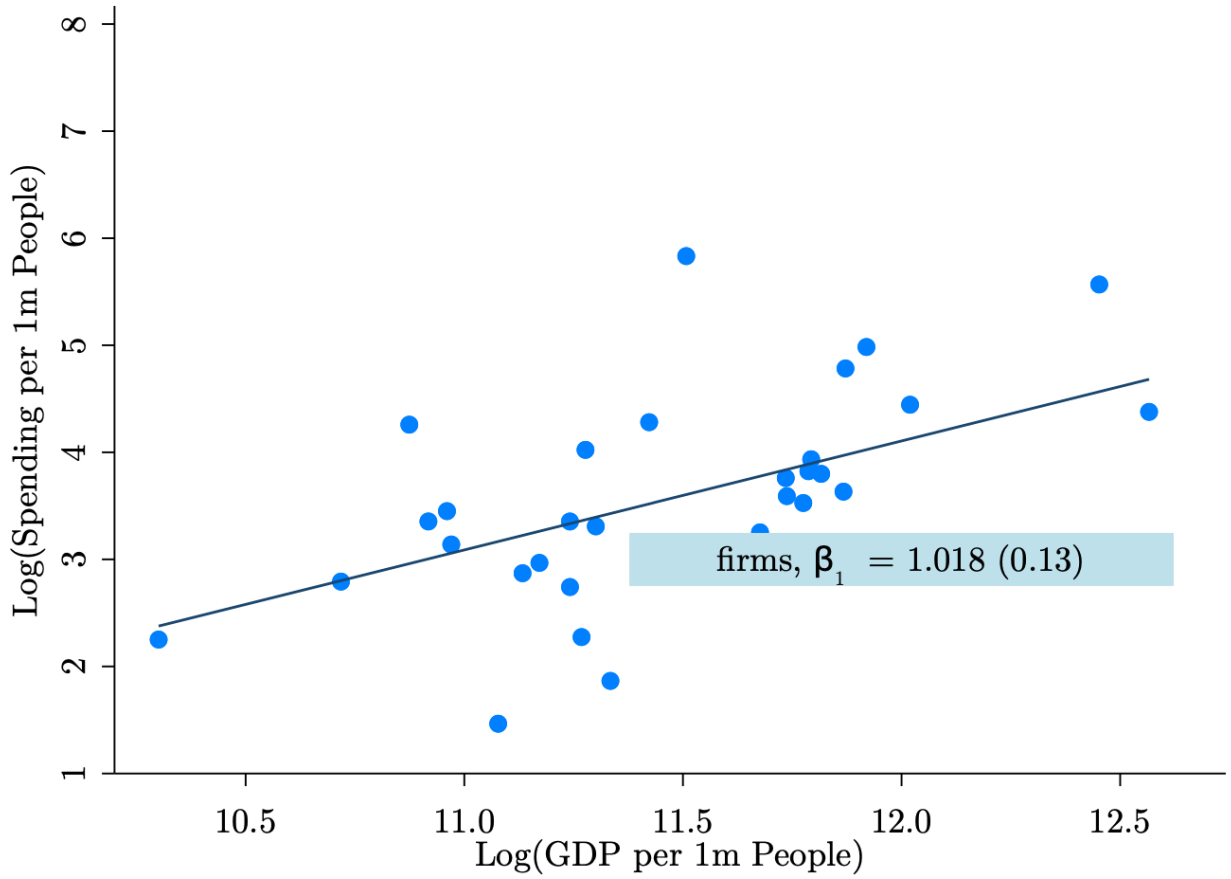


(b) Government Spending



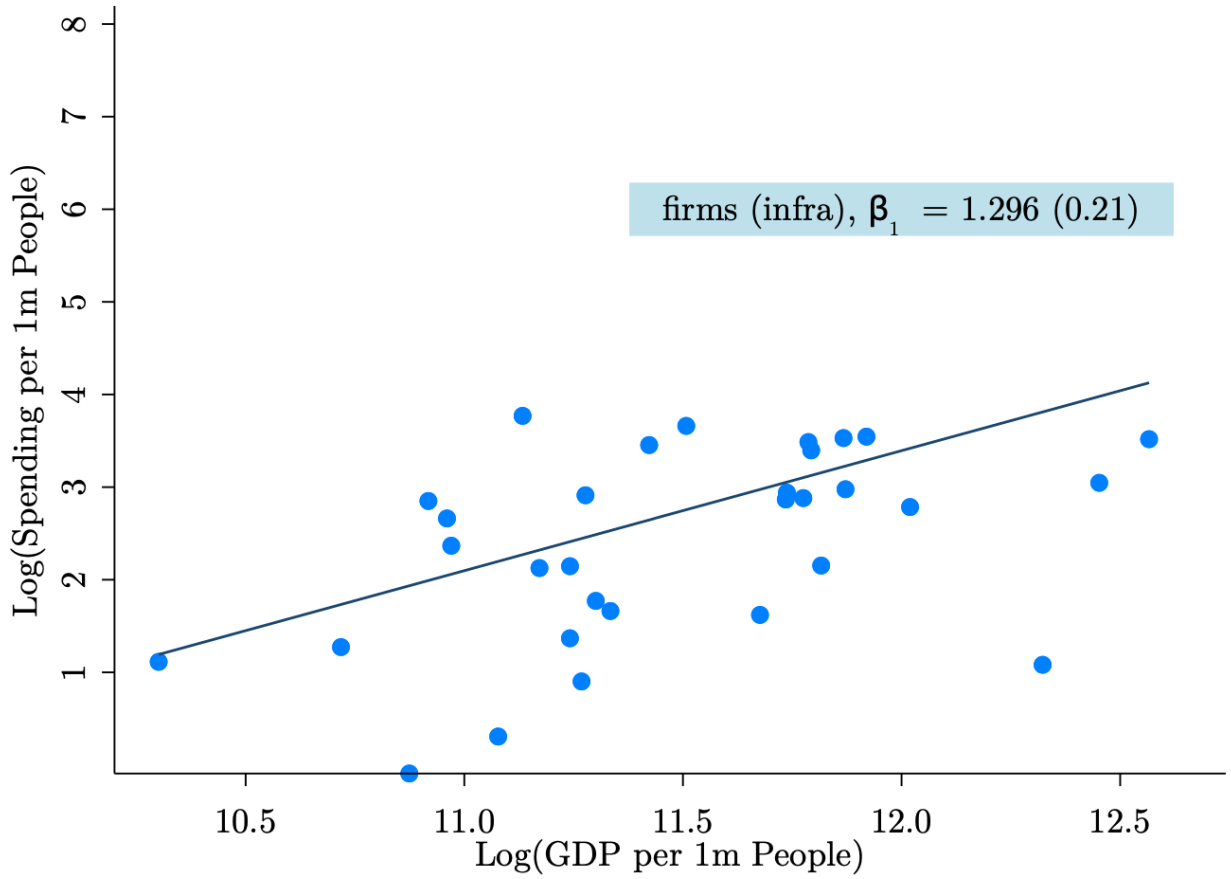
Notes: This figure depicts the relationship between a state-level multidimensional poverty index and firm (CSR) and government spending. The scattered dots indicate state-level observations, blue for firm spending and green for government spending. The lines indicate fitted linear approximations, blue for firm spending and green for government spending. The unit of observation is at the state level. The dependent variables are the log of spending (in millions, denominated in 2015 INR) per one million people by firms (CSR) and the government, aggregated from 2015 to 2019. The independent variable is the multidimensional poverty index in 2015/2016, multiplied by minus one (Government of India, 2024). Observations are weighted by the 2011 population.

Figure A.10: CSR Spending and State-Level GDP, Excluding Firm Location States



Notes: This figure depicts the relationship between state-level GDP as well as firm (CSR) spending, excluding spending in states in which firms are headquartered or have an establishment. The scattered dots indicate state-level observations. The line indicates a fitted linear approximation. The unit of observation is at the state level. The dependent variable is the log of spending (in millions, denominated in 2015 INR) per one million people by firms (CSR), aggregated from 2015 to 2019. The independent variable is the log of state-level GDP (in millions, denominated in 2015 INR) per one million people in 2013. Observations are weighted by the 2011 population.

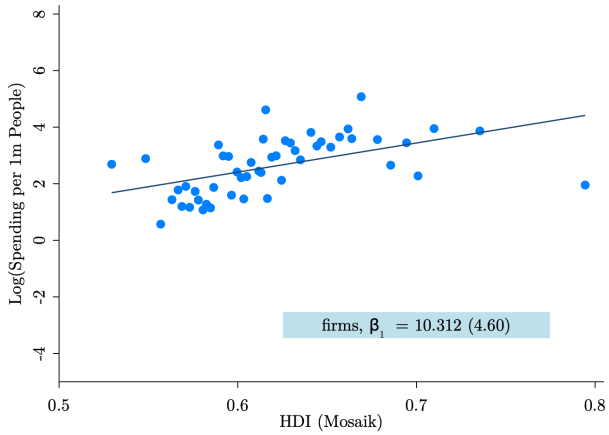
Figure A.11: CSR Spending and State-Level GDP, Firms' Rural Infrastructure Projects Only



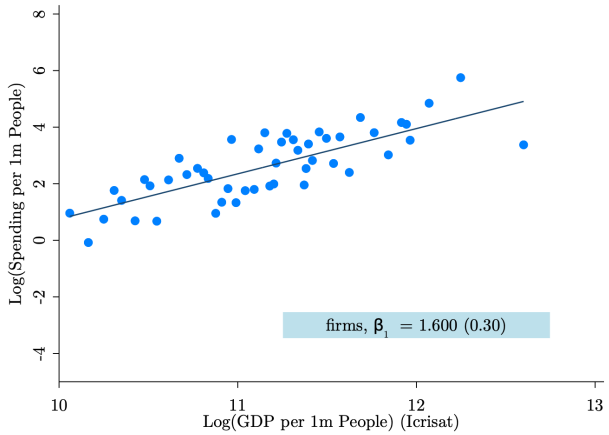
Notes: This figure depicts the relationship between state-level GDP as well as firm (CSR) spending, only focusing on projects in the rural infrastructure topic. The scattered dots indicate state-level observations. The line indicates a fitted linear approximation. The unit of observation is at the state level. The dependent variable is the log of CSR spending (in millions, denominated in 2015 INR) per one million people by firms on rural infrastructure projects, aggregated from 2015 to 2019. The independent variable is the log of state-level GDP (in millions, denominated in 2015 INR) per one million people in 2013. Observations are weighted by the 2011 population.

Figure A.12: CSR Spending and District-Level Measures

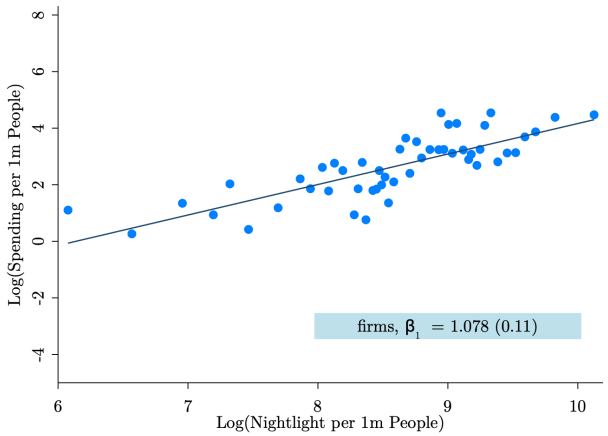
(a) Human Development Index (Mosaik)



(b) District GDP (Icrisat)

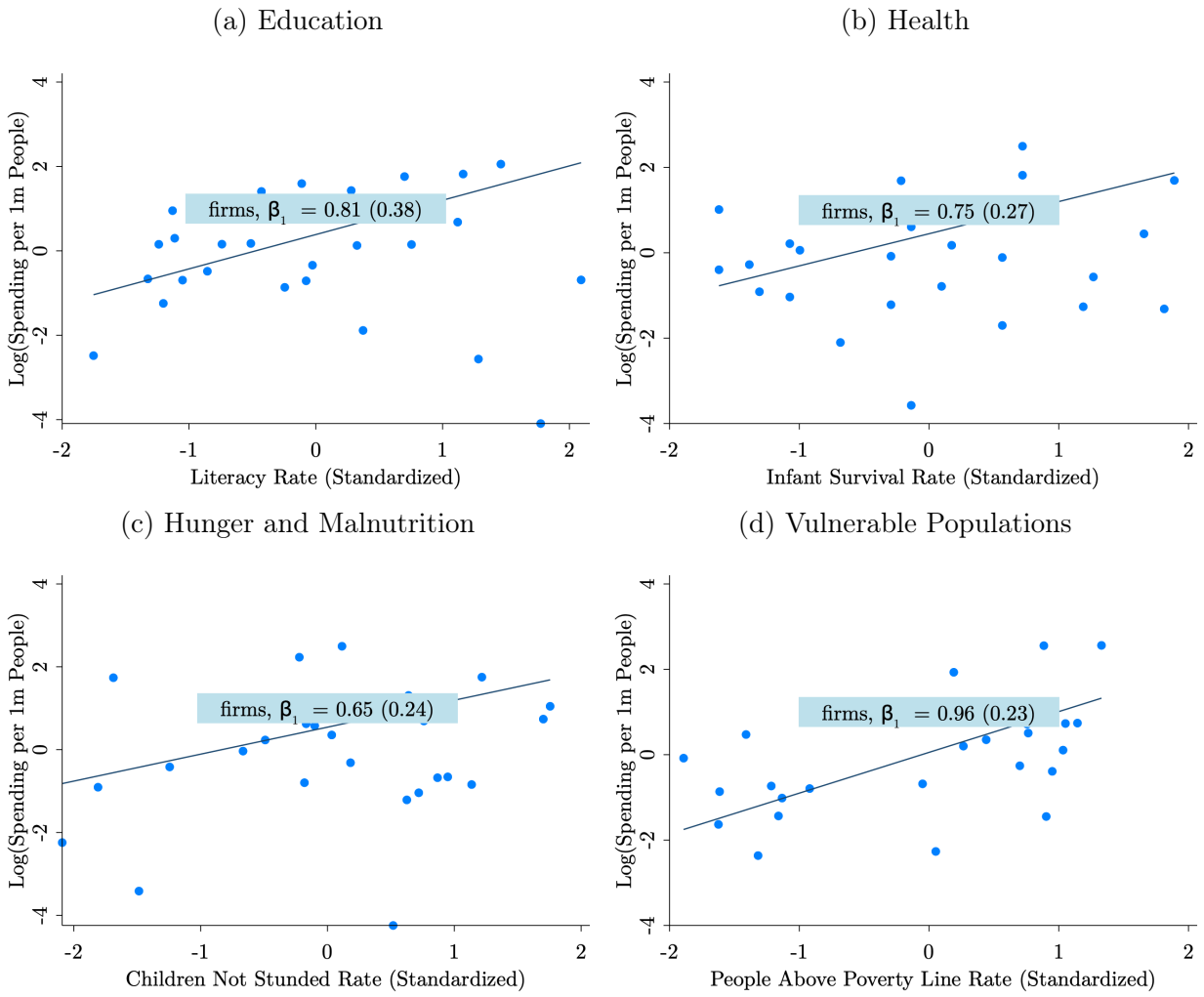


(c) Nightlight (SHRUG)



Notes: This figure depicts the relationship between CSR spending and district-level proxies for local development. The unit of observation is at the district level, and observations are binned into 50 equal-sized bins. The lines indicate fitted linear approximations. The dependent variables are the log of spending (in millions, denominated in 2015 INR) per one million people by firms (CSR), aggregated from 2015 to 2019. The independent variables are a human development index in Figure A.12(a) (Mosaik, 2023), a GDP measure in Figure A.12(b) (ICRISAT, 2020), and nightlight as a proxy for local economic conditions in Figure A.12(c) (Henderson et al., 2011; Asher et al., 2021; SHRUG, 2024). Observations are weighted by the 2011 population.

Figure A.13: CSR Topic Spending and Development Indicators



Notes: This figure depicts the relationship between CSR spending on given topics and the respective state-level development indicators. The scattered dots indicate state-level observations. The lines indicate fitted linear approximations. The unit of observation is at the state level. The dependent variable is the log of CSR spending (in millions, denominated in 2015 INR) per one million people on a given topic, aggregated from 2015 to 2019. Each panel plots this variable residualized on topic fixed effects, and the vertical scale is held common across panels. In Figure A.13(a), the independent variable is the literacy rate in 2011 (Government of India, 2021b). In Figure A.13(b), the independent variable is the survival rate for infants in 2013 (Government of India, 2021a). In Figure A.13(c), the independent variable is the rate of children not stunted in 2016 (Government of India, 2020). In Figure A.13(d), the independent variable is the rate of people living above the poverty line in 2012 (Government of India, 2014). Each development indicator is standardized across states to mean zero and unit variance, so the slope reported in each panel is in log points per standard deviation of the indicator. The fitted lines and the reported coefficients are weighted by the 2011 population; the plotted points are unweighted.

Figure A.14: CSR Spending and State-Level GDP



Notes: This figure describes the heterogeneity of the relationship between CSR spending and the log of state-level GDP, derived from equation (9). The unit of observation is at the firm×state level. The dependent variable is the share of CSR spending that firm (f) spends in state (s). In the first group, the sample is split into firms that spend voluntarily and those that do not. Firms are defined as spending voluntarily if they spend more than 1% of their profits on CSR in 2014 before the policy or if they spend more than 2.5% of their profits on average in the years 2015 to 2019 after the policy. In the second group, the sample is split by how exposed firms are to investors, measured by whether the firms are listed on stock exchanges and whether the equity share of controlling shareholders is below or above median. In the third group, the sample is split by how exposed firms are to employees, measured by the average wage per worker, the total wage bill over total income, and employee training expenses over total expenses. In the fourth group, the sample is split by how exposed firms are to consumers, measured by advertisement expenses over total expenses, and a downstreamness index of the industry (obtained from Antràs et al. (2012)). In the fifth group, the sample is split by how exposed firms are to the government, measured by whether the industry is heavily regulated (based on Awasthi et al. (2019)) and whether government firms are present in the Prowess sample for that industry. In the sixth group, interest coverage is defined as EBITDA over interest expense and business group affiliation follows Bertrand et al. (2002). If not otherwise specified, heterogeneity variables are obtained from the 2013 Prowess accounting data. The figure shows 95% confidence intervals.

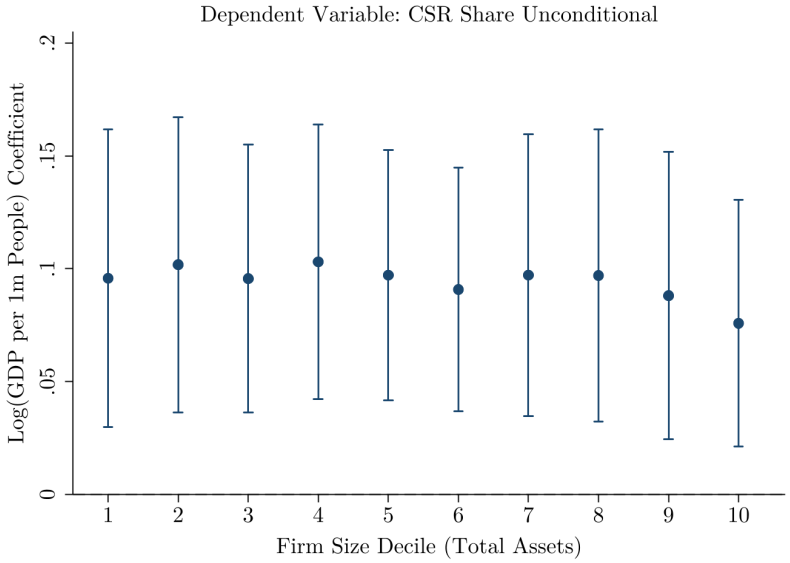
Figure A.15: CSR Spending and Firm Headquarter State



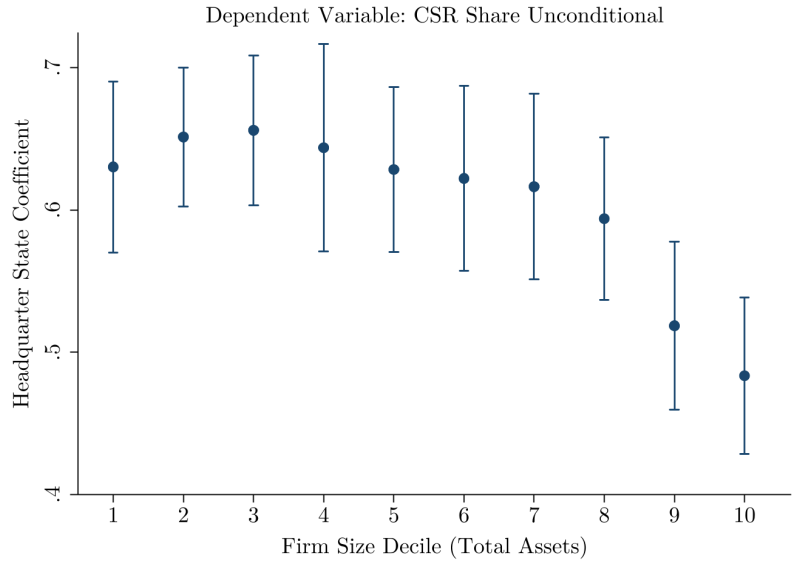
Notes: This figure describes the heterogeneity of the relationship between CSR spending and an indicator of the firm headquarter state, similar to equation (9). The unit of observation is at the firm×state level. The dependent variable is the share of CSR spending that firm (f) spends in state (s). In the first group, the sample is split into firms that spend voluntarily and those that do not. Firms are defined as spending voluntarily if they spend more than 1% of their profits on CSR in 2014 before the policy or if they spend more than 2.5% of their profits on average in the years 2015 to 2019 after the policy. In the second group, the sample is split by how exposed firms are to investors, measured by whether the firms are listed on stock exchanges and whether the equity share of controlling shareholders is below or above median. In the third group, the sample is split by how exposed firms are to employees, measured by the average wage per worker, the total wage bill over total income, and employee training expenses over total expenses. In the fourth group, the sample is split by how exposed firms are to consumers, measured by advertisement expenses over total expenses, and a downstreamness index of the industry (obtained from Antràs et al. (2012)). In the fifth group, the sample is split by how exposed firms are to the government, measured by whether the industry is heavily regulated (based on Awasthi et al. (2019)) and whether government firms are present in the Prowess sample for that industry. In the sixth group, interest coverage is defined as EBITDA over interest expense and business group affiliation follows Bertrand et al. (2002). If not otherwise specified, heterogeneity variables are obtained from the 2013 Prowess accounting data. The figure shows 95% confidence intervals.

