WIDER Working Paper 2023/4

Behind the numbers: exploring caste inequities in entrepreneurial success

Rajesh Raj Natarajan[1] and Kunal Sen[2]

January 2024

United Nations University World Institute for Development Economics Research

wider.unu.edu

Abstract: The documented under-representation of marginalized groups in business ownership and the labour market is a concerning issue. This study explores how caste disparities in small-firm entrepreneurship impact on firm performance in India, focusing on the informal sector. Our examination shows a significant productivity gap between firms owned by disadvantaged castes and others, including Other Backward Classes and Forward Castes, across the productivity distribution. The results of our decomposition exercise provide evidence for the importance of both differences in observables and returns to these observables in explaining the caste gap in productivity. This implies that even with improvements in firm attributes for Scheduled Caste or Scheduled Tribe businesses, the productivity disadvantage of firms owned by marginalized groups may persist. Pervasive market and non-market discrimination against marginalized groups suggests that the significant caste disparities in entrepreneurship and business performance will continue to impede the economy, presenting serious challenges for public policy.

Key words: caste, productivity gap, informal-sector firms, discrimination, India

JEL classification: J16, L25, L26, L60

Acknowledgements: This paper uses unit-level data on informal firms from the National Sample Survey Office, access to which is gratefully acknowledged. The usual disclaimers apply.

[1] Sikkim University, Gangtok, India, corresponding author: rajeshraj.natarajan@gmail.com [2] UNU-WIDER, Helsinki, Finland, and University of Manchester, UK

This study has been prepared within the UNU-WIDER project Structural transformation – old and new paths to economic development.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-462-5

https://doi.org/10.35188/UNU-WIDER/2024/462-5

Typescript prepared by Merl Fluin.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

The entrenched presence of ethnic and racial discrimination in developed labour markets is extensively documented (Borowczyk-Martins et al. 2017; Dahl and Krog 2018; Heath and Di Stasio 2019; Lang and Lehmann 2012; Lang and Spitzer 2020). In India, caste discrimination persists as a long-standing issue, particularly for disadvantaged castes such as Dalits (Scheduled Castes, SCs) and Adivasis (Scheduled Tribes, STs), defining the labour market landscape (Agrawal 2014; Arabsheibani et al. 2018; Gang et al. 2017; Thorat et al. 2007; Thorat and Newman 2010). Studies highlight pervasive discrimination against SCs and STs, spanning access to formal-sector jobs, discriminatory recruitment practices, and wage disparities. While existing research on caste discrimination in the Indian labour market largely focuses on waged employment, the persistent discrimination faced by these groups extends beyond wages to the domain of entrepreneurship. In a country such as India, where approximately half the workforce is self-employed, discrimination operates across various markets, including credit, land, inputs, and outputs (State of Working India 2023). Despite an extensive literature examining the uneven distribution of entrepreneurial ambitions, participation, and outcomes across social groups in developed countries (for a review, see Carter et al. 2015), studies within the Indian context are limited (Deshpande and Sharma 2016).

This study is pivotal in bridging the literature gap by scrutinizing caste discrimination in small businesses within the informal sector. The significance of this work is multifaceted. Firstly, it addresses a critical void because a substantial portion of entrepreneurial activities among SCs and STs is geared towards survival and is predominantly concentrated in the lower echelons of business ownership (Harriss-White et al. 2014; Harriss-White and Vidyarthee 2010; Mosse 2018). Moreover, businesses owned by disadvantaged castes, being easily identifiable, are more prone to caste discrimination from customers, suppliers, and lenders (Deshpande and Sharma 2016). This discrimination further complicates the ability of low-caste owners to secure loans, hire labour, and negotiate terms for inputs and outputs, effectively pushing them away from the large-business sector and compelling them to operate predominantly in the microenterprise sector (State of Working India 2023). Additionally, there exists ethnographic and qualitative evidence underscoring the persistent presence of caste-based discrimination against businesses owned by disadvantaged castes (Jodhka 2010; Prakash 2015).

This study, to the best of our knowledge, is one of the first to analyse caste disparities in small-business performance.[1] Leveraging extensive nationwide survey data on small businesses from the National Sample Survey Organization (NSSO), we delve into understanding these caste gaps in firm performance. Our approach goes beyond conventional ordinary least square (OLS) regressions, incorporating quantile regressions to provide a nuanced perspective on caste differences that extends beyond average productivity. Furthermore, employing the recentred influence function (RIF)-based decomposition method, we meticulously dissect the productivity gap into composition (explained) and structural (unexplained) effects at every decile across the productivity distribution.

Our findings reveal significant productivity gaps for both SC and ST businesses compared with non-SCST businesses. Notably, the productivity disparity is more pronounced for ST businesses,

[1] Among existing studies on caste discrimination, the research carried out by Deshpande and Sharma (2016) stands out as particularly relevant to this line of enquiry. Their study analyses caste differences in earnings among household businesses in India, using data from the India Human Development Survey for the period 2004–05.

where average productivity is approximately half that of non-SCST firms. While SC firms exhibit a relatively uniform productivity disadvantage compared with non-SCST firms across the productivity distribution, ST-owned firms show a decreasing pattern in productivity differences, with the productivity gap being particularly acute at the lower end of the distribution. The productivity decompositions indicate that although observable factors, especially firm characteristics, explain a substantial portion of the productivity gap, differences in coefficients and unobservable factors predominantly account for this gap. This implies that even with improvements in firm attributes for SC and ST businesses, the productivity disadvantage of firms owned by marginalized groups is likely to persist.

Our contribution to the literature is multifaceted. Firstly, we provide a unique perspective from a developing country, enriching the broader discourse on racial and ethnic disparities in small-business ownership. Secondly, while the impact of caste discrimination on education, income, and wealth is well documented, our study sheds light on the often-overlooked role of caste in business ownership and performance, offering insights into the underlying factors. This oversight is noteworthy given the substantial differences and the pivotal role of entrepreneurship in achieving income and social mobility (Kurian 2007; Omvedt 2017). The ongoing dialogue on 'Dalit capitalism in India' underscores the importance of Dalits entering the business sector for upwards social and economic mobility (Deshpande and Sharma 2016). However, the majority of enterprises owned by disadvantaged castes are microenterprises operating within households, limiting scalability and growth possibilities (Deshpande and Sharma 2013; Iyer et al. 2013). Our study provides comprehensive evidence for the first time regarding factors hindering the upwards progression of firms owned by marginalized groups. Thirdly, we present a new compilation of estimates regarding business ownership rates and outcomes by social group, utilizing an all-India data set covering small businesses in the manufacturing, trade, and service sectors. Finally, from a broader perspective, our study offers valuable insights into how caste operates in the economy as a structure of discrimination, curtailing opportunities and exacerbating inequalities.

The remainder of this paper is broken down into sections. Section 2 discusses the data and descriptive statistics. The section also presents a discussion of important correlates of firm productivity. The decomposition method employed in this study is outlined in section 3. In section 4, we thoroughly discuss the decomposition results and robustness test outcomes. The last section concludes with some policy implications.

2 Data and descriptive statistics

2.1 Data

We utilize unit-level data on informal-sector firms obtained from the latest two rounds (67th and 73rd) of the NSSO surveys on unincorporated non-agricultural enterprises for the years 2010–11 and 2015–16. These comprehensive nationwide surveys are conducted every five years, capturing the operational and economic attributes of informal-sector firms in India. The surveys cover firms in the manufacturing, trade, and service sectors. The firms covered are not registered under the Factories Act of 1948, exempting them from the industrial licensing and labour laws that apply to firms in the formal sector.[2] The NSSO employs a multistage, stratified random sampling method,

[2] We found a few firms to have employed more than 20 workers, meaning they should ideally have been part of the formal sector. We do not know whether these firms were operating illegally or had experienced size expansion after a

with census villages in rural areas and urban frame survey blocks in urban areas as the first-stage units (for more detail, see NSSO 2012, 2017). The sample firms or ultimate-stage units are selected from these first-stage units.[3] In the 67th round, 334,474 firms were surveyed (162,375 rural and 172,099 urban), while the 73rd round covered 290,113 firms (143,179 rural and 146,934 urban). The sampled firms account for roughly 0.5 to 0.6 per cent of the estimated population of firms in the informal sector. Taking this into consideration, we ensure the representativeness of our estimates by reweighting the firm-level observations using inverse sampling multipliers.

In the past decade or two, there has been a surge in studies leveraging this data set to analyse various facets of informal-sector firms in India. Explored topics include growth and productivity (Chatterjee et al. 2021; Hsieh and Klenow 2014; Raj and Sen 2016), firm transition (Kathuria et al. 2013; Kesar and Bhattacharya 2019; Mazumdar and Sarkar 2013; Raj and Sen 2015), and entrepreneurship (Banerji et al. 2016; Gang, Raj et al. 2022; Ghani et al. 2014a, 2014b), among others. A recent study has also examined the productivity gap by gender of owner and determinants using the same data set (Gang, Natarajan et al. 2022). However, studies examining caste gaps in productivity among small businesses in India are scarce. The data set's inclusion of crucial information about the caste of firm owners enables us to address this conspicuous gap in the existing literature.

In the survey, the firm owners self-identified with a specific caste, permitting us to categorize firms into three distinct, non-overlapping groups: SC-owned firms, ST-owned firms, and firms owned by others (non-SCSTs). After combining Other Backward Class (OBC)-owned and Forward Caste-owned firms into one category, given the absence of a significant productivity gap between them (see Figure A1 in the Appendix), our primary focus revolves around three overarching definitions of firm ownership: SC-owned firms, ST-owned firms, and non-SCST-owned firms.

The combined data set originally had a sample of 624,587 firms.[4] Some filters were applied to arrive at the working sample for this study. We restricted our analysis to firms with a solitary owner, excluding partnership firms—constituting around five per cent of the total sample—as the caste of the owner could not be ascertained. We further excluded firms with missing observations for the variables crucial to our study, and we identified outlier firms, which might potentially contaminate our results. Our data-filtering culminated in a final sample of 583,972 firms.[5]

Our main dependent variable is firm productivity, which is proxied using labour productivity.[6] We calculate labour productivity by dividing gross value-added by the number of workers, with the natural logarithm of labour productivity employed in our analyses (see Table 1 for variable definitions).[7]

---

period as part of the informal sector, or whether the discrepancy arose from data entry errors. Although we consider them in our estimations, our results are robust to their exclusion.

[3] In the case of large first-stage units, substrata known as second-stage strata (SSS) are formed, from which the sample firms are selected.

[4] The combined data set pools the cross-sectional data sets for the periods 2010–11 and 2015–16.

[5] The data-cleaning process led us to drop about 6.5 per cent of firms from the combined data set.

[6] Total factor productivity is also used as an alternative measure of productivity. We discuss this in subsection 4.4.

[7] Needless to say, we use the real values of gross value-added to compute labour productivity. The relevant deflator is employed to convert the nominal values to real values.

Table 1: Description of variables

Variables	Definition
Dependent variable	
Labour productivity	Real gross value-added/number of workers. Log of labour productivity is used in our empirical analysis
Caste ownership	
Non-SCST	Binary variable for firms owned by general or OBC entrepreneurs
SC	Binary variable for firms owned by SC entrepreneurs
ST	Binary variable for firms owned by ST entrepreneurs
Independent variables	
Firm characteristics	
Size	Logarithm of employment
Urban firms	Binary variable for firms in urban areas
Age	
Below 2 years	Binary variable for firms aged 2 years or below
3-9 years	Binary variable for firms aged between 3 and 9 years
Above 9 years	Binary variable for firms above 9 years
Assistance	Binary variable for enterprises that received government assistance
Registration	Binary variable for registered firms
Linkage	Binary variable for firms with subcontracting work
Account maintenance	Binary variable for firms maintaining accounts
Firm constraints	
Financial constraint	Binary variable for firms with borrowing constraint
Electricity constraint	Binary variable for firms with electricity constraint
Gender of firm owner	Binary variable for women-owned firms

Note: for the age variable, the benchmark category is under 2 years.

Source: authors' compilation.

The data also provides information on several potential determinants of firm productivity. All variables used in the analysis are binary variables, except firm size, which is measured in number of workers. The variables are grouped into three sets, namely, firm characteristics, firm constraints, and gender of firm owner. Besides firm size, firm characteristics include two dummy variables for age (one for three to nine years, and one for above nine years, with below two years as the reference category), and a dummy variable each for whether the firm has received government support, is registered, maintains accounts, and undertakes work on a contract basis. The set also includes a dummy variable that distinguishes between rural and urban firms. The firm constraints set includes one binary variable for borrowing constraint and another for electricity constraint. Our model also incorporates a dummy variable for gender, distinguishing between male-owned and female-owned firms. Two sets of controls, region and sector, are also included. The region set groups together dummies for six Indian regions, and the sector set accommodates dummies for three broad sectors of industrial activity. The vector of observed covariates used in our analysis is thus given by equation 1:

$$X = [Firm\ characteristics, Firm\ constraints, Gender, Region, Sector\ ] \quad (1)$$

While certain explanatory variables such as firm size and age are potentially endogenous, Fortin et al. (2011) argue that decompositions should be viewed as accounting exercises. These exercises facilitate a quantitative assessment of the factors contributing to outcome disparities between groups, without implying causal connections between these factors and outcomes.[8]

2.2 Descriptive statistics and correlates of firm productivity

Table A1 in the Appendix reports the summary statistics for the sample of firms owned by SCs, STs, and non-SCSTs separately. Of the total 583,962 firms, 58,428 are owned by SCs (ten per cent), 32,543 are owned by STs (6.6 per cent), and 492,991 are owned by non-SCSTs (84.4 per cent).

Between 2010–11 and 2015–16, overall productivity exhibits a notable rise, irrespective of the caste of firm owners (Table 2). However, significant productivity disparities exist among social groups. SC- and ST-owned firms consistently report lower productivity compared with those owned by non-SCST entrepreneurs across various percentiles. These productivity gaps are evident at the mean, 25th, 50th, and 75th percentiles of labour productivity, with statistical significance at the five per cent level. These differences remain broadly stable during the survey periods in 2010 and 2016. Figure A2, depicting the kernel density distribution of productivity, illustrates that non-SCST firms consistently outperform SC and ST firms, positioned distinctly to the right in the distribution. This confirms our observation of a sizeable caste gap in labour productivity, with non-SCST firms enjoying a considerable productivity advantage over SC and ST firms.

This significant gap in firm productivity might be attributed to various characteristics, most of which exhibit sharp differences between SC and ST businesses and non-SCST businesses. Table A1, comparing descriptive statistics by caste of firm owner, broadly supports this observation. In our sample, the size of an average firm is roughly the same across caste groups, employing around 1.5 workers per firm. Regarding female ownership, we notice a significant variation across castes, with SC and ST businesses having a higher proportion of women-owned firms compared with non-SCST firms.

In the breakdown of firms between rural and urban areas, a notable disparity emerges among social groups. While urban firms constitute over half of non-SCST firms in our sample, their presence is notably lower among SC and ST businesses, accounting for 37 per cent and 17 per cent respectively. Additionally, there is a distinct age-wise distribution pattern across castes. Old firms (nine years and above) constitute a dominant share among firms owned by SC and ST entrepreneurs. Conversely, a significant majority of non-SCST firms fall into the middle-aged (two to nine years) and old categories.

Differences across caste groups are also evident in registration, account maintenance, and government support. Only 29 per cent of firms in our sample are registered, and this proportion differs across social groups. The proportion of registered firms among SC and ST businesses is nearly half that among non-SCST businesses. The practice of maintaining accounts seems to be very low among firms in the informal sector. Only nine per cent of firms practise the maintenance of proper accounts. For SC and ST businesses, this practice is almost non-existent. Only four per cent of firms owned by SC and ST entrepreneurs report record-keeping practices.

[8] In any case, we ran a specification after dropping these possible endogenous variables. Our results are robust to this alternative specification too. The results are discussed in subsection 4.4.

