Azim Premji University

CSE Working Paper

#51

Estimating the productivity gap between organised and unorganised small-scale units in India's manufacturing sector

Amit Basole, Dimple Chopde and Paaritosh Nath

June 2023

Centre for Sustainable Employment

cse@azimpremjiuniversity.edu.in

Estimating the productivity gap between organised and unorganised small-scale units in India's manufacturing sector*

Amit Basole, Dimple Chopde and Paaritosh Nath

Azim Premji University, Bengaluru, India

May 28, 2023

Abstract

Small manufacturing firms are considered to be engines of growth and job creation. While most research on small firms focuses on formal sector units, in India informal sector units far outnumber the formal. This is true even for manufacturing units employing 5 to 49 workers, which constitute only 5% of all unorganised units, but in absolute numbers are nine times more numerous than organised units in the same size class. Such firms have the potential to contribute to structural transformation but their capacities vis-a-vis formal firms are not well understood. To address this, we create a unit-level dataset combining Annual Survey of Industries data for organised (formal) units with the National Sample Survey data on unorganised (informal) units. We also discuss problems involved in this exercise and some ways to deal with them. We find that matching organised and unorganised units on observable characteristics reduces the labour productivity differences between them to around 25 percent. We discuss some policy implications of our results.

Keywords: India, manufacturing, firm size, productivity, Blinder-Oaxaca decomposition, Propensity Score Matching.

JEL classification: O14, O17

*We thank Prof R. Nagaraj, Dr. Radhicka Kapoor and all the participants of the Delhi workshop as well as the Indian Economy coference at CDS, Trivandrum for valuable comments. Thanks are also due to Anand Shrivastava for guidance. Errors remain authors' responsibility.

1 Introduction

Employment and productivity growth in the manufacturing sector are key drivers of structural transformation. In India, despite high levels of aggregate GDP growth, the manufacturing sector has failed to expand its share of employment and output significantly raising concerns over the pace of structural transformation (Basole et al., 2018; Kannan and Raveendran, 2009; Thomas, 2013). While a debate rages on the role of modern services in aiding this transformation (Amirapu and Subramanian, 2015; Ghani and Kharas, 2010; Gupta and Basole, 2020), it is relatively uncontroversial that manufacturing has a key role to play (Ghose, 2020).

As in many other developing economies, the manufacturing sector in India is dual in nature, both from an analytical and a regulatory/legal perspective. The legal or regulatory divide is the one between organised and unorganised units based on whether or not the unit is registered under the Factories Act of 1948. This is a criterion based on firm size (number of workers).[1] It corresponds roughly to the formal-informal divide in other developing economies. In this paper we use the terms organised/formal and unorganised/informal interchangeably.[2]

The regulatory dualism has analytical counterparts. For example, a large productivity gap is usually observed between registered (organised/formal) and unregistered (unorganised/informal) units. Arguably a more significant analytical dualism, from the point of view of structural transformation and growth, is the one between nano/microenterprises that are largely subsistence oriented and employ no wage labour and minimal capital versus the rest of the manufacturing units which employ wage workers and are driven by profitability concerns (Lewis 1954). From this perspective, there is a dualism within the unorganised sector. The vast majority of units can be said to fall in the subsistence category with no hired workers (84 percent)(Mehrotra and Giri, 2019). But a minority does employ hired workers and represents the profit-oriented segment of the unorganised sector. We are concerned with this segment in the present paper.

Academic as well as policy work has recognised the importance of this dynamic segment of the informal sector in enabling job creation. One example is the 'cluster approach" to industrialisation

[1] Registration is required for units employing 10 or more workers operating with power or 20 or more workers operating without power.

[2] Though a distinction is usefully made between types of firms (organised or unorganised) and types of workers (formal or informal), this distinction is not relevant to the present study. The international literature uses the term "informal" to refer to firms as well as workers.

(R. Nagaraj, 2021; Schmitz and Nadvi, 1999). This literature has emphasised the widespread prevalence of informal industrial clusters in developing countries. While it cannot be said that all clusters are dynamic and job creating, the literature is helpful in understanding the conditions under which such clusters become competitive and expand to create industrial employment which is mostly small-unit based and informal in nature.

Another strand of research focuses on subcontracting arrangements between organised firms and the more dynamic units of the unorganised sector (Basole et al., 2015; Shekar and Nataraj, 2022). Here again there are two possibilities. The sub-contracting arrangement may resemble a classical "putting-out" system which leaves the unorganised unit with no surplus to invest and grow. Or it may act as a route for technology transfer and upgrading of unorganised units, thereby increasing their productivity and scope for expansion.

In order to inform policy-making that enables the growth of the profit-oriented part of the unorganised sector, we need detailed unit-level comparative information on how this sector performs vis-a-vis its organised counterpart. For example, while it is well-understood that organised sector units are more productive than unorganised ones, it is still worth asking how much of a productivity gap remains after units are matched on observable characteristics such as labour and capital endowments, geographies and industries. The answer to this question can help guide policy intervention that aim to increase productivity in the unorganised sector.