Figure A.16: CSR Spending, State-Level GDP, and Firm Headquarter State

(a) Tilt Toward Richer States

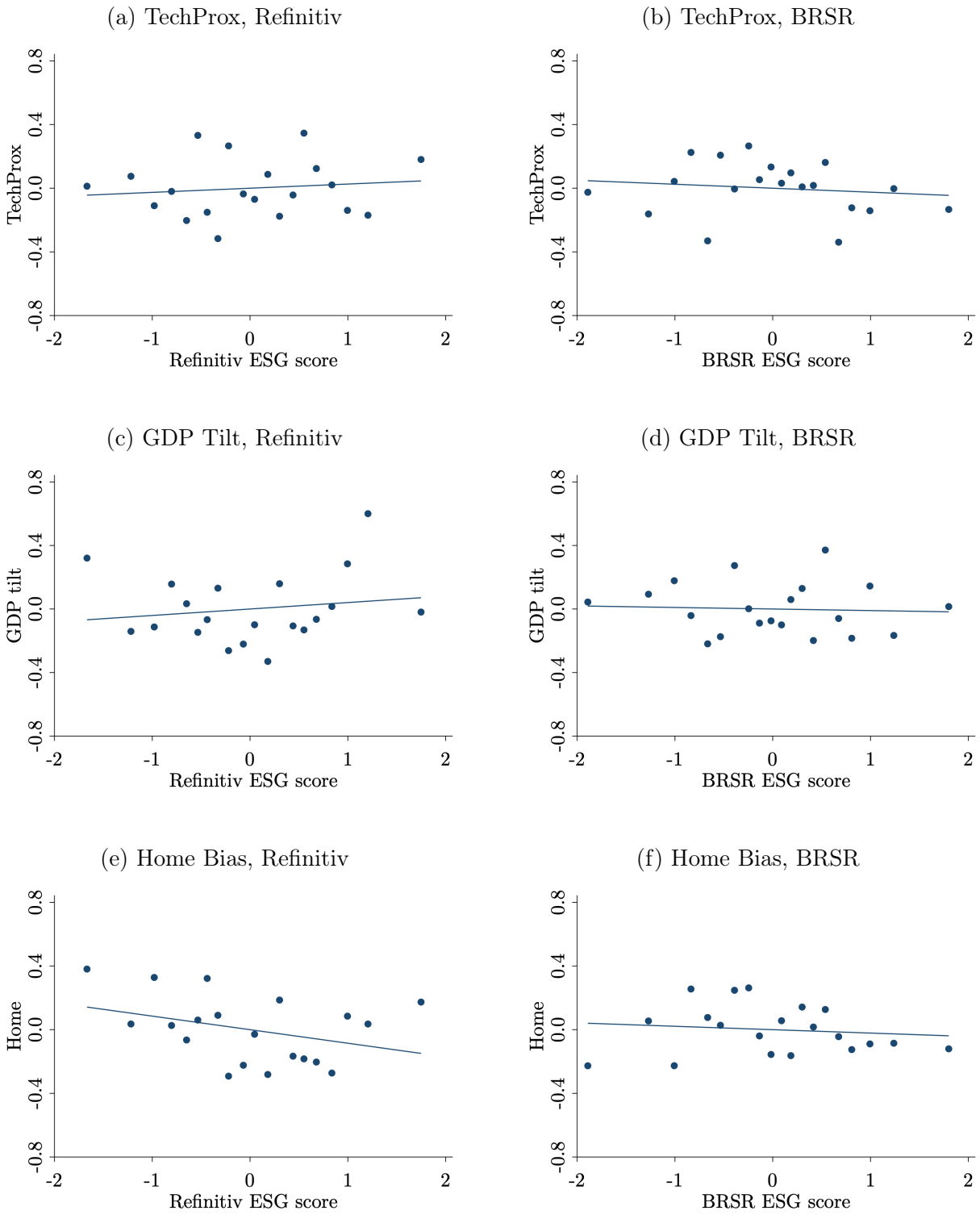


(b) Home Bias



Notes: Figure A.16(a) plots the coefficient on log state GDP per capita obtained by running equation (9) separately for each decile of the firm size distribution, where firm size is measured by total assets in 2013 in the Prowess data. Figure A.16(b) plots the coefficient on an indicator for the firm's headquarter state, obtained by running the same specification with that indicator in place of state GDP, again separately by decile. Both regressions are weighted by state population and cluster standard errors at the state level, as in equation (9). The figure shows 95% confidence intervals.

Figure A.17: ESG Scores and Firm-Level Allocation



Notes: This figure describes the relationship between each firm's ESG score and each of the three allocation measures of Table 4. Each figure has 20 equal-sized bins. The variables on the vertical axes are the firm-specific slopes *TechProx*, *GDP tilt*, and *Home*. All variables are residualized on industry fixed effects and log total assets, so that each panel corresponds to the even-numbered column of Table 4 for the same measure and score.

Table A.1: Summary Statistics

	Mean	SD	Median
Firm-level			
Income (m INR)	15,137	96,366	3,008
Voluntary CSR (yes/no)	0.244	0.429	0.000
Firm-topic-level			
CSR share unconditional	0.062	0.189	0.000
Any CSR spending (yes/no)	0.223	0.416	0.000
CSR share conditional	0.280	0.314	0.142
Firm-state-level			
CSR share unconditional	0.029	0.151	0.000
Any CSR spending (yes/no)	0.067	0.250	0.000
CSR share conditional	0.437	0.401	0.290
Observations			
Unique firms (nr)	6,573		
Unique firm×topics (nr)	105,168		
Unique firm×states (nr)	196,415		

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

Table A.2: Firm CSR Share Proximity and Firm Product Share Proximity

	Firm CSR Share Proximity$_{ff'}$	
	(1)	(2)
Firm Product Share Proximity$_{ff'}$	0.03***	0.03**
	(0.01)	(0.01)
Firm Clustering		✓
R-squared	0.00	0.00
Observations	188,805	188,805

Notes: This table describes the relationship between the firm CSR share proximity and the firm product share proximity across all pairs of firms. The unit of observation is at the firm-pair-level. Firm Product Share Proximity$_{ff'}$ is the cosine similarity of the vectors of product shares of firms f and f'. Let $\boldsymbol{\theta_f}$ be the vector of product sales shares $\theta_{fj} = (\text{Sales}_{fj}/\text{Sales}_f)$, the share of product j in the sales of firm f. Then,

$$\text{Firm Product Share Proximity}_{ff'} = \cos(\boldsymbol{\theta_f}, \boldsymbol{\theta_{f'}}) \tag{A.1}$$

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

$$\text{Firm CSR Share Proximity}_{ff'} = \cos(\boldsymbol{\psi_f}, \boldsymbol{\psi_{f'}}) \tag{A.2}$$

We keep pairs of firms (f, f') such that $f < f'$ to avoid duplicates. Standard errors are not clustered in column 1 and clustered at the firm-level in column 2. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.3: Firm Semantic Proximity Within and Across Industries

	Firm Semantic Proximity$_{ff'}$		
	2-digit	3-digit	4-digit
	(1)	(2)	(3)
Same Industry$_{ff'}$	0.118***	0.146***	0.155***
	(0.000)	(0.000)	(0.001)
Constant	0.679***	0.681***	0.681***
	(0.000)	(0.000)	(0.000)
R-squared	0.03	0.02	0.02
Observations	5,808,936	5,808,936	5,808,936

Notes: This table describes the average textual proximity of firm level descriptions between pairs of firms in the same industry, relative to firms in different industries. The unit of observation is at the firm-pair-level. Firm Semantic Proximity$_{ff'}$ is defined as the cosine similarity of the embeddings characterizing firms f and f', obtained from the text in SEC 10-K filings. Same Industry$_{ff'}$ is a dummy equal to 1 if firms f and f' belong to the same SIC industry, at the 2, 3, and 4-digit levels in columns 1 to 3, respectively. We keep pairs of firms (f, f') such that $f < f'$ to avoid duplicates. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.4: Firms' Production Across Industries and Industry Semantic Proximity

	Product Sales Share$_{f\ i(prod)}$		Any Product$_{f\ i(prod)}$	
	(1)	(2)	(3)	(4)
Industry Semantic Proximity$_{i(f)i(prod)}$	0.037**	0.055**	0.100***	0.131***
	(0.015)	(0.022)	(0.020)	(0.027)
Avg dep var	0.004	0.004	0.013	0.013
Industry(Product) FE		✓		✓
R-squared	0.00	0.02	0.01	0.03
Observations	38,409	38,409	38,409	38,409

Notes: This table describes whether firm f in industry $i(f)$ is more likely to produce in industry $i(prod)$ if industries $i(f)$ and $i(prod)$ have similar embeddings. Industry Semantic Proximity$_{i(f)i(prod)}$ is the proximity between industry i of the firm and industry i of the product as defined by the cosine similarity of their respective embedding vectors. In columns 1 and 2, the dependent variable is firm's sales value of products in a given industry $i(prod)$, calculated over the entire sales value of the firm f. In columns 3 and 4, the dependent variable is an indicator that equals one if firm f produces in $i(prod)$. Columns 2 and 4 include fixed effects for the industry of the products. The unit of observation is at the firm×industry(product)-level. We exclude observations for which the firm industry equals the product industry. Standard errors are clustered at the industry level of the firm. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.5: Industries' Input-Output Features and Industry Semantic Proximity

	IO Proximity$_{ii'}$		
	(1)	(2)	(3)
Industry Semantic Proximity$_{ii'}$	0.069***	0.434***	0.595***
	(0.007)	(0.056)	(0.058)
Benchmark	Leontief	Input	Output
R-squared	0.08	0.06	0.10
Observations	990	990	990

Notes: This table describes the relationship of textual proximity between pairs of industries and other benchmarks that capture proximity across industries. For each pair of industry ii', Industry Semantic Proximity$_{ii'}$ is the proximity between industry i and i' as defined by the cosine similarity of their respective embedding vectors. IO Proximity$_{ii'}$ are metrics for proximity between industries based on the input-output matrix for India (obtained from the OECD). Leontief$_{ii'}$ is the ii' entry of the Leontief inverse of the input-output matrix. Input$_{ii'}$ is the cosine similarity of the input shares of industries i and i'. Output$_{ii'}$ is the cosine similarity of the output shares of industries i and i'. The unit of observation is at the industry-pair-level. We keep pairs of industries (i, i') such that $i < i'$ to avoid duplicates. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.6: CSR Spending and Proximity, Disaggregated and Aggregated Topics

Panel A: Firm × Disaggregated topic			
	CSR Share Unconditional$_{fd}$	Any CSR Spending$_{fd}$	CSR Share Conditional$_{fd}$
	(1)	(2)	(3)
Proximity$_{i(f)d}$	0.001***	0.006***	0.005**
	(0.000)	(0.001)	(0.002)
Avg dep var	0.009	0.056	0.166
SD effect (of mean dep var)	0.137	0.108	0.029
Firm FE	✓	✓	✓
Disaggregated topic FE	✓	✓	✓
R-squared	0.05	0.13	0.41
Observations	481,234	481,234	25,177
Panel B: Firm × Aggregated topic			
	CSR Share Unconditional$_{fd}$	Any CSR Spending$_{fd}$	CSR Share Conditional$_{fd}$
	(1)	(2)	(3)
Proximity$_{i(f)d}$	0.017***	0.028***	0.026***
	(0.005)	(0.005)	(0.006)
Avg dep var	0.100	0.321	0.311
SD effect (of mean dep var)	0.169	0.087	0.082
Firm FE	✓	✓	✓
Aggregated topic FE	✓	✓	✓
R-squared	0.21	0.36	0.33
Observations	65,730	65,730	19,264

Notes: This table describes the relationship between CSR spending and proximity, derived from equation (8), on the firm×disaggregated topic level (Panel A) and the firm×aggregated topic level (Panel B). In Panel A, the dependent variables are the share of CSR spending in firm (f) over disaggregated topics (d) relative to the total firm CSR spending, an indicator for any CSR spending by firm (f) in a given disaggregated topic (d), and the share of CSR spending conditional on any spending. Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a disaggregated topic. Disaggregated topics are defined in Table C.2. In Panel B, the dependent variables are the share of CSR spending in firm (f) over aggregated topics (d), an indicator for any CSR spending by the firm (f) in a given aggregated topic (d), and the share of CSR spending conditional on any spending. Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and an aggregated topic. Aggregated topics are defined in Table C.3. Standard errors are clustered at the industry×topic level, where the topic is disaggregated in Panel A and aggregated in Panel B. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.7: CSR Spending and Proximity, Industry × Topic Level

	CSR Share Unconditional$_{id}$ (1)	Any CSR Spending$_{id}$ (2)	CSR Share Conditional$_{id}$ (3)	Log(CSR Share Conditional)$_{id}$ (4)
Proximity$_{id}$	0.019***	0.046***	0.020***	0.340***
	(0.005)	(0.013)	(0.006)	(0.063)
Avg dep var	0.063	0.867	0.072	.
Industry FE	✓	✓	✓	✓
Topic FE	✓	✓	✓	✓
R-squared	0.61	0.57	0.66	0.53
Observations	1,136	1,136	983	983

Notes: This table describes the relationship between CSR spending and proximity, derived from equation (8), on the industry×topic level. The dependent variables are the share of CSR spending in industry (i) over topics (d), an indicator for any CSR spending by the industry (i) in a given topic (d), the share of CSR spending conditional on any spending, and the log of the share of CSR spending conditional on any spending. Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. Standard errors are robust. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.8: Comparison Proximity and Government Spending

	CSR Share Unconditional$_{fd}$			
	(1)	(2)	(3)	(4)
Proximity$_{i(f)d}$	0.017***	0.009**	0.019**	0.020***
	(0.005)	(0.005)	(0.009)	(0.007)
Government Share$_d$		0.093***		0.073***
		(0.004)		(0.006)
Weight	None	None	CSR	CSR
SD effect (of mean dep var) prox	0.169		0.195	
SD effect (of mean dep var) gov		0.934		0.728
Firm FE	✓	✓	✓	✓
Topic FE	✓		✓	
R-squared	0.21	0.16	0.25	0.17
Observations	65,730	65,730	65,730	65,730

Notes: This table compares proximity and government spending shares in the allocation of CSR. Topics are aggregated at a level consistent between CSR topics and government expenditures, as indicated in Table C.3. Proximity and the government share are each standardized to mean zero and unit variance before estimation, so every coefficient is the effect of a one-standard-deviation increase in the regressor. We report that effect as a proportion of the mean dependent variable, that is, the coefficient divided by the mean of the dependent variable: for proximity in columns 1 and 3, and for the government share in columns 2 and 4. The unit of observation is at the firm×topic level. The dependent variable is the share of CSR spending of a firm (f) over topics (d). Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and an aggregated topic. Government Share$_d$ is the share of government spending on a given topic d. Columns 3 and 4 weight observations by CSR spending. Standard errors are clustered at the industry×topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.9: Indirect Project Implementation and Proximity

	Indirect Project Implementation$_{pfd}$	
	Excl Text Data	Incl Text Data
	(1)	(2)
Proximity$_{i(f)d}$	-0.007**	-0.009**
	(0.004)	(0.004)
Avg dep var	0.514	0.550
Firm FE	✓	✓
Topic FE	✓	✓
R-squared	0.62	0.57
Observations	82,201	82,201

Notes: This table describes the relationship between indirect project implementation and proximity, based on equation (A.4). The unit of observation is at the firm×project level. The dependent variable is an indicator that is one if the project is implemented indirectly. In column 1, we define a project to be implemented indirectly if firms report this to be the case based on an MCA question. In column 2, we additionally count projects as indirect if their textual description contains a token or bigram indicative of third-party implementation or an NGO (see Appendix C.1.5). Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. Observations are weighted by the inverse of the number of projects by topic to give equal weight to all topics, as in the main proximity regressions. Standard errors are clustered at the topic×industry level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

$$\text{Indirect Project Implementation}_{pfd} = \beta_0 + \text{Proximity}_{i(f)d} + \alpha_f + \alpha_d + \varepsilon_{pfd} \tag{A.4}$$

Table A.10: CSR Spending, State-Level Characteristics, and Firm Locations

	CSR Share Unconditional$_{fs}$			Any CSR Spending$_{fs}$			CSR Share Conditional$_{fs}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log(GDP per 1m People)$_s$	0.096***	0.014***	0.010**	0.156***	0.058***	0.050***	0.145***	-0.001	-0.006
	(0.031)	(0.004)	(0.004)	(0.040)	(0.008)	(0.008)	(0.022)	(0.011)	(0.010)
1(Firm HQ State)$_{fs}$		0.601***	0.558***		0.722***	0.651***		0.333***	0.307***
		(0.026)	(0.025)		(0.022)	(0.016)		(0.023)	(0.022)
1(Firm Establishment State)$_{fs}$			0.146***			0.242***			0.135***
			(0.010)			(0.014)			(0.010)
Avg dep var	0.029	0.029	0.029	0.067	0.067	0.067	0.437	0.437	0.437
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.08	0.52	0.54	0.18	0.44	0.46	0.42	0.57	0.58
Observations	196,415	196,415	196,415	196,415	196,415	196,415	9,736	9,736	9,736

Notes: This table describes the relationship between state-level characteristics as well as firm headquarters, firm establishments as proxied by large capital investments, and CSR spending, derived from equation (9). The unit of observation is at the firm×state level. The dependent variables are the share of CSR spending of a firm (f) over states (s), an indicator for any CSR spending by firm (f) in a given state (s), and the share of CSR spending conditional on any spending. The independent variables are the log of state-level GDP per 1 million people, an indicator that equals one if the firm is headquartered in the state as per government records, and an indicator that equals one if the firm has an establishment in the state, proxied by a large past capital investment. We control for the log of population in millions. Observations are weighted by the 2011 population. Standard errors are clustered at the state-level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.11: ESG Scores and Allocation, Coverage and Main Results

	Full sample (1)	Refinitiv sample (2)	BRSR sample (3)
Panel A: Coverage			
Firms	6,573	449	722
Share of total CSR spending	100.0%	58.8%	60.9%
Panel B: CSR Spending and Proximity, Equation (8)			
	CSR Share Unconditional$_{fd}$		
Proximity$_{i(f)d}$	0.010***	0.012***	0.013***
	(0.003)	(0.004)	(0.003)
Avg dep var	0.062	0.063	0.063
Firm FE	✓	✓	✓
Topic FE	✓	✓	✓
R-squared	0.24	0.29	0.30
Observations	105,168	7,184	11,552
Panel C: CSR Spending and State Characteristics, Equation (9)			
	CSR Share Unconditional$_{fs}$		
Log(GDP per 1m People)$_s$	0.014***	0.015***	0.018***
	(0.004)	(0.003)	(0.004)
1(Firm Headquarter State)$_{fs}$	0.601***	0.450***	0.502***
	(0.026)	(0.033)	(0.027)
Avg dep var	0.029	0.025	0.027
Firm FE	✓	✓	✓
R-squared	0.52	0.46	0.49
Observations	196,415	13,413	21,592

Notes: This table describes the two samples of firms for which an ESG score is available and shows that the paper's main results hold within them. Column 1 uses all firms in the CSR regression sample; columns 2 and 3 restrict to the firms entering Panels A and B of Table 4, namely those with a Refinitiv score and those with a BRSR score. Panel A reports the number of firms and the share of total CSR expenditure in the regression sample that they account for. Panels B and C re-estimate equations (8) and (9) on each sample. The unit of observation is the firm × topic in Panel B and the firm × state in Panel C. Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. Panel C weights observations by state population and controls for log state population, and reports the specification that includes both the headquarter indicator and state GDP per capita, so that column 1 corresponds to column 2 of Table 3 rather than to its column 1. Standard errors are clustered by industry × topic in Panel B and by state in Panel C, as in Tables 2 and 3. ∗∗∗, ∗∗ and ∗ indicate significance at the 1%, 5% and 10% levels.