Table 2: Productivity levels by social group

Year	ST					SC					Non-SCST			
	Mean	25	50	75		Mean	25	50	75		Mean	25	50	75
2010-11	9.10	8.46	9.16	9.85	*	9.35	8.78	9.51	10.03	*	9.67	9.09	9.81	10.39
2015-16	9.35	8.59	9.55	10.26	*	9.65	9.02	9.84	10.42	*	10.03	9.43	10.18	10.75
2010-11 to 2015-16	9.28	8.53	9.44	10.15	*	9.55	8.94	9.71	10.30	*	9.92	9.30	10.06	10.66

Note: 25, 50, and 75 correspond to 25th, 50th and 75th percentiles. Sample weights supplied by NSSO are used in estimations. * indicates that the differences in means and selected percentiles between SC and ST firms and non-SCST firms are significant at the 5 per cent level.

Source: authors' calculations.

In our sample, the proportion of firms that have received any kind of support from the government is very low across the board (only 0.8 per cent). The prevalence of subcontracting is remarkably low, with only nine per cent of all firms reporting any prior marketing agreement with other units. However, the participation of SC and ST businesses in such subcontracting relationships is below average (five per cent and eight per cent respectively).

The constraints small firms face in accessing credit and electricity also vary by caste of firm owner. The SC firms record a higher share—higher than the all-India share—of firms with borrowing constraints as compared with ST firms and non-SCST firms. In the case of firms reporting power outages, the share of non-SCST businesses is higher than the all-India share (four per cent), as compared with SC and ST firms (0.25 per cent).

There is substantial variation across regions with respect to both the location of small firms and the representation of caste groups in entrepreneurship. The most dynamic regions are south, east, and north, with an almost 25 per cent share of small firms. The share of firms in central and north-east regions is very low at six per cent and three per cent respectively. The regional representation of small businesses varies by caste of entrepreneur too. The share of SC businesses exceeds the all-India share in northern and eastern regions, while the share of ST businesses surpasses the all-India share in western, eastern, and central regions. However, the regional distribution of non-SCST businesses roughly mimics the pattern for the all-India sample.

The most important activity for small businesses is trading, which accounts for 37 per cent of India's small firms. This seems to be true for ST and non-SCST businesses: 40 per cent of ST businesses and 38 per cent of non-SCST businesses operate in this sector. SC-owned businesses, on the other hand, are more dominant in services, which include activities such as transport and storage, accommodation and food, finance and insurance, education, and health and social work, among others. Thirty-five per cent of SC entrepreneurs are engaged in this sector, while ST entrepreneurs constitute 25 per cent.

To provide some initial insights into the role of these covariates in the caste gap in productivity, we analyse average labour productivity by selected covariates for ST, SC, and non-SCST businesses. The results are presented in Table 3. Caste differences in productivity persist among both rural and urban firms. The significant edge in productivity for OBC and upper-caste entrepreneurs is not confined to firms owned by any particular gender: the differences are apparent for both male-run and female-run firms. Caste differences in productivity are visible for firms of all ages. Across all age groups, productivity levels are substantially higher for firms owned by Forward Castes. The superiority of non-SCST firms in firm productivity extends to other characteristics such as government assistance, registration, subcontracting, and account maintenance. This difference is stark and widespread: the average productivity of SC and ST firms consistently lags behind that of non-SCST firms, irrespective of sector and state.

Table 3: Average labour productivity by selected covariates

Variable	ST	SC	Non-SCST
All firms	9.276	9.551	9.920
Location			
Rural	9.141	9.369	9.620
Urban	9.943	9.854	10.201
Gender			
Male	9.355	9.762	10.121
Female	8.997	8.806	9.074
Age			
Below two years	8.947	9.049	9.494
3-9 years	9.558	9.647	9.977
Above 9 years	9.070	9.599	9.980
Assistance			
Yes	9.740	9.920	10.139
No	9.271	9.549	9.918
Registration			
Yes	10.196	10.212	10.485
No	9.088	9.403	9.663
Linkage			
Yes	8.664	8.755	9.106
No	9.314	9.629	10.003
Account maintenance			
Yes	10.183	10.177	10.685
No	9.239	9.523	9.833
Electricity constraint			
Yes	9.456	9.563	10.030
No	9.271	9.551	9.915
Financial constraint			
Yes	9.126	9.562	9.841
No	9.287	9.455	9.926
Region			
North	9.659	9.550	9.954
West	9.534	9.799	10.139
East	8.832	9.415	9.682
South	9.434	9.678	9.958
Central	8.792	9.176	9.690
North-east	10.073	10.070	10.155
Activity			
Manufacturing	8.917	9.060	9.473
Trade	9.201	9.730	10.128
Services	9.897	9.842	10.113
Year			
2010-11	9.098	9.354	9.673
2015-16	9.354	9.647	10.030

Notes: sample weights supplied by NSSO are used in estimations.

Source: authors' calculations.

Our study further estimates a multivariate regression model to understand the potential factors associated with productivity. The model takes the form:

$$logLP = \beta_0 + \beta_1 Caste + \beta_2 Size + \beta_3 U + \beta_4 Age + \beta_5 Assistance + \beta_6 Regis + \beta_7 linkage + \beta_8 Acmaint + \beta_9 Fin + \beta_{10} Elec + \beta_{11} Female + \gamma_s + \delta_j + \theta_t + \varepsilon \quad (2)$$

Equation 2 is estimated using OLS and unconditional quantile regression methods. Table 4 reports the OLS estimates and the unconditional quantile regression estimates for selected quantiles for the pooled sample.

Table 4: OLS and unconditional quantile regression estimations, pooled sample

Dependent variable: log labour productivity	OLS	p10	p50	p90
Firm characteristics				
ST	-0.3954***	-0.5403***	-0.3393***	-0.2099***
	(0.0699)	(0.0545)	(0.0290)	(0.0296)
SC	-0.1940***	-0.2385***	-0.1532***	-0.1843***
	(0.0175)	(0.0247)	(0.0118)	(0.0174)
Size	-0.2392***	-0.3184***	-0.2348***	-0.1344***
	(0.0218)	(0.0165)	(0.0094)	(0.0103)
Urban firms	0.4151***	0.4942***	0.3809***	0.3637***
	(0.0316)	(0.0178)	(0.0094)	(0.0118)
3-9 years	0.5010***	1.0144***	0.3981***	0.1698***
	(0.0264)	(0.0371)	(0.0197)	(0.0197)
Above 9 years	0.4700***	0.9552***	0.3734***	0.1342***
	(0.0268)	(0.0347)	(0.0184)	(0.0180)
Assistance	0.1771***	0.1935**	0.1500***	0.1863***
	(0.0348)	(0.0908)	(0.0137)	(0.0117)
Registration	0.4454***	0.5699***	0.3946***	0.3316***
	(0.0181)	(0.0209)	(0.0115)	(0.0140)
Linkage	-0.2625***	-0.0093	-0.3359***	-0.3970***
	(0.0362)	(0.0345)	(0.0140)	(0.0247)
Account maintenance	0.4690***	0.4630***	0.4504***	0.4177***
	(0.0188)	(0.0319)	(0.0154)	(0.0194)
Firm constraints				
Financial constraint	-0.0803***	-0.0774**	-0.0620***	-0.0632***
	(0.0250)	(0.0362)	(0.0158)	(0.0183)
Electricity constraint	0.1230***	0.1401***	0.1037***	0.0675***
	(0.0296)	(0.0339)	(0.0132)	(0.0137)
Gender of firm owner				
Female	-0.7590***	-0.9055***	-0.7571***	-0.5874***
	(0.0221)	(0.0248)	(0.0132)	(0.0187)
State effects?	Y	Y	Y	Y
Sectoral effects?	Y	Y	Y	Y
Year effects?	Y	Y	Y	Y
Constant	8.9885***	7.3754***	9.2030***	10.3052***
	(0.0643)	(0.0423)	(0.0218)	(0.0237)
R^2/pseudo R^2	0.3612	0.2207	0.2231	0.1893
Number of observations	583,962	583,962	583,962	583,962

Note: in all estimations, we use sample weights provided by NSSO. Robust standard errors (clustered by district) in parentheses.

Source: authors' calculations.

Our findings unequivocally highlight the pivotal role of the firm owner's caste in determining productivity. The coefficients of SC and ST dummy variables, with statistical significance at the one per cent level, reveal that all else being equal, SC and ST businesses consistently exhibit lower levels of labour productivity compared with non-SCST businesses. In terms of magnitude, ST businesses demonstrate a 32.6 per cent lower productivity than their non-SCST counterparts, while SC businesses are 17.6 per cent less productive.[9] It is noteworthy that the conditional productivity gaps (32.6 per cent for ST and 17.6 per cent for SC businesses) are narrower than the unconditional gaps (64 per cent for ST and 37 per cent for SC businesses), indicated in Table 3. The size variable demonstrates an inverse relationship with labour productivity, suggesting that larger firms experience lower productivity. As anticipated, urban-based businesses exhibit higher labour productivity, potentially due to proximity to markets and easier access to information. Furthermore, firms aged between three and nine years and those older than nine years display higher labour productivity compared with those below three years, indicating a positive correlation between firm age and productivity.

Our findings underscore that registered, government-supported, and well-documented firms tend to exhibit higher productivity levels. Conversely, businesses engaged in contractual work show lower productivity levels compared with those not on contracts. A significant productivity disparity also emerges between male-run and female-run businesses. The negative and significant coefficient on the female dummy suggests that that all else being equal, women-owned businesses are negatively correlated with labour productivity. Specifically, female-owned firms demonstrate a 53 per cent lower productivity than their male-owned counterparts, aligning with prior empirical evidence (Gang, Natarajan et al. 2022).

Beyond OLS estimates, Table 4 presents the quantile regression estimates for the 10^{th}, 50^{th}, and 90^{th} percentiles. When the OLS results are juxtaposed against the results for the various percentiles, we observe that limiting the discussion to OLS alone misses a large part of the picture. Undoubtedly, SC and ST businesses exhibit lower productivity than non-SCST firms across the distribution, indicated by consistently negative and significant coefficients for SC and ST dummies at all deciles. However, the difference is more pronounced at the 10^{th} percentile, gradually diminishing as we ascend the productivity distribution. The coefficients of other variables maintain the same sign and significance across the percentiles. As with the SC and ST variables, the impact of these factors is more pronounced at lower deciles and diminishes at upper deciles. For instance, aspects such as urban location and the business's operational years exert a more substantial influence at the lower end of the productivity distribution. Similarly, the productivity contrast between male and female entrepreneurs is more conspicuous in the lower percentiles compared with the upper percentiles.

Pooled regressions rest on the presumption that returns to observed variables are the same across social groups. Recognizing the impracticality of this assumption, especially in the context of India, we also perform the OLS estimations for each caste group. Table 5 presents the results disaggregated by caste. Reassuringly, estimates from the caste-specific OLS regressions confirm findings based on pooled regressions. The relationship between selected variables and firm productivity remains consistent across caste groups, although the magnitude of influence varies among them.

[9] The coefficient estimates are -0.3954 and -0.1940 respectively for SC and ST dummy variables. Since the dependent variable is log-transformed, we calculate the percentage difference in labour productivity as `(exp(-0.3954)-1)*100` for ST businesses and `(exp(-0.1940)-1)*100` for SC businesses, following Halvorsen and Palmquist (1980).

Table 5: OLS and unconditional quantile regression coefficients on log labour productivity

Variables	ST firms				SC firms				Non-SCST firms			
	OLS	p10	p50	p90	OLS	p10	p50	p90	OLS	p10	p50	p90
Firm characteristics												
Size	-0.6772***	-0.9986***	-0.6750***	-0.3659***	-0.4740***	-0.5875***	-0.4763***	-0.2940***	-0.1928***	-0.2512***	-0.1918***	-0.1125***
	(0.0885)	(0.1150)	(0.0532)	(0.0534)	(0.0344)	(0.0251)	(0.0382)	(0.0277)	(0.0202)	(0.0175)	(0.0102)	(0.0083)
Urban firms	0.5150***	0.4975***	0.4437***	0.5326***	0.3833***	0.4985***	0.3184***	0.3472***	0.4113***	0.4791***	0.3821***	0.3572***
	(0.0615)	(0.0429)	(0.0268)	(0.0372)	(0.0414)	(0.0267)	(0.0222)	(0.0193)	(0.0319)	(0.0213)	(0.0118)	(0.0114)
3-9 years	0.5450***	0.8516***	0.5156***	0.3883***	0.5715***	1.1292***	0.5019***	0.1946***	0.4881***	0.9876***	0.3791***	0.1616***
	(0.0737)	(0.2852)	(0.0966)	(0.0598)	(0.0441)	(0.0809)	(0.0483)	(0.0555)	(0.0302)	(0.0399)	(0.0226)	(0.0135)
Above 9 years	0.3604***	0.7523***	0.3293***	0.1580**	0.5675***	1.1751***	0.4752***	0.1409**	0.4653***	0.9257***	0.3658***	0.1333***
	(0.1135)	(0.2858)	(0.0978)	(0.0641)	(0.0481)	(0.0844)	(0.0497)	(0.0558)	(0.0299)	(0.0371)	(0.0207)	(0.0105)
Assistance	0.2116	0.3533***	0.0377	0.3590***	0.2487***	0.2763**	0.1240	0.2286**	0.1529***	0.1272	0.1439***	0.1838***
	(0.1304)	(0.1328)	(0.1504)	(0.0507)	(0.0868)	(0.1166)	(0.1130)	(0.1066)	(0.0380)	(0.0785)	(0.0244)	(0.0198)
Registration	0.5327***	0.5195***	0.4816***	0.3661***	0.4587***	0.5075***	0.4251***	0.3791***	0.4297***	0.5665***	0.3762***	0.3215***
	(0.0553)	(0.0759)	(0.0312)	(0.0365)	(0.0318)	(0.0344)	(0.0292)	(0.0197)	(0.0182)	(0.0234)	(0.0139)	(0.0135)
Linkage	-0.0633	0.4318**	-0.1717	-0.1822**	-0.1430*	0.1241**	-0.1953***	-0.2495***	-0.2844***	-0.0419	-0.3538***	-0.4102***
	(0.1354)	(0.1734)	(0.1068)	(0.0822)	(0.0801)	(0.0566)	(0.0452)	(0.0316)	(0.0380)	(0.0383)	(0.0251)	(0.0179)
Account maintenance	0.5461***	0.4000***	0.5144***	0.5530***	0.4453***	0.5113***	0.2850***	0.5118***	0.4513***	0.4381***	0.4347***	0.4101***
	(0.0840)	(0.0775)	(0.0374)	(0.0434)	(0.0667)	(0.0736)	(0.0655)	(0.0288)	(0.0188)	(0.0277)	(0.0173)	(0.0195)
Firm constraints												
Financial constraint	-0.0604	0.0004	-0.0901**	-0.1007*	-0.1134*	-0.1755**	-0.0842**	-0.0468***	-0.0746***	-0.0603*	-0.0585***	-0.0655***
	(0.1066)	(0.1433)	(0.0442)	(0.0524)	(0.0641)	(0.0880)	(0.0425)	(0.0168)	(0.0238)	(0.0358)	(0.0170)	(0.0232)
Electricity constraint	0.1459	0.3998	0.1300	0.2023***	0.1505	0.1584***	0.0561	0.0837***	0.1128***	0.1382***	0.1002***	0.0682***
	(0.1074)	(0.2852)	(0.0900)	(0.0459)	(0.0922)	(0.0405)	(0.0768)	(0.0305)	(0.0306)	(0.0440)	(0.0187)	(0.0085)
Gender of firm owner												
Female	-0.3921***	-0.3596***	-0.4039***	-0.2195***	-0.7904***	-0.8709***	-0.7661***	-0.6581***	-0.7816***	-0.9371***	-0.7758***	-0.6006***
	(0.0523)	(0.0625)	(0.0760)	(0.0612)	(0.0468)	(0.0317)	(0.0352)	(0.0215)	(0.0225)	(0.0313)	(0.0186)	(0.0189)
State effects?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sectoral effects?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year effects?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Constant	8.8535***	7.3141***	8.9415***	9.9425***	8.7059***	6.8974***	8.9174***	10.1320***	9.0043***	7.4194***	9.2305***	10.3177***
	(0.1613)	(0.3551)	(0.1070)	(0.0908)	(0.0736)	(0.0802)	(0.0560)	(0.0609)	(0.0667)	(0.0460)	(0.0244)	(0.0182)

R^2/pseudo R^2	0.3849	0.2346	0.2571	0.2047	0.3437	0.2150	0.2118	0.1890	0.3531	0.2161	0.2181	0.1806
Number of observations	32,543	32,543	32,543	32,543	58,428	58,428	58,428	58,428	492,991	492,991	492,991	492,991

Note: in all estimations, we use sample weights provided by NSSO. Robust standard errors (clustered by district) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations.