The paper aims to answer the above question by comparing the performance of small manufacturing units, defined as those employing 5 to 49 workers, in the organised and the unorganised sectors. The lower bound is motivated by a desire to exclude nano (one person or own-account) and micro-units that run largely with family labour. The upper-limit of 49 workers is chosen because a negligible proportion of unorganised units exceed this size. Of course, there is some arbitrariness in using firm size or number of workers as a proxy for the analytical distinction between subsistence-oriented and profit-oriented units. But we have verified that the results are not substantially altered with slightly different size-based classifications.

We combine unit-level data from the ASI for the year 2015-16 with NSS data on Unincorporated Enterprises for the same year. We take advantage of the fact that even though the ASI and NSSO survey frames are exclusive of each other, in practice, due to reasons such as non-registration of

eligible units (see D. R. Nagaraj, 2018) or failure of earlier eligible units to de-register, units with similar number of workers can be found in both datasets.[3] The problems associated with combining the datasets at the unit level and our approaches to solving them, are discussed in Section Three. While there are several studies that undertake a comparative analysis of the two segments of the Indian manufacturing sector using Annual Survey of Industries (ASI) and National Sample Survey (NSS) data, most are either at the industry level (Allen et al., 2018; Hasan and Jandoc, 2010; Kumari, 2022) or have focused mainly on size distribution of enterprises (Kapoor, 2022; Mehrotra and Giri, 2019). One very recent study has used a combined ASI-NSSO dataset to study productivity differences across the size distribution (Rabbani and Raj SN, 2023).

We find that small units in the organised sector display levels of productivity (gross value added per unit) around 50 percent higher than units in the unorganised sector, controlling for labour and capital endowments as well as other relevant factors such as share of unpaid workers, share of women workers, rural or urban location and state as well as industry of operation. A Blinder-Oaxaca decomposition of the productivity gap reveals that 63 percent of the gap is explained by differences in endowments and another 23 percent by differences in returns to these endowments. We also validate the results of the decomposition using Propensity Score Matching to more closely match small units in both sectors. Matched units differ in their labour productivity levels by 24 percent which agrees well with the returns component of the Blinder-Oaxaca decomposition.

The decomposition also shows that, among the observable characteristics, as expected, labour and capital account for most of the gap. Interestingly, returns to labour contribute positively to the gap while returns to capital stock contribute negatively. Two important implications emerge from our analysis. First, the fact that the returns component is much smaller than the endowments component indicates that given similar resources, unorganised units can significantly bridge the productivity gap. This is important because, though a small minority within the unorganised sector, in absolute terms, such units are numerous.[4] Thus, from the perspective of structural transformation, this profit-oriented minority can be a significant source of growth and job creation. Second, the decomposition results imply that while unorganised sector units are capital constrained, organised sector units are

[3]The ASI frame comes from Inspector of Factories list and only includes plants registered under the Factories Act. NSSO unorganised manufacturing surveys are based on an area frame and they exclude manufacturing units registered under the Factories Act.

[4]For example, the 2015 NSSO data estimates just over 900,000 manufacturing units employing between 5 and 49 total workers (working owners, family workers and hired workers). In contrast, the ASI data estimates just over 100,000 such units in the organised sector for the same year.

labour constrained. We discuss these implications in the Section 5.

The remainder of this paper is organised as follows. The next section briefly reviews the literature on firm size, informality and productivity. Sections 3 discuss the data and methods. Section 4 presents the results. We discuss the results in Section 5 and Section 6 concludes.

2 Literature review

2.1 Dualism within dualism

The dualist tradition going back in Economics to Lewis (1954) posits that informal firms exist in an economic space distinct from that occupied by formal firms, the former being governed by the logic of subsistence and characterised by surplus labour (in the sense of "wage" exceeding marginal product), and the latter by profit maximization (wages equal to or less than marginal product). Over the years, the dualist perspective has moved on to identifying heterogeneities within the informal sector. Ranis and Stewart (1999) argue that the urban informal economy can be divided into two distinct segments- a traditional sector and a modernising informal sector. The traditional component of the urban informal economy serves as a low productivity sponge for surplus labour leaving agriculture, while the modern informal sector is a dynamic contributor to economic growth. This dualism within the informal sector has been noted in the Indian manufacturing context as well (Bhattacharya et al., 2013).

The significance of this dualism can be understood in two ways. First, the dynamic informal firms, in principle, may have the potential to grow in size and productivity if provided the right infrastructural and regulatory environment. And second, these firms, though a small minority within the informal sector, are numerous in absolute terms and therefore represent an opportunity to create a large number of productive manufacturing jobs.

It is relevant to note here that Kesar and Bhattacharya (2020) find that there is a tendency for the dualism within the unorganised manufacturing sector to reproduce itself. That is, subsistence firms do not grow to become dynamic ones. This suggests that relatively larger firms in the unorganised sector may be born large. What implications this has for further scaling up of the dynamic firms, remains to be explored. We take this issue up in the Discussion section.