Table A.12: ESG Scores and Allocation, Robustness by ESG Score Construction

	TechProx (std)$_f$			GDP tilt (std)$_f$		Home (std)$_f$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline measure, plus indicators divided by Prowess turnover (23 indicators)							
ESG score (std)$_f$	-0.064	-0.063	-0.080	-0.052	-0.057	-0.030	-0.038
	(0.055)	(0.056)	(0.055)	(0.056)	(0.056)	(0.056)	(0.055)
Log(assets)$_f$		0.017	0.033		-0.112***		-0.188***
		(0.023)	(0.023)		(0.024)		(0.025)
Fixed assets/assets$_f$			-0.083				
			(0.279)				
Panel B: Baseline measure, plus indicators divided by filed turnover (23 indicators)							
ESG score (std)$_f$	-0.044	-0.041	-0.040	0.041	0.020	0.060	0.025
	(0.043)	(0.044)	(0.044)	(0.042)	(0.041)	(0.042)	(0.041)
Log(assets)$_f$		0.014	0.031		-0.110***		-0.186***
		(0.023)	(0.023)		(0.024)		(0.026)
Fixed assets/assets$_f$			0.010				
			(0.278)				
Panel C: Baseline measure, ranked within industry × year (19 indicators)							
ESG score (std)$_f$	-0.010	-0.014	-0.030	-0.022	0.008	-0.071	-0.021
	(0.054)	(0.055)	(0.054)	(0.053)	(0.052)	(0.054)	(0.053)
Log(assets)$_f$		0.018	0.036		-0.112***		-0.187***
		(0.023)	(0.023)		(0.024)		(0.026)
Fixed assets/assets$_f$			0.002				
			(0.278)				
Panel D: Baseline measure, plus all remaining Core indicators (44 indicators)							
ESG score (std)$_f$	-0.093**	-0.101*	-0.082	0.103**	0.002	0.192***	0.028
	(0.046)	(0.057)	(0.057)	(0.047)	(0.053)	(0.045)	(0.057)
Log(assets)$_f$		-0.010	0.010		-0.111***		-0.181***
		(0.028)	(0.029)		(0.027)		(0.032)
Fixed assets/assets$_f$			-0.055				
			(0.274)				
Industry FE	✓	✓	✓	✓	✓	✓	✓
Size control		✓	✓		✓		✓
Tangibility control			✓				
Observations	722	722	713	722	722	722	722

Notes: This table repeats Panel B of Table 4 under four alternative constructions of the BRSR score, described in Appendix C.2.6. Panels A and B add the greenhouse-gas, water, energy and waste footprints per rupee of turnover, measuring turnover by Prowess total income and as filed respectively. Panel C ranks the baseline indicators within industry and year rather than within year. Panel D adds all remaining Core indicators, 25 in total, 17 of them absolute quantities. A percentile rank on such a quantity orders firms by the scale of their operations rather than by their conduct. Panel D retains the filed-turnover intensities used in Panel B and adds the quantities alongside them, together with four intensities that Panels A and B exclude because they are scaled by a firm-chosen unit of physical output rather than by turnover. Column 3 adds the share of fixed assets in total assets, a further proxy for physical scale. Asset measures are from the 2013 accounting data. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table A.13: ESG Scores and Allocation, Robustness Excluding Governance (G)

	TechProx (std)$_f$		GDP tilt (std)$_f$		Home (std)$_f$	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Refinitiv, environmental and social attributes						
ESG score (std)$_f$	0.074	0.053	-0.080*	-0.012	-0.233***	-0.136**
	(0.052)	(0.057)	(0.047)	(0.055)	(0.055)	(0.057)
Log(assets)$_f$		0.031		-0.103***		-0.146***
		(0.041)		(0.038)		(0.037)
Industry FE	✓	✓	✓	✓	✓	✓
Size control		✓		✓		✓
R-squared	0.13	0.13	0.19	0.21	0.17	0.20
Observations	449	449	449	449	449	449
Panel B: BRSR, environmental and social attributes (11 indicators)						
ESG score (std)$_f$	-0.036	-0.039	-0.023	-0.007	-0.054	-0.028
	(0.043)	(0.044)	(0.043)	(0.042)	(0.044)	(0.042)
Log(assets)$_f$		0.019		-0.111***		-0.187***
		(0.023)		(0.024)		(0.026)
Industry FE	✓	✓	✓	✓	✓	✓
Size control		✓		✓		✓
R-squared	0.07	0.07	0.14	0.17	0.08	0.15
Observations	722	722	722	722	722	722

Notes: This table repeats Table 4 using only the environmental and social components of each ESG score, excluding governance. Panel A uses the simple average of the environmental and social pillar scores published by London Stock Exchange Group (2024); LSEG weights its categories by an industry-specific materiality matrix that is not public, so the two pillars are weighted equally here. Panel B uses the six BRSR Core attributes that are not governance — the renewable share of energy, the waste recovery rate, employee wellbeing and safety, gender diversity, inclusive development, and fairness in engaging with customers and suppliers — comprising 11 of the 19 indicators in the baseline score, aggregated exactly as in the baseline: signed so that a higher value is better, converted to percentile ranks among that year's filers, averaged within attribute, and then weighted equally across attributes. The eight indicators of attribute 9, openness of business, which measure concentration of purchases, sales, and related-party transactions, are the governance component and are excluded. All other elements of the construction and the specification are identical to Table 4. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

B Results for Alternative Samples

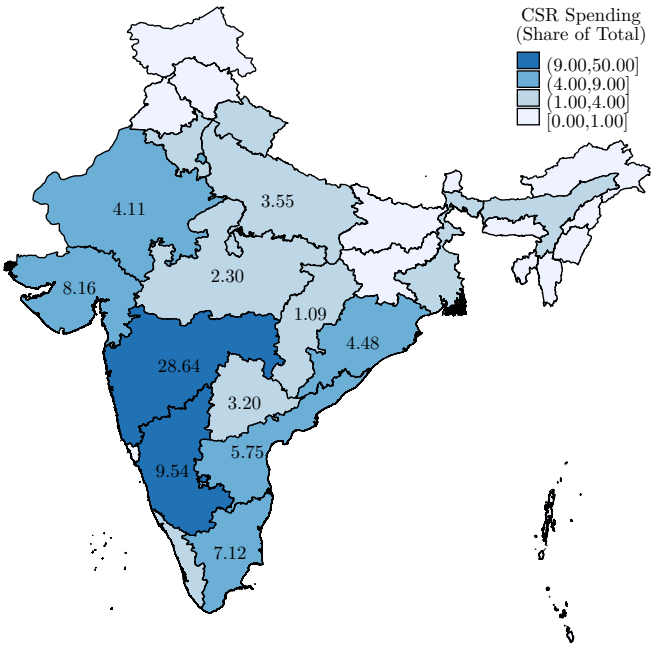
This appendix reproduces our key findings, when feasible, for two alternative samples. First, the sample of all 11,487 firms in the MCA data. By definition, we do not have industry information for firms that are only in the MCA data but are not in the accounting data. Thus, we cannot replicate our results in section 5 for this sample. Second, the voluntary CSR sample of 1,602 firms, defined in section 2.

Table B.1 presents the allocation of CSR expenditures across topics (see Table 1 for results on the main sample). We see that the allocation is very similar in these two samples compared to our main sample. Consequently, both our fact 1 (CSR spending is concentrated on health and education) and fact 2 (firms' allocation across topics correlates with the allocation of other public good providers) hold in all samples. Fact 3 (firms specialize in topics) is also similar in those samples. In the full sample of all firms in the MCA data, we observe an average Herfindahl-Hirschman Index of 0.64, and 38% of firms allocate more than 90% of their spending to only one topic. Subsetting on the firm $\times$ year observations where firms report multiple projects, 25% of firms allocate more than 90% of their spending to only one topic. In the sample of firms that spend voluntarily, we observe an average Herfindahl-Hirschman Index of 0.60, and 29% of firms allocate more than 90% of their spending to only one topic. Subsetting on the firm $\times$ year observations where firms report multiple projects, 19% of firms allocate more than 90% of their spending to only one topic. Finally, Figure B.1 demonstrates that fact 4 (CSR spending is highly concentrated geographically) also looks similar in all three samples.

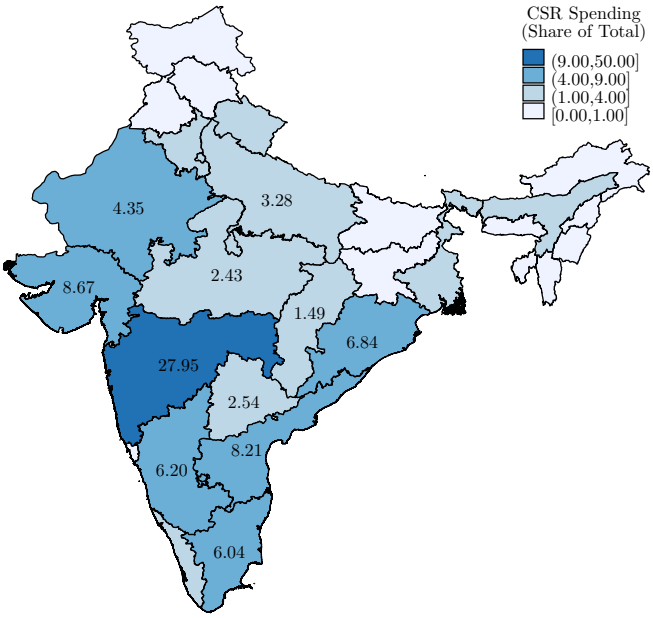
Table B.2 demonstrates that the results on technological proximity are similar if we restrict the sample to firms in the voluntary CSR sample. In terms of the relationship between state-level GDP and CSR spending, Figure B.2 draws a picture for all firms in the MCA data and firms that spend voluntarily that is similar to our main sample. Table B.3 provides similar evidence at the firm-level. To summarize, both the descriptive facts and findings that we demonstrate on our main sample hold for the sample of all firms in the MCA data and firms that spend voluntarily.

Figure B.1: Allocation of Spending Across States (Full MCA Sample and Voluntary CSR Sample)

(a) Full MCA Sample



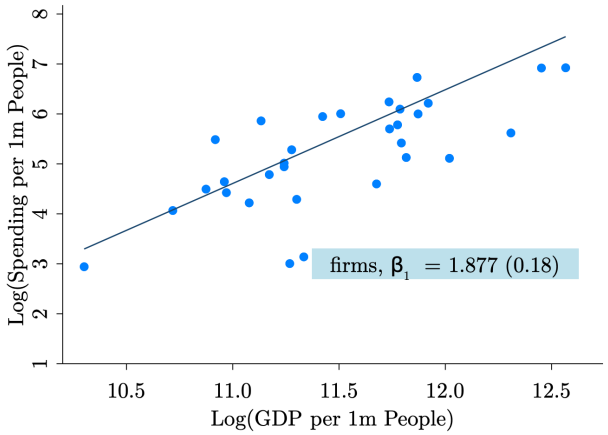
(b) Voluntary CSR Sample



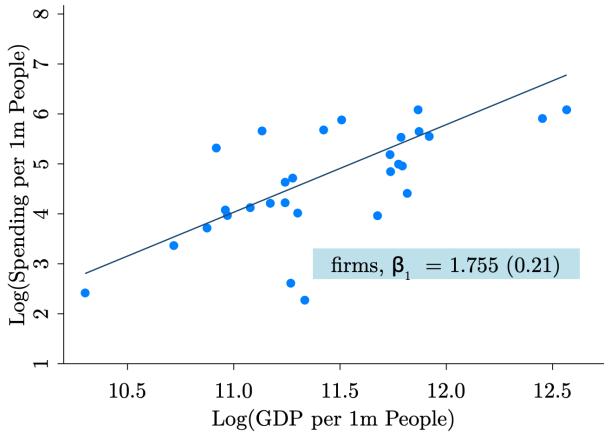
Notes: This figure depicts CSR spending shares by state, for all firms in the MCA data and for the voluntary CSR sample.

Figure B.2: CSR Spending, Government Spending, and State-Level GDP (Full MCA Sample and Voluntary CSR Sample)

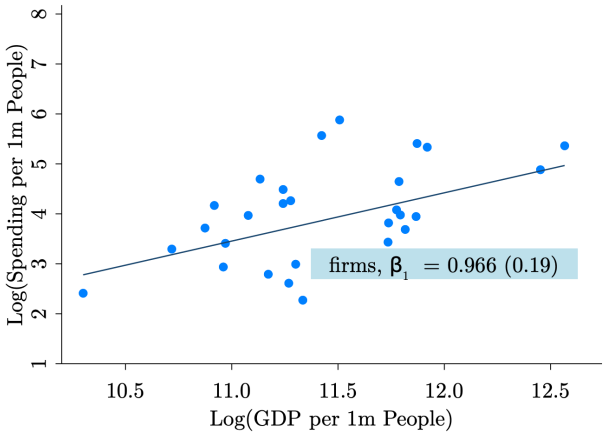
(a) CSR Spending (Full MCA Sample)



(b) CSR Spending (Voluntary CSR)



(c) CSR Outside HQ States (Voluntary CSR)



Notes: This figure presents the relationship between state-level GDP and firm (CSR) spending, for all firms in the MCA data and for the voluntary CSR sample. CSR spending outside of headquarter states is not provided for the full MCA sample since data on headquarters is obtained from Prowess. The scattered dots indicate state-level observations. The lines indicate fitted linear approximations. The unit of observation is at the state level. The dependent variable is the log of spending (in millions, denominated in 2015 INR) per one million people by firms (CSR), aggregated from 2015 to 2019. The independent variable is the log of state-level GDP (in millions, denominated in 2015 INR) per one million people in 2013. Observations are weighted by the 2011 population.

Table B.1: Spending Share by Topic (Full MCA Sample and Voluntary CSR Sample)

Topic	CSR Share Full MCA Sample	CSR Share Voluntary CSR Sample
Education	32%	30%
Health	18%	18%
Rural Infrastructure	8%	8%
Environmental sustainability	8%	8%
Vocational skills	5%	6%
Technology	5%	6%
Livelihood enhancement	5%	4%
Sanitation	4%	5%
Hunger and malnutrition	4%	4%
Safe drinking water	2%	2%
Vulnerable populations	2%	2%
Emergency relief	2%	1%
Sports	2%	2%
Women empowerment	1%	1%
Agroforestry	1%	0%
Animal welfare	0%	0%

Notes: This table displays the share of total CSR spending by topic for all firms in the MCA data and for the voluntary CSR sample.

Table B.2: CSR Spending and Proximity (Voluntary CSR Sample)

	CSR Share Unconditional$_{fd}$ (1)	Any CSR Spending$_{fd}$ (2)	CSR Share Conditional$_{fd}$ (3)
	Panel A: Not Weighted		
Proximity$_{i(f)d}$	0.010***	0.021***	0.018***
	(0.002)	(0.004)	(0.006)
Avg dep var	0.062	0.261	0.239
Firm FE	✓	✓	✓
Topic FE	✓	✓	✓
R-squared	0.27	0.36	0.36
Observations	25,632	25,632	6,360
	Panel B: Weighted by Total CSR Spending		
Proximity$_{i(f)d}$	0.012**	0.047***	0.005
	(0.006)	(0.016)	(0.011)
Avg dep var	0.062	0.506	0.124
Firm FE	✓	✓	✓
Topic FE	✓	✓	✓
R-squared	0.29	0.38	0.34
Observations	25,632	25,632	6,360

Notes: This table describes the relationship between CSR spending and proximity, derived from equation (8), for the voluntary CSR sample. The unit of observation is at the firm×topic level. The dependent variables are the share of CSR spending of a firm (f) over topics (d), an indicator for any CSR spending by firm (f) in a given topic (d), and the share of CSR spending conditional on any spending. Proximity$_{i(f)d}$ is the textual measure of closeness between an industry and a topic defined in section 5.1. In Panel A, observations are unweighted. In Panel B, observations are weighted by the total CSR spending of each firm, winsorized at the 1st and 99th percentile. Standard errors are clustered at the industry×topic level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

Table B.3: CSR Spending, State-Level Characteristics, and Firm Headquarter (Full MCA Sample and Voluntary CSR Sample)

	CSR Share Unconditional$_{fs}$		Any CSR Spending$_{fs}$		CSR Share Conditional$_{fs}$	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Full MCA Sample					
Log(GDP per 1m People)$_s$	0.097***		0.152***		0.137***	
	(0.030)		(0.040)		(0.020)	
Avg dep var	0.029		0.061		0.481	
Firm FE	✓		✓		✓	
R-squared	0.09		0.17		0.41	
Observations	333,240		333,240		13,991	
	Panel B: Voluntary CSR Sample					
Log(GDP per 1m People)$_s$	0.087***	0.011**	0.143***	0.051***	0.102***	-0.026
	(0.024)	(0.005)	(0.031)	(0.010)	(0.020)	(0.017)
1(Firm Headquarter State)$_{fs}$		0.585***		0.709***		0.359***
		(0.034)		(0.027)		(0.024)
Avg dep var	0.030	0.030	0.077	0.077	0.386	0.386
Firm FE	✓	✓	✓	✓	✓	✓
R-squared	0.07	0.49	0.20	0.43	0.41	0.57
Observations	47,947	47,947	47,947	47,947	2,890	2,890

Notes: This table describes the relationship between state-level characteristics as well as firm headquarters and CSR spending, derived from equation (9), for all firms in the MCA data and firms in the voluntary sample. Headquarter information is not available for all firms in the MCA data since it is obtained from Prowess. The unit of observation is at the firm×state level. The dependent variables are the share of CSR spending of a firm (f) over states (s), an indicator for any CSR spending by firm (f) in a given state (s), and the share of CSR spending conditional on any spending. The independent variables are the log of state-level GDP per 1 million people and an indicator that equals one if the firm is headquartered in the state as per government records. We control for the log of population in millions. Observations are weighted by the 2011 population. Standard errors are clustered at the state-level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

C Data and Variable Construction

C.1 CSR Data

C.1.1 Sample Overview

The raw data contains 176,988 projects from 15,632 firms. To derive the main sample for our analysis, we take the following steps. First, we exclude projects where the text indicates an issue for classification across topics (see flags detailed in section C.1.2). This leaves 124,813 projects by 11,487 firms. Second, we match 61% of these firms and 91% of the CSR expenditure in the CSR data to the accounting data, resulting in 92,180 projects of 7,064 firms.[36] Third, we apply an additional set of cleaning steps, excluding: holding companies, firms without industry information, firms with zero CSR spending, and projects that cannot be assigned to a topic. This results in our final dataset of 86,334 projects by 6,573 firms. In addition, when we consider allocation across locations, we exclude four states with a population that is lower than one million, as well as the small state of Chandigarh, which does not have government spending data, and firms without headquarter locations. For our analysis at the firm×topic and firm×state levels, we fill the data, meaning that even if a firm does not spend on a topic or in a state, we include an observation with zero spending.

C.1.2 Cleaning of Project Descriptions Data

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

1. Convert the text to lowercase, removing special characters and numbers
2. Tokenize the text (splitting strings into tokens), lemmatize, and stem the tokens
3. Translate Hindi tokens to English
4. Remove uninformative tokens:
    - Create a list of common uninformative token sequences: it includes common token sequences found in the project descriptions but unrelated to CSR projects, e.g., "CSR overheads", "project not found", "administration expenditures", "detailed in report", etc.
    - Remove tokens flagged as uninformative

Flags for uninformative descriptions. In addition, we flag observations with uninformative descriptions for the purpose of our textual analysis, defined as satisfying at least one of the following criteria:

1. No project description

[36] The accounting data is obtained from https://prowessdx.cmie.com.