3 Distributional decomposition analysis

Our main objective in this study is to understand the factors that explain the gap in productivity between SC and ST businesses and non-SCST businesses. The question of interest is to see not only how the covariates affect average productivity but also other facets of the productivity distribution, including measures of inequality. Taking cognizance of this, the study employs a decomposition approach developed to link changes in the distribution of independent characteristics to statistics defined on the productivity distribution, namely, percentiles, the Gini coefficient, and percentile ratios, among others. This approach is known as the RIF decomposition approach. Originally proposed by Firpo et al. (2018), this approach employs RIF regressions, in conjunction with a reweighting strategy, to decompose the differences in distributional statistics beyond the mean.[10] The decomposition based on RIF regressions is an improved extension of the Oaxaca-Blinder decomposition approach, and it has three main advantages over other existing approaches: (1) ease of implementation, (2) scope for deriving the detailed contribution of each individual covariate in the aggregate decomposition, and (3) the possibility of extending the decomposition to any distributional measure for which RIF can be defined (Firpo et al. 2018; Rios-Avila 2020). The approach is explained in detail below.

Let us assume that we have a joint distribution function, $f_{Y,Z,c}(y_i, z_i, C_i)$, which explains the relationship between labour productivity (Y), firm and owner attributes (Z), and the caste identity of the firm owner (C). The joint probability distribution function and cumulative distribution of Y conditional on C can then be stated:

$$f^{c}{}_{Y,Z}(y,z) = f^{c}{}_{Y|Z}(Y|Z) f^{c}{}_{Z}(Z) \tag{3a}$$

$$F^{c}{}_{Y}(y) = \int F^{c}{}_{Y|Z}(Y|Z)\, dF^{c}{}_{Z}(Z) \tag{3b}$$

where ^{c} denotes that the density is conditional on $\mathrm{C} = \mathrm{c}$ with $c\epsilon[0,1]$. To examine the productivity gap between SC and ST businesses ($\mathrm{c} = 1$) and non-SCST businesses ($\mathrm{c} = 0$) for a given distributional statistic v, the cumulative conditional distribution of Y can be used:

$$\Delta v = v_1 - v_0 = v(F^{1}{}_{Y}) - v(F^{0}{}_{Y}) \tag{4a}$$

$$\Delta v = v\left(\int F^{1}{}_{Y|Z}\,(Y|Z) dF^{1}{}_{Z}\right) - v\left(\int F^{0}{}_{Y|Z}\,(Y|Z) dF^{0}{}_{Z}\right) \tag{4b}$$

Equation 4b clarifies that variations in the statistic of interest Δv result from disparities in the distribution of Z ($dF^{1}{}_{Z}(Z) \neq dF^{0}{}_{Z}(Z)$) and the relationship between Y and Z ($dF^{1}{}_{Y|Z}(Y|Z) \neq dF^{0}{}_{Y|Z}(Y|Z)$). This mirrors the standard Oaxaca-Blinder decomposition, comparing distinctions in average characteristics and coefficients. To gauge the significance of differences in characteristics (composition effect) and differences in returns to these characteristics (structural effect) in explaining the overall gap in the distributional statistic v, it is crucial to establish a counterfactual statistic v_{cf}:

$$v_{cf} = v(F^{cf}{}_{Y}) = v\left(\int F^{0}{}_{Y|Z}\,(Y|Z) dF^{1}{}_{Z}(Z)\right) \tag{5}$$

[10] Influence functions were developed to present a measure of how a particular feature of the distribution (such as the median, or an inequality measure such as Gini) is influenced by each data point in the distribution.

Using the counterfactual scenario in equation 5, the gap in distribution statistic v can be decomposed into two components, namely, the composition effect (Δv_X) and the structure effect (Δv_s):

$$\Delta v = \underbrace{(v_1 - v_{cf})}_{\Delta v_s} + \underbrace{(v_{cf} - v_0)}_{\Delta v_X} \tag{6}$$

The foremost challenge, however, is the identification of a counterfactual statistic v_{cf} because outcomes and covariates are not observed for the same firm owner in two states. We therefore rely on the semiparametric reweighting procedure proposed by DiNardo et al. (1996) to identify the counterfactual scenario ${F^0}_{Y|Z}(Y|Z)d{F^1}_Z(Z)$ from the observed data. Although the distribution of outcomes and covariates for the counterfactual scenario ${F^{cf}}_{Y|Z}$ is not observed, an approximation of the counterfactual distribution can be obtained by multiplying the distribution of observed characteristics $d{F^0}_Z(Z)$ by a reweighting factor $\omega(Z)$, thereby representing the distribution $d{F^1}_Z(Z)$ (Rios-Avila 2020). Following this, the counterfactual scenario in equation 5 can be revised:

$${F^{cf}}_Y = \int {F^0}_{Y|Z}(Y|Z)d{F^1}_Z(Z) \cong \int {F^0}_{Y|Z}(Y|Z)d{F^0}_Z(Z)\omega(Z) \tag{7}$$

The reweighting factor $\omega(Z)$ is identified as follows:

$$\omega(Z) = \frac{d{F^1}_Z(Z)}{d{F^0}_Z(Z)} = \frac{dF_{Z|C}(Z|C=1)}{dF_{Z|C}(Z|C=0)} = \frac{dF_{Z|C}(C=1|Z)}{dF_C(C=1)}\frac{dF_C(C=0)}{dF_{Z|C}(C=0|Z)}$$
$$= \frac{1-P}{P}\frac{P(C=1|Z)}{1-P(C=1|Z)} \tag{8}$$

where p denotes the share of firms in group $\mathrm{C}=1$, and $P(C=1|Z)$ stands for the conditional probability of a firm with characteristics Z of group $\mathrm{C}=1$. Effectively, this suggests that the reweighting factor $\omega(Z)$, which is essential to identify the counterfactual distribution ${F^{cf}}_{Y|Z}$, can be arrived at by estimating the conditional probability $P(C=1|Z)$ using parametric methods. As shown by Firpo et al. (2018), logit or probit models can be used for the purpose. Upon obtaining the reweighting factor, the RIF regression is estimated separately for each group and the counterfactual as follows:

$$v_1 = E[RIF\{y_i, v(F_Y^1)\}] = \bar{X}^{1\prime}\hat{\beta}^1 \tag{9a}$$

$$v_0 = E[RIF\{y_i, v(F_Y^0)\}] = \bar{X}^{0\prime}\hat{\beta}^0 \tag{9b}$$

$$v_{cf} = E\left[RIF\left\{y_i, v\left(F_Y^{cf}\right)\right\}\right] = \bar{X}^{cf\prime}\hat{\beta}^{cf} \tag{9c}$$

This strategy results in four final decomposition components:

$$\Delta v = \underbrace{\underbrace{(\bar{X}^{cf} - \bar{X}^0)'\hat{\beta}^0}_{\Delta v_X^p} + \underbrace{\bar{X}^{cf\prime}(\hat{\beta}^{cf} - \hat{\beta}^0)}_{\Delta v_X^e}}_{\Delta v_X} + \underbrace{\underbrace{\bar{X}^{1\prime}(\hat{\beta}^1 - \hat{\beta}^{cf})}_{\Delta v_S^p} + \underbrace{(\bar{X}^1 - \bar{X}^{cf})'\hat{\beta}^{cf}}_{\Delta v_S^e}}_{\Delta v_S} \tag{10}$$

The components Δv_X and Δv_S correspond to the aggregate composition effect and aggregate structural effect in the Oaxaca-Blinder decomposition. In the RIF decomposition, the aggregate composition effect breaks down into pure composition effect (Δv_X^p) and specification error (Δv_X^e),

while the aggregate structural effect is dissected into pure structural effect (Δv_S^p) and reweighting error (Δv_S^e). Both specification and reweighting errors reflect the overall model fitness. Researchers rely on the specification error to assess model quality and RIF approximation. Ideally, a minimal specification error close to zero is preferred. The reweighting error, indicating reweighting strategy quality, is expected to approach zero in large samples.

In this study, we implement this empirical strategy to understand the sources of the productivity gaps between SC businesses and non-SCST businesses, and between ST businesses and non-SCST businesses. As the model specifications include categorical variables, the choice of the omitted category is likely to influence the estimated effects. This is an issue most often highlighted in the literature on Oaxaca-Blinder decompositions (Fortin et al. 2011; Yun 2005), and even the RIF decomposition is liable to suffer from it. Several solutions are offered in the literature to address this issue. One among them is to normalize all categorical regressors using the averaging approach (Yun 2005). We follow this procedure in our study.[11] As discussed earlier, we group all covariates included in our model specifications into six different covariate sets to facilitate an interpretation of the results. These covariate sets are: (1) firm characteristics, (2) firm constraints, (3) gender, (4) period, (5) region, and (6) sector. The contribution of each covariate set is merely the sum of the contributions of the individual covariates included in the set.

4 Results

We present the results in four subsections. In the first three subsections, we decompose the gap in productivity between SC businesses and non-SCST businesses, and between ST businesses and non-SCST businesses, in the three main statistics of interest (average productivity, a set of selected percentiles, and inequality measures such as the Gini coefficient, interquartile ratios, and variance), as per equation 10, into composition and structural effects. For each statistic, we also conduct a detailed decomposition, breaking down the composition and structural effects into contributions attributable to each covariate. Subsection 4.1 focuses on average productivity and its decomposition; subsection 4.2 decomposes the productivity gap in selected percentiles; subsection 4.3 considers inequality measures. Lastly, we assess the sensitivity of our findings to alternative samples, regression specifications, and estimation without weights in subsection 4.4. on robustness tests.

4.1 Average productivity

Table 6 outlines the outcomes of the RIF decomposition for the caste gap in average labour productivity, breaking it down into characteristics and coefficient effects, both overall and for each covariate.[12] The results reveal notable disparities for both SC and ST businesses compared with non-SCST businesses. On average, SC-owned firms exhibit 30.84 per cent (0.3688 log points) lower labour productivity than non-SCST-owned firms (Table 6, column 1). The productivity gap is even more pronounced for ST businesses, with their average labour productivity being approximately half that of non-SCST businesses (Table 6, column 2). What contributes to the

[11] The crux of Yun's (2005) solution rests on obtaining the estimates for all possible reference groups and then computing the average.

[12] We have also performed the decomposition using the standard Oaxaca decomposition method. The results are presented in Table A2 in the Appendix. The results are similar to those obtained using the RIF decomposition method.

significant productivity advantage favouring non-SCST businesses? The aggregate and detailed decomposition results in Table 6 provide insights into the drivers of this productivity gap.[13]

Differences both in observed firm attributes and in the returns to these attributes contribute in the same direction to the observed gap in productivity between SC businesses and non-SCST businesses. The contributions of both differences are more or less of the same magnitude. The composition effect accounts for about 47 per cent of the gap in productivity, whereas 53 per cent of the gap originates through the structural effect. This means that non-SCST entrepreneurs benefit more from endowments than their SC counterparts and also exhibit a clear structural advantage in the returns to observable characteristics. It can be inferred from the results that convergence in returns to covariates would help in bridging the productivity gap between SC businesses and non-SCST businesses by 53 per cent. Similarly, achieving parity in endowments between SC businesses and non-SCST businesses would reduce the productivity gap by 47 per cent.

The difference in firm characteristics accounts for the bulk of the contribution of the composition effect to the productivity advantage of non-SCST businesses over SC businesses. About 70 per cent of the composition effect is attributable to differences in firm characteristics. Other significant contributors include the regional location of the firm and the gender of the entrepreneur, which contribute approximately 13 per cent each to the composition effect. In relative terms, differences in firm obstacles and time effects contribute only minimally to the explained gap in productivity. Although the structural effect accounts for more than half of the productivity gap, none of the covariates show significant effects. The size of the coefficients, however, indicates that the largest contribution to the structural effect originates from firm characteristics. The differences in returns to firm characteristics account for 61 per cent of the structural effect. The higher contribution from the intercept term emphasizes the role of omitted factors favouring non-SCST businesses via the structural effect.

What fuels the productivity gap between ST and non-SCST businesses? The decomposition results reveal that about 38 per cent of the mean productivity gap is explained by the composition effect. In terms of magnitude, a large part of the productivity differences is accounted for by the structural effect, whose coefficient, however, is found to be non-significant. Both composition and structural effects favour businesses owned by non-SCST entrepreneurs, suggesting that non-SCST firms possess superior endowments linked to higher productivity and consequently derive greater returns from these endowments compared with their ST counterparts. As seen from the detailed decompositions, the majority of the explained productivity gap emerges from the caste differences in firm attributes. We have already observed that ST businesses lag behind non-SCST businesses in all firm characteristics (Table A1), contributing significantly to the large productivity differential between ST and non-SCST firms. This contribution is so substantial that it alone explains 85 per cent of the explained gap in productivity. Turning to other covariates, we find little or insignificant contribution to the composition effect.

[13] Our estimations returned a non-significant specification and reweighting error, indicating that the model chosen was appropriate and the counterfactual was correctly identified.

Table 6: Productivity decomposition at the mean

	SC vs non-SCST		ST vs non-SCST	
	(1)		(2)	
Mean productivity (SC or ST)	9.5509***		9.2756***	
	(0.0379)		(0.0665)	
Mean non-SCST productivity	9.9196***		9.9196***	
	(0.0302)		(0.0309)	
Counterfactual	9.7469***		9.6765***	
	(0.0240)		(0.0247)	
Gap in productivity	0.3688***		0.6440***	
	(0.0288)		(0.0674)	
Composition effect				
Total	0.1727***	46.83	0.2431***	37.75
	(0.0315)		(0.0428)	
Firm characteristics	0.1193***	69.08	0.2055***	84.53
	(0.0103)		(0.0234)	
Firm constraints	0.0031**	1.80	0.0008	0.33
	(0.0013)		(0.0009)	
Gender of owner	0.0222***	12.85	0.0202***	8.31
	(0.0038)		(0.0074)	
Period	0.0062***	3.59	-0.0008	-0.33
	(0.0013)		(0.0032)	
Region	0.0229*	13.26	0.0202	8.31
	(0.0130)		(0.0268)	
Sector of activity	0.0022	1.27	0.0048**	1.97
	(0.0014)		(0.0023)	
Specification error	-0.0031	-1.80	-0.0075	-3.09
	(0.0355)		(0.0455)	
Structural effect				
Total	0.1961*	53.17	0.4009	62.25
	(0.1044)		(0.4341)	
Firm characteristics	0.1193	60.84	0.3497	87.23
	(0.2936)		(1.1826)	
Firm constraints	-0.0221	-11.27	-0.0027	-0.67
	(0.2265)		(0.5554)	
Gender of owner	0.0080	4.08	0.1015	25.32
	(0.0600)		(0.1172)	
Period	0.0040	2.04	0.0164	4.09
	(0.0245)		(0.1023)	
Region	0.0135	6.88	0.0587	14.64
	(0.0518)		(0.1941)	
Sector of activity	-0.0009	-0.46	0.0356	8.88
	(0.0027)		(0.0405)	
Constant	0.0749	38.19	-0.1634	-40.76
	(0.3843)		(1.3529)	
Reweighting error	-0.0007	-0.36	0.0050	1.25
	(0.0205)		(0.0292)	

Note: estimations are performed using non-SCST entrepreneurs' coefficients as reference. We normalized the category regressors using an averaging approach (Yun 2005). In all estimations, we used sample weights provided by NSSO. Robust standard errors (clustered by district) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations.