2.2 Regulations and firm size

The question of firm size is analytically distinct from the regulatory question of formality-informality, though in practice the two are intimately related. While the firm size distribution is a continuum (barring regulation-induced discontinuities, for which see Amirapu and Gechter (2020)), the legal apparatus imposes a dichotomy on this continuum via requirements of compliance and registration.

How large a unit is in terms of workers employed and whether it is compliant with the laws that apply to its size are both outcomes of a complex process. The unit's performance (productivity and profitability) is endogenous to its size as well its organised-unorganised status. What this means is that whether a unit is registered or unregistered under a particular Act, such as the Factories Act may be determined by its productivity and profitability which decide its ability to meet the costs of compliance. But equally, its registration status may determine its productivity or profitability, say via access to formal credit. Similarly, the scale of the unit may be a consequence of the legal regime due to size-based increases in costs of compliance. The scale in turn may affect productivity and thereby its ability to comply with regulations.

Empirical work on Mexican firms shows that if entry costs to formality are lowered alongside capital being more readily supplied through microfinance, informal firms are in a position to register, borrow, and eventually expand, sparking growth in the process (Fajnzylber et al., 2009). Similar results are also seen among Brazilian firms, highlighting that lowering the costs of entry into the formal sector is an effective policy to reduce informality even in employment arrangements, particularly in comparison with options such as reducing payroll taxes and an increase in unemployment benefits (Ulyssea, 2010).

But it is also possible that firms voluntarily remain small in order to avoid detection and regulation. There are two different cost considerations at play here. As firms expand in size and move towards formalisation, they are subject to transformation costs, which include extensive preparation, increased record keeping, registration fees and taxes (Nelson and De Bruijn, 2005). On the other hand, there are also evasion costs which increase with firm size. According to Amin and Islam (2015), evasion costs play a potentially more significant role in determining the low size of informal firms. These include paying penalties on being caught, or announcing bankruptcy on being discovered, only to reappear as another firm altogether (Gelb et al., 2009). Thus, in this scenario, while

informal firms might lack the scale to produce efficiently and lose out on the benefits accruing from formalisation (Straub, 2005), the cost advantages incurred from avoiding taxes and regulations more than compensates for their low scale and efficiency (Farrell, 2004; Levy, 2010).

2.3 Firm size, productivity and formality

Finally, we come to how the existing literature sees the relationship between firm size, formality, and productivity. There exists a large body of work documenting the positive correlation between firm size and total factor productivity within the organised sector (Baldwin et al., 2002; Van Ark and Monnikhof, 1996; Van Biesebroeck, 2005). This has been attributed to larger sized firms experiencing economies of scale and greater allocative efficiency (Rand and Torm, 2012). Bartelsman et al. (2013) note, however, that this relationship varies significantly across countries, being stronger in more advanced economies. Others such as McKenzie and Sakho (2010) have argued that the premium to formality has a heterogeneous effect dependent upon the size of the formal firm. Registration increases the profitability of medium sized firms, but lowers profits for both the smaller and larger firms.

When it comes to cross-sector comparisons, it is usually observed that informal firms are less productive compared to their formal counterparts (Dabla-Norris et al., 2008; La Porta and Shleifer, 2008). The reasons include lack of legislation (Perry, 2007), lower access to credit (Straub, 2005) and lack of access to productive public goods (Djankov et al., 2003; Marcouiller and Young, 1995). Productivity differentials might also arise out of workers of differing skill levels self selecting themselves into formal and informal sectors, rather than due to any intrinsic characteristics of the types of firms located in either sectors (Amaral and Quintin, 2006; Dimova et al., 2008).

Taymaz et al. (2009) compares the productivity levels of around 5000 formal and informal small and medium sized firms in Turkey and finds that the labour productivity gap between the two segments is around 107 percent. Taking the dynamic aspect of the relationship into account, Perry (2007) looking at seven Latin American and Caribbean countries, find that firms that started out informal tend to have a labour productivity difference of 30 percent when compared to firms that were formal to begin.

Using data from World Bank firm-level surveys, La Porta and Shleifer (2008) find that informal

firms are unproductive not just relative to large formal firms but also when compared with smaller sized formal firms. Formal firms are characterised by better educated managers, tend to use more capital and use external finance to a greater extent than their informal counterparts. They conclude that removing barriers to formalisation, would not allow the productivity gap between formal and informal firms to be reduced substantially.

More closely aligned to the present study, Báez-Morales (2015) look at the manufacturing sector in Mexico restricting analysis to formal and informal micro-firms (leaving out the larger sized formal firms). They find that formal micro firms have both greater output and efficiency than informal micro firms. Others have challenged this stylized relationship, noting that informal firms are at par or perform better than formal forms when it comes to efficiency as illustrated by the fact that they continue to persist in rapidly modernising economies (De Soto et al., 1989). A study of Bolivian firms notes that the "romantic view" of informal firm owners having comparable or higher entrepreneurial ability than formal ones holds true but only for large sized informal firms (McKenzie and Sakho, 2010).