2. "Word salads": project description lists the titles of many different topics
3. Project description where more than 60% of original bigrams correspond to uninformative token sequences defined above
4. Project description Word2Vec embedding is empty (e.g., if the only remaining token following the cleaning is a proper noun)
5. Project description equal (or highly similar) to the title of the topic or groupings of topics on the CSR portal: this avoids using project descriptions that are just a repetition of the title of the topic.

The raw data contains 176,988 projects. This procedure flags 68,992 observations as having uninformative descriptions. Observations flagged in 5 are not used in our textual corpus, but can be reliably classified across topics and hence enter our firm×topic CSR spending data.

C.1.3 Vectorization of the Textual Data

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

This method involves encoding words as numerical vectors, known as word embeddings, which capture semantic meaning based on the context in which they appear. There are two primary ways to generate these embeddings: (1) training them on a custom corpus containing domain-specific documents or (2) utilizing pre-trained embeddings derived from a large, general-purpose corpus that includes the words of interest. We opt for the latter approach because (a) we lacked access to a sufficiently large and diverse collection of documents to train reliable embeddings, and (b) existing research supports the effectiveness of pre-trained word embeddings (Rios and Lwowski, 2020). Specifically, we use the 300-dimensional embeddings from the Word2vec model of Mikolov (2013), which were trained on Google News data encompassing approximately 3 million words.[37]

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

$$\mathbf{v}_p = \frac{1}{N_p} \sum_{j=1}^{N_p} w_{jp} * \mathbf{v}_{jp} \tag{C.1}$$

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

The weights w_{jp} are determined using the Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF assigns importance to each word based on how frequently it

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

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

Unless specified, $\mathbf{v}_p$ is normalized to have norm 1.[38]

C.1.4 Classification Across Topics

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

Step 1: Manually classify the 28 original topics into 19 aggregated topics. We reclassify all topics with less than 2,000 observations unless there is no sufficiently close topic. This mapping is detailed in Table C.1, which also shows the number of observations by initial topic. Figure C.1 reports the average pairwise similarity between the initial topics.

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

1. Compute average embedding of projects $\mathbf{v}_{pd(p)}$ (not normalized) in topic d:

$$\mathbf{v}_d = \frac{\sum_{p=1}^{N_d} \mathbf{v}_{pd(p)}}{N_d} \tag{C.2}$$

2. For each project p in topic d, compute:
   - The similarity between p and its own topic $d(p)$: $\text{OwnProximity}_p = \cos(\mathbf{v}_p, \mathbf{v}_{d(p)})$

[38] For a robustness test, text embeddings were obtained from the OpenAI API using the text-embedding-3-small model (released 25 January 2024; OpenAI does not version embedding models by snapshot date). The API was accessed in April 2024 (project-level embeddings on 9 April 2024, industry-level embeddings and proximity scores on 19–24 April 2024) via the OpenAI Python client, version 1.3.8, under Python 3.10.

- The largest similarity between p and any topic d': $\text{Max1Proximity}_p = \max_{d'} \cos(\mathbf{v}_p, \mathbf{v}_{d'})$
- The topic d' with the maximum similarity: $\text{IsMax}_p = \arg\max_{d'} \cos(\mathbf{v}_p, \mathbf{v}_{d'})$
- The second-largest similarity between p and any topic d': $\text{Max2Proximity}_p = \max_{d' \neq \text{IsMax}_p} \cos(\mathbf{v}_p, \mathbf{v}_{d'})$

3. If $\text{Max1Proximity}_p > \lambda^{miscl} \times \text{OwnProximity}_p$, we say observation p is misclassified
   - If $\text{Max1Proximity}_p > \lambda^{recl} \times \text{Max2Proximity}_p$, we reclassify observation p to topic IsMax_p
   - Otherwise, we say observation p cannot be classified
4. Discard observations that cannot be classified
5. Repeat until the topic assigned in iteration n is the same as the topic assigned in iteration $n + 1$.

In our implementation, we use $\lambda^{miscl} = \lambda^{recl} = 1.2$. We impose a number of additional rules:

1. Observations in agroforestry contain projects related to farming as well as projects related to environmental sustainability. Because of the small number of observations in agroforestry, the algorithm tends to reclassify environmental projects in agroforestry. We manually assign the projects containing the tokens ['tree plantation', 'protection flora fauna', 'maintenance flora fauna', 'biodiversity protection', 'environmental sustainability'] to environmental sustainability.
2. Rural development contains a mix of infrastructure projects and projects corresponding to the other topics but implemented in rural areas. We reclassify observations in this topic using $\lambda^{miscl} = \lambda^{recl} = 1.05$ so that observations related to other topics are reclassified "aggressively", and observations remaining in rural development mostly consist of rural infrastructure projects.
3. 'Slum area development' and 'Other central government funds' consist of highly heterogeneous projects. We force the reclassification of projects in these topics into the closest topic.

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

Looking at the word clouds for each topic, the procedure performs well. A caveat of our methodology is that all the projects with tokens in Hindi that we could not translate (proper nouns, spelling mistakes) get bundled in the topic "Art and culture" which has the initial largest share of such tokens. In the absence of a solution to this caveat, we remove this topic from our baseline analysis. Since this topic is small, this does not materially affect our results.

C.1.5 Definition of Third-Party Implementation

To define third-party implementation, we initially rely on the original MCA variable. This variable is self-reported by the firm and likely contains measurement error. In particular, the mandate does not detail what extent of outsourcing implies that firms should declare their CSR projects as implemented via a third-party. We find, for example, that the probability of indirect implementation is only weakly correlated with the likelihood that the project description contains an NGO name. We consider a version of the variable where we relabel projects as indirectly implemented if the project description contains text possibly referring to indirect implementation or an NGO name as a robustness check.

- Check for tokens and bigrams indicative of third-party implementation: ['gift', 'offering', 'grant', 'endowment', 'input', 'participation', 'support', 'contribution', 'donation', 'partnership', 'ngo', 'benefaction', 'collaboration', 'alliance', 'association', 'joint venture', 'cooperation', 'affiliation', 'organization', 'agency', 'implemented by', 'bureau', 'department', 'authority', 'office', 'establishment', 'advisory', 'counsel', 'guidance', 'expert advice', 'consulting', 'foundation', 'in collaboration', 'professional services', 'fund', 'executed by', 'carried out by', 'enacted by', 'put into effect by', 'non-profit organization', 'operated through', 'voluntary organization', 'nonprofit group', 'together', 'jointly', 'in partnership', 'working together', 'donations to']
- Check for mention of NGOs using a list of the 23 NGOs most frequently present in our dataset

C.1.6 Description of the Final Dataset

Figures C.2 to C.4 show word clouds of the project descriptions by topic.

C.2 Other Data Sources

C.2.1 Government Expenditures

We obtain government expenditures data from the Reserve Bank of India for 2010–2021, retrieved on 8th of December 2022.[39] We use state-level expenditures, which include expenditures from central government transfers to states and cover most expenditures in India that correspond to the topics in the CSR data, except some food subsidies implemented directly by the central government. We use the 'revised' expenditure variables and consider all regular and capital expenditures labeled 'development expenditures'. We exclude from our analysis expenditures that have no equivalent in the CSR data: energy and transport, tax

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

collection expenditures, interest payments, 'organs of state' (this includes police and judiciary), grants to lower levels of governments (these represent 1% of state expenditures), and other (which includes items like tourism expenditures). The mapping between government expenditures and CSR topics is described in Table C.3.

C.2.2 NGO Activity

To compare CSR spending with NGO activity by topic, we exploit a report prepared by the state of Haryana that collects data on the 150 highest-capability NGOs operating in the state and classifies them by sustainable development goal (which we map to our topics).[40] The mapping between sustainable development goals and CSR topics is described in Table C.3.

C.2.3 Indian Industry Descriptions

The industry descriptions come from the National Industry Classification report from 2008. We extract the descriptions from its 'Detailed Structure' and 'Explanatory Notes' sections. We manually clean the text for typos and mentions of products/services that should be excluded from each industry. Table C.4 shows an example of the full text for one industry i. We obtain the word embeddings associated with the description of each industry i. We clean the text as described in steps 1–4 of section C.1.2. We obtain word embeddings using the same Word2vec model, again applying TF-IDF. For each industry i, we obtain an embeddings vector $\mathbf{v}_i$.

C.2.4 US 10-K filings

We construct an alternative version of the proximity metric where the industry corpus uses firm-level descriptions of firm activity using the US SEC 10-K filings. This is the same data as the one used by Hoberg and Phillips (2016) to define product similarity between pairs of firms.

First, we construct a crosswalk from US SIC codes to Indian NIC 2008 2-digit industry codes by combining published concordances from the Forum for Research in Empirical International Trade with manual adjustments for ambiguous cases.

Second, we merge the SEC 10-K data with Compustat (by Central Index Key) to obtain firms' SIC codes. We use our mapping to obtain firm's industries as per the NIC 2-digit classification. We extract only section 1 ("Business") of each 10-K filing. The text is sometimes very long, and the end of the business section often contains text not related to the business description (often it contains sections on the competition, regulation, etc.). To keep the text

[40] https://sdgcc.in/wp-content/uploads/2020/10/SDG-NGO-LINKAGES.pdf

most closely related to the firm's activity, we retain the first half of the text. We then apply the same text cleaning pipeline as for the Indian NIC textual descriptions.

When constructing the proximity metric, we want the text corpus to be balanced across industries. This ensures that the industry embeddings are as representative for each industry, and avoids TF-IDF from discounting words related to specific, overrepresented industries. Hence, we select up to 400 firm×year observations per NIC industry using random subsampling or oversampling for balance.[41]

C.2.5 Firm Establishments

To proxy for firms' establishments beyond their headquarters, we utilize CMIE's CapEx database. This database tracks firms' investment projects in India — including new plants, capacity expansions, and major upgrades — with project-level detail on location, recording the state and, where available, the district in which each project is undertaken. The database covers projects with investment above approximately INR 10 million (approx. 160,000 USD) based on media reports, company announcements, and filings. We match firms across the CapEx and Prowess databases using company identifiers. For each matched firm, we code an indicator equal to one for any state in which the firm has undertaken at least one capital investment project since 1995. Firms with no record in the CapEx database are coded as having no establishment presence outside their headquarters state, following the treatment of the CapEx database in the literature (e.g., Alok and Ayyagari (2020)).

C.2.6 ESG Scores

Sources and extraction. The Refinitiv scores are obtained through the LSEG Data Library for Python, querying the fields (TR.TRESGScore) (the overall ESG score), (TR.EnvironmentPillarScore), (TR.SocialPillarScore) and (TR.GovernancePillarScore) (the three pillar scores). The universe is the set of firms in our regression sample, queried at annual frequency.

The BRSR filings are obtained from the National Stock Exchange of India. The index of filings is retrieved from the NSE corporate-filings endpoint nseindia.com/companies-listing/corporate-filings-bussiness-sustainabilitiy-reports, which returns, for each filing, the NSE symbol, company name, fiscal year, and a link to the XBRL instance document. We then download each XBRL file and parse it into a long table of indicators and values. The taxonomy governing these filings, used to match indicators to the BRSR Core indicators, is published by NSE alongside the filings.

[41] We keep all observations if there are fewer than or equal to 400 per industry; if there are more, we use proportional subsampling within firms or, when there are more than 400 unique firms, randomly select one observation per firm (prioritizing those with non-missing employment and revenue data) until reaching the target; for industries with fewer than 400 observations, we use random oversampling with replacement.

The indicator set. SEBI's BRSR Core (Securities and Exchange Board of India, 2023) groups nine ESG attributes and specifies, for each, the indicators a firm must report and the unit in which to report them. The circular lists 49 such indicators. We match each to an element of the XBRL taxonomy that governs the filings, using the cross-reference the circular gives to a question in the reporting form together with the element's label. Two indicators cannot be constructed: a count of permanent disabilities, for which the reporting format has no field, and the number of days of accounts payable, which no firm in our sample reports. Where the circular asks for a rate whose components are filed separately — the renewable share of energy, the waste recovery rate, and the share of harassment complaints upheld — we form the rate from its components. This leaves 44 indicators.

We use 19 of the 44 indicators, following two restrictions. First, we exclude indicators that are absolute quantities — the tonnes of plastic, e-waste, battery and other waste a firm generates, the volume of water it discharges, the energy it consumes, and counts of fatalities and of counterparties. A percentile rank on such a quantity orders firms by the scale of their operations: a firm that doubles its output and doubles its waste has not become less responsible. Ranking carries this quality, because a percentile rank is a monotone transformation and so preserves exactly the ordering it is given.

Second, among the remaining indicators we keep those that are free of units, being either already a percentage or a defined rate, or a ratio of two quantities reported in the same unit, so that the unit cancels. This excludes the remaining eight indicators, which the circular measures as intensities — per rupee of turnover, or per unit of physical output. They span four attributes. Energy and circularity each retain a unit-free indicator, the renewable share and the recovery rate, so the attribute survives without its intensity; greenhouse gases and water have none, so those two attributes are absent from the baseline altogether. The difficulty is not with the physical quantities themselves but with the denominator. Turnover is filed on inconsistent monetary scales. Firms appear to report in rupees, lakhs and crores interchangeably.

The 19 indicators that remain span seven of the nine attributes. *Energy footprint* contributes the renewable share of energy consumption; *circularity* the waste recovery rate; *employee wellbeing* spending on wellbeing measures as a share of revenue and the lost-time injury frequency rate; *gender diversity* the female share of gross wages, complaints under the prevention of sexual harassment framework as a share of female employees, and the share of those complaints upheld; *inclusive development* the share of inputs sourced from small producers, the share sourced from within the district, and the share of the wage bill paid in smaller towns; *fairness* the share of data breaches involving customer information; and *openness of business* eight measures of concentration in purchases, sales and related-party transactions.

Aggregation. We sign each indicator so that a higher value is better, and convert it to a percentile rank among the firms reporting that indicator in that year. We average the ranks within each attribute and then weight the seven attributes equally, rescaling to run from 0 to 100. We then average each firm's score over the years in which it files.

Relation to the commercial rating. The construction follows London Stock Exchange Group (2024) in three respects: each indicator carries a polarity so that a higher value is always better, indicators are converted to percentile ranks, and scores are aggregated in stages, from indicators to categories and from categories to an overall score. It differs in three. LSEG benchmarks its environmental and social categories within TRBC industry group and its governance categories within country of incorporation, whereas we rank firms against all filers in a given year; since every firm in our sample is Indian, our convention coincides with theirs for governance but not for the environmental and social categories. LSEG weights categories using a proprietary materiality matrix, whose weights vary by industry for the environmental and social categories and are constant across industries for governance, whereas we weight attributes equally. And LSEG selects 186 of the more than 870 measures it collects on grounds of comparability and materiality, whereas our indicator set is the regulator's.

Robustness. Tables A.12 and A.13 examine the robustness of the results in Table 4. Table A.12 reports the results under four alternative constructions of the BRSR-based score: adding the turnover-scaled intensities with turnover taken from Prowess rather than as filed; adding them as filed; ranking firms within industry and year rather than within year alone; and including all 44 indicators. In the first three, all coefficients remain insignificant. The last is instructive in its own right. Of the 25 indicators it adds, 17 are absolute quantities, such as the tonnes of waste a firm generates, and eight are further intensities.[42] Ranking firms on absolute quantities orders them by the scale of their operations and not by their conduct: a firm that doubles its output and its waste has not become less responsible. Because less counts as better, a high score on these indicators marks a firm with a small physical footprint. Accordingly, the score built from all 44 indicators correlates –0.38 with log assets, against 0.01 for the baseline score. Without the size control, it is significantly related to all three allocation measures. With the size control, the associations with the tilt toward richer states and with home bias disappear. The association with technological advantage remains negative and significant at the 10% level (–0.101); it falls to –0.082 and is no longer significant once we also control for the share of fixed assets in total assets (column

[42]The eight intensities are the four used in Panels A and B — the greenhouse-gas, water, energy and waste footprints per rupee of turnover, one for each footprint attribute — and footprints scaled instead by a unit of physical output. That denominator is chosen by each firm and not recorded in the filings, so those four are not comparable across firms and enter only in Panel D.

3). The absolute quantities thus largely proxy for the scale of a firm's operations rather than for its conduct, which is why the baseline score excludes them.

Table A.13 repeats the analysis without the governance dimension, which is unlikely to relate to the allocation we study. For Refinitiv, we use the average of the environmental and social pillar scores. For the BRSR-based score, we drop the attribute on openness of business, which leaves 11 of the 19 indicators. The coefficients on technological advantage and on the tilt toward richer states remain insignificant for both scores, with and without the size control, except that the Refinitiv coefficient on the tilt toward richer states is negative and significant at the 10% level without the size control. The one association that holds with the size control is with home bias: firms with higher Refinitiv scores concentrate less spending in their headquarter state (–0.136 standard deviations, significant at the 5% level). These firms do not spend more in poorer states, however, and the BRSR-based score shows no such association.

C.3 Construction of CSR Project Clusters

For descriptive purposes, we implement k-means clustering to partition CSR projects into thematically coherent groups within each development topic. The k-means algorithm iteratively assigns observations to clusters by minimizing the within cluster sum of distances to centroids, then updates the centroids based on cluster membership until convergence (MacQueen, 1967; Lloyd, 1982). We employ the cosine distance as our distance measure: $d(p, p') = 1 - \cos(\mathbf{v_p}, \mathbf{v_{p'}})$.

For each of the 16 development topics, we perform cosine k-means clustering in a systematic grid search with $k \in \{3, \ldots, 7\}$ clusters, selecting the optimal number of clusters based on the highest silhouette score. The silhouette score of each project p assigned to cluster C_p is defined as follows:

$$s(p) = \frac{b(p) - a(p)}{\max\{a(p), b(p)\}}$$

where: $a(p) = \frac{1}{|C_p|-1} \sum_{p' \in C_p, p' \neq p} d(p, p')$ is the average distance between point p and all other points in the same cluster C_p; $b(p) = \min_{C \neq C_p} \frac{1}{|C|} \sum_{p' \in C} d(p, p')$ is the minimum average squared distance between point p and all points in another cluster $C \neq C_p$. The overall silhouette score for the clustering is then computed as the average over all points: $S = \frac{1}{N} \sum_{i=p}^{N} s(p)$. The silhouette score measures both cluster cohesion (how close observations are to their own cluster centroid) and separation (how distant they are from neighboring clusters), providing a robust metric for cluster quality that ranges from -1 to 1, with higher values indicating better-defined clusters (Rousseeuw, 1987). This optimization process ensures that each social topic achieves its most natural thematic subdivision.

Table C.2 summarizes the clustering results across development topics, where each row corresponds to a cluster within a specific social topic. The first column lists the social topic.

The second column provides a cluster name. Column 3 lists the number of observations in each cluster. Column 4 lists the silhouette score of each cluster. Finally, columns 5 to 9 report the top five bigrams in project descriptions. The cluster name is based on manual inspection of the top five bigrams and trigrams and the project descriptions closest to the cluster centroid.

C.4 Construction of the Proximity Variable

This procedure requires that the CSR project description is sufficiently informative and that we could reliably classify the observation in a CSR topic. We therefore restrict the sample to observations not flagged as uninformative (see section C.1.2) and with a valid topic assignment (see section C.1.4). We thus work with 92,979 observations.

Define $\mathbf{v}_d$ as the average embedding of projects $\mathbf{v}_{pd(p)}$ in topic d:

$$\mathbf{v}_d = \frac{1}{N_d} \sum_{p \in \mathcal{P}_d} \mathbf{v}_{pd(p)} \tag{C.3}$$

Define

$$\text{Proximity}_{id} = \cos(\mathbf{v}_d, \mathbf{v}_i) \tag{C.4}$$

We also construct Proximity_{id} by taking the median or the mean across projects weighted by token count. The correlation coefficient between these different variants exceeds 0.98.