Shifting the focus to the unexplained gap in productivity, the results highlight that none of the observable factors significantly explain the structural disadvantage of ST firms compared with non-SCST firms. However, when we examine the size of the coefficients, the structural effect is largely driven by a substantial differential in returns to firm characteristics. The other covariates play a minor role in explaining the productivity gap via the structural effect. The gap that could be explained by differential returns to firm characteristics might have been partially offset by better returns to unobserved firm attributes for ST firms, as indicated by a large and negative intercept effect.

4.2 Productivity at selected percentiles

We now shift our focus to the decomposition of the caste gap in productivity across various points in the productivity distribution.[14] The aggregate decompositions are visually depicted in Figure 1.[15] The detailed decomposition results for specific percentiles are presented in Table 7 for the SC versus non-SCST comparison and in Table 8 for the ST versus non-SCST comparison.

We do not observe substantial variation in the productivity gap between SC and non-SCST businesses across the productivity distribution. At the median, SC businesses are 30 per cent (35.78 percentage points) less productive than non-SCST businesses. The gap ranges between 29.8 per cent and 32.2 per cent at the lower end, and between 28.7 per cent and 30.3 per cent at the upper end. Clearly, even when we consider firm heterogeneity along the distribution, SC businesses seem to be consistently disadvantaged. The decomposition results suggest an almost identical contribution from composition and structural effects to the productivity gap along the productivity distribution.[16] In terms of magnitude, the structural effect is marginally dominant at the two ends of the productivity distribution, and the composition effect in the middle parts. These results underscore the importance of addressing differences in both the levels and returns to firm characteristics between SC and non-SCST businesses to reduce the productivity gap between them.

The detailed decomposition results reveal that the factors identified as crucial in the mean decomposition are mostly relevant across all deciles (Table 7).[17] Differences in firm characteristics play a major role in the contribution of the composition effect to the productivity gap at all deciles of the productivity distribution. The coefficient is positive and significant, indicating the advantage of non-SCST businesses in firm characteristics over SC businesses throughout the productivity distribution. Nevertheless, the coefficient does not vary monotonically. At the very bottom, it takes the minimum value of 9.01 percentage points. It then rises steeply to the 40th percentile, where it reaches the maximum value of 13.62 percentage points, dropping two percentage points thereafter. Regional effects, an important contributor to the composition effect in the mean decomposition, are non-significant at the upper and lower deciles. Regarding the gender of owners, its effect is small but significant across the entire distribution. However, the role of gender in the composition effect declines as we move up the distribution. The effects of firm constraints and the time dummy are significant in most if not all deciles, but their effect is negligible throughout.

---

[14] The RIF regression estimates for median productivity are reported in Table A3.

[15] We summarize the specification and reweighting errors for the SC-non-SCST and ST-non-SCST comparisons in Figure A3. The specification error displayed in the figure is small in both sets of comparisons, signifying that the model is well specified.

[16] The structural effect is, however, significant at the 60th, 80th, and 90th percentiles only.

[17] See Figure A4 for the share of covariates in the contribution of the composition effect to the productivity gap at each decile.

Table 7: Productivity decompositions at selected percentiles: SC versus non-SCST firms

	p10	p20	p30	p40	p50	p60	p70	p80	p90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean SC productivity	8.1670***	8.7046***	9.1206***	9.4458***	9.7058***	9.9465***	10.1937***	10.4329***	10.7632***
	(0.0541)	(0.0709)	(0.0502)	(0.0413)	(0.0401)	(0.0359)	(0.0364)	(0.0308)	(0.0352)
Mean non-SCST productivity	8.5249***	9.0936***	9.4889***	9.7998***	10.0637***	10.2924***	10.5325***	10.7795***	11.1241***
	(0.0460)	(0.0428)	(0.0381)	(0.0335)	(0.0293)	(0.0283)	(0.0272)	(0.0261)	(0.0276)
Counterfactual	8.3498***	8.9059***	9.3068***	9.6168***	9.8820***	10.1236***	10.3578***	10.6240***	10.9509***
	(0.0324)	(0.0347)	(0.0339)	(0.0285)	(0.0254)	(0.0216)	(0.0207)	(0.0182)	(0.0174)
RIF productivity gap	0.3579***	0.3890***	0.3683***	0.3540***	0.3578***	0.3459***	0.3387***	0.3466***	0.3609***
	(0.0470)	(0.0588)	(0.0423)	(0.0347)	(0.0332)	(0.0295)	(0.0300)	(0.0236)	(0.0243)
Composition effect									
Total	0.1751***	0.1877***	0.1821***	0.1830***	0.1817***	0.1688***	0.1746***	0.1555***	0.1732***
	(0.0519)	(0.0429)	(0.0361)	(0.0314)	(0.0292)	(0.0308)	(0.0313)	(0.0327)	(0.0355)
Firm characteristics	0.0901***	0.1180***	0.1315***	0.1362***	0.1305***	0.1282***	0.1228***	0.1158***	0.1154***
	(0.0154)	(0.0134)	(0.0109)	(0.0099)	(0.0093)	(0.0108)	(0.0116)	(0.0125)	(0.0143)
Firm constraints	0.0033	0.0034	0.0041**	0.0031*	0.0032**	0.0031**	0.0027**	0.0021**	0.0020
	(0.0026)	(0.0025)	(0.0020)	(0.0016)	(0.0013)	(0.0012)	(0.0012)	(0.0011)	(0.0013)
Gender of owner	0.0383***	0.0382***	0.0323***	0.0261***	0.0198***	0.0160***	0.0126***	0.0096***	0.0073***
	(0.0077)	(0.0071)	(0.0056)	(0.0045)	(0.0034)	(0.0028)	(0.0023)	(0.0019)	(0.0019)
Period	0.0065***	0.0062***	0.0070***	0.0069***	0.0068***	0.0064***	0.0062***	0.0052***	0.0049***
	(0.0018)	(0.0015)	(0.0016)	(0.0015)	(0.0014)	(0.0013)	(0.0013)	(0.0011)	(0.0011)
Region	0.0460	0.0314	0.0219	0.0210*	0.0209*	0.0223**	0.0209**	0.0171*	0.0167
	(0.0282)	(0.0204)	(0.0163)	(0.0126)	(0.0110)	(0.0108)	(0.0102)	(0.0101)	(0.0109)
Sector of activity	0.0010	-0.0005	-0.0003	0.0008	0.0011	0.0014	0.0025	0.0030	0.0062***
	(0.0026)	(0.0023)	(0.0021)	(0.0019)	(0.0017)	(0.0016)	(0.0015)	(0.0019)	(0.0018)
Specification error	-0.0102	-0.0091	-0.0144	-0.0111	-0.0007	-0.0085	0.0069	0.0025	0.0207
	(0.0729)	(0.0570)	(0.0467)	(0.0351)	(0.0294)	(0.0281)	(0.0262)	(0.0242)	(0.0242)
Structural effect									
Total	0.1828	0.2013	0.1862	0.1710	0.1762	0.1771*	0.1641	0.1912**	0.1878*
	(0.1672)	(0.2108)	(0.1420)	(0.1127)	(0.1129)	(0.1038)	(0.1109)	(0.0891)	(0.1062)
Firm characteristics	0.4405	0.3546	0.2775	0.1794	0.1420	0.0225	0.0265	0.0105	-0.1764
	(0.4944)	(0.6479)	(0.4173)	(0.3405)	(0.3279)	(0.3761)	(0.3430)	(0.3544)	(0.4660)
Firm constraints	0.0256	-0.0540	-0.0048	0.0361	0.0566	0.0398	0.0034	0.0144	-0.0133
	(0.3709)	(0.5065)	(0.2791)	(0.1926)	(0.1894)	(0.1737)	(0.1989)	(0.1742)	(0.1496)
Gender of owner	0.0966	-0.0176	0.0314	0.0243	-0.0022	-0.0011	-0.0094	-0.0082	-0.0113
	(0.1513)	(0.1413)	(0.1040)	(0.0712)	(0.0610)	(0.0429)	(0.0433)	(0.0414)	(0.0456)
Period	0.0261	-0.0091	0.0034	0.0044	-0.0002	-0.0022	-0.0082	0.0007	-0.0029
	(0.0466)	(0.0548)	(0.0374)	(0.0338)	(0.0308)	(0.0278)	(0.0269)	(0.0214)	(0.0256)
Region	-0.0369	-0.0121	-0.0070	0.0134	0.0152	0.0441	0.0327	0.0287	0.0134
	(0.0785)	(0.1144)	(0.0925)	(0.0721)	(0.0623)	(0.0702)	(0.0681)	(0.0562)	(0.0721)
Sector of activity	0.0001	-0.0027	-0.0011	-0.0022	-0.0026	-0.0013	-0.0009	0.0003	0.0006

	(0.0046)	(0.0057)	(0.0042)	(0.0042)	(0.0042)	(0.0036)	(0.0041)	(0.0031)	(0.0039)
Constant	-0.3691	-0.0576	-0.1130	-0.0840	-0.0321	0.0760	0.1210	0.1459	0.3789
	(0.6169)	(0.8677)	(0.5102)	(0.3967)	(0.3760)	(0.4186)	(0.4055)	(0.4004)	(0.4919)
Reweighting error	-0.0001	-0.0002	-0.0003	-0.0004	-0.0006	-0.0008	-0.0009	-0.0010	-0.0012
	(0.0274)	(0.0307)	(0.0302)	(0.0253)	(0.0221)	(0.0184)	(0.0171)	(0.0148)	(0.0141)

Note: estimations are performed using non-SCST entrepreneurs' coefficients as reference. We normalized the category regressors using an averaging approach (Yun 2005). In all estimations, we used sample weights provided by NSSO. Robust standard errors (clustered by district) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations.

Table 8: Productivity decompositions at selected percentiles: ST versus non-SCST firms

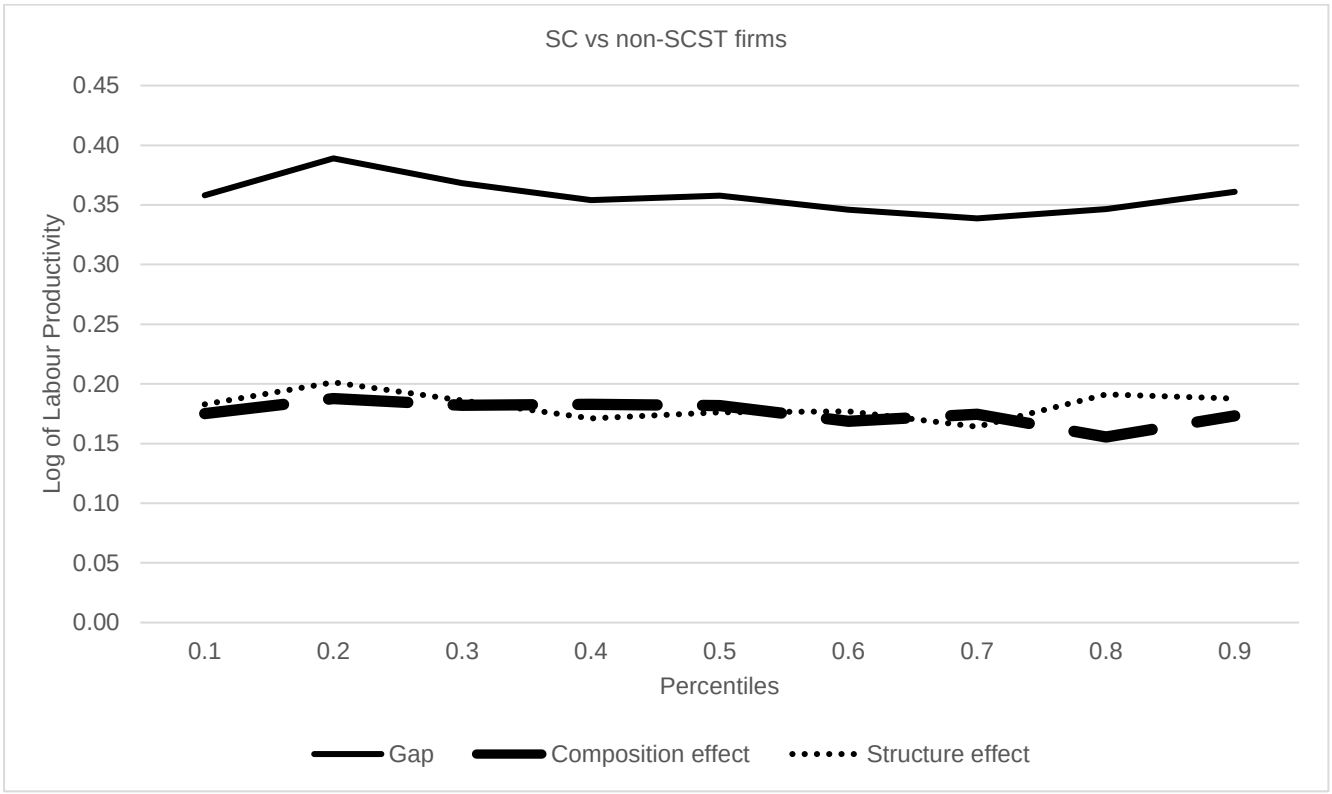
	p10	p20	p30	p40	p50	p60	p70	p80	p90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean ST productivity	7.7472***	8.2978***	8.7667***	9.1108***	9.4387***	9.7111***	10.0053***	10.3005***	10.6201***
	(0.1231)	(0.1070)	(0.0838)	(0.0753)	(0.0651)	(0.0611)	(0.0542)	(0.0479)	(0.0372)
Mean non-SCST productivity	8.5249***	9.0936***	9.4889***	9.7998***	10.0637***	10.2924***	10.5325***	10.7795***	11.1241***
	(0.0473)	(0.0437)	(0.0388)	(0.0341)	(0.0299)	(0.0289)	(0.0279)	(0.0269)	(0.0285)
Counterfactual	8.3187***	8.8792***	9.2541***	9.5493***	9.8059***	10.0289***	10.2577***	10.5137***	10.8408***
	(0.0431)	(0.0388)	(0.0397)	(0.0297)	(0.0255)	(0.0192)	(0.0160)	(0.0149)	(0.0133)
RIF productivity gap	0.7777***	0.7957***	0.7222***	0.6891***	0.6249***	0.5813***	0.5272***	0.4790***	0.5040***
	(0.1280)	(0.1053)	(0.0840)	(0.0755)	(0.0658)	(0.0623)	(0.0560)	(0.0502)	(0.0411)
Composition effect									
Total	0.2062***	0.2144***	0.2348***	0.2505***	0.2577***	0.2635***	0.2748***	0.2658***	0.2833***
	(0.0717)	(0.0589)	(0.0528)	(0.0442)	(0.0403)	(0.0397)	(0.0398)	(0.0407)	(0.0445)
Firm characteristics	0.1645***	0.2080***	0.2251***	0.2344***	0.2267***	0.2210***	0.2101***	0.1966***	0.1955***
	(0.0394)	(0.0327)	(0.0267)	(0.0227)	(0.0206)	(0.0217)	(0.0228)	(0.0259)	(0.0300)
Firm constraints	0.0028	0.0024*	0.0015	0.0008	0.0003	0.0001	-0.0000	-0.0001	-0.0004
	(0.0023)	(0.0014)	(0.0013)	(0.0010)	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0012)
Gender of owner	0.0348***	0.0347***	0.0293***	0.0237***	0.0180***	0.0145***	0.0115***	0.0087***	0.0066**
	(0.0134)	(0.0131)	(0.0108)	(0.0087)	(0.0067)	(0.0054)	(0.0043)	(0.0033)	(0.0028)
Period	-0.0008	-0.0008	-0.0009	-0.0009	-0.0009	-0.0008	-0.0008	-0.0007	-0.0006
	(0.0034)	(0.0033)	(0.0036)	(0.0036)	(0.0035)	(0.0033)	(0.0032)	(0.0027)	(0.0025)
Region	0.0020	-0.0076	-0.0037	0.0112	0.0308	0.0397	0.0449*	0.0411	0.0490*
	(0.0458)	(0.0379)	(0.0315)	(0.0298)	(0.0254)	(0.0252)	(0.0251)	(0.0270)	(0.0292)
Sector of activity	0.0109**	0.0105***	0.0091***	0.0063***	0.0046**	0.0038*	0.0018	0.0008	-0.0007