A study of informal firms in West Africa while admitting to significant productivity differentials between formal and informal firms also notes that the productivity gap is much smaller for large informal firms than for small informal firms (Benjamin and Mbaye, 2012). The former, according to the authors, lie at the fringes of the formal and informal divide and are much closer to formal sector firms in terms of productivity. Seemingly the opposite conclusion is reached by Amin and Islam (2015) whose study of 500 firms across seven African countries suggests that within the informal setup, smaller sized firms have higher labour productivity than larger sized firms. However, as we show later, this result (negative relationship between firm size and labour productivity) only holds when we control for capital stock.

Coming to India, exiting studies confirm the patterns noted in the global trends above. For one, the organised manufacturing segment in India outperforms the unorganised in terms of productivity (both total and partial measures). Krishna et al. (2018) find that, between 2003-04 to 2007-08, relative Total Factor Productivity (TFP) for unorganised industry was 33 percent lower than that of the organised sector. Similarly, Kathuria et al. (2010) note that between 1994-2005, labour productivity in the organised sector was on an average 4.4 times higher than that in the unorganised

sector. Kumari (2022) is the most recent paper that estimates the determinants of the efficiency differential between formal and informal industries at the 2-digit level. The author finds, using a Blinder-Oaxaca decomposition, that differences in capital endowments contribute significantly to the gap. Finally, another very recent study, conducted at the unit level with a combined ASI-NSSO dataset, documents a rising productivity gap between small and large firms in Indian manufacturing (Rabbani and Raj SN, 2023).

Given the importance of the unorganised manufacturing sector in terms of employment as well as number of firms, more studies are needed examining productivity differentials between organised and unorganised sectors, ideally at the unit level. The present study is an effort in this direction.

3 Data and methods

3.1 Constructing a combined manufacturing dataset

We construct a unit-level dataset for the entire Indian manufacturing sector by combining ASI factory-level data for the organised sector with the NSSO Unincorporated Enterprises Survey for 2015-16. Only manufacturing units (NIC 2008 10 to 33) are used for the analysis.

While earlier studies have used both datasets to generate firm size distributions and other characteristics across the organised-unorganised divide, concerns regarding the comparability and combined use of these two datasets still remain.[5] Thus, before we can proceed with the analysis, these concerns need to be addressed. In this section we discuss issues of sampling frame and multipliers, as well as comparability of key variables across the two surveys.

3.1.1 Survey frames and multipliers

The ASI and NSSO survey frames are exclusive of each other. The ASI frame comes from Inspector of Factories list and only includes plants registered under the Factories Act. NSSO unorganised sector surveys are based on an area frame and they exclude manufacturing units registered under the Factories Act. In practice, due to reasons such as non-registration of eligible units or failure of earlier eligible units to de-register, units with similar number of workers can be found in both

[5]Earlier studies have either constructed industry-level aggregates from both datasets or compared the two sectors without combining them. In contrast we combine at unit-level. See Allen et al. (2018), Hasan and Jandoc (2010), and Raj SN and Sen (2019).

datasets. But despite this the two datasets remain exclusive in the sense that no particular unit is found in both (Bedi and Banerjee, 2007).

ASI data contains a census sector, where all listed factories above a certain size are surveyed and the survey multiplier equals 1. For all other factories, the multiplier takes the form E_{is}/e_{is}, where E is the total number of factories in the sample sector in a stratum, e is the number of factories surveyed out of total number of factories in the sample sector in a stratum, i indexes states, s indexes stratum in the i-th state.

The NSSO uses an area frame based on Census villages and Urban Frame Surveys from which First Stage Units (FSUs) are selected followed by listing and sampling of Ultimate Stage Units. Here the multipliers take the form $(Z/n)x(1/z)x(E/e)$, where once again, E is the total number of enterprises listed in sample FSU, e is the number of enterprises surveyed in sample FSU, Z is the total size of a rural or urban sub-stratum, z is the size of sample village/block used for selection and n is the number of sample villages/blocks surveyed.

In both cases the multiplier gives the number of manufacturing units in the population represented by the sample unit. As may be expected, ASI weights are smaller, i.e. a sample unit represents a smaller number of population units compared to NSSO. The average multiplier in ASI is 4.1, while that in NSSO is 88.6. Since the universes are non-overlapping, pooling unit-level data together from both universes with existing weights gives entire manufacturing universe.

3.1.2 Comparing the two datasets

Table 1 summarises the similarities and differences between the two surveys. The reference period for ASI is the accounting year while that for NSSO is either the last 30 days or the year preceding the survey date, depending on whether the surveyed unit keeps written accounts. We standardise the reference period to a year by converting monthly flows into annual ones.[6]

[Table 1 here]

The surveyed unit in case of ASI is the factory while for NSSO it is the establishment. This distinction

[6]This can be done either by multiplying the monthly numbers by 12 or by the number of months of the year that the unit operated. Choosing one or the other does not make a significant difference to the results, so we have retained the simpler method of multiplying monthly values by 12.

is important in cases where a single firm may have multiple factories and firm level characteristics are relevant to the study. But it matters less for our purposes since our primary variables of interest are at the factory or unit level. Hereafter we refer consistently to "units" in both datasets, rather than "firms".