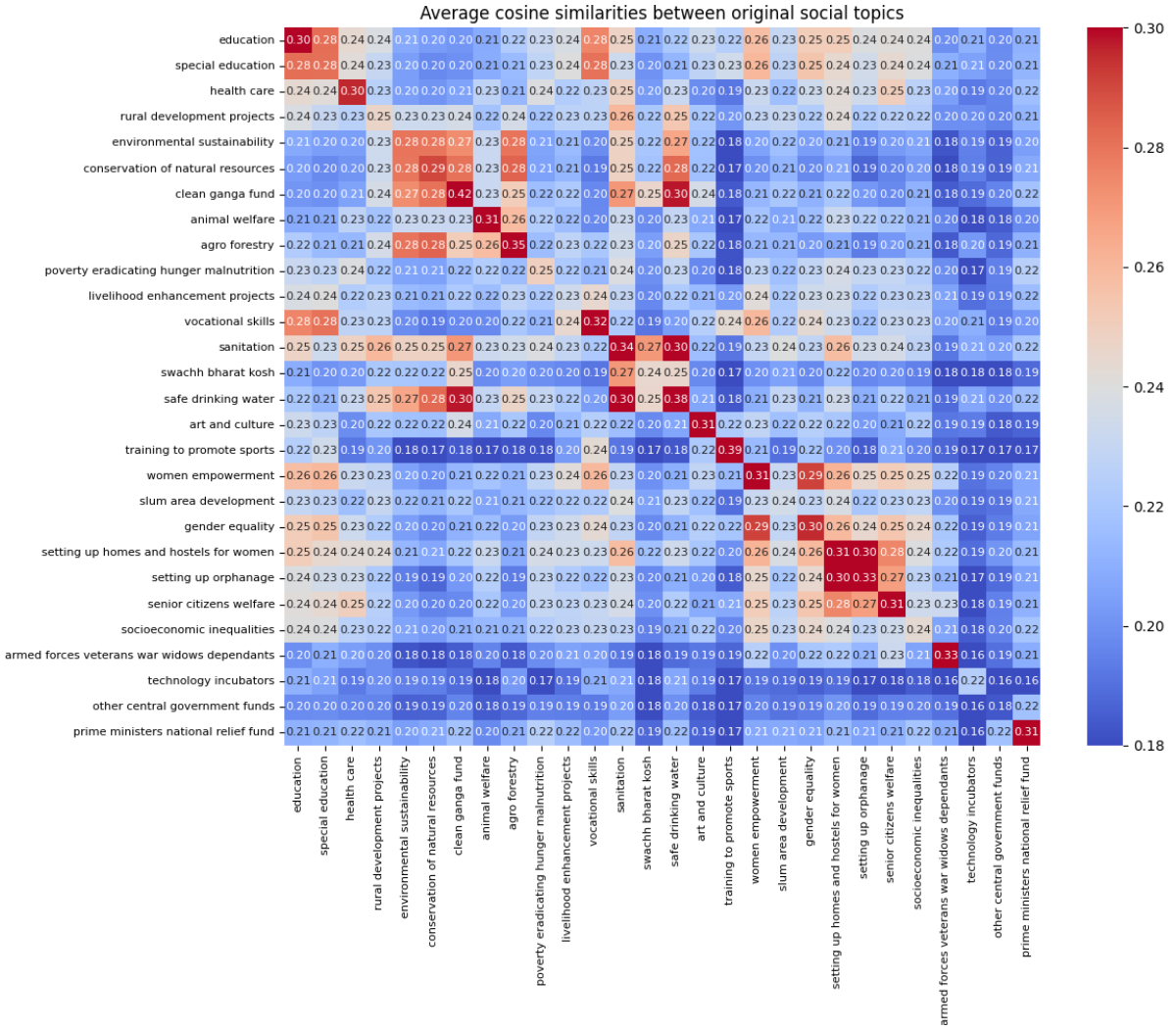
Analysis of the closest (furthest) CSR topics and industries. To interpret the semantic proximity between development topics and industries, we conduct an analysis of the token-level drivers underlying the proximity scores.

For each industry i, we identify the development topic d with the highest and lowest proximity, as defined in equation (7). Similarly, for each topic d, we identify the industry i with the highest and lowest proximity.

To understand the textual basis for these relationships, we extract the most representative tokens for each industry and topic. For each industry i and topic d, we select the 20 tokens in the text that are closest to the group centroid, and call these 'core tokens'. For each industry×topic pair, we compute the cosine similarity between all possible pairs of core tokens, and report the 10 core token pairs with the highest similarity.

Table C.5 presents, for each industry i among the top 20 industries, the closest social topic d, and the closest token pairs associated to this pair (i, d) as defined above. Table C.6 repeats this exercise, but shows the furthest social topic d for each industry i. Tables C.7 and C.8 present the same results starting from social topics d and selecting the closest and furthest industry i, respectively.

Figure C.1: Average Cosine Similarity Between Initial Topics



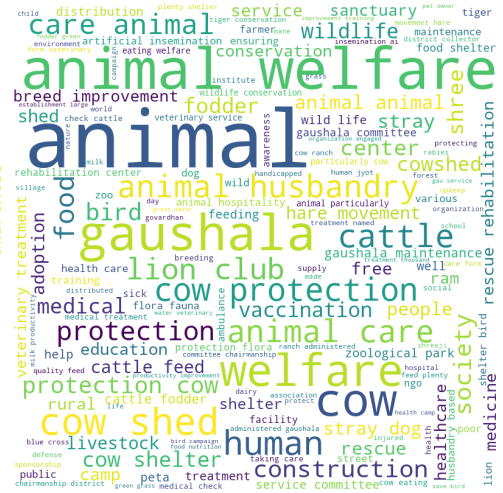
Notes: This figure depicts the average cosine similarity of projects for each pair of topics. The diagonal elements report the average cosine similarity of projects within a topic.

Figure C.2: Project Descriptions by Topics: Word Clouds

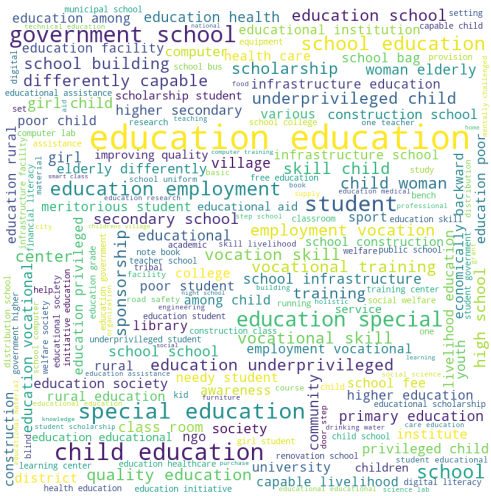
(a) Agroforestry



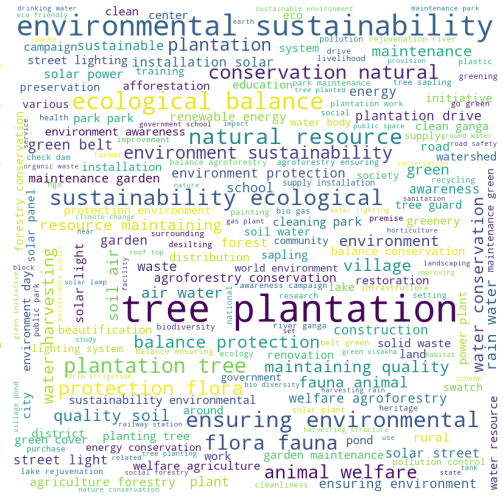
(b) Animal Welfare



(c) Education



(d) Environmental Sustainability



(e) Health



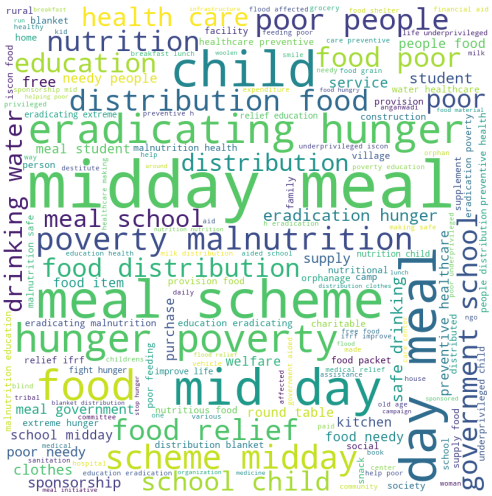
(f) Livelihood Enhancement



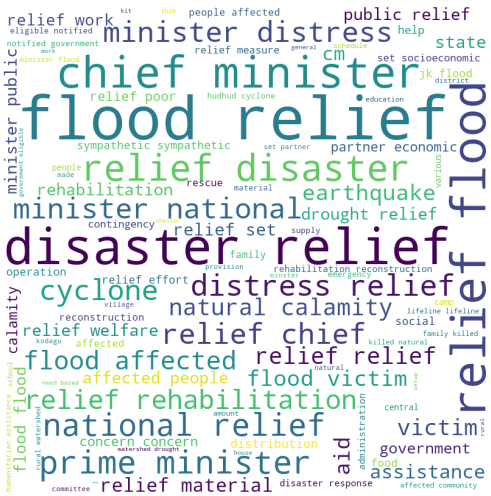
Notes: This figure shows word clouds of CSR project descriptions by topic.

Figure C.3: Project Descriptions by Topics: Word Clouds

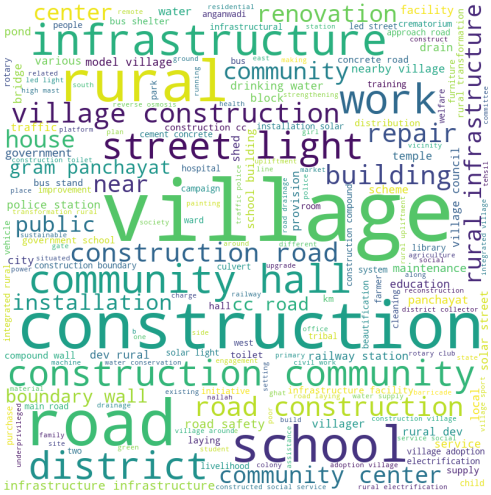
(a) Hunger and Malnutrition



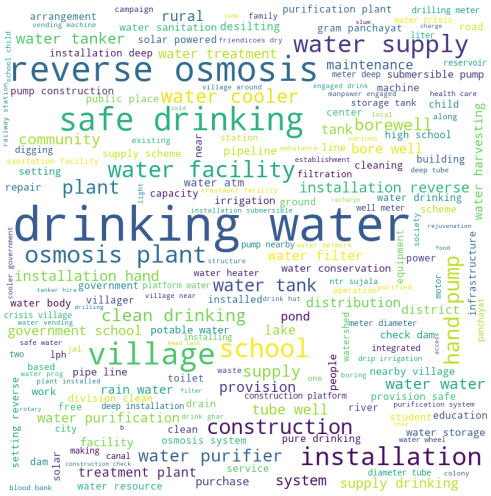
(b) Emergency Relief



(c) Rural Infrastructure



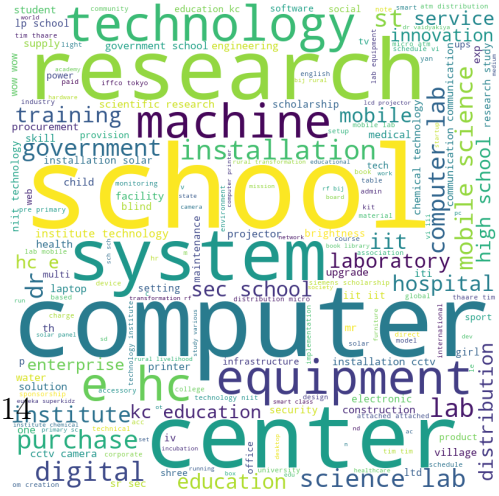
(d) Safe Drinking Water



(e) Sanitation



(f) Technology



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

Figure C.4: Project Descriptions by Topics: Word Clouds

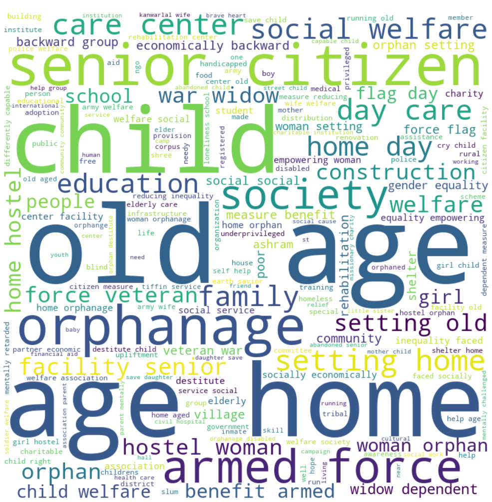
(a) Sports



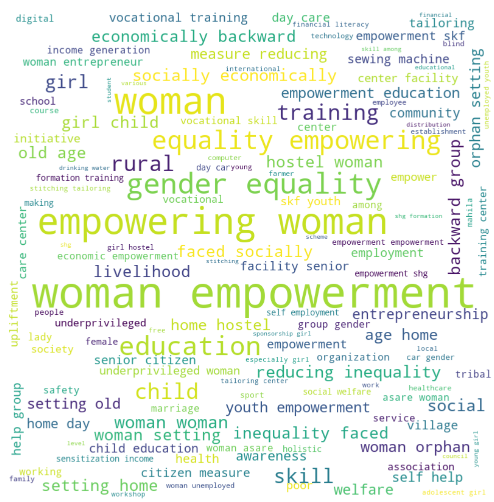
(b) Vocational Skills



(c) Vulnerable Populations



(d) Women Empowerment



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

Table C.1: Definition of Topics

Initial	Obs #	Final
Education	36,962	Education
Special education	1,761	Education
Health care	20,787	Health
Rural development projects	7,785	Rural Infrastructure
Environmental sustainability	5,878	Environmental sustainability
Conservation of natural resources	728	Environmental sustainability
Clean Ganga Fund	47	Environmental sustainability
Animal welfare	1,334	Animal welfare
Agro forestry	175	Agroforestry
Poverty eradicating hunger malnutrition	4,716	Hunger and malnutrition
Livelihood enhancement projects	3,772	Livelihood enhancement
Vocational skills	3,189	Vocational skills
Sanitation	2,881	Sanitation
Swachh bharat kosh	701	Sanitation
Safe drinking water	2,000	Safe drinking water
Art and culture	1,903	Removed
Training to promote sports	2,076	Sports
Women empowerment	2,176	Women empowerment
Slum area development	304	Reallocated to closest topic
Gender equality	461	Vulnerable populations
Setting up homes and hostels for women	472	Vulnerable populations
Armed forces veterans war widows dependents	465	Vulnerable populations
Senior citizens welfare	911	Vulnerable populations
Setting up orphanage	472	Vulnerable populations
Socioeconomic inequalities	1,653	Vulnerable populations
Technology incubators	221	Technology
Other central government funds	1,421	Reallocated to closest topic
Prime Ministers National Relief Fund	631	Emergency relief

Notes: This table reports the mapping between initial topics and the 16 topics in the main sample. It also records the number of observations by initial topic.

Table C.2: Optimal K-Means Cosine Clusters

Topic	Cluster Theme	N of Projects	Silhouette Score	Bigram #1	Bigram #2	Bigram #3	Bigram #4	Bigram #5
Education	Special education	11698	0.091	child_education	special_education	education_special	education_employment	education_education
Education	Scholarships & support for meritorious students	8102	-0.069	government_school	student_education	meritorious_student	school_education	scholarship_student
Education	School construction & infrastructure	4423	0.075	government_school	school_building	school_infrastructure	school_construction	high_school
Health	Preventive health	5454	0.215	health_care	preventive_health	care_health	care_preventive	preventive_healthcare
Health	Medical treatment and patient care	4802	-0.113	cancer_patient	medical_treatment	health_care	palliative_care	poor_patient
Health	Medical infrastructure and equipment	3171	-0.026	medical_aid	hospital_medical	medical_equipment	medical_facility	medical_camp
Health	Mobile health camps	1364	0.104	health_camp	medical_camp	camp_health	health_checkup	blood_camp
Rural Infrastructure	Construction (roads, walls, street light)	1904	0.016	road_construction	street_light	construction_road	boundary_wall	village_construction
Rural Infrastructure	Construction (community centers)	1443	0.052	community_hall	construction_community	gram_panchayat	community_center	village_construction
Rural Infrastructure	Rural transformation	966	0.057	rural_infrastructure	rural_dev	dev_rural	rural_transformation	infrastructure_rural
Environmental sustainability	Ecology/conservation projects	1914	0.185	environmental_sustainability	ecological_balance	sustainability_ecological	ensuring_environmental	natural_resource
Environmental sustainability	Tree plantation	1711	0.112	tree_plantation	plantation_tree	plantation_drive	maintenance_garden	planting_tree
Environmental sustainability	Green energy	869	0.038	solar_street	installation_solar	street_light	green_belt	solar_power
Environmental sustainability	Water harvesting & conservation	730	0.013	water_harvesting	rain_water	water_conservation	clean_ganga	harvesting_rain
Vocational skills	Vocational training	1761	0.117	vocational_training	vocational_skill	employment_vocational	skill_vocational	skill_training
Vocational skills	Skill acquisition	798	0.035	skill_training	education_skill	skill_center	skill_education	skill_skill
Vocational skills	Staff training	385	-0.093	driver_training	training_center	staff_salary	skill_training	salary_staff
Technology	Education technology	1443	-0.019	e_hc	kc_education	sec_school	hc_e	high_school
Technology	Mobile science labs & support to labs	609	0.135	mobile_science	science_lab	product_initiative	initiative_marketing	marketing_product
Technology	Computers, cameras and other equipment donations	596	-0.023	installation_cctv	cctv_camera	computer_lab	eureka_superkidz	installation_solar
Livelihood enhancement	Sustainable livelihood and education	1264	0.105	sustainable_livelihood	education_livelihood	livelihood_education	livelihood_livelihood	social_welfare
Livelihood enhancement	Disability inclusion	414	0.019	differently_capable	capable_livelihood	capable_child	person_disability	mentally_retarded
Livelihood enhancement	Economic development	405	-0.018	world_vision	financial_inclusion	quality_life	social_economic	partner_economic
Livelihood enhancement	Donations of vehicles	156	-0.054	tractor_traveller	selected_organization	product_selected	vehicle_product	traveller_vehicle
Sanitation	Sanitation infrastructure	1666	-0.042	sanitation_facility	drinking_water	school_sanitation	sanitation_hygiene	sanitary_napkin
Sanitation	Toilet construction	1454	0.237	construction_toilet	toilet_construction	toilet_block	school_construction	toilet_school
Sanitation	Cleanliness campaigns	254	0.150	swachh_campaign	campaign_swachh	swach_campaign	cleanliness_campaign	construction_toilet
Hunger and malnutrition	Eradicating hunger	957	0.073	eradicating_hunger	hunger_poverty	poverty_malnutrition	health_care	eradication_hunger
Hunger and malnutrition	Midday meal scheme	890	0.284	midday_meal	meal_scheme	day_meal	mid_day	scheme_midday
Hunger and malnutrition	Food distribution	858	0.049	distribution_food	food_poor	food_distribution	poor_people	food_relief
Hunger and malnutrition	Money gifts	182	-0.081	round_table	joy_giving	giving_week	table_round	specified_money
Safe drinking water	Water supply infrastructure (tanks, pumps, wells...)	1287	-0.136	drinking_water	hand_pump	water_supply	water_tank	tube_well
Safe drinking water	Water safety programs	939	0.212	drinking_water	safe_drinking	water_facility	water_safe	water_supply
Safe drinking water	Water purification (notably reverse osmosis equ...	644	0.053	reverse_osmosis	osmosis_plant	drinking_water	osmosis_water	water_purifier
Vulnerable populations	Hostels for old age, widows & orphans	2164	0.005	old_age	age_home	setting_home	home_hostel	hostel_woman
Vulnerable populations	Social welfare programs	748	0.159	social_welfare	welfare_social	social_service	child_welfare	backward_group
Vulnerable populations	Veterans support	644	-0.039	armed_force	war_widow	force_veteran	benefit_armed	flag_day
Vulnerable populations	Senior citizen care	321	0.109	senior_citizen	facility_senior	old_age	care_center	age_home
Emergency relief	Flood relief	618	0.338	flood_relief	relief_flood	flood_affected	flood_victim	relief_rehabilitation
Emergency relief	Disaster relief	511	0.077	disaster_relief	relief_disaster	natural_calamity	relief_rehabilitation	relief_relief
Emergency relief	Contribution to Prime Minister Relief Fund	316	0.397	chief_minister	relief_chief	distress_relief	minister_distress	prime_minister
Emergency relief	Cyclone relief	234	-0.081	cyclone_relief	cm_relief	relief_cyclone	disaster_relief	relief_material
Emergency relief	Contribution to eligible projects notified by g...	98	-0.123	notified_government	eligible_notified	government_eligible	national_relief	prime_minister
Sports	Sport competition organization	897	-0.159	sport_game	game_sport	sport_tournament	cricket_tournament	rural_sport
Sports	Support to olympic / nationally-recognized sports	505	-0.061	olympic_sport	rural_sport	recognized_sport	sport_olympic	nationally_recognized
Sports	Support to sport clubs	295	0.053	rotary_club	club_rotary	cricket_club	cricket_association	club_nigdi
Sports	Rural sport	233	0.335	rural_sport	sport_rural	sport_sport	sport_art	art_culture
Sports	Sponsorship	224	-0.126	sponsorship_sport	sport_event	sport_sponsorship	event_sponsorship	event_sport
Sports	Support to sport training, incl. rural	115	0.462	recognized_sport	nationally_recognized	sport_nationally	sport_training	rural_sport
Sports	Sport equipment	64	0.032	sport_equipment	sport_material	sport_item	equipment_sport	distribution_sport
Women empowerment	Education and young women empowerment	874	0.142	woman_empowerment	empowerment_woman	empowerment_education	education_woman	youth_empowerment
Women empowerment	Gender equality	509	0.070	empowering_woman	gender_equality	equality_empowering	woman_empowering	woman_gender
Women empowerment	Disadvantaged women and girls	258	0.044	girl_child	asare_woman	woman_asare	woman_child	girl_woman
Women empowerment	Textile industry training for women	153	0.045	sewing_machine	center_woman	training_center	woman_empowerment	tailoring_center
Women empowerment	Women associations, councils, and civic partici...	71	-0.095	state_council	council_woman	woman_state	mahila_dakshata	service_sedan
Agroforestry	Sustainable agriculture	232	0.029	organic_farming	agriculture_farm	national_agriculture	water_resource	improved_agriculture
Agroforestry	Farmer training	172	0.146	training_farmer	farmer_training	farmer_farmer	farmer_agriculture	farmer_education
Agroforestry	Agronomy	149	-0.073	guar_seed	seed_distribution	agronomic_practice	distribution_guar	extension_service
Animal welfare	Animal care	444	0.176	animal_welfare	animal_husbandry	animal_care	welfare_animal	care_animal
Animal welfare	Animal shelters	383	-0.148	lion_club	animal_welfare	hare_movement	stray_dog	welfare_animal
Animal welfare	Cow sheds	316	0.096	cow_shed	cow_protection	protection_cow	shed_cow	welfare_cow

Notes: This table summarizes the CSR project clusters obtained by k-means clustering. Details are provided in the text.