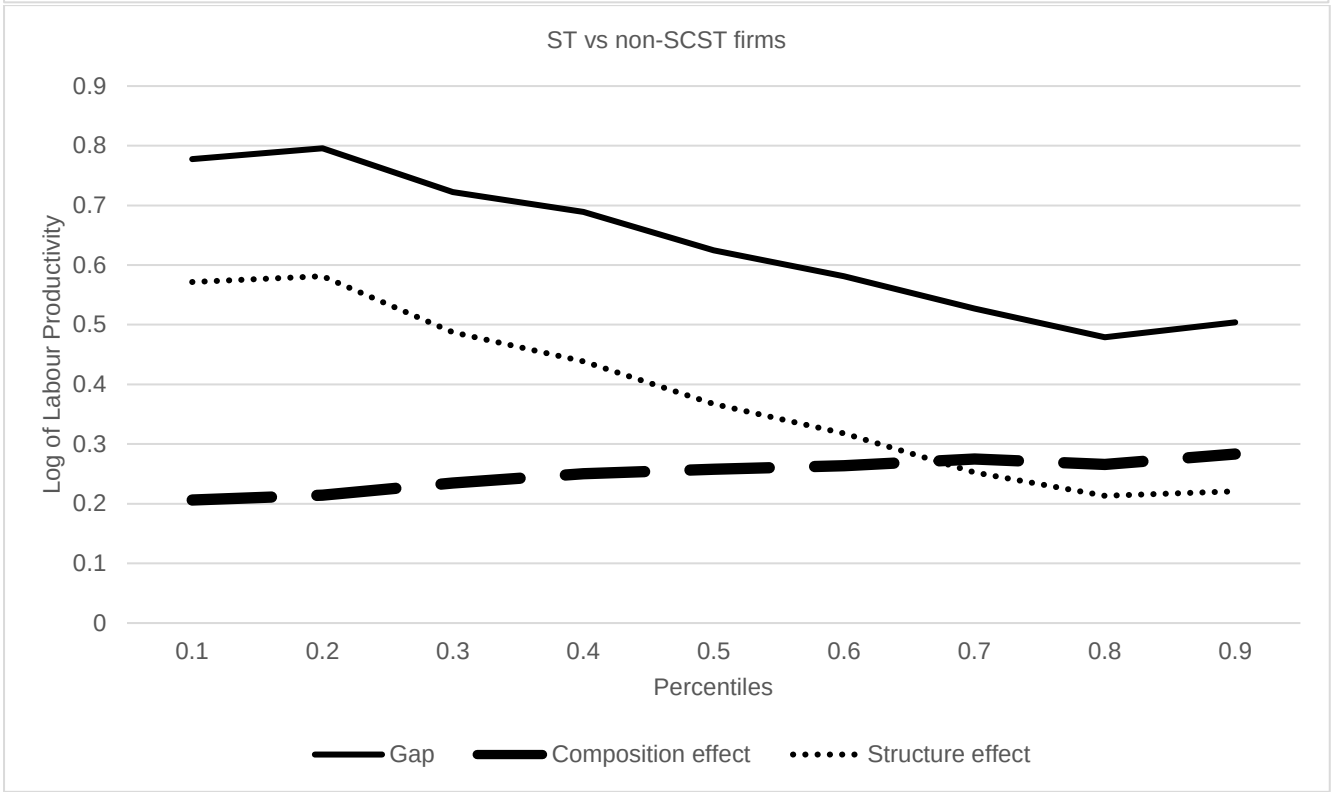
	(0.0053)	(0.0036)	(0.0025)	(0.0024)	(0.0022)	(0.0021)	(0.0021)	(0.0020)	(0.0027)
Specification error	-0.0080	-0.0329	-0.0257	-0.0250	-0.0219	-0.0148	0.0073	0.0193	0.0339
	(0.1051)	(0.0768)	(0.0607)	(0.0462)	(0.0363)	(0.0316)	(0.0274)	(0.0242)	(0.0244)
Structural effect									
Total	0.5715	0.5813	0.4874	0.4386	0.3672	0.3178	0.2524	0.2133	0.2207
	(1.1303)	(0.7889)	(0.5377)	(0.4536)	(0.3471)	(0.3056)	(0.2598)	(0.2277)	(0.2099)
Firm characteristics	1.2242	0.7980	0.4367	0.2696	0.1903	0.1617	0.1116	-0.0308	0.0157
	(3.3578)	(2.4059)	(1.4514)	(1.2941)	(1.1354)	(1.1213)	(1.1995)	(0.9919)	(1.2717)
Firm constraints	0.1202	0.0298	0.0645	-0.0711	-0.0276	-0.1002	-0.1298	-0.0835	-0.0495
	(1.5085)	(1.1611)	(0.7275)	(0.7555)	(0.5773)	(0.6065)	(0.5303)	(0.4702)	(0.5643)
Gender of owner	0.1728	0.1341	0.1406	0.0989	0.0879	0.0575	0.0449	0.0305	0.0427
	(0.4634)	(0.3557)	(0.2305)	(0.1897)	(0.1688)	(0.1557)	(0.1256)	(0.1042)	(0.1062)
Period	0.0441	0.0512	0.0168	-0.0047	-0.0063	-0.0086	-0.0043	-0.0011	0.0073
	(0.2697)	(0.1869)	(0.1495)	(0.1305)	(0.1048)	(0.0966)	(0.0749)	(0.0699)	(0.0628)
Region	0.0393	0.1056	0.0872	0.0855	0.0722	0.0647	0.0587	0.0472	0.0271
	(0.4900)	(0.3282)	(0.2465)	(0.2148)	(0.1608)	(0.1787)	(0.1568)	(0.1767)	(0.1747)
Sector of activity	0.0379	0.0400	0.0386	0.0452	0.0486	0.0425	0.0375	0.0309	0.0128
	(0.1148)	(0.0764)	(0.0636)	(0.0579)	(0.0495)	(0.0554)	(0.0466)	(0.0457)	(0.0377)
Constant	-1.0762	-0.5851	-0.3047	0.0093	-0.0031	0.0964	0.1311	0.2181	0.1634
	(3.3687)	(2.5248)	(1.6541)	(1.6546)	(1.3667)	(1.3279)	(1.5508)	(1.2429)	(1.4579)
Reweighting error	0.0091	0.0078	0.0078	0.0059	0.0051	0.0037	0.0028	0.0021	0.0013
	(0.0445)	(0.0404)	(0.0416)	(0.0336)	(0.0302)	(0.0255)	(0.0230)	(0.0223)	(0.0201)

Note: estimations are performed using non-SCST entrepreneurs' coefficients as reference. We normalized the category regressors using an averaging approach (Yun 2005). In all estimations, we used sample weights provided by NSSO. Robust standard errors (clustered by district) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations.

Figure 1: RIF aggregate decomposition





Note: we use the coefficients of non-SCST firms as reference in all decompositions. Specification and reweighting errors are not reported, but together they add up to the difference between the productivity gap and the sum of composition and structure effects. The specifications are the same as in Tables 3 and 4.

Source: authors' calculations.

The contribution of the structural effect to the productivity gap is positive along the entire distribution, but it is significant only at the 60th, 80th, and 90th percentiles. However, along the productivity distribution, none of the covariates make a statistically significant contribution to the structural effect. Yet, if we look at the size of the coefficient, from the 10th percentile to the median, the structural effect is mainly driven by higher returns of firm characteristics to non-SCST businesses. The differences in returns to characteristics shrink substantially at the upper end of the productivity distribution. The unidentified firm attributes, as captured by the large size of the intercept term, also contribute significantly to the structural effect, especially at the top deciles.

The decomposition analysis that compares ST businesses to non-SCST businesses produces some interesting results (Table 8). Firstly, we find that the gap in productivity varies significantly along the productivity distribution. At the same time, the productivity gap is positive and significant at every decile, indicating that ST businesses are less productive than non-SCST businesses along the entire distribution. Secondly, the gap is more pronounced at the lower end of the distribution, ranging between 50 per cent and 55 per cent—much higher than the gap observed at the mean (47 per cent). Thirdly, there is a discernible reduction in the gap as we ascend the productivity distribution: it decreases from 54 per cent at the 10th percentile level to 39 per cent at the 90th percentile level. This analysis highlights the crucial observation that productivity differences are more substantial among less productive ST businesses and comparatively minor among highly productive ST businesses.

Elucidating the productivity gap, Figure 1 illustrates that the structural effect takes precedence over the composition effect at the lower and middle tiers of the productivity distribution, while the inverse holds true at the top. As we traverse the productivity spectrum, the proportion of the gap explained by observable differences increases from 26.5 per cent at the 10th percentile to 56.2 per cent at the 90th percentile. On the other hand, the proportion of the gap explained by returns to observables witnesses a decline along the distribution. Based on these observations, we can say that the productivity disadvantage of ST businesses at the lower and middle deciles is due to lower returns to firm attributes, and at the upper deciles it is due to differences in observables. Alternatively, it might mean that among the more productive ST businesses, endowments (rather than the return to endowments) matter more.

The detailed decomposition results show that the sign and magnitude of the impact of observables across the deciles by and large mimic the results for mean decomposition (Table 8). The composition effect is primarily driven by the differences in characteristics of firms owned by ST and non-SCST entrepreneurs. Its contribution, however, does not vary monotonically across deciles. At the 10th percentile, firm characteristics account for approximately 80 per cent of the composition effect. It then increases to 96 per cent at the 30th percentile. We then see a consistent drop in its share to 69 per cent at the top decile. Still, it can be safely concluded that the bulk of the composition effect at any decile comes from firm characteristics. The gender of owners and the sector of activity are the other two important contributors to the composition effect, although both variables play a far less important role compared with firm characteristics. Of the two, the contribution of gender to the composition effect sees a consistent decline along the distribution. The structural effect, although positive and larger in terms of magnitude, is not statistically significant along the productivity distribution. None of the covariates seem to have a significant effect on the structural effect across deciles. Yet, if we merely focus on the magnitude of impact, we find that the structural effect at the lower and middle deciles is conditioned by firm characteristics, gender of owner, and regional location. At the upper deciles, the structural effect is almost entirely explained by unidentified firm attributes, as indicated by the large and positive intercept term.

4.3 Inequality measures: Gini coefficient and percentile ratios

A basic conclusion emerging from our previous analysis is that the productivity gap exhibits a different pattern along the productivity distribution in our two comparisons. For SC businesses, the productivity gap is more or less constant throughout the distribution, whereas for ST businesses, the productivity gap narrows down almost monotonically along the productivity distribution.

The RIF decomposition applied to five inequality measures confirms these findings for both comparisons. Results are shown in Table A4 in the Appendix. Column 1 presents the results for the difference in productivity between the 90th and 10th percentiles, a measure of the top-to-bottom gap, and columns 2 and 3 provide the estimates for the 50–10 and 90–50 differences, the gap at the lower and upper ends of the distribution respectively. The Gini values of labour productivity are presented in column 4. The variance of log labour productivity is used as another measure of inequality in column 5. The summary measure of inequality (the 90–10 gap) is nearly zero and insignificant in the comparison of SC businesses against non-SCST businesses, suggesting that the labour productivity gap remains largely stable along the productivity distribution. Further, the 90–50 and 50–10 measures, which are very small and statistically insignificant, reveal that the difference in productivity remains the same at both ends of the productivity distribution. The stability in the productivity gap along the distribution is also confirmed by other inequality measures, namely the Gini coefficient and the variance of logs.

The aggregate decomposition of inequality measures reveals consistent contributions from composition and structural effects across the productivity gap, showing minimal variation along the distribution. Delving into the composition effect, disparities in firm characteristics notably amplify productivity inequality at the lower end of the distribution (the 50–10 measure). Conversely, the gender of entrepreneurs contributes to a reduction in productivity inequality across the distribution, a trend confirmed by all inequality measures. Additionally, composition effects related to the sector of activity elucidate the ascending productivity differential, particularly in the top half of the distribution.

The comparison of ST businesses with non-SCST businesses presents a distinct scenario. The inequality measures are negative and statistically significant, suggesting that the productivity differences are not the same along the distribution. The overall measure of productivity inequality (the 90–10 gap) suggests that productivity differences are the lowest at the top end of the distribution. The estimates for the 90–50 and 50–10 measures indicate that the lower-end productivity differential (the 50–10 gap) constitutes 56 per cent of the overall productivity gap between ST and non-SCST businesses. Aggregate decompositions do not suggest a statistically significant contribution from composition and structural effects. However, when we look at the detailed decompositions of inequality measures, it can be seen that the owner's gender has a gap-narrowing effect across all inequality measures. The much larger and statistically significant coefficient of firm characteristics at the 50–10 measure suggests that it has a gap-widening effect on the composition effect. Caste disparities in sectoral choices notably contribute to the widening of the productivity gap, especially at the upper end of the productivity distribution.

4.4 Robustness tests

This subsection discusses the results of the robustness tests. We check whether the findings are sensitive to (1) changes in the sample (e.g., restricting it to male-run firms, or excluding younger firms), (2) decomposition without sample weights, (3) decomposition after dropping the possible endogenous variables, and (4) alternative measures of productivity. For reasons of space, we exclusively present the decomposition results at the mean. Table A5 in the Appendix displays the

outcomes of the initial three robustness tests separately for SC versus non-SCST businesses and ST versus non-SCST businesses. Column 1 involves re-estimating the RIF regression without using sampling weights. Column 2 provides the RIF regression estimates for male-run firms in the sample. Column 3 excludes younger firms from the sample.[18] In column 4, we truncate the specification by omitting potential endogenous variables.[19]

The results are to a large extent stable across specifications. Our key finding—that non-SCST businesses are more productive than SC businesses and ST businesses—survives in all four specifications in Table A5. The finding that ST businesses have a larger productivity gap vis-à-vis SC-owned firms is also consistent across specifications, except for re-estimation without weights. The evenly balanced role of the composition effect and structural effect in the productivity gap for SC versus non-SCST businesses is also validated by the robustness tests, with the exception of re-estimation without endogenous variables. The significant role of the composition effect in the ST businesses versus non-SCST businesses comparison is also confirmed by the robustness tests. Another important finding on the vital role of firm characteristics in driving the contribution of the composition effect to the productivity gap is robust across specifications for both comparisons.