The geographical indicators are the same in both surveys as are the National Classification of Industry (NIC) codes. Value-added is measured in both surveys as receipts less non-wage expenditures. Wages are measured as payments to hired labour. The number of workers is defined similarly in both surveys as a position (and not a person) to be filled by a directly or indirectly employed person who is either a paid or unpaid (family) worker.[7]

The most significant difference between the two surveys concerns the measurement of capital stock. While the ASI records the book value of plant, machinery and other fixed assets, the NSSO records the market value.[8] To the extent that these diverge over time, the value of capital stock across the two surveys will not be comparable.

We address this problem in the following ways. First, we only consider the value of plant and machinery and leave out land and buildings since the value of the latter may be subject to more fluctuations that are unrelated to the manufacturing process itself. Second, we draw on the Prowess IQ database from the Centre for Monitoring the Indian Economy to arrive at a book to market value converter. Prowess supplies the market value to book value ratio of investments for firms in its database. We calculate the median ratio for each industry at the 3-digit level (NIC-2008 101 to 339). For this exercise, only firms that have fixed capital and total output in the range observed in the ASI data for factories with 5 to 49 workers are used, to improve comparability across the two datasets.[9]

[7] As per the ASI Instruction Manual,

> A worker includes all persons employed directly, informally or formally or through contractor on payment of wages or salaries and engaged in any manufacturing process or its ancillary activities like cleaning any part of the machinery or any premises used for manufacturing or storing materials or any kind of work incidental to or connected with the manufacturing process.

As per the NSSO concepts and definitions document,

> A worker is defined as one who participates either full time or part time in the activity of the enterprise. The worker may serve the enterprise in any capacity-primary or supervisory. He/she may or may not receive wages /salaries in return to his/her work incidental to or connected with the enterprise activity. A worker need not mean that the same person is working continuously; it only refers to a position.

.

[8] We are indebted to Prof Ravi Shrivastava for alerting us to this issue. See Krishna et al. (2018) as well as ASI and NSSO instruction manuals for more details.

[9] 99th percentile values in the ASI data 2015-16 once we account for size= 5-49 were as follows: fixed capital- INR

Since there are a large number of missing values for this variable in Prowess, any firm supplying this information at least once between 2005 and 2015 is retained. The ratio is multiplied with the capital stock (excluding land and buildings) as reported in ASI. This converts the book value to the market value and makes it comparable to the NSSO numbers. We find that all but one NIC 3-digit industries display a market to book value ratio close to 1. The complete distribution (without the one outlier mentioned above) is shown in Appendix Figure 1.

Going forward our main results are presented with the adjusted capital stock variable where the Prowess-derived ratio has been used for converting book value to market value. But we have also provided results with the unadjusted variable in Appendix Table 1 for comparison. We find that the adjustment reduces the organised premium slightly, but it remains economically meaningful and statistically significant.

Finally, as one more check, we perform our analysis on a restricted sample of unorganised sector units that are up to three years old. The justification is that for relatively newer firms, the market value and the book value of assets should correspond more closely to each other and hence the capital stock values for these relatively younger unorganised units should be more comparable to ASI book values. The results are not substantially altered by doing this and are provided in Appendix Table 2.

3.2 Basic descriptive statistics

Some basic descriptive statistics for the combined dataset for 2015 are given in Table 2. As expected the median unit in the ASI data is much larger (21 workers) than the median unit in the NSS data (1 worker).[10] But the total number of estimated workers in the unorganised sector is 2.5 times larger than those working in the organised sector (35 million versus 14 million).[11]

[Table 2 here]

As mentioned earlier, we are interested in comparing the performance of small manufacturing units in both sectors. Hence we restrict the scope of further analysis to units employing at least 5 but less

19,70,00,000, total output per year- INR 86,60,00,000.

[10] Note that the mean at 89 workers per factory, is much higher than the median in ASI data, indicating the presence of a few very large factories or a right skew in the distribution.

[11] We consider total persons engaged in an organised unit and not only production workers, to make it compatible with unorganised units where working owners are included among workers.

than 50 total workers (including working owners and unpaid workers). Such units, employing 5-49 total workers represent 66% of ASI factories and 5% of NSSO enterprises.

Coming to the small unit combined dataset - our pooled cross-sectional unit level data contains 21,572 ASI units and 9,453 NSSO units. The weighted statistics for the small unit segment of both datasets are given in Table 2. Note, first of all, that even though units in the 5 to 49 worker range are only 5% of the NSSO universe, in absolute terms they vastly outnumber the ASI units in the same segment (approximately 900,000 as opposed to 100,000). Hence the total employment in the unorganised small unit segment is also much larger at 7.4 million as compared to that in the organised segment (1.8 million). This difference remains high even if we restrict the analysis only to hired workers (excluding unpaid workers and working owners).

Of course, the fact that the difference in estimated units is much larger than estimated workers, indicates that the median unorganised sector small unit is much smaller at 6 workers, compared to the median organised sector small unit (15 workers). The median organised sector small unit is also more than twice as productive as measured by value added per worker. Thus, even though the organised sector only accounts for 20% of total workers in this segment, it accounts for 46% of value-added and 54% of capital stock.