Table C.3: Mapping of CSR Topics to Government Spending and NGO Activity

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

Notes: This table reports the mapping of CSR topics to government spending and NGO activity.

Table C.4: Example of Text from NIC Handbook (Division 16)

Panel A: Text from Explanatory Notes

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

161 Sawmilling and planning of wood.

162 Manufacture of products of wood, cork, straw and plaiting materials. This group includes the manufacture of products of wood, cork, straw or plaiting materials, including basic shapes as well as assembled products.

Panel B: Text from Further Industry Breakdown

Saw milling and planing of wood

Saw milling and planing of wood

Sawing and planing of wood

Manufacture of unassembled wooden flooring including parquet flooring

Manufacture of wooden railway sleepers

Activities related to saw milling and planing of wood n.e.c.

Manufacture of products of wood, cork, straw and plaiting materials

Manufacture of veneer sheets; manufacture of plywood, laminboard, particle board and other panels and board

Manufacture of ply wood and veneer sheets

Manufacture of particle board and fibreboard including densified wood

Manufacture of flush doors and other boards or panels

Manufacture of other plywood products n.e.c.

Manufacture of builders' carpentry and joinery

Manufacture of structural wooden goods [intended to be used primarily in the construction industry such as beams, rafters, roof struts, glue-laminated and metal connected, prefabricated wooden roof trusses, doors, windows, shutters and their frames, whether or not containing metal fittings, stairs, railings, wooden beadings and mouldings, shingles and shakes etc.]

Manufacture of prefabricated buildings, or elements thereof, predominantly of wood

Manufacture of builders' carpentry and joinery n.e.c.

Manufacture of wooden containers

Manufacture of wooden boxes, barrels, vats, tubs, packing cases etc.

Manufacture of plywood chests

Manufacture of market basketry, grain storage bins and similar products made of bamboo or reed

Manufacture of other wooden containers and products entirely or mainly of cane, rattan, bamboo, willow, fibre, leaves and grass n.e.c.

Manufacture of other products of wood; manufacture of articles of cork, straw and plaiting materials

Manufacture of wooden industrial goods

Manufacture of cork and cork products

Manufacture of wooden agricultural implements

Manufacture of various articles made of bamboo, cane and grass

Table C.5: Closest Development Topic by Industry

Industry Name	Topic	Closest 10 Token Pairs
IT and consultancy (62)	Vocational skills	[skill,expertise: 0.433], [training,testing: 0.388], [preservice,communication: 0.337], [preservice,technical: 0.333], [learning,communication: 0.331], [vocational,technical: 0.323], [training,technical: 0.322], [aptitude,expertise: 0.316], [carpentry,designing: 0.311], [skilled,expertise: 0.308]
Financial (64)	Livelihood enhancement	[welfare,pension: 0.409], [employment,income: 0.383], [employment,investment: 0.350], [farming,business: 0.342], [disadvantaged,income: 0.340], [sustainable,investment: 0.335], [employment,business: 0.335], [community,business: 0.325], [welfare,income: 0.317], [agriculture,banking: 0.314]
Petroleum products (19)	Safe drinking water	[electricity,gas: 0.508], [electricity,fuel: 0.455], [electricity,coal: 0.448], [electricity,propane: 0.403], [electricity,kerosene: 0.399], [groundwater,mineral: 0.384], [water,gas: 0.381], [sullage,liquid: 0.376], [reservoir,gas: 0.371], [water,liquid: 0.365]
Power and utilities (35)	Safe drinking water	[water,water: 1.000], [electricity,electricity: 1.000], [sewage,water: 0.662], [groundwater,water: 0.659], [potable,water: 0.625], [electricity,electric: 0.614], [electricity,energy: 0.603], [lake,water: 0.590], [electricity,power: 0.581], [river,water: 0.577]
Oil and gas extraction (6)	Safe drinking water	[electricity,gas: 0.508], [filtration,separator: 0.496], [electricity,coal: 0.448], [water,sand: 0.446], [bunding,separator: 0.423], [filtration,gasification: 0.413], [reservoir,hydrocarbon: 0.407], [bunding,sand: 0.402], [effluent,gasification: 0.400], [reservoir,drilling: 0.399]
Chemicals (20)	Safe drinking water	[checkdams,agarbatti: 0.458], [filtration,coating: 0.420], [filtration,deodorizing: 0.416], [potability,nitric: 0.399], [filtration,reagent: 0.395], [filtration,detergent: 0.391], [filtration,polyurethane: 0.387], [filtration,chemical: 0.382], [filtration,spray: 0.379], [sullage,agarbatti: 0.378]
Basic metals (24)	Safe drinking water	[groundwater,manganese: 0.448], [potability,manganese: 0.414], [effluent,manganese: 0.366], [potable,manganese: 0.362], [water,manganese: 0.330], [effluent,smelting: 0.313], [electricity,smelting: 0.309], [filtration,smelting: 0.308], [filtration,manganese: 0.308], [sewage,manganese: 0.305]
Pharmaceuticals (21)	Health	[healthcare,pharmaceutical: 0.540], [medical,unani: 0.424], [medical,pharmaceutical: 0.409], [medical,allopathic: 0.407], [doctor,homoeopathic: 0.400], [outpatient,allopathic: 0.396], [helath,homoeopathic: 0.372], [treatment,antibiotic: 0.367], [physician,allopathic: 0.354], [doctor,allopathic: 0.352]
Motor vehicles (29)	Safe drinking water	[electricity,electrical: 0.573], [sewer,electrical: 0.372], [sewage,electrical: 0.364], [sewerage,electrical: 0.361], [bunding,radiator: 0.340], [electricity,alternator: 0.340], [filtration,exhaust: 0.338], [irrigation,tractor: 0.338], [irrigation,electrical: 0.336], [drainage,electrical: 0.333]
Coal mining (5)	Safe drinking water	[bunding,belowground: 0.507], [electricity,fuel: 0.455], [electricity,coal: 0.448], [filtration,cleaning: 0.412], [checkdams,belowground: 0.410], [groundwater,mineral: 0.384], [sullage,belowground: 0.382], [sewerage,quarrying: 0.373], [checkdams,quarrying: 0.372], [drainage,grading: 0.369]
Wholesale trade (46)	Agroforestry	[agricultural,agricultural: 1.000], [dairy,dairy: 1.000], [agriculture,agricultural: 0.873], [farming,agricultural: 0.718], [ag,agricultural: 0.711], [livestock,agricultural: 0.645], [farm,agricultural: 0.638], [dairying,dairy: 0.631], [horticulture,agricultural: 0.625], [agrarian,agricultural: 0.619]
Civil engineering (42)	Infrastructure	[construction,construction: 1.000], [highway,highway: 1.000], [road,road: 1.000], [sewerage,sewerage: 1.000], [sewerage,sewage: 0.786], [sewerage,sewer: 0.725], [highway,road: 0.697], [drainage,sewer: 0.654], [drainage,sewerage: 0.631], [drainage,sewage: 0.613]
Other transport (30)	Safe drinking water	[lake,pontoon: 0.477], [river,pontoon: 0.458], [lake,boat: 0.445], [river,boat: 0.442], [lake,watercraft: 0.398], [creek,pontoon: 0.391], [river,barge: 0.382], [sullage,transport: 0.372], [sullage,pontoon: 0.368], [sewerage,transport: 0.367]
Industrial machinery (28)	Safe drinking water	[electricity,electrical: 0.573], [filtration,compressor: 0.467], [electricity,turbine: 0.424], [filtration,coating: 0.420], [filtration,electroplating: 0.417], [filtration,hydraulic: 0.409], [electricity,furnace: 0.398], [bunding,conveyor: 0.393], [filtration,turbine: 0.391], [bunding,hydraulic: 0.390]
Non-metallic minerals (23)	Safe drinking water	[bunding,rockwool: 0.493], [checkdams,rockwool: 0.484], [sullage,rockwool: 0.468], [bunding,aluminous: 0.452], [irrigation,rockwool: 0.417], [filtration,nonwoven: 0.402], [checkdams,aluminous: 0.393], [filtration,rockwool: 0.391], [water,rockwool: 0.388], [potable,calcined: 0.382]
Data and info (63)	Technology	[psn,internet: 0.355], [psn,web: 0.346], [edu,internet: 0.344], [brail,library: 0.334], [edu,web: 0.319], [edu,portal: 0.302], [pwd,database: 0.300], [pwd,searchable: 0.298], [pwd,portal: 0.293], [pwd,web: 0.293]
Logistics and warehousing (52)	Safe drinking water	[sullage,lighterage: 0.500], [sewerage,infrastructure: 0.450], [drainage,maintenance: 0.408], [checkdams,lighterage: 0.392], [sewerage,railway: 0.389], [bunding,lighterage: 0.380], [sullage,transport: 0.372], [sewerage,transport: 0.367], [drainage,infrastructure: 0.364], [electricity,infrastructure: 0.359]
Food and beverages (10)	Agroforestry	[dairy,milk: 0.732], [dairy,meat: 0.549], [dairy,vegetable: 0.480], [pisciculture,oilcake: 0.472], [farmer,vegetable: 0.468], [livestock,meat: 0.466], [grower,potato: 0.463], [dairy,potato: 0.463], [agricultural,oilcake: 0.452], [farmer,potato: 0.450]
Tobacco (12)	Agroforestry	[agricultural,agricultural: 1.000], [agriculture,agricultural: 0.873], [farming,agricultural: 0.718], [ag,agricultural: 0.711], [livestock,agricultural: 0.645], [farm,agricultural: 0.638], [horticulture,agricultural: 0.625], [agrarian,agricultural: 0.619], [dairy,agricultural: 0.603], [irrigation,agricultural: 0.570]
Rubber and plastics (22)	Safe drinking water	[water,hose: 0.447], [reservoir,pipe: 0.425], [sewer,pipe: 0.415], [bunding,conveyor: 0.393], [sewage,pipe: 0.391], [sullage,carboy: 0.381], [filtration,conveyor: 0.378], [creek,tub: 0.372], [water,tub: 0.370], [bunding,hose: 0.362]

Notes: This table presents, for each industry, the closest social topic based on the semantic proximity score (see equation (7)). Results are restricted to the top 20 industries. For each pair, we report the 10 most similar token pairs (from the 20 tokens closest to each group's centroid in the word embedding space), along with their cosine similarity scores.

Table C.6: Furthest Development Topic by Industry

Industry Name	Topic	Closest 10 Token Pairs
IT and consultancy (62)	Emergency relief	[assistance,expertise: 0.334], [assistance,service: 0.300], [emergency,communication: 0.254], [reconstruction,maintenance: 0.250], [assistance,technical: 0.249], [reconstruction,installation: 0.243], [assistance,communication: 0.221], [emergency,maintenance: 0.214], [assistance,facility: 0.198], [emergency,service: 0.197]
Financial (64)	Technology	[ict,intermediation: 0.222], [motorola,intermediation: 0.205], [edu,intermediation: 0.202], [tecnology,intermediation: 0.202], [dx,intermediation: 0.193], [pmc,intermediation: 0.188], [pwd,intermediation: 0.183], [psn,mortgage: 0.183], [engg,banking: 0.182], [pwd,corporation: 0.179]
Petroleum products (19)	Sports	[kabbadi,kerosene: 0.216], [hockey,bitumen: 0.185], [cycling,refining: 0.176], [football,refinery: 0.152], [cycling,bitumen: 0.149], [gymnastics,refining: 0.147], [judo,butane: 0.146], [kabbadi,briquette: 0.145], [hockey,gas: 0.145], [judo,kerosene: 0.143]
Power and utilities (35)	Sports	[cycling,diesel: 0.190], [cycling,energy: 0.179], [athletics,energy: 0.167], [sport,diesel: 0.164], [sport,generation: 0.155], [cycling,thermal: 0.153], [cycling,solar: 0.150], [tennis,water: 0.146], [hockey,gas: 0.145], [triathlon,diesel: 0.140]
Oil and gas extractions (6)	Sports	[hockey,shale: 0.183], [cycling,liquefaction: 0.172], [tennis,sand: 0.168], [cycling,mining: 0.167], [boxing,mining: 0.160], [triathlon,sand: 0.160], [football,shale: 0.158], [rugby,shale: 0.156], [cycling,pyrolysis: 0.154], [sporting,mining: 0.153]
Chemicals (20)	Emergency relief	[catastrophe,chemical: 0.204], [disaster,chemical: 0.178], [diaster,chemical: 0.149], [aid,dye: 0.146], [flood,chemical: 0.139], [emergency,chemical: 0.137], [quake,chemical: 0.129], [disater,chemical: 0.122], [catastrophe,reagent: 0.118], [reconstruction,chemical: 0.117]
Basic metals (24)	Emergency relief	[cyclone,alumina: 0.229], [cyclone,tin: 0.209], [diaster,alumina: 0.203], [catastrophe,alumina: 0.184], [flooding,alumina: 0.183], [flooding,manganese: 0.179], [reconstruction,alumina: 0.173], [reconstruction,metallurgic: 0.166], [earthquake,copper: 0.161], [reconstruction,steel: 0.160]
Pharmaceuticals (21)	Emergency relief	[catastrophe,chemical: 0.204], [emergency,allopathic: 0.190], [disaster,chemical: 0.178], [relief,homoeopathic: 0.163], [humanitarian,pharmaceutical: 0.160], [emergency,unani: 0.153], [cyclone,unani: 0.152], [diaster,chemical: 0.149], [diaster,suger: 0.145], [tsunami,unani: 0.144]
Motor vehicles (29)	Emergency relief	[emergency,electrical: 0.240], [earthquake,electrical: 0.224], [flooding,electrical: 0.209], [cyclone,lorry: 0.201], [emergency,brake: 0.195], [emergency,passenger: 0.184], [devastation,trailer: 0.182], [quake,electrical: 0.179], [tsunami,airbags: 0.177], [emergency,ignition: 0.167]
Coal miningg (5)	Sports	[cycling,transportation: 0.305], [athletics,transportation: 0.209], [rugby,opencast: 0.205], [athlete,hard: 0.188], [sport,hard: 0.186], [tournament,operation: 0.179], [cycling,agglomeration: 0.171], [cycling,mining: 0.167], [boxing,mining: 0.160], [football,hard: 0.160]
Wholesale trade (46)	Emergency relief	[aid,food: 0.331], [flood,agricultural: 0.255], [assistance,food: 0.254], [humanitarian,agricultural: 0.245], [humanitarian,food: 0.241], [displaced,food: 0.240], [emergency,electrical: 0.240], [relief,food: 0.238], [earthquake,electrical: 0.224], [flooding,agricultural: 0.222]
Civil engineering (42)	Technology	[engg,engineering: 0.367], [cnc,engineering: 0.285], [tecnology,engineering: 0.277], [cnc,repair: 0.263], [pwd,railway: 0.243], [cnc,industrial: 0.235], [ict,industrial: 0.225], [engg,industrial: 0.221], [cnc,maintenance: 0.215], [hyundai,railway: 0.211]
Other transport (30)	Emergency relief	[rescue,helicopter: 0.437], [emergency,helicopter: 0.334], [rescue,hovercraft: 0.301], [humanitarian,cargo: 0.269], [rescue,boat: 0.266], [rescue,crane: 0.263], [cyclone,vessel: 0.234], [rescue,pontoon: 0.231], [diaster,barge: 0.231], [emergency,transport: 0.229]
Industrial machinery (28)	Emergency relief	[emergency,electrical: 0.240], [emergency,equipment: 0.238], [assistance,equipment: 0.232], [earthquake,electrical: 0.224], [flooding,electrical: 0.209], [reconstruction,machinery: 0.198], [emergency,hydraulic: 0.195], [aid,equipment: 0.190], [quake,electrical: 0.179], [rescue,hydraulic: 0.177]
Non-metallic minerals (23)	Emergency relief	[flooding,rockwool: 0.198], [reconstruction,cement: 0.192], [flood,rockwool: 0.190], [cyclone,rockwool: 0.166], [earthquake,tile: 0.165], [quake,earthenware: 0.162], [cyclone,nonwoven: 0.158], [earthquake,earthenware: 0.158], [catastrophe,graphite: 0.154], [quake,tile: 0.150]
Data and info (63)	Emergency relief	[reconstruction,infrastructure: 0.383], [rescue,search: 0.367], [assistance,service: 0.300], [assistance,infrastructure: 0.240], [flooding,infrastructure: 0.235], [disaster,infrastructure: 0.233], [flood,infrastructure: 0.223], [assistance,agency: 0.216], [diaster,infrastructure: 0.212], [disaster,agency: 0.209]
Logistics and warehousing (52)	Sports	[cycling,transportation: 0.305], [cycling,transport: 0.284], [cycling,railway: 0.250], [paralympic,railway: 0.210], [athletics,transportation: 0.209], [tournament,facility: 0.196], [paralympic,airfield: 0.181], [kabbadi,railway: 0.179], [badminton,berthing: 0.177], [cycling,freight: 0.170]
Food and beverages (10)	Sports	[kabbadi,papads: 0.369], [kabbadi,sweetmeat: 0.364], [kabbadi,curd: 0.322], [kabbadi,oilcake: 0.287], [kabbadi,rice: 0.249], [badminton,tapioca: 0.248], [kabbadi,khandsari: 0.242], [badminton,papads: 0.217], [kabbadi,iodized: 0.211], [kabbadi,tapioca: 0.207]
Tobacco (12)	Emergency relief	[flood,agricultural: 0.255], [humanitarian,agricultural: 0.245], [flooding,agricultural: 0.222], [aid,agricultural: 0.215], [assistance,agricultural: 0.181], [reconstruction,agricultural: 0.178], [emergency,agricultural: 0.171], [cyclone,bidi: 0.162], [assistance,suitable: 0.162], [displaced,agricultural: 0.153]
Rubber and plastics (22)	Emergency relief	[emergency,hose: 0.183], [rescue,hose: 0.179], [disater,carboy: 0.170], [relief,rubber: 0.136], [catastrophe,material: 0.136], [diaster,container: 0.135], [assistance,material: 0.134], [aid,material: 0.127], [flood,material: 0.125], [emergency,tube: 0.122]

Notes: This table presents, for each industry, the furthest social topic based on the semantic proximity score (see equation (7)). Results are restricted to the top 20 industries. For each pair, we report the 10 most similar token pairs (from the 20 tokens closest to each group's centroid in the word embedding space), along with their cosine similarity scores.