We also conduct a robustness test where we replace labour productivity with total factor productivity. We compute total factor productivity as a residual from a production function.[20] The findings are essentially unchanged when a different measure of performance is used. The only change we notice is that the magnitude of the productivity gap between non-SCST businesses and SC and ST businesses is relatively small when we perform decompositions on total factor productivity. Results for the remaining variables are virtually the same as those in the baseline specification, where labour productivity is used as a measure of productivity.[21]

In the last robustness test, we perform the decomposition on a matched sample. The lack of common support is a problem in any decomposition procedure (Fortin et al. 2011). It can lead to biased results if the firms that are being compared lack comparable attributes (Nopo 2008). One solution is to create a more comparable category of SC and ST firms that bear a resemblance to firms in the non-SCST category and then perform the decomposition exercise. To see whether our results are biased due to the lack of common support, we create a matched sample using a non-parametric approach proposed by Nopo (2008).[22] We then employ the Oaxaca-Blinder decomposition procedure on the matched sample. Results are tabulated in Table A6. The covariates used to generate the matched sample are the same as those included in earlier decompositions. As is evident from the table, the results are largely consistent with those from the unmatched sample. Except for a marginal decline in the productivity gap, the magnitude and percentage contributions of aggregate components to the productivity differential and of covariates to these components are qualitatively similar.

---

[18] We dropped firms below the median age. Instead of categorical variables, we introduced logarithm of age in this specification.

[19] The trimmed model specification does not include the vector firm constraints and the variables firm size, assistance, registration, and linkage in the vector firm characteristics.

[20] Two alternative measures of total factor productivity are used. One is derived from real gross value-added, and another uses real gross output. For the former, labour and capital appear as inputs, and for the latter, labour, capital, and raw materials. The results are not tabulated here but are available from the authors on request.

[21] We do not present the results here, but they are available on request.

[22] This non-parametric matching procedure is implemented in Stata using the nopomatch command (Atal et al. 2010).

5 Conclusions and policy implications

The stark under-representation of marginalized groups in both business ownership and the labour market is extensively documented. In light of this, a critical empirical question arises: how does the enduring disparity in entrepreneurial activity across caste categories impact on economic outcomes? This paper rigorously addresses this question, delving into the impact of caste disparities in small-firm entrepreneurship on firm performance. Concentrating on businesses within the informal sector in India, we scrutinize the presence of a substantial productivity gap between firms owned by marginalized groups (SCs and STs) and others. Furthermore, we meticulously identify the factors driving this productivity gap.

The results reveal pronounced productivity gaps for both SC and ST businesses compared with businesses owned by OBCs and Forward Castes. Particularly striking is the severe productivity disadvantage for ST-owned firms, where productivity is only half that of non-SCST-owned firms. In the case of SC-owned firms, the productivity disadvantage compared with non-SCST firms persists across almost the entire productivity distribution. For ST entrepreneurs, the productivity differences exhibit a decreasing pattern along the distribution, with the productivity gap being most acute at the bottom. Notably, a substantial productivity gap exists between firms owned by SC and ST entrepreneurs and others across the entire distribution. These findings affirm prior studies indicating that the owner's caste significantly influences occupation choice, firm performance, and earnings (Audretsch et al. 2013; Deshpande and Sharma 2016; Iyer et al. 2013; Thorat and Sadana 2009).

The substantial and persistent productivity gap prompts crucial questions regarding why it has not equalized and what underlies its persistence. Our decomposition analysis yields compelling evidence highlighting the significance of disparities in both observable factors and the returns to these factors in elucidating the caste gap in productivity. Non-SCST businesses derive greater benefits from their endowments compared with their SC and ST counterparts, revealing a distinct structural advantage in returns to observable characteristics. Notably, differences in coefficients and unobservable factors predominantly contribute to the productivity gap. This signifies that even with improvements in firm attributes for SC and ST businesses, the productivity disadvantage for firms owned by marginalized groups is likely to endure. Moreover, the Firpo et al. (2018) quantile decompositions, particularly for ST businesses, underscore that the penalty for belonging to a marginalized group is more conspicuous at lower levels of the productivity distribution, indicating that less productive firms face greater penalties in the Indian informal sector.

This study, which is at the forefront of examining the link between caste identity and firm performance in India's small-business sector, also has significance for the broader literature on racial and ethnic disparities in business performance, particularly in the West, such as in the United States and United Kingdom (Brown et al. 2022; Carter et al. 2015; Fairlie and Robb 2007). While evidence from these countries points to substantial strides in achieving equal racial representation in firm ownership (Perry et al. 2023), the contrasting scenario in India is noteworthy. The persistent under-representation of marginalized groups in business ownership, coupled with their confinement to non-entrepreneurial and survival-oriented sectors, underscores a divergent trend. The prevalent market and non-market discrimination against marginalized groups signals that significant caste disparities in entrepreneurship and business performance will persist, posing formidable challenges for the economy and public policy.

From a policy standpoint, two critical issues emerge. Firstly, our results underscore the indispensability of policies aimed at reducing identity-based disparities if inclusive growth is to be genuinely achieved. The primary hurdle for businesses owned by marginalized entrepreneurs is the

insufficient availability of working capital, yet the reach and effectiveness of credit support programmes remain limited. Addressing this necessitates the development of policy measures specifically targeting the diverse constraints faced by disadvantaged castes in product and credit markets. Secondly, our findings emphasize the urgency of confronting the pre-market discrimination experienced by marginalized groups, which contributes to lower-quality educational and skills attainment. This challenge is exacerbated by the reduction of various remedial protections, budget allocations, and programmes previously extended to SCs (Mosse 2020). Consequently, policies that address their social and educational challenges, facilitating their catch-up with the rest of the population, are imperative.

The study has noteworthy limitations that offer avenues for future research. Firstly, the focus on informal-sector firms, which are largely driven by supplementary income goals rather than growth motives, may result in an underestimated productivity gap, not fully reflecting the actual disparity. Secondly, the decomposition exercise, while shedding light on potential driving factors, does not establish causality between these factors and outcomes. Thirdly, the intricate interplay of measurable and unobservable factors, including discrimination (indicated by a substantial intercept term), remains challenging to disentangle. Critical factors such as family background and entrepreneurs' risk-taking ability, which are known determinants of firm performance, lack representation in our data. Additionally, immeasurable owner attributes such as ability or motivation, which are acknowledged influencers of firm performance, are not captured. The role of discrimination, a potential constraint on the performance of SC and ST businesses, adds another layer of complexity.

References

Agrawal, T. (2014). 'Gender and Caste-Based Wage Discrimination in India: Some Recent Evidence'. *Journal of Labour Market Research*, 47(4): 329–40. https://doi.org/10.1007/s12651-013-0152-z

Arabsheibani, G.R., P. Gupta, T. Mishra, and M. Parhi (2018). 'Wage Differential Between Caste Groups: Are Younger and Older Cohorts Different?' *LSE Research Online*. Available at: http://eprints.lse.ac.uk/90510/1/Arabsheibani__wage-differential.pdf (accessed 23 January 2024).

Atal, J.P., A. Hoyos, and H. Nopo (2010). 'NOPOMATCH: Stata Module to Implement Nopo's Decomposition'. Statistical Software Components S457157. Boston, MA: Boston College Department of Economics. Available at: https://ideas.repec.org/c/boc/bocode/s457157.html (accessed 23 January 2024).

Audretsch, D.B., W. Bonté, and J.P. Tamvada (2013). 'Religion, Social Class, and Entrepreneurial Choice'. *Journal of Business Venturing*, 28(6): 774–89. https://doi.org/10.1016/j.jbusvent.2013.06.002

Banerji, S., R.S.N. Raj, and K. Sen (2016). 'Monitoring Costs, Credit Constraints and Entrepreneurship'. *Manchester School*, 84(5): 573–99. https://doi.org/10.1111/manc.12122

Borowczyk-Martins, D., J. Bradley, and L. Tarasonis (2017). 'Racial Discrimination in the U.S. Labor Market: Employment and Wage Differentials by Skill'. *Labour Economics*, 49: 106–27. https://doi.org/10.1016/j.labeco.2017.09.007

Brown, J., J. Earle, M. Kim, K. Lee, and J. Wold (2022). 'Black-Owned Firms, Financial Constraints, and the Firm Size Gap'. *American Economic Association Papers & Proceedings*, 112: 282–86. https://doi.org/10.1257/pandp.20221027

Carter, S., S. Mwaura, M. Ram, K. Trehan, and T. Jones (2015). 'Barriers to Ethnic Minority and Women's Enterprise: Existing Evidence, Policy Tensions and Unsettled Questions'. *International Small Business Journal*, 33(1): 49–69. https://doi.org/10.1177/0266242614556823

Chatterjee, S., T. Lebesmuehlbacher, and A. Narayanan (2021). 'How Productive Is Public Investment? Evidence from Formal and Informal Production in India'. *Journal of Development Economics*, 151: 102625. https://doi.org/10.1016/j.jdeveco.2021.102625

Dahl, M., and N. Krog (2018). 'Experimental Evidence of Discrimination in the Labour Market: Intersections Between Ethnicity, Gender, and Socio-Economic Status'. *European Sociological Review*, 34(4): 402–17. https://doi.org/10.1093/esr/jcy020

Deshpande, A., and S. Sharma (2013). 'Entrepreneurship or Survival? Caste and Gender of Small Business in India'. *Economic and Political Weekly*, 28(28): 38–49.

Deshpande, A., and S. Sharma (2016). 'Disadvantage and Discrimination in Self-Employment: Caste Gaps in Earnings in Indian Small Businesses'. *Small Business Economics*, 46: 325–46. https://doi.org/10.1007/s11187-015-9687-4

DiNardo, J., N.M. Fortin, and T. Lemieux (1996). 'Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach'. *Econometrica*, 64(5): 1001–44. https://doi.org/10.2307/2171954

Fairlie, R., and A. Robb (2007). 'Why Are Black-Owned Businesses Less Successful Than White-Owned Businesses? The Role of Families, Inheritances, and Business Human Capital'. *Journal of Labour Economics*, 25(2): 289–323. https://doi.org/10.1086/510763

Firpo, S., N. Fortin, and T. Lemieux (2018). 'Decomposing Wage Distributions Using Recentered Influence Function Regressions'. *Econometrics*, 6(2): 28. https://doi.org/10.3390/econometrics6020028

Fortin, N., T. Lemieux, and S. Firpo (2011). 'Decomposition Methods in Economics'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics, Volume 4A*. Amsterdam: Elsevier. https://doi.org/10.1016/S0169-7218(11)00407-2

Gang, I.N., R.R. Natarajan, K. Sen, and M.S. Yun (2022). 'Does the Gender of the Owner Affect the Productivity of Enterprises in India's Informal Economy?' *Journal of Income Distribution*, 31(3–4). https://doi.org/10.25071/1874-6322.40548

Gang, I.N., R.S.N. Raj, and K. Sen (2022). 'Finance, Gender, and Entrepreneurship: India's Informal Sector Firms'. *Journal of Development Studies*, 58(7): 1383–402. https://doi.org/10.1080/00220388.2022.2061855

Gang, I.N., K. Sen, and M.S. Yun (2017). 'Is Caste Destiny? Occupational Diversification Among Dalits in Rural India'. *European Journal of Development Research*, 29(2): 476–92. https://doi.org/10.1057/s41287-016-0011-1

Ghani, E., W.R. Kerr, and S.D. O'Connell (2014a). 'Political Reservations and Women's Entrepreneurship in India'. *Journal of Development Economics*, 108: 138–53. https://doi.org/10.1016/j.jdeveco.2014.01.008

Ghani, E., W.R. Kerr, and S.D. O'Connell (2014b). 'Spatial Determinants of Entrepreneurship in India'. *Regional Studies*, 48(6): 1071–89. https://doi.org/10.1080/00343404.2013.839869

Halvorsen, R., and R. Palmquist (1980). 'The Interpretation of Dummy Variables in Semilogarithmic Equations'. *American Economic Review*, 70(3): 474–75.

Harriss-White, B., and K. Vidyarthee (2010). 'Stigma and Regions of Accumulation: Mapping Dalit and Adivasi Capital in the 1990s'. In B. Harriss-White and J. Heyer (eds), *The Comparative Political Economy of Development: Africa and South Asia*. Abingdon: Routledge.

Harriss-White, B., K. Vidyarthee, and A. Dixit (2014). 'Regions of Dalit and Adivasi Discrimination in India's Business Economy'. In B. Harriss-White with others (eds), *Dalits and Adivasis in India's Business Economy: Three Essays and an Atlas*. Gurgaon: Three Essays Collective.

Heath, A.F., and V. Di Stasio (2019). 'Racial Discrimination in Britain, 1969–2017: A Meta-Analysis of Field Experiments on Racial Discrimination in the British Labour Market'. *British Journal of Sociology*, 70(5): 1774–98. https://doi.org/10.1111/1468-4446.12676

Hsieh, C., and P.J. Klenow (2014). 'The Life Cycle of Plants in India and Mexico'. *Quarterly Journal of Economics*, 129(3): 1035–84. https://doi.org/10.1093/qje/qju014

Iyer, L., T. Khanna, and A. Varshney (2013). 'Caste and Entrepreneurship in India'. *Economic and Political Weekly*, 48(6): 52–60.

Jodhka, S. (2010). 'Dalits in Business: Self-Employed Scheduled Castes in North-West India'. *Economic and Political Weekly*, 45(11): 41–48.

Kathuria, V., R.S.N. Raj, and K. Sen (2013). 'The Effects of Economic Reforms on Manufacturing Dualism: Evidence from India'. *Journal of Comparative Economics*, 41(4): 1240–62. https://doi.org/10.1016/j.jce.2012.10.003

Kesar, S., and S. Bhattacharya (2019). 'Dualism and Structural Transformation: The Informal Manufacturing Sector in India'. *European Journal of Development Research*, 50: 560–82. https://doi.org/10.1057/s41287-019-00228-0

Kurian, N.J. (2007). 'Widening Economic and Social Disparities: Implications for India'. *Indian Journal of Medical Research*, 126(4): 374–80.

Lang, K., and J.K. Lehmann (2012). 'Racial Discrimination in the Labor Market: Theory and Empirics'. *Journal of Economic Literature*, 50(4): 1–48. https://doi.org/10.1257/jel.50.4.959

Lang, K., and A.K. Spitzer (2020). 'Race Discrimination: An Economic Perspective'. *Journal of Economic Perspectives*, 34(2): 68–89. https://doi.org/10.1257/jep.34.2.68

Mazumdar, D., and S. Sarkar (2013). *Manufacturing Enterprises in Asia: Size Structure and Economic Growth*. London: Routledge.

Mosse, D. (2018). 'Caste and Development: Contemporary Perspectives on a Structure of Discrimination and Advantage'. *World Development*, 110: 422–36. https://doi.org/10.1016/j.worlddev.2018.06.003

Mosse, D. (2020). 'The Modernity of Caste and the Market Economy'. *Modern Asian Studies*, 54(4): 1225–71. https://doi.org/10.1017/S0026749X19000039

Nopo, H. (2008). 'Matching as a Tool to Decompose Wage Gaps'. *Review of Economics and Statistics*, 90(2): 290–99. https://doi.org/10.1162/rest.90.2.290

NSSO (2012). 'Key Results of Survey on Unincorporated Non-Agricultural Enterprises (Excluding Construction) in India, NSS KI (67/2.34), NSS 67th Round (July 2010 – June 2011)'. New Delhi: Government of India, National Sample Survey Office.

NSSO (2017). 'Key Indicators of Unincorporated Non-Agricultural Enterprises (Excluding Construction) in India, NSS KI (73/2.34), NSS 73rd Round (July 2015 – June 2016)'. New Delhi: Government of India, National Sample Survey Office.

Omvedt, G. (2017). 'Dalits and Economic Policy: Contributions of Dr B.R. Ambedkar'. In S.M. Michael (ed.), *Dalits in Modern India*. Los Angeles: Sage.

Perry, A.M., M. Donoghoe, and H. Stephens (2023). 'Who Is Driving Black Business Growth? Insights from the Latest Data on Black-Owned Businesses'. *Brookings*. Available at: www.brookings.edu/articles/who-is-driving-black-business-growth-insights-from-the-latest-data-on-black-owned-businesses/ (accessed 23 January 2024).