The question that we explore is how far these raw productivity differences are accounted for by difference in endowments or observed characteristics.

3.3 Measuring productivity differences

The main outcome variable is gross value added (GVA) per unit. We use this measure rather than labour productivity (gross value added per worker) because we control separately for labour endowment. However, replacing GVA per unit with GVA per worker does not change the basic results, except for the coefficient on labour as we discuss later.

The main independent variable is a dummy variable that characterises a unit as "organised" or "unorganised" based on the dataset it belongs to. But, in addition, we also exploit the variation within unorganised sector units to construct a second independent variable called "degree of organisedness". This variables draws on the fact that the NSSO data contains units that are registered under some government Act as well as those that are not registered under any Act. Similarly there

are units that maintain books of account and those that do not. The variable takes four values as follows:

1. NSSO units that did not maintain accounts & were not registered under any Act
2. NSSO units that maintained accounts but were not registered or did not maintain accounts but were registered under some Act
3. NSSO units that maintained accounts & were registered
4. ASI units

One caveat here is that we cannot determine the exact nature of accounts maintained by unorganised units. And as with the organised dummy, the degree of organisedness is also endogenous. That is, a unit's productivity may determine its ability to register or maintain accounts as much as the other way around. Given these caveats, without making causal claims, we show that these questions (on whether a firm was registered with any Act or maintained account) are indeed significantly correlated with the ability of the unit to produce value.

We employ two approaches to estimate GVA/unit differences between organised and unorganised firms. First, we perform an OLS regression followed by a Blinder-Oaxaca Decomposition. Second, we undertake a Propensity Score Matching (PSM) exercise to estimate differences for more closely matched units. PSM is usually performed to estimate causal average treatment effects (ATEs or ATTs), but our results cannot be given a causal interpretation, since, as discussed in Section Two, units are not randomly assigned to the organised or unorganised sector. Rather characteristics endogenously determine status and the status in terms determines their characteristics. Given this, the (non-causal) interpretation of the PSM estimate is only that units matched as closely as possible still differ in performance by the estimated amount.

3.3.1 OLS regression and Blinder-Oaxaca decomposition

The regression model is as follows:

$$Y_i = \alpha + \beta_1 * Org_i + \beta_2 * L_i + \beta_3 * K_i + \beta_4 * [X_i] + \epsilon$$

Y_i = log GVA per unit or ln GVA per worker

Org_i = Organised dummy or degree of organisedness

L_i = log number of workers

K_i = log capital stock

$[X_i]$ = Controls

Controls include whether a firm is home-based, rural/urban location, the proportion of women workers, the type of ownership (proprietorship, partnership, private company and others such as self-help group, trust, NGO, and public firm), along with industry (NIC 3-digit) and state fixed effects.

Blinder-Oaxaca decomposition and its variations have been widely used to decompose gender and racial wage gaps as well for estimating treatment effects in various contexts (Fortin et al., 2011). In cases where the treatment variable is exogenous the decomposition gives causal treatment effects. But in cases such as the present once where "treatment" is endogenous, the results can only be interpreted in an accounting sense. That is, the decomposition allows us to quantify the extent to which the outcome gap between organised and unorganised units can be accounted for by differences in observed characteristics. This gap is usually separated into three components: endowments (observable characteristics), returns to these characteristics, and an interaction term:

$$\triangle Y = [E(X_A) - E(X_B)]'\beta_B + E(X_B)'(\beta_A - \beta_B) + [E(X_A) - E(X_B)]'(\beta_A - \beta_B)$$

In our case, Group A are the organised sector manufacturing units while Group B are the unorganised sector units. The above decomposition equation from Jann et al. (2008) is from the viewpoint of Group B. For the endowments effect, group differences in covariates are weighted by the coefficients of Group B. In other words, this component measures the expected change in the unorganised sector unit outcome, if these units had the same endowments and characteristics (capital, labour, location etc) as organised units. For the second component (returns), the differences in coefficients for the two sectors are weighted by the unorganised units' characteristics. That is, the second component measures the expected change in unorganised sector mean outcome, if these units had organised

sector coefficients. The third (interaction) term accounts for the fact that differences in endowments and coefficients exist simultaneously between the two groups.

3.3.2 Propensity score matching

Even after making the two types of units more comparable to each other by restricting the number of workers, the median units in both sectors remain substantively different from each other. In this context, a limitation of the Blinder-Oaxaca approach is that the measured effects could rely on counter-factuals that do not exist. One way around this is to create a more comparable set of "control" units from the unorganised sector that closely resemble the "treated" units in the organised sector using Propensity Score Matching (PSM). Note that the language of "treatment" and "control" is used here only for convenience since PSM is a technique for causal inference and most literature on it uses this terminology.

The basic principle underlying the technique is to use observed covariates to find matching observations from one group for each observation of the other group. A propensity score (that ranges from 0 to 1) is estimated for each observation in the "control" group using a logit model. In our case, this score captures the degree to which a unit in the unorganised sector resembles a unit in the organised ("treated") sector with respect to the same characteristics that are used for the decomposition exercise described in the previous section. The next step is to calculate the outcome variables for both groups. The estimate for the effect is given as the mean of the differences between the values of the matched observations.