Table C.7: Closest Industry by Development Topic

Topic	Industry Name	Closest 10 Token Pairs
Environmental sustainability	Forestry and logging (2)	[forestry,forestry: 1.000], [forest,forest: 1.000], [forestry,silviculture: 0.690], [forestry,forest: 0.676], [forestry,timber: 0.647], [reforestation,replanting: 0.646], [forest,timber: 0.608], [agriculture,forestry: 0.604], [agricultural,forestry: 0.586], [afforestation,replanting: 0.545]
Education	Education (85)	[education,education: 1.000], [educational,educational: 1.000], [school,school: 1.000], [vocational,vocational: 1.000], [literacy,literacy: 1.000], [academic,academic: 1.000], [classroom,classroom: 1.000], [student,student: 1.000], [eduction,education: 0.798], [elementary,school: 0.787]
Rural Infrastructure	Civil engineering (42)	[construction,construction: 1.000], [highway,highway: 1.000], [road,road: 1.000], [sewerage,sewerage: 1.000], [sewerage,sewage: 0.786], [sewerage,sewer: 0.725], [highway,road: 0.697], [drainage,sewer: 0.654], [drainage,sewerage: 0.631], [drainage,sewage: 0.613]
Hunger and malnutrition	Food and beverage services (56)	[food,food: 1.000], [meal,meal: 1.000], [eat,meal: 0.559], [nutritious,food: 0.547], [food,meal: 0.543], [nutrition,food: 0.520], [food,beverage: 0.510], [eat,drink: 0.507], [nutritious,meal: 0.504], [eat,food: 0.504]
Vocational skills	Education (85)	[vocational,vocational: 1.000], [training,training: 1.000], [education,education: 1.000], [educational,educational: 1.000], [education,educational: 0.678], [vocational,education: 0.611], [mentoring,tutoring: 0.594], [education,literacy: 0.594], [teaching,education: 0.571], [teaching,classroom: 0.563]
Technology	Computer and electronics (26)	[tecnology,technology: 0.466], [lenovo,modem: 0.453], [lcd,modem: 0.438], [tft,optical: 0.432], [pda,device: 0.424], [lenovo,computer: 0.421], [lcd,diode: 0.413], [lcd,optical: 0.404], [pda,modem: 0.397], [tecnology,electromedical: 0.392]
Health	Health care (86)	[health,health: 1.000], [medical,medical: 1.000], [hospital,hospital: 1.000], [patient,patient: 1.000], [dental,dental: 1.000], [clinic,clinic: 1.000], [outpatient,outpatient: 1.000], [treatment,treatment: 1.000], [doctor,doctor: 1.000], [outpatient,inpatient: 0.820]
Sanitation	Specialized construction (43)	[construction,construction: 1.000], [cleaning,cleaning: 1.000], [construction,excavation: 0.571], [construction,demolition: 0.564], [construction,building: 0.533], [cleaning,sanding: 0.479], [sewer,plumbing: 0.454], [drainage,plumbing: 0.448], [construction,installation: 0.436], [cleanliness,cleaning: 0.434]
Sports	Sports and recreation (93)	[sport,sport: 1.000], [sporting,sporting: 1.000], [athlete,athlete: 1.000], [sportsman,sportsman: 1.000], [athlete,sportsman: 0.583], [athletics,athlete: 0.565], [sport,racing: 0.562], [sport,sporting: 0.560], [athletics,sport: 0.552], [boxing,sport: 0.533]
Vulnerable populations	Public administration and defence (84)	[education,education: 1.000], [welfare,health: 0.506], [education,health: 0.473], [welfare,social: 0.473], [welfare,education: 0.468], [education,agriculture: 0.433], [family,community: 0.426], [education,social: 0.365], [welfare,agriculture: 0.361], [welfare,government: 0.357]
Safe drinking water	Water collection, treatment and supply (36)	[water,water: 1.000], [irrigation,irrigation: 1.000], [river,canal: 0.722], [sewage,water: 0.662], [groundwater,water: 0.659], [creek,canal: 0.627], [potable,water: 0.625], [lake,canal: 0.610], [drainage,irrigation: 0.600], [lake,water: 0.590]
Women empowerment	Education (85)	[education,education: 1.000], [literacy,literacy: 1.000], [education,educational: 0.678], [education,vocational: 0.611], [education,literacy: 0.594], [literacy,educational: 0.534], [education,academic: 0.519], [education,school: 0.506], [education,diploma: 0.481], [education,instruction: 0.474]
Animal welfare	Crops and animals (1)	[livestock,agricultural: 0.645], [livestock,farming: 0.545], [husbandry,farming: 0.494], [cattle,wheat: 0.479], [cattle,agricultural: 0.462], [livestock,wheat: 0.438], [cattle,farming: 0.422], [livestock,crop: 0.419], [pig,potato: 0.417], [husbandry,agricultural: 0.415]
Emergency relief	Public administration and defence (84)	[reconstruction,infrastructure: 0.383], [humanitarian,military: 0.334], [assistance,service: 0.300], [humanitarian,social: 0.299], [assistance,education: 0.293], [aid,education: 0.291], [aid,government: 0.284], [humanitarian,economic: 0.262], [humanitarian,security: 0.256], [emergency,health: 0.256]
Livelihood enhancement	Public administration and defence (84)	[education,education: 1.000], [community,community: 1.000], [social,social: 1.000], [agriculture,agriculture: 1.000], [farming,agriculture: 0.711], [educational,education: 0.678], [vocational,education: 0.611], [literacy,education: 0.594], [welfare,health: 0.506], [rural,agriculture: 0.486]
Agroforestry	Crops and animals (1)	[farming,farming: 1.000], [agricultural,agricultural: 1.000], [agriculture,agricultural: 0.873], [farm,farming: 0.731], [farming,agricultural: 0.718], [ag,agricultural: 0.711], [agriculture,farming: 0.711], [farmer,farming: 0.687], [dairying,farming: 0.675], [livestock,agricultural: 0.645]

Notes: This table presents, for each topic, the closest industry based on the semantic proximity score (see equation (7)). For each pair, we report the 10 most similar token pairs (from the 20 tokens closest to each group's centroid in the word embedding space), along with their cosine similarity scores.

Table C.8: Furthest Industry by Development Topic

Topic	Industry Name	Closest 10 Token Pairs
Environmental sustainability	Postal and courier (53)	[wetland,parcel: 0.312], [agriculture,transport: 0.308], [forestry,transport: 0.284], [forestry,local: 0.264], [sustainable,transport: 0.263], [environmental,local: 0.260], [agricultural,transport: 0.253], [conservation,local: 0.250], [forestry,postal: 0.238], [water,transport: 0.227]
Education	Metal ores mining (7)	[lifeskills,beneficiating: 0.300], [nonformal,beneficiating: 0.247], [vocational,beneficiation: 0.226], [curriculum,beneficiating: 0.226], [vocational,beneficiating: 0.226], [eduction,beneficiating: 0.222], [lifeskills,beneficiation: 0.210], [eduction,beneficiation: 0.198], [elementary,molybdenum: 0.194], [education,beneficiating: 0.191]
Rural Infrastructure	Publishing (58)	[infrastructure,software: 0.293], [sarai,encyclopaedia: 0.248], [chaupal,encyclopaedia: 0.237], [locality,periodical: 0.229], [rural,internet: 0.222], [sarai,periodical: 0.219], [building,publishing: 0.219], [kutcha,encyclopaedia: 0.207], [locality,encyclopaedia: 0.206], [infrastructure,directory: 0.206]
Hunger and malnutrition	Metal ores mining (7)	[nutritional,mineral: 0.276], [food,mineral: 0.263], [anganwadis,bauxite: 0.249], [malnourished,manganese: 0.237], [anganwadi,bauxite: 0.235], [malnutrition,manganese: 0.234], [anganwadis,beneficiation: 0.222], [nutritious,mineral: 0.222], [nutrition,mineral: 0.222], [sanitation,bauxite: 0.213]
Vocational skills	Metal ores mining (7)	[skilling,beneficiation: 0.402], [upskilling,beneficiation: 0.362], [enterpreneurship,beneficiating: 0.338], [reskilling,beneficiation: 0.328], [reskilling,beneficiating: 0.312], [upskilling,beneficiating: 0.303], [enterpreneurship,beneficiation: 0.302], [employability,beneficiation: 0.284], [vocationally,beneficiating: 0.283], [skilling,beneficiating: 0.281]
Technology	Metal ores mining (7)	[amu,ilmenite: 0.323], [cnc,beneficiating: 0.315], [tecnology,beneficiating: 0.314], [amu,niobium: 0.289], [ict,beneficiating: 0.283], [amu,tungsten: 0.282], [amu,cobalt: 0.278], [cnc,tungsten: 0.261], [tft,ilmenite: 0.251], [tft,tantalum: 0.250]
Health	Metal ores mining (7)	[dental,manganese: 0.201], [education,beneficiating: 0.191], [healthcare,mining: 0.189], [health,manganese: 0.185], [education,beneficiation: 0.185], [pediatric,antimony: 0.183], [treatment,beneficiation: 0.179], [obstetric,ore: 0.179], [treatment,cobalt: 0.178], [outpatient,molybdenum: 0.176]
Sanitation	Metal ores mining (7)	[construction,mining: 0.350], [water,mineral: 0.350], [water,manganese: 0.330], [sewage,manganese: 0.305], [sewerage,manganese: 0.291], [drainage,manganese: 0.275], [sewerage,bauxite: 0.273], [anganwadis,bauxite: 0.249], [construction,beneficiating: 0.245], [water,uranium: 0.243]
Sports	Metal ores mining (7)	[paralympic,gold: 0.262], [judo,gold: 0.261], [kabbadi,gold: 0.232], [gymnastics,gold: 0.229], [badminton,gold: 0.229], [sporting,gold: 0.195], [football,nickel: 0.179], [rugby,beneficiation: 0.177], [hockey,gold: 0.170], [boxing,gold: 0.169]
Vulnerable populations	Metal ores mining (7)	[destitute,bauxite: 0.199], [education,beneficiating: 0.191], [education,beneficiation: 0.185], [orphange,beneficiating: 0.163], [welfare,beneficiation: 0.151], [orphange,bauxite: 0.149], [destitute,beneficiation: 0.145], [orphaned,beneficiation: 0.139], [orphange,rutile: 0.138], [child,mineral: 0.133]
Safe drinking water	Publishing (58)	[checkdams,atlas: 0.217], [potability,content: 0.211], [electricity,internet: 0.202], [checkdams,periodical: 0.195], [checkdams,encyclopaedia: 0.190], [lake,atlas: 0.168], [potability,software: 0.167], [drinking,content: 0.163], [electricity,directory: 0.163], [electricity,newspaper: 0.159]
Women empowerment	Metal ores mining (7)	[empowerment,beneficiation: 0.445], [upliftment,beneficiation: 0.439], [upliftment,beneficiating: 0.347], [empowerment,beneficiating: 0.322], [upliftment,bauxite: 0.281], [upliftment,rutile: 0.275], [disadvantaged,beneficiation: 0.264], [entrepreneurship,beneficiation: 0.254], [equality,beneficiation: 0.245], [upliftment,ilmenite: 0.240]
Animal welfare	Metal ores mining (7)	[gaushalas,beneficiating: 0.308], [livestock,mineral: 0.268], [gaushala,beneficiation: 0.261], [cattle,ore: 0.260], [gaushalas,beneficiation: 0.252], [cattle,uranium: 0.249], [husbandry,beneficiating: 0.239], [gaushalas,bauxite: 0.236], [husbandry,beneficiation: 0.236], [gaushala,beneficiating: 0.227]
Emergency relief	Basic metals (24)	[cyclone,alumina: 0.229], [cyclone,tin: 0.209], [diaster,alumina: 0.203], [catastrophe,alumina: 0.184], [flooding,alumina: 0.183], [flooding,manganese: 0.179], [reconstruction,alumina: 0.173], [reconstruction,metallurgic: 0.166], [earthquake,copper: 0.161], [reconstruction,steel: 0.160]
Livelihood enhancement	Metal ores mining (7)	[upliftment,beneficiation: 0.439], [farming,mining: 0.384], [upliftment,beneficiating: 0.347], [agriculture,mining: 0.331], [agriculture,mineral: 0.296], [agriculture,beneficiating: 0.286], [employability,beneficiation: 0.284], [upliftment,bauxite: 0.281], [upliftment,rutile: 0.275], [sustainable,beneficiation: 0.265]
Agroforestry	Postal and courier (53)	[agriculture,transport: 0.308], [agricultural,transport: 0.253], [orchard,parcel: 0.235], [irrigation,transport: 0.232], [pisciculture,local: 0.229], [farming,transport: 0.227], [pisciculture,various: 0.226], [orchard,pickup: 0.224], [agricultural,local: 0.220], [livestock,transport: 0.219]

Notes: This table presents, for each topic, the furthest industry based on the semantic proximity score (see equation (7)). For each pair, we report the 10 most similar token pairs (from the 20 tokens closest to each group's centroid in the word embedding space), along with their cosine similarity scores.

D Implementation of the CSR Mandate

D.1 Additional Details on the CSR Mandate

This appendix provides more detail on the implementation of the CSR mandate defined in Section 135 of the Companies Act, 2013. The mandate stipulates that firms above a certain size must form a CSR committee, 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; these correspond to the 'initial' topics listed in Table C.1. It states that CSR activities should be undertaken by the companies in "project/programme mode": one-off events such as awards, sponsorships, etc. would not qualify as CSR expenditure under the Act. It also states that activities undertaken in pursuance of the 'normal course of business' of a company, including expenditures that benefit only the company's employees, sponsorships done for marketing purposes, or contributions to a political party, do not qualify as CSR expenditures.

The amount spent on CSR cannot be claimed as business expenditure and hence is not tax-deductible. While no specific tax exemption has been extended to CSR, spending on several activities (e.g., rural development projects, contribution to Prime Minister National Relief Fund) already enjoys exemptions under different sections of Tax Act, 1961, subject to fulfillment of specified conditions.

Firms can implement their CSR activities directly, or via implementing agencies. Implementing agencies can be an entity established by the company itself, a government entity, or any other entity with a track record of undertaking similar activities. The mandate does not detail what extent of outsourcing implies that firms should declare their CSR projects as implemented via a third-party.

D.2 Change in CSR Expenditures After the Mandate was Implemented

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

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

where f indexes the firm and y the year, and the outcome variable is CSR spending scaled by average profits in the preceding three years. Treated_{fy} is equal to one if the firm is liable under the CSR regulation in year y, Post_y is a dummy equal to one every year from 2015 onwards, γ_y are year fixed effects and γ_f are firm fixed effects. γ_g are group fixed effects, which indicate the liability status of a given firm in a given year. Standard errors are clustered at the firm level.

We define CSR spending in Prowess as the sum of two variables: "social and community expenses (incl. CSR)" and "donations". Social and community expenses are expenses incurred by firms for the benefit of society in general. Donations include donations for social causes, religious purposes, or political parties. Both social and community expenses and donations are reported in firms' financial statements as part of "miscellaneous expenditures". In 2015, in alignment with the introduction of the policy, Prowess began to ask explicitly, and separately, about CSR expenditures. Since this variable was not available before, we do not utilize it to estimate the policy impact, which requires pre- and post-policy data.

Table D.1 describes the results. In column 1, we observe that the share of CSR spending over average profits increases by 1.1% for liable firms relative to non-liable firms after the mandate is implemented. This effect remains stable if we replace the year fixed effects with year×industry×state fixed effects in column 2. The effect is smaller than 2% because liable firms already spent 0.7% of profits on CSR before the mandate, so reaching the threshold required an increase of 1.3 percentage points rather than 2 percentage points. Non-liable firms spend 0.8% both before and after (see also Figure 1(b) for raw trends).

The key identification assumption is that the CSR expenditures of firms liable and non-liable under the CSR mandate would have evolved in the same way in the absence of the mandate. While this assumption is untestable, Figure D.1 documents parallel pre-trends in an event study analysis.

Table D.2 also separately shows the effect on social and community expenses, as well as donations. We observe a small decline in donations (column 3), offset by a larger increase in social expenses (column 2). This might reflect a reclassification of spending that was previously reported as donations into the social and community expenses category, which maps more directly onto CSR activities. Some donations may also not qualify as CSR, leading firms to substitute toward eligible activities. Our definition of CSR as the sum of social expenses and donations is immune to this substitution effect.

D.3 Are Firms Manipulating Their CSR Spending?

We next discuss evidence of whether firms manipulate their CSR spending. During our period, the government relied on provisions such as mandatory disclosures, board and CSR committee accountability with an independent director, and audit of accounts to ensure correct accounting of CSR expenditures. In 2019, the government also introduced fines for failing to meet the mandate's requirements.

Firms have two possible options to manipulate their level of CSR spending. First, firms might manipulate their accounting variables to change their treatment status. We initially investigate which threshold is most binding: income, net worth, or profit (Figure D.2(a) to D.2(c)). We observe that for income and net worth, only 6% of firm-years have values higher than the respective threshold. In contrast, 32% of firm-years have values higher than the respective threshold for profits. This suggests that profit is the binding threshold for the majority of firms. Figure D.2(d) depicts the distribution of profits before the policy, between 2007 and 2014. Figure D.2(e) depicts it after the policy, between 2015 and 2019. Zooming in on the part of the distribution around the threshold, visual inspection shows only minor bunching below the threshold. This evidence suggests that only a small minority of firms manipulated their liability status. To address this dimension of manipulation, we further test a version of the difference-in-difference specification in equation (D.1) in columns 3 and 4 of Table D.1, but instrument the treatment status Treated_{fy} with a pre-policy variable Treated_{f}. The latter is an indicator equal to one if the firm is liable under the CSR regulation in the year 2014, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years (2011–2013). Results are quantitatively similar under this specification.

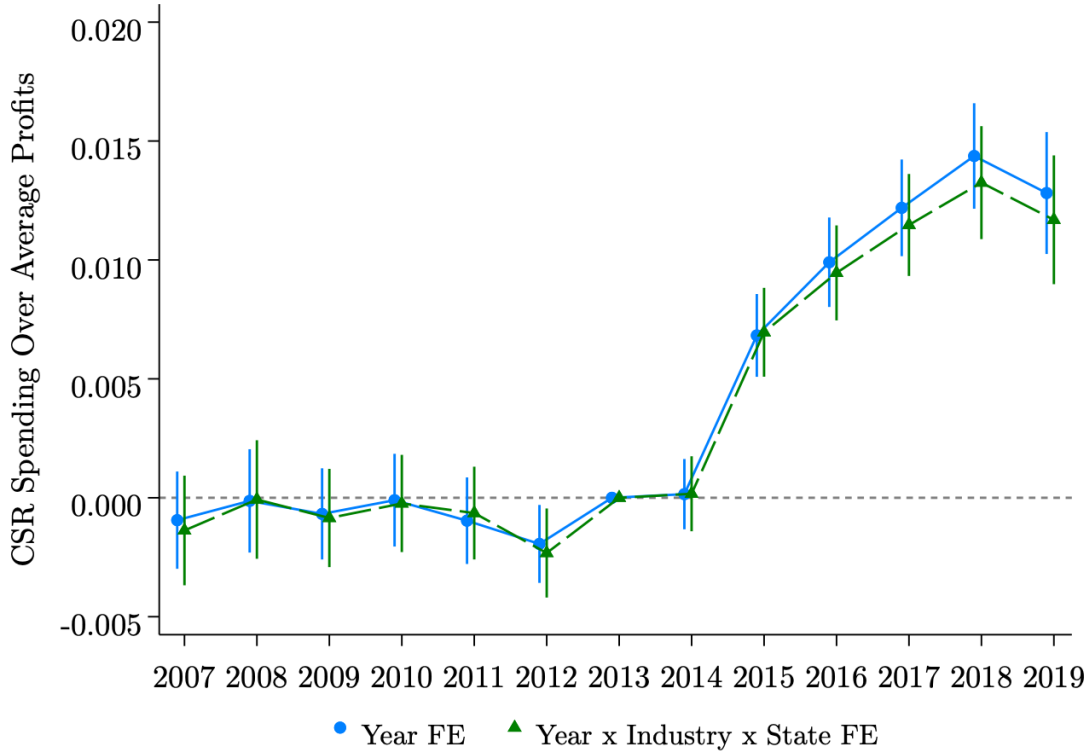
Second, firms might manipulate by wrongly classifying some of their expenditures as CSR to increase their total CSR spending (section D.1 above describes which expenses are excluded from CSR spending by the policy). Such behavior would not necessarily be a concern for our study, unless reclassified expenditures are systematically assigned to topics that are technologically close to the firms' industries, or to locations close to where the firm operates, but would nevertheless cast some doubt on the plausibility of the CSR data. To test whether firms are doing so, we run equation (D.1), using as dependent variables expenditures reported in Prowess that firms could plausibly re-classify as CSR. In Table D.3, we see little or no effect on expenditures on the (work) environment, employee welfare or training, social amenities, or advertisement. There appears to be a decrease in marketing expenses, which could indicate some reclassification, but might also be consistent with firms spending less on this dimension because the policy might have a negative effect on their private valuation.

D.4 Does CSR Crowd-Out Government Expenditures?