Prakash, A. (2015). *Dalit Capital: State, Markets and Civil Society in Urban India*. New Delhi: Routledge.

Raj, R.S.N., and K. Sen (2015). 'Finance Constraints and Firm Transition in the Informal Sector: Evidence from Indian Manufacturing'. *Oxford Development Studies*, 43(1): 123–43. https://doi.org/10.1080/13600818.2014.972352

Raj, R.S.N., and K. Sen (2016). *Out of the Shadows? The Informal Manufacturing in Post-Reform India*. New Delhi: Oxford University Press.

Rios-Avila, F. (2020). 'Recentered Influence Functions (RIFs) in Stata: RIF Regression and RIF Decomposition'. *Stata Journal*, 20(1): 51–94. https://doi.org/10.1177/1536867X20909690

State of Working India (2023). 'Social Identities and Labour Market Outcomes'. Bengaluru: Azim Premji University, Centre for Sustainable Employment.

Thorat, S., Aryama, and P. Negi (2007). *Reservation and Private Sector: Quest for Equal Opportunity and Growth*. New Delhi: Rawat Publications.

Thorat, S., and S. Newman (2010). *Blocked by Caste: Economic Discrimination in Modern India*. New Delhi: Oxford University Press.

Thorat, S,. and N. Sadana (2009). 'Caste and Ownership of Private Enterprises'. *Economic and Political Weekly*, 44(23): 13–16.

Yun, M.S. (2005). 'A Simple Solution to the Identification Problem in Detailed Wage Decompositions'. *Economic Inquiry*, 43: 766–72. https://doi.org/10.1093/ei/cbi053

Appendix

Table A1: Summary statistics

	All enterprises		SC enterprises		ST enterprises		Non-SCST enterprises	
	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.
Log of labour productivity	9.846	1.082	9.551	1.072	9.276	1.198	9.920	1.061
Log of employment	0.348	0.520	0.244	0.420	0.338	0.453	0.365	0.535
Location	0.483	0.500	0.374	0.484	0.168	0.374	0.515	0.500
Gender of owner	0.197	0.398	0.221	0.415	0.222	0.415	0.192	0.394
Age of firm								
Age of firm, below 2 years	0.121	0.326	0.124	0.330	0.096	0.295	0.122	0.327
Age of firm, 3-9 years	0.446	0.497	0.413	0.492	0.445	0.497	0.451	0.498
Age of firm, above 9 years	0.433	0.496	0.463	0.499	0.458	0.498	0.427	0.495
Assistance from government?	0.008	0.090	0.006	0.079	0.009	0.095	0.008	0.091
Registered under act/authority?	0.290	0.454	0.183	0.386	0.169	0.375	0.312	0.463
Undertake work on contract basis?	0.091	0.288	0.089	0.285	0.059	0.235	0.093	0.290
Accounts maintained?	0.092	0.289	0.043	0.203	0.039	0.194	0.102	0.303
Faced any power constraint?	0.036	0.187	0.027	0.163	0.026	0.159	0.038	0.191
Any borrowing constraint?	0.082	0.274	0.104	0.305	0.071	0.257	0.079	0.269
Region								
North	0.228	0.419	0.267	0.443	0.041	0.199	0.231	0.421
West	0.179	0.383	0.135	0.342	0.193	0.395	0.185	0.388
East	0.244	0.429	0.341	0.474	0.326	0.469	0.224	0.417
South	0.261	0.439	0.168	0.374	0.150	0.358	0.281	0.450
Central	0.057	0.232	0.061	0.238	0.137	0.344	0.053	0.224
North-east	0.031	0.173	0.028	0.164	0.152	0.359	0.026	0.158

Activity								
Manufacturing	0.315	0.464	0.326	0.469	0.349	0.477	0.311	0.463
Trade	0.369	0.483	0.321	0.467	0.402	0.490	0.375	0.484
Services	0.316	0.465	0.353	0.478	0.250	0.433	0.314	0.464
Number of firms	583,962		58,428		32,543		492,991	

Note: estimates are calculated using sample weights provided by NSSO.

Source: authors' calculations.

Table A2: Oaxaca-Blinder decomposition of productivity gaps

	SC vs non-SCST firms		ST vs non-SCST firms	
	(1)		(2)	
Total gap	0.3688***		0.6440***	
	(0.0261)		(0.0753)	
Composition effect	0.1750***	47.45	0.2552***	39.63
	(0.0219)		(0.0365)	
Structure effect	0.1938***	52.55	0.3888***	60.37
	(0.0179)		(0.0640)	
Contributions to the composition effect by component				
Firm characteristics	0.1182***	32.05	0.2058***	31.96
	(0.0120)		(0.0206)	
Firm constraints	0.0031***	0.84	0.0008	0.12
	(0.0009)		(0.0011)	
Gender of owner	0.0225**	6.10	0.0232*	3.60
	(0.0104)		(0.0140)	
Period	0.0061	1.65	-0.0010	-0.16
	(0.0040)		(0.0077)	
Region	0.0228***	6.18	0.0216	3.35
	(0.0083)		(0.0174)	
Sector of activity	0.0023	0.62	0.0048*	0.75
	(0.0017)		(0.0028)	

Note: estimations are performed using non-SCST entrepreneurs' coefficients as reference. The parentheses next to the estimated output report the percentage contribution to the total gap. In all estimations, we used sample weights provided by NSSO. Standard errors (clustered at district level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations.

Table A3: Median RIF regression coefficients on labour productivity

Variables	Labour productivity		
	ST	SC	Non-SCST
	(1)	(2)	(3)
Log of employment	-0.6551***	-0.5120***	-0.1877***
	(0.0623)	(0.0360)	(0.0195)
Location	0.4357***	0.4279***	0.4472***
	(0.0680)	(0.0425)	(0.0299)
Gender of owner	-0.4338***	-0.7900***	-0.6986***
	(0.0772)	(0.0458)	(0.0227)
Age of firm, 3-9 years	0.5263***	0.3799***	0.3630***
	(0.0783)	(0.0431)	(0.0251)
Age of firm, above 9 years	0.3597***	0.3429***	0.3159***
	(0.0816)	(0.0442)	(0.0232)
Assistance	0.1279	0.2113**	0.1025*
	(0.1382)	(0.0961)	(0.0555)
Registered	0.5911***	0.5155***	0.4988***
	(0.0798)	(0.0462)	(0.0201)
Work on contract	-0.2652	-0.2503***	-0.3231***
	(0.2250)	(0.0666)	(0.0371)
Account maintenance	0.4370***	0.3775***	0.3875***
	(0.0985)	(0.1061)	(0.0197)
Power constraint	0.2168*	0.2592***	0.0791**
	(0.1122)	(0.0684)	(0.0312)
Financial constraint	-0.1595	-0.0742	-0.0924***
	(0.1260)	(0.0703)	(0.0285)
Year (2015-16)	0.4088***	0.4078***	0.3938***
	(0.0691)	(0.0379)	(0.0259)
West	-0.1338	0.0841	0.1064
	(0.1143)	(0.0784)	(0.0647)
East	-0.5222***	-0.0495	-0.0798
	(0.1377)	(0.0745)	(0.0657)
South	-0.2147	0.1151	0.0342
	(0.1601)	(0.0747)	(0.0629)
Central	-0.6834***	-0.3421***	-0.3137***
	(0.1908)	(0.0803)	(0.0628)
North-east	0.4616***	0.3343***	0.0490
	(0.1132)	(0.0843)	(0.0698)

Trade	0.1524*	0.2048***	0.1067***
	(0.0824)	(0.0521)	(0.0244)
Services	0.7337***	0.2655***	0.1167***
	(0.0907)	(0.0548)	(0.0217)
Constant	8.9108***	9.0046***	9.2309***
	(0.1348)	(0.0770)	(0.0542)
Observations	32,543	58,428	492,991
R^2	0.3207	0.2624	0.2663

Note: the data and set of covariates are the same as in Table 3. In all estimations, we used sample weights provided by NSSO. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations.

Table A4: RIF decomposition for selected inequality measures

Variables	SC vs non-SCST firms					ST vs non-SCST firms				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	iqr9010	iqr5010	iqr9050	Gini	Variance	iqr9010	iqr5010	iqr9050	Gini	Variance
Mean productivity (SC or ST)	2.5992***	1.5388***	1.0604***	0.4790***	1.1250***	2.5992***	1.5388***	1.0604***	0.4790***	1.1250***
	(0.0403)	(0.0304)	(0.0169)	(0.0051)	(0.0315)	(0.0419)	(0.0318)	(0.0175)	(0.0053)	(0.0329)
Mean non-SCST productivity	2.5962***	1.5388***	1.0573***	0.4683***	1.1502***	2.8730***	1.6916***	1.1813***	0.5242***	1.4359***
	(0.0522)	(0.0420)	(0.0303)	(0.0075)	(0.0438)	(0.0983)	(0.0790)	(0.0448)	(0.0108)	(0.1076)
Total gap	0.0031	-0.0000	0.0031	0.0108	-0.0252	-0.2737***	-0.1528*	-0.1209***	-0.0452***	-0.3109***
	(0.0463)	(0.0391)	(0.0286)	(0.0066)	(0.0368)	(0.1049)	(0.0848)	(0.0459)	(0.0111)	(0.1107)
Composition effect	-0.0019	0.0066	-0.0085	0.0006	0.0044	0.0772	0.0516	0.0256	0.0125	0.0820
	(0.0535)	(0.0428)	(0.0225)	(0.0066)	(0.0435)	(0.0722)	(0.0544)	(0.0318)	(0.0093)	(0.0584)
Structure effect	0.0050	-0.0066	0.0116	0.0102	-0.0296	-0.3509	-0.2044	-0.1465	-0.0577	-0.3929
	(0.1713)	(0.1414)	(0.1013)	(0.0242)	(0.1428)	(0.9795)	(0.8050)	(0.3539)	(0.0943)	(1.0599)
Contribution to the composition effect by components										
Firm characteristics	0.0254	0.0404***	-0.0151	0.0011	0.0181	0.0310	0.0622*	-0.0312	-0.0020	0.0227
	(0.0170)	(0.0131)	(0.0099)	(0.0029)	(0.0127)	(0.0433)	(0.0326)	(0.0217)	(0.0060)	(0.0348)
Firm constraints	-0.0013	-0.0001	-0.0011	-0.0003	-0.0027	-0.0032	-0.0025	-0.0007	-0.0004*	-0.0028**
	(0.0029)	(0.0026)	(0.0016)	(0.0003)	(0.0029)	(0.0025)	(0.0022)	(0.0012)	(0.0002)	(0.0014)
Gender of owner	-0.0310***	-0.0185***	-0.0125***	-0.0043***	-0.0244***	-0.0282**	-0.0168**	-0.0114**	-0.0039***	-0.0222**
	(0.0071)	(0.0055)	(0.0028)	(0.0008)	(0.0053)	(0.0114)	(0.0077)	(0.0046)	(0.0015)	(0.0088)
Period	-0.0017	0.0002	-0.0019**	-0.0004*	-0.0016	0.0002	-0.0000	0.0002	0.0001	0.0002
	(0.0016)	(0.0012)	(0.0009)	(0.0002)	(0.0013)	(0.0009)	(0.0002)	(0.0010)	(0.0002)	(0.0009)
Region	-0.0293	-0.0251	-0.0043	-0.0028	-0.0133	0.0469	0.0288	0.0181	0.0077	0.0535
	(0.0281)	(0.0239)	(0.0087)	(0.0030)	(0.0224)	(0.0469)	(0.0365)	(0.0213)	(0.0058)	(0.0370)
Sector of activity	0.0051*	0.0001	0.0051**	0.0018***	0.0039	-0.0115**	-0.0062	-0.0053*	-0.0019***	-0.0076*
	(0.0030)	(0.0025)	(0.0023)	(0.0006)	(0.0028)	(0.0054)	(0.0049)	(0.0027)	(0.0006)	(0.0044)

Note**:** iqr9010 is the difference between the 90th percentile and the 10th percentile of log labour productivity. The series iqr5010 and iqr9050 are computed analogously. The Gini coefficient is expressed in percentage points and ranges from 0 (perfect equality) to 100 (perfect inequality). Variance captures the impact of covariates on the variance of the distributions of log labour productivity. We normalized the category regressors using an averaging approach (Yun 2005). In all the estimations, we used weights supplied by NSSO. Standard errors (clustered at district level) in parentheses. *** p<0.01, ** p<0.05, *p<0.10.

Source: authors' calculations.

Table A5: RIF regression decomposition of the productivity gap at the mean, robustness tests

	SC vs Non-SCST				ST vs Non-SCST			
	Without weights	Male-run firms	Old firms	Truncated specification	Without weights	Male-run firms	Old firms	Truncated specification
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean productivity (SC or ST)	9.6843***	9.7622***	9.6121***	9.5509***	9.8021***	9.3553***	9.1577***	9.2756***
	(0.0217)	(0.0351)	(0.0411)	(0.0389)	(0.0442)	(0.0747)	(0.0883)	(0.0676)
Mean non-SCST productivity	10.0741***	10.1208***	9.9895***	9.9196***	10.0741***	10.1208***	9.9895***	9.9196***
	(0.0218)	(0.0256)	(0.0329)	(0.0300)	(0.0218)	(0.0263)	(0.0334)	(0.0308)
Counterfactual	9.8899***	9.9704***	9.8016***	9.7762***	9.8937***	9.8537***	9.6590***	9.7155***
	(0.0103)	(0.0147)	(0.0278)	(0.0285)	(0.0199)	(0.0161)	(0.0320)	(0.0313)
RIF productivity gap	0.3898***	0.3586***	0.3774***	0.3688***	0.2720***	0.7655***	0.8319***	0.6440***
	(0.0133)	(0.0257)	(0.0367)	(0.0292)	(0.0466)	(0.0736)	(0.0880)	(0.0679)
Characteristics								
Total	0.1856***	0.1553***	0.1915***	0.1420***	0.2124***	0.2728***	0.3320***	0.2092***
	(0.0109)	(0.0164)	(0.0197)	(0.0189)	(0.0479)	(0.0437)	(0.0461)	(0.0450)
Firm characteristics	0.1300***	0.1273***	0.1237***	0.0633***	0.1320***	0.2404***	0.2512***	0.1613***
	(0.0060)	(0.0106)	(0.0100)	(0.0089)	(0.0153)	(0.0237)	(0.0243)	(0.0231)
Firm constraints	0.0033***	0.0034**	0.0039***		0.0004	0.0007	0.0005	
	(0.0008)	(0.0014)	(0.0013)		(0.0010)	(0.0009)	(0.0011)	
Gender of owner	0.0266***		0.0178***	0.0266***	0.0654***		0.0270***	0.0228**
	(0.0018)		(0.0048)	(0.0064)	(0.0150)		(0.0086)	(0.0100)
Period	0.0125***	0.0089***	0.0093***	0.0064***	0.0023	0.0028	0.0042	-0.0001
	(0.0010)	(0.0013)	(0.0017)	(0.0014)	(0.0033)	(0.0033)	(0.0042)	(0.0032)
Region	0.0100	0.0152	0.0307**	0.0399***	0.0112	0.0240	0.0417	0.0154
	(0.0097)	(0.0126)	(0.0139)	(0.0137)	(0.0362)	(0.0276)	(0.0271)	(0.0287)
Sector of activity	0.0033***	0.0006	0.0062***	0.0058**	0.0011*	0.0050*	0.0074*	0.0099**
	(0.0006)	(0.0014)	(0.0020)	(0.0025)	(0.0006)	(0.0028)	(0.0045)	(0.0039)
Coefficients (unexplained)								
Total	0.2071***	0.2082*	0.1885*	0.2344**	0.0938	0.4955	0.4965	0.4354
	(0.0755)	(0.1109)	(0.1021)	(0.1095)	(0.1779)	(0.5453)	(0.6023)	(0.5172)
Firm characteristics	0.1359	0.1596	0.0296	0.0298	0.1831	0.3452	0.3216	0.2100
	(0.1424)	(0.3792)	(0.5032)	(0.1875)	(0.4299)	(1.8525)	(2.1715)	(1.1507)
Firm constraints	0.0547	0.0212	0.0480		-0.0215	-0.0077	0.0072	

	(0.0623)	(0.1535)	(0.1730)		(0.2113)	(0.6691)	(0.7787)	
Gender of owner	-0.0114		0.0250	0.0098	0.0712		0.0938	0.1374
	(0.0305)		(0.0937)	(0.0528)	(0.0639)		(0.1712)	(0.1282)
Period	0.0012	0.0022	0.0069	0.0033	0.0026	0.0155	0.0214	0.0165
	(0.0049)	(0.0222)	(0.0246)	(0.0257)	(0.0072)	(0.1166)	(0.1315)	(0.1201)
Region	0.0007	0.0207	0.0086	0.0098	-0.0980	0.0656	0.0703	0.0768
	(0.0374)	(0.0623)	(0.0590)	(0.0570)	(0.1388)	(0.2062)	(0.2759)	(0.2233)
Sector of activity	-0.0024	-0.0085	0.0014	-0.0003	-0.0049	0.0278	0.0598	0.0486
	(0.0036)	(0.0233)	(0.0066)	(0.0028)	(0.0067)	(0.0733)	(0.0704)	(0.0610)
Specification error	-0.0015	-0.0049	-0.0035	0.0014	-0.0320	-0.0056	-0.0014	-0.0050
	(0.0342)	(0.0351)	(0.0360)	(0.0371)	(0.0308)	(0.0448)	(0.0586)	(0.0442)

Note: in truncated specification, we drop possible endogenous variables. The estimations are performed using non-SCST entrepreneurs' coefficients as reference. The reweighting error is not reported, but it can be computed as the difference between the total gap and the sum of the characteristics, coefficients, and specification error. We normalized the category regressors using an averaging approach (Yun 2005). In all estimations, we used sample weights provided by NSSO. Robust standard errors (clustered at district level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: authors' calculations.

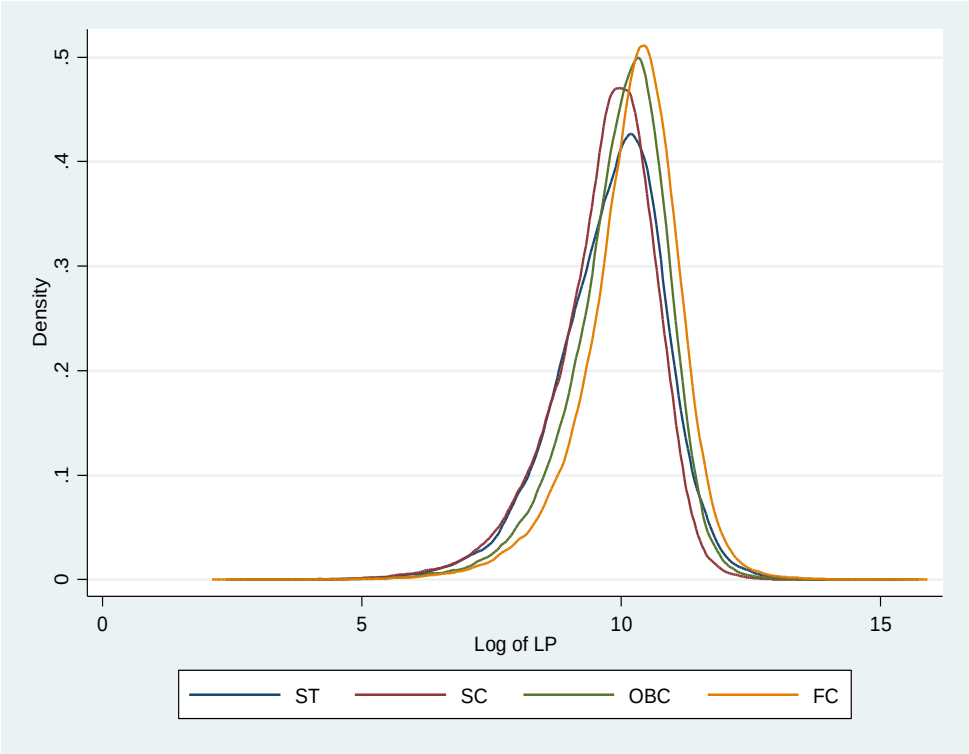
Table A6: Oaxaca-Blinder decomposition on matched sample

	SC vs Non-SCST firms		ST vs non-SCST firms	
	(1)		(2)	
Total gap	0.3345***		0.6108***	
	(0.0267)		(0.0757)	
Composition effect	0.1480***	44.25	0.2364***	38.70
	(0.0228)		(0.0352)	
Structure effect	0.1865***	55.75	0.3744***	61.30
	(0.0178)		(0.0638)	
Contributions to the composition effect by component				
Firm characteristics	0.0941***	28.13	0.1830***	29.96
	(0.0119)		(0.0206)	
Firm constraints	0.0027**	0.81	0.0005	0.08
	(0.0011)		(0.0013)	
Gender of owner	0.0197*	5.89	0.0259*	4.24
	(0.0113)		(0.0156)	
Period	0.0064	1.91	-0.0012	-0.20
	(0.0042)		(0.0079)	
Region	0.0214***	6.40	0.0225	3.68
	(0.0078)		(0.0161)	
Sector of activity	0.0036**	1.08	0.0056*	0.92
	(0.0018)		(0.0030)	

Note: matched sample is created using Nopo (2008). Parentheses next to the estimated output report the percentage contribution to the total gap. In all estimations, we used sample weights provided by NSSO. Standard errors (clustered at district level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

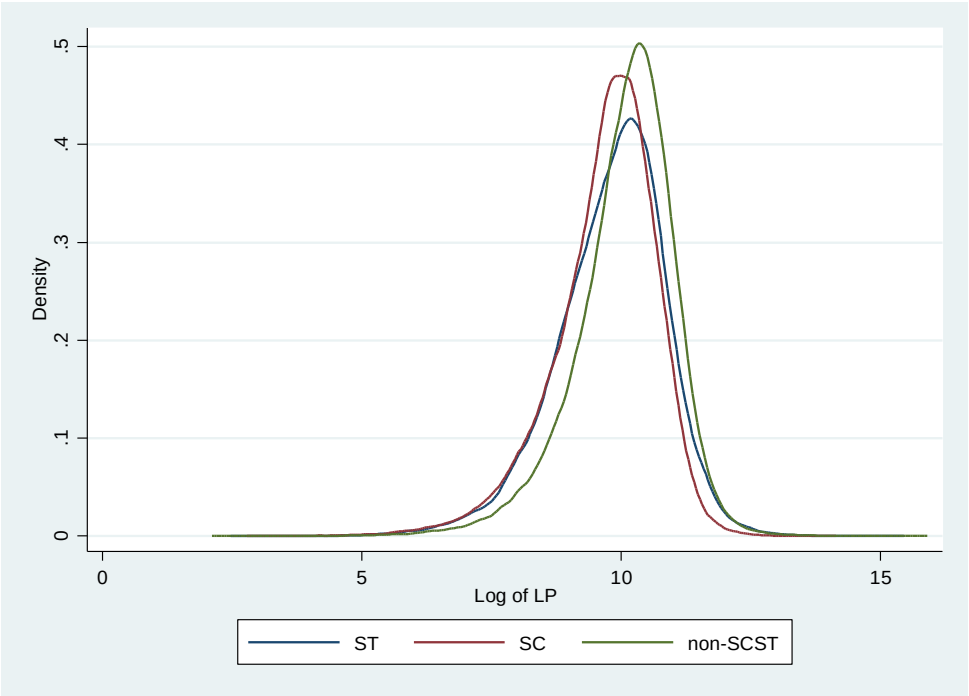
Source: authors' calculations.

Figure A1: Kernel density of labour productivity across gender of firm owner



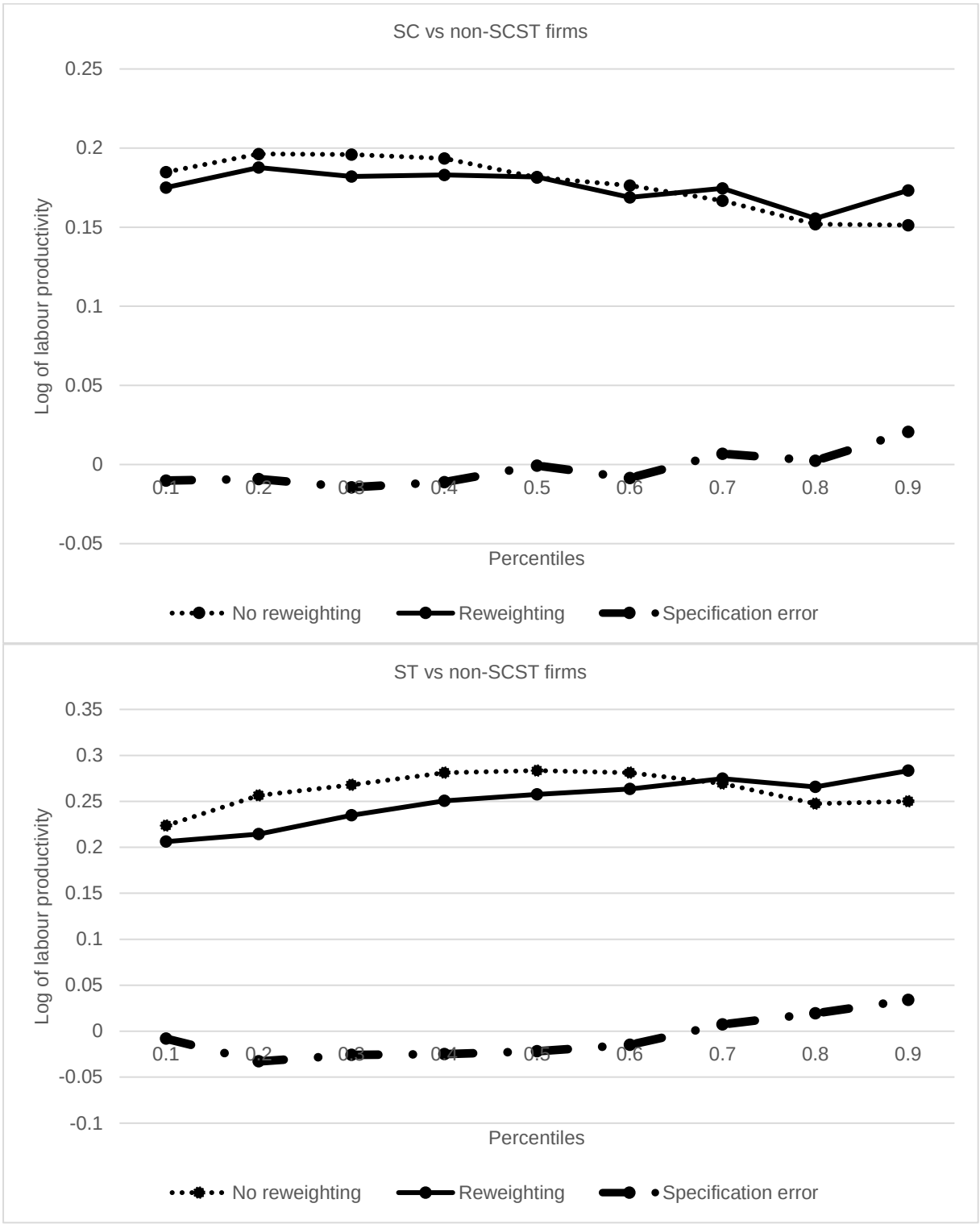
Source: authors' illustration.

Figure A2: Kernel density of labour productivity: SC and ST versus non-SCST



Source: authors' illustration.

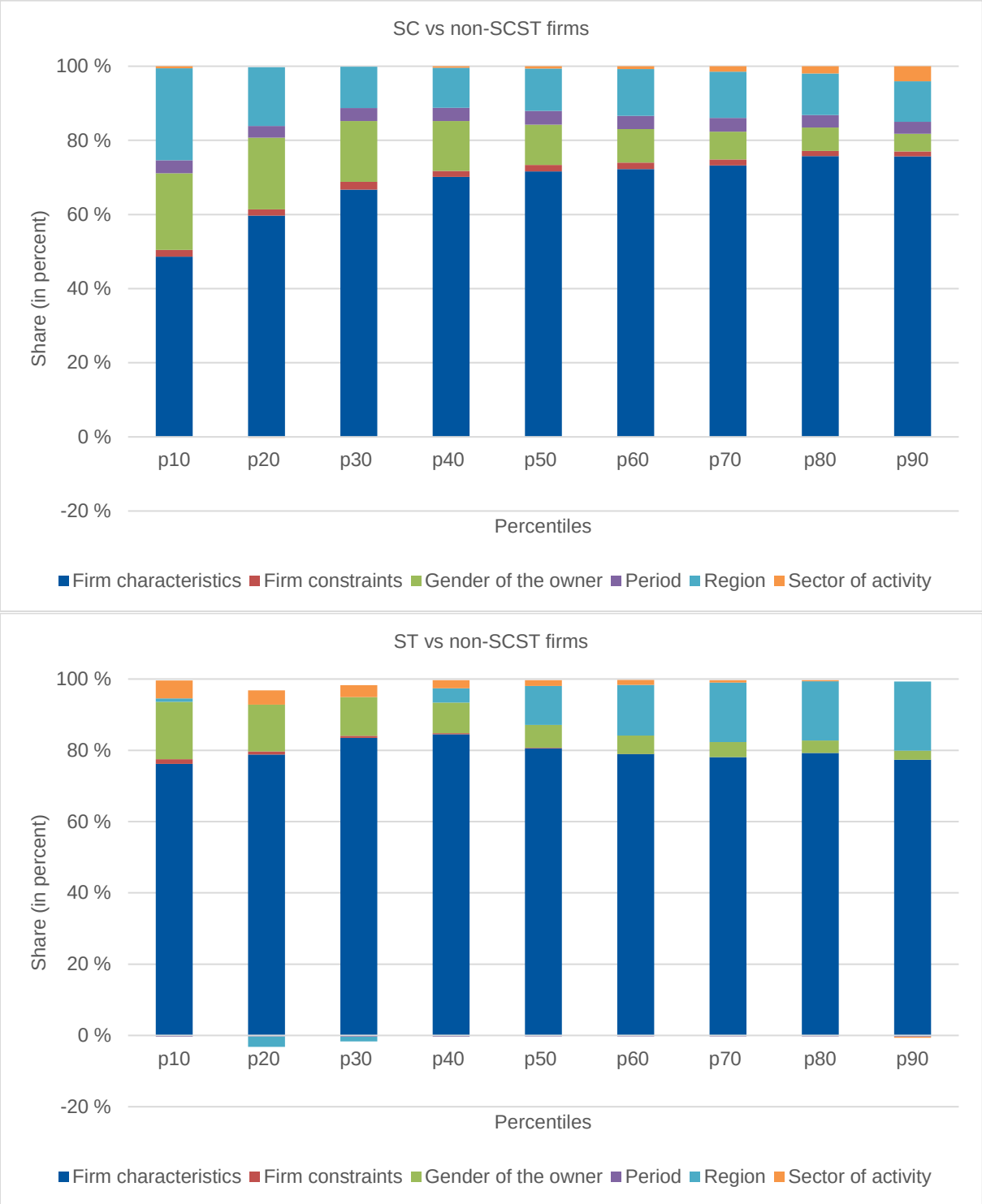
Figure A3: Specification error in RIF decomposition



Note: specification error is the difference between the total composition effect obtained by the standard and the reweighted RIF regression.

Source: authors' illustration.

Figure A4: RIF decomposition: share of components in characteristics



Source: authors' calculations.