The PSM exercise is implemented in Stata using the *kmatch* command. The "average treatment effect on the treated" (ATT) calculated here is the amount by which organised units differ in per unit or per worker productivity from their matched unorganised sector counterparts.

4 Results

4.1 Raw differences in productivity across sectors

We start by showing the distributions of GVA per unit and GVA per worker for the full uncensored ASI-NSSO combined datasets as well as the small units dataset. All subsequent results pertain to the small units dataset. Figure 1 shows the distribution of GVA per unit in the organised (blue)

and unorganised (red) sectors for the full dataset and Figure 2 shows the same only for the censored small unit dataset. As expected, the distributions move closer after limiting the dataset to units employing 5 to 49 workers. The overlap is stronger for labour productivity (Figure 3 and Figure 4. Indeed, looking at the two distributions in Figure 4 it is clear that small units in the organised and the unorganised sector do look substantially similar to each other at least in terms of value-added per worker.

[Figures 1 to 4 here]

Table 3 provides a wider range of descriptive statistics for organised and unorganised units. As expected, the average small unit in the organised sector is larger (18 workers versus 8 workers), better endowed with capital (nearly ten times larger) and just over three times more productive than its unorganised counterpart in terms of value added per worker. It also pays significantly higher wages (around 1.8 times). The fact that the jump in labour productivity is much higher than the jump in wages indicates that profitability is higher in the organised sector.

[Table 3 here]

Going beyond the organised-unorganised dichotomy, Figure 5 shows that even within the unorganised sector, there is substantial heterogeneity in performance. NSSO units that report maintaining accounts and being registered under some Act, look more similar to ASI units at least as far as the raw distributions are concerned. In fact, one sees a clear ordering as the distributions move from left to right with increasing degree of "organised-ness." Table 4 presents the same set of descriptive statistics as shown in the previous table, this time by degree of organisedness. Two points are worth noting. First, there is significant variation within the unorganised sector as seen in the increase in unit size, capital stock, and labour productivity across degrees 1, 2, and 3. But there is also a sharper increase associated with the movement to ASI units (degree 4). For example, organised degree 3 units (i.e. those that maintain accounts and are registered under some Act) are nearly twice as productive in terms of GVA per worker than those that do not do either (degree 1), while the average ASI unit is 2.4 times more productive than the degree 3 unit. As before, the increase in the wage rate with degree of organisedness is proportionately smaller than the increase in labour productivity, indicating that profitability grows with organisedness.

Intriguingly, the proportion of units located in urban areas rises steadily with organised degree from less than 60% for degree 1 to 84% for degree 3, but then falls again to 62% for the organised sector units. One possible explanation for this is that maintenance of accounts or registration may require unorganised unit entrepreneurs to have some basic level of education that is more widespread in urban areas while the location of ASI units follows a different logic altogether. We do not explore this issue further, but it merits a closer look.

[Figure 5 here]

[Table 4 here]

4.2 OLS estimates of the organised premium

Table 5 presents the OLS results. The main coefficient of interest is the one for the organised dummy, which we call the organised premium. The outcome variable is log GVA per unit. As expected, adding controls reduces the raw productivity gap till the premium stabilises around 50%.

[Table 5 here]

The other covariates have expected signs. Column 9 presents the final model with all the controls including industry and state fixed effects. The elasticity of value-added with respect to capital is around 0.2 to 0.25 while that for labour is around 0.85 to 0.9. Note that the fact that the labour elasticity is less than 1 means that a regression of value-added per worker on these same covariates gives a negative coefficient for log workers while keeping all other cofficients the same (see Appendix Table 3).[12] This result has been reported in the literature as a negative relationship between firm size and productivity (Amin and Islam, 2015). That is, controlling for capital stock, firms with more workers display lower levels of labour productivity.

Note also that, as expected, home-based units are less productive than non-home based ones (by around 20%). The urban premium is substantial at 17%. Finally, units with a higher proportion of women workers face a significant penalty of over 70%. Whether this is because units that employ a larger proportion of women are women-owned and hence suffer from discrimination in factors and product markets or because of other factors such as labour time and skill differences is difficult to determine given data limitations (for e.g. ASI does not report gender of the owner and neither

[12] This is a mathematical consequence of the fact that ln(GVA/workers) = ln(GVA) - ln(workers).

dataset reports educations levels etc for workers).

The above analysis allows us to estimate the difference in per unit value added across the two sectors controlling for some obvious factors. However, it has a few limitations. First the binary variable treats all unorganised and organised units as identical. Second, it forces the coefficients such as capital and labour elasticities to be the same for both types of units. We address the first issue here and the next issue in the subsequent section.

We address the issue of diversity within the unorganised sector by bringing in the variable "degree of organisedness." See Section 3 for the construction of this variable. Table 6 reproduces the full model for the organised dummy and in addition also shows the results for a regression with the degree of organisedness as the main covariates of interest. Note that each increase in the degree of organisedness is accompanied by a statistically significant and economically meaningful jump in value-added per unit, controlling for the usual factors. So unorganised sector manufacturing units that either maintained some accounts or were registered under some Act (but not both) were 6% more productive than those that did not do either, controlling for capital stock, labour and other variables. Further, those unorganised manufacturing units that did both (maintained accounts and registered) were 12% more productive than those that did neither. This exercise points us to the fact that the average organised premium hides substantial variation within the unorganised sector.

[Table 6 here]

4.3 Decomposing the organised-unorganised gap

The foregoing analysis took the approach of assuming that returns to various factors were identical for organised and unorganised units, and then estimated the size of the organised premium. We now turn to the results of the Blinder-Oaxaca decomposition which allows us to answer the question, what part of the productivity gap is accounted for by endowment differences between organised and unorganised units?

The outcome variable is log GVA per unit and the covariates used are the same as in the OLS regression. The Blinder-Oaxaca approach stratifies the OLS model by the organised dummy thereby allowing all the coefficients to vary by type of unit. Before moving to the full model, we note that a basic decomposition with only labour and capital as endowments shows 72% of the gap being

explained by endowment differences and 28% by returns to labour and capital (results not shown). Table 7 shows the results for the ful model where in addition to labour and capital endowments we also account for the unit's location (rural/urban and State), worker composition (unpaid and female share) and industry of operation. Of the productivity gap of 1.7 log units, the major portion (63%) is still accounted for by observable characteristics (endowments). But this leaves significant room for returns and the interaction terms (23% and 13.5% respectively).

[Table 7 here]

Recall that our decomposition is from the viewpoint of unorganised sector units. That is, the endowments component measures the expected change in the unorganised sector unit outcome, if these units had the same characteristics as organised units. The returns component measures the expected change in unorganised sector mean outcome, if these units had organised sector coefficients.

It is also instructive to examine the main covariates individually. We see that, as expected, labour and capital differences account for the major portion of the endowments component (76%). Labour also contributes significantly and positively to the returns component. But note that the capital stock variable has a negative sign indicating that returns to capital are much higher for unorganised units while returns to labour are higher for organised units. This is reminiscent of the literature on farm size and productivity which generally finds that land productivity is higher on smaller farms while labour productivity is higher on larger farms (Griffin et al., 2002). We discuss the implications of this findings in the next section.

It may seem unusual to consider State or industry of operation as an "endowment" on par with labour or capital. The logic for including them is that we would like to know how far observable characteristics go in explaining the performance gaps. Including the industry, for example, controls for the fact that unorganised sector units are distributed differently across different industries as compared to organised sector units. Thus we would like to compare both type of units within an industry, rather than comparing units in different industries. The same goes for State of operation and other controls.

Finally, note that the returns component can include a number of important but unobserved unit-level characteristics such as managerial ability and quality of labour or entrepreneurship as well

as social factors such as discrimination. Thus the results of our decomposition can be interpreted from two different perspectives. The fact that differences in endowments explains 63% of the gap leaves significant room for the above factors to play a role. But it also means that the larger part of the productivity gap is explained by the fact that unorganised units are more poorly endowed than organised units. We discuss the policy implications of both viewpoints in the next section.

4.4 Productivity differentials between matched units

The foregoing analysis has compared small units employing between 5 and 49 workers in the organised and unorganised manufacturing sectors. As mentioned earlier, a limitation of this analysis is that, since the median units in both sectors remain substantively different from each other, there is a danger that the measured effects rely on counterfactuals that do not exist. One way around this is to create a more comparable set of "control" units from the unorganised sector that closely resemble the "treated" units in the organised sector using Propensity Score Matching (PSM). The quotation marks are there to remind us that there is no exogenous treatment here and hence the measured effects are not causal.

Figures 6 and 7 depict the consequences of deploying a PSM approach on the distributions of two key variables, capital and labour. The covariates used to generate the propensity scores are the same as the ones used in the decomposition described earlier. As can be seen, the matched units are much closer together than they were prior to matching. The average difference in GVA per unit between organised and unorganised units after propensity score matching is 0.43 log units. This accounts for 24% of the raw difference.

[Figures 6 and 7 here]

This compares well in magnitude to the returns component of the Blinder-Oaxaca decomposition reported in the earlier section. The similarity in the two estimates indicates that missing counterfactuals may not be a significant concern. Thus from both the methods, Blinder-Oaxaca as well as PSM, we come to the conclusion that around 23-24 percent of the gap in gross value added per unit between organised and unorganised units is not explained by the differences in their endowments or observed characteristics.

5 Discussion

We now discuss the wider implications of the results presented here and also deal with some shortcomings of the study.

Several prior studies have addressed the question of productivity (both partial and total) and performance gaps between the formal and informal or, in the Indian context, organised and unorganised, sectors. In the Indian case, where firm level panel datasets are usually not available, TFP studies have taken the form of industry-level analyses. However, such analyses necessitate the combining of potentially very different units into industry-level average characteristics and outcomes. Conversely, firm or factory level studies circumvent