Finally, we investigate the impact of CSR spending on government spending in Table D.4. A simple OLS regression of the log of government spending on the log of CSR spending on the state×year level yields an insignificant coefficient of 0.012. To further explore this relationship, we employ an instrumental variables approach. The instrument is constructed as the share of aggregated state-level firm profit subject to the policy (measured prior to 2013), interacted with a post-policy indicator. While this instrument positively predicts CSR spending, the first-stage coefficient is not statistically significant. In the second stage, the estimated effect of CSR on government spending remains similar in magnitude to the

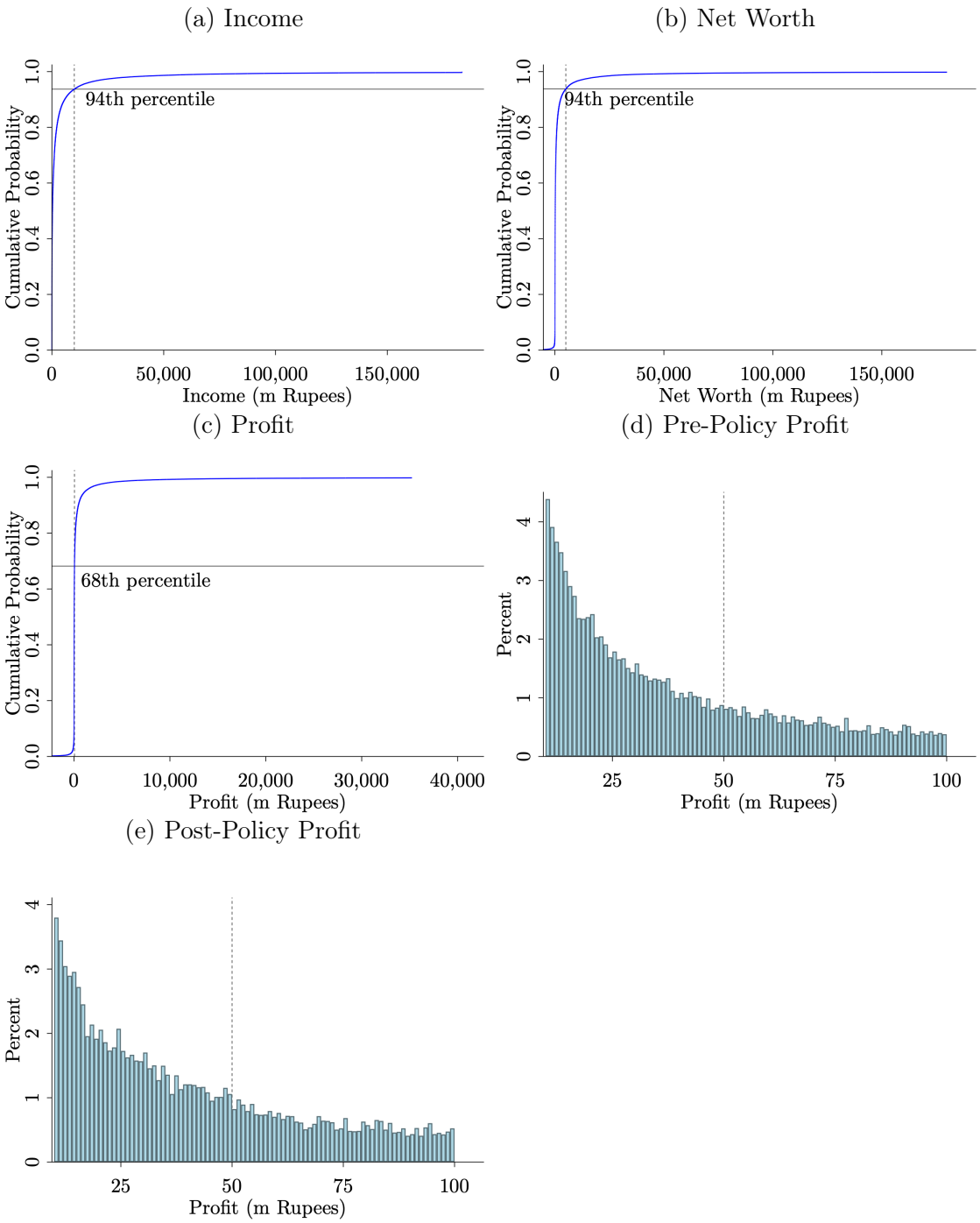
OLS result and likewise insignificant. Given the limitations of the research design, we refrain from making any causal claims about the impact of CSR spending on government spending.

Figure D.1: Effect of the Policy on CSR Spending



Notes: This figure describes the effect of the policy on CSR spending, derived from equation (D.1). The unit of observation is at the firm×year level. The dependent variable is the CSR spending of a given firm (f) in a given year (y) over average profits in the past three years. The independent variables are the interactions of Treated_{fy} with year indicators. Treated_{fy} is an indicator equal to one if the firm is liable under the CSR regulation in year y, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years. The green dashed line replaces the year fixed effects in the regression with year×industry×state fixed effects. Variables are winsorized at the 99th percentile. Standard errors are clustered at the firm level. The figure shows 95% confidence intervals.

Figure D.2: Manipulation of Liability Status



Notes: This figure tests for manipulation of the liability status. The unit of observation is at the firm×year level. Figures D.2(a) to D.2(c) show the cumulative probability for the three size thresholds. Figure D.2(d) shows the profit distribution between 2007 and 2014. Figure D.2(e) shows the profit distribution between 2015 and 2019.

Table D.1: Effect of the Policy on CSR Spending

	$CSR_{fy}/\overline{Profit}^{3y}_{fy}$			
	DID	DID	DID-IV	DID-IV
	(1)	(2)	(3)	(4)
$Treated_{fy} \times Post_y$	0.011***	0.011***		
	(0.001)	(0.001)		
$Treated_f \times Post_y$			0.012***	0.012***
			(0.001)	(0.001)
Firm FE	✓	✓	✓	✓
Group FE	✓	✓	✓	✓
Year FE	✓		✓	
Year×Ind.×State FE		✓		✓
F Statistic			231	174
R-squared	0.31	0.34	0.00	0.00
Observations	197,729	197,729	197,729	197,729

Notes: This table describes the effect of the policy on CSR spending, derived from equation (D.1). The unit of observation is at the firm×year level. In columns 1 and 2, the dependent variable is the CSR spending of a given firm (f) in a given year (y) over average profits in the past three years. The independent variable is the interaction of $Treated_{fy}$ with $Post_y$. $Treated_{fy}$ is an indicator equal to one if the firm is liable under the CSR regulation in year y, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years. $Post_y$ is a dummy equal to one every year from 2015 onwards. In columns 3 and 4, we instrument the time-varying liability variable $Treated_{fy}$ with $Treated_f$, which is an indicator equal to one if the firm is liable under the CSR regulation in the year 2014, that is, if either profits, income, or net worth, are above their respective thresholds in any of the three preceding financial years (2011–2013). 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 D.2: Accounting for CSR

	CSR (Social + Donations) (1)	Social (2)	Donations (3)
$\text{Treated}_{fy} \times \text{Post}_y$	0.44**	0.49***	-0.05**
	(0.19)	(0.18)	(0.03)
Avg dep var	0.18	0.07	0.11
Firm FE	✓	✓	✓
Group FE	✓	✓	✓
Year×Ind.×State FE	✓	✓	✓
R-squared	0.30	0.29	0.35
Observations	197,729	197,729	197,729

Notes: This table describes CSR spending in Prowess as the sum of social and community expenses as well as donations, derived from equation (D.1). The unit of observation is at the firm×year level. All expenses are divided by total expenses and multiplied by 100. The independent variable is the interaction of Treated_{fy} with Post_y. Treated_{fy} is an indicator equal to one if the firm is liable under the CSR regulation in year y, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years. Post_y is a dummy equal to one every year from 2015 onwards. 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 D.3: Test for Manipulation by Relabeling

	Environment	Employee Welfare	Employee Training	Social Amenities	Advertising	Marketing
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Treated}_{fy} \times \text{Post}_y$	-0.033	0.508	0.014*	-0.000	0.054	-0.155***
	(0.026)	(0.439)	(0.008)	(0.000)	(0.034)	(0.049)
Avg dep var	0.02	0.53	0.03	0.00	0.60	1.14
Firm FE	✓	✓	✓	✓	✓	✓
Group FE	✓	✓	✓	✓	✓	✓
Year×Ind.×State FE	✓	✓	✓	✓	✓	✓
R-squared	0.31	0.24	0.39	0.18	0.42	0.55
Observations	197,729	197,729	197,729	197,729	197,729	197,729

Notes: This table tests for manipulation by relabeling regular business expenses, derived from equation (D.1). The unit of observation is at the firm×year level. All expenses are divided by total expenses and multiplied by 100. The dependent variables are expenses that the firm could have relabeled as CSR. The independent variable is the interaction of Treated_{fy} with Post_y. Treated_{fy} is an indicator equal to one if the firm is liable under the CSR regulation in year y, that is, if either profits, income, or net worth are above their respective thresholds in any of the three preceding financial years. Post_y is a dummy equal to one every year from 2015 onwards. 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 D.4: CSR and Government Spending

	Log(Government Spending)$_{sy}$		Log(CSR)$_{sy}$
	OLS	Second Stage	First Stage
	(1)	(2)	(3)
Log(CSR)$_{sy}$	0.012	0.010	
	(0.009)	(0.028)	
Instrument (Treated Profit Share×Post)$_{sy}$			0.828
			(1.518)
State FE	✓	✓	✓
Year FE	✓	✓	✓
R-squared	1.00	0.60	0.96
Observations	326	326	326

Notes: This table tests for the effect of CSR spending on government spending. The unit of observation is the state×year level. Column 1 provides an OLS regression (equation (D.2)). Column 2 provides the second stage (equation (D.3)), and column 3 the first stage (equation (D.4)). In columns 1 and 2, the dependent variable is the log of government spending in a given state and year. In column 3, the dependent variable is the log of CSR spending in a given state and year. The independent variable in column 1 is the log of CSR spending in a given state and year, and in column 2 the same variable instrumented by the share of treated profits in a given state times a Post variable that is one after the policy introduction. Controls include state GDP, the number of firms in Prowess, the number of treated firms in Prowess, the average income and profit of firms in Prowess, as well as state-level linear time trends and region×year fixed effects. Standard errors are clustered at the state level. $***$, $**$ and $*$ indicate significance at the 1%, 5% and 10% levels.

$$\text{Log(Government Spending)}_{sy} = \beta \text{Log(CSR)}_{sy} + \phi X_{sy} + \gamma_s + \gamma_y + \epsilon_{sy} \tag{D.2}$$

$$\text{Log(Government Spending)}_{sy} = \beta \widehat{\text{Log(CSR)}}_{sy} + \phi X_{sy} + \gamma_s + \gamma_y + \epsilon_{sy} \tag{D.3}$$

$$\text{Log(CSR)}_{sy} = \beta \text{Treated Profit Share} \times \text{Post}_{sy} + \phi X_{sy} + \gamma_s + \gamma_y + \epsilon_{sy} \tag{D.4}$$

E Model

E.1 Model Derivations

Production. Production occurs across projects, indexed by $p \in \mathcal{P}$. $\mathcal{P}$ can be partitioned into $\mathcal{P}^{CSR}$ for CSR projects and $\mathcal{P}^{pri}$ for for-profit goods sold in competitive markets. Projects are produced by combining tasks $\tau \in \mathcal{T}$. The relative importance of tasks varies across projects as characterized by the vector $\boldsymbol{\phi_p} = [\phi_{p\tau}]_{\tau\in\mathcal{T}}$ such that $\forall p$, $\sum_\tau \phi_{p\tau} = 1$. To perform tasks, firms hire workers and are endowed with a task-specific productivity vector $\mathbf{z_f} = [z_{f\tau}]_{\tau\in\mathcal{T}}$. Firm f's output in project p is:

$$y_{fp} = \left[\prod_{\tau\in\mathcal{T}} x_{fp\tau}^{\phi_{p\tau}}\right]^{\rho}$$

where $\rho < 1$ and $x_{fp\tau}$ is the amount of task τ used in production of project p. In addition, $x_{fp\tau} = \exp(z_{f\tau})\ell_{fp\tau}$, where $\ell_{fp\tau}$ is the labor assigned to task τ and project type p. The wage w is taken to be exogenous.

Project-specific productivity. Conditional on producing project p, the firm allocates labor ℓ_{fp} across tasks in a way that maximizes project-specific returns:

$$\pi_{fp} = \max_{\{\ell_{fp\tau}\}} \zeta_{fp} y_{f,p} - w\ell_{fp}$$

subject to: $\sum_{\tau\in\mathcal{T}} \ell_{fp\tau} = \ell_{fp}$. When p is a for-profit good, ζ_{fp} is the market price of good p. When p is a CSR project ζ_{fp} captures how much firm f values project type p, as above. The optimal labor allocation satisfies: $\frac{\ell_{fp\tau}}{\ell_{fp}} = \phi_{p\tau}$. Project-level profit maximization yields an expression for project output as a function of firm-project productivity α_{fp} :

$$y_{fp} = \exp(\alpha_{fp})\ell_{fp}^{\rho}, \tag{E.1}$$

with:

$$\alpha_{fp} = \rho||\boldsymbol{\phi_p}||||\mathbf{z_f}||\cos(\boldsymbol{\phi_p}, \mathbf{z_f}) + \kappa_p \tag{E.2}$$

where $\kappa_p = \rho\sum_\tau \phi_{p\tau}\log(\phi_{p\tau})$ is a product-level constant.

$\cos(\boldsymbol{\phi_p}, \mathbf{z_f})$ is the cosine similarity between the vector of returns to tasks for that project type $\boldsymbol{\phi_p}$ and the firms' vector of task-specific technologies $\mathbf{z_f}$. In the main text, we assume that $||\boldsymbol{\phi_p}|| = \phi, \forall p$ and $||\mathbf{z_f}|| = 1, \forall f$. Then,

$$\alpha_{fp} = \rho\phi\cos(\boldsymbol{\phi_p}, \mathbf{z_f}) + \kappa_p \tag{E.3}$$

Privately and socially optimal allocations. The firm allocates labor across projects to maximizes total returns:

$$\max_{\{\ell_{fp}\}} \sum_{p \in \mathcal{P}^{pri}} (\zeta_{fp} y_{fp} - w\ell_{fp}) + \sum_{p \in \mathcal{P}^{CSR}} (\zeta_{fp} y_{fp} - w\ell_{fp})$$

where we now define the exogenous CSR expenditure requirement as $\sum_{p \in \mathcal{P}^{CSR}} w\ell_{fp} \geq E$. We obtain that firms' privately optimal CSR shares follow the expression in equation (3), with α_{fp} defined in (2). The social planner maximizes the social welfare function given in (5) above. This yields the same socially optimal allocation across CSR project types p as above (equation (6)).

Implications for CSR productivities across industries. Our empirical tests exploit variation in firms' technologies at the level of their industry. We define an industry as a set of private goods $\mathcal{P}_i \subset \mathcal{P}^{pri}$ centered around a technological vector $\boldsymbol{\phi_i}$: $\forall p \in \mathcal{P}_i, \boldsymbol{\phi_p} = \boldsymbol{\phi_i} + \boldsymbol{\epsilon_p}$. We assume that $\boldsymbol{\epsilon_p}$ is mean-zero and i.i.d. That is, for all p, i, $\boldsymbol{\phi_i} \perp \boldsymbol{\epsilon_p}$.

In addition, we assume that firms belonging to industry i have a productivity vector centered around $\boldsymbol{\phi_i}$: $\mathbf{z_f} = \boldsymbol{\phi_i} + \boldsymbol{\epsilon_f}$. Likewise, we assume that $\boldsymbol{\epsilon_f}$ is mean-zero and i.i.d. That is, for all f, i, $\boldsymbol{\phi_i} \perp \boldsymbol{\epsilon_f}$.

This ensures that, on average, firms in industry i have $p \in \mathcal{P}_i$ as their main product, in line with how firms are classified across industries in standard datasets. That is, firms' industry choice reflects the proximity between their vector of task-specific productivity and industries' technological requirements.

Proof. Let us characterize the sales share of product p for firm f. We assume that for private goods, the firms' valuation corresponds to the price of the good, i.e. $\zeta_{fp} = p_{fp}$. In addition, we assume that the price of good p is determined by a standard demand curve:

$$p_{fp}^{-\sigma} = \frac{y_{fp}}{Y}$$

where Y is the total output of the economy. Labor employed in the production of good p for firm f is determined by the following first-order condition:[43]

$$p_{fp} \rho \exp(\alpha_{fp}) \ell_{fp}^{\rho-1} = w, \forall p \in \mathcal{P}^{pri}$$

[43]Here, firms act as price takers, and take the price of good p_{fp} as given. The implications below would hold if we instead assumed that firms internalize the demand curve and set a constant markup over marginal cost.

Plugging the demand curve and using the production function, we obtain:

$$y_{fp} = \exp\left(\frac{\sigma\alpha_{fp}}{\sigma(1-\rho)+\rho}\right)\left(\left(\frac{w}{\rho}\right)^{\frac{-\rho\sigma}{(1-\rho)\sigma+\rho}} Y^{\frac{\rho}{(1-\rho)\sigma+\rho}}\right)$$

which is increasing in α_{fp}.

$$p_{fp}y_{fp} = \exp\left(\frac{\alpha_{fp}(\sigma-1)}{\sigma(1-\rho)+\rho}\right)\left(\frac{w}{\rho}\right)^{\frac{-\rho(\sigma-1)}{(1-\rho)\sigma+\rho}} Y^{\frac{1}{(1-\rho)\sigma+\rho}}$$

Then, with $\sigma > 1$, the sales share of product p, denoted $\tilde{s}_{fp}$, is strictly increasing in α_{fp}. That is, firms' main product in terms of sales share is their highest productivity product.

We now show that firms in industry i have higher productivity in products in industry i. Take product p in industry i and product p' in industry i'. We have:

$$\alpha_{fp} = \rho\ \boldsymbol{\phi_i}\cdot\boldsymbol{\phi_i} + \kappa_p$$
$$\alpha_{fp'} = \rho\ \boldsymbol{\phi_i}\cdot\boldsymbol{\phi_{i'}} + \kappa_{p'}$$

Assuming that $\kappa_p = \rho\sum_\tau \phi_{p\tau}\log(\phi_{p\tau})$ is approximately constant across products, and using the Cauchy-Schwarz inequality, we obtain the desired results. □

For a firm f in industry i, we can then write:

$$\alpha_{fp} = \rho\phi\cos(\boldsymbol{\phi_p}, \boldsymbol{\phi_i}) + \kappa_p \tag{E.4}$$

E.2 Model-Based Tests of Measurement Approach

Test 1. Industry predicts semantic proximity across firms. Let $i(f)$ denote the industry of firm f. If $i(f) = i(f')$ and $i(f) \neq i(f")$, then:

$$\cos(\mathbf{z_f}, \mathbf{z_{f'}}) \geq \cos(\mathbf{z_f}, \mathbf{z_{f''}}) \tag{E.5}$$

Proof. From our definition of industries,

$$\cos(\mathbf{z_f}, \mathbf{z_{f'}}) = \mathbf{z_f}\cdot\mathbf{z_{f'}} = \boldsymbol{\phi_i}\cdot\boldsymbol{\phi_i}$$

Following the same steps, $\cos(\mathbf{z_f}, \mathbf{z_{f''}}) = \boldsymbol{\phi_i}\cdot\boldsymbol{\phi_{i''}}$ for $i \neq i''$. $\boldsymbol{\phi_i}\cdot\boldsymbol{\phi_i} \geq \boldsymbol{\phi_i}\cdot\boldsymbol{\phi_{i''}}$ by the Cauchy-Schwarz inequality. □

Test 2. Semantic proximity across industries predicts firms' sales shares across industries. Assume that the valuation of private goods is equalized across goods: $\zeta_{fp} = \zeta_f$. Take a product p in industry i. Let f' (f'') be a firm in industry i' (i''). Then,

$$\cos(\boldsymbol{\phi}_{\mathbf{i}}, \boldsymbol{\phi}_{\mathbf{i'}}) \geq \cos(\boldsymbol{\phi}_{\mathbf{i}}, \boldsymbol{\phi}_{\mathbf{i''}}) \Rightarrow s_{f'p} \geq s_{f''p} \tag{E.6}$$

That is, fixing product p in industry i, a firm in an industry closer to industry i will have a higher sales share on product p.

Proof. Using the assumption that $\zeta_{fp} = \zeta_f$, s_{fp} is strictly increasing in $\alpha_{fp} = \rho\phi \cos(\boldsymbol{\phi}_{\mathbf{p}}, \mathbf{z}_{\mathbf{f}}) + \kappa_p$ for any firm f and product p. For firm f' in industry i',

$$\alpha_{f'p} = \rho\phi \cos(\boldsymbol{\phi}'_{\mathbf{i}}, \boldsymbol{\phi}_{\mathbf{i}}) + \kappa_p \tag{E.7}$$

and similarly for firm f'' in industry i''. The conclusion immediately follows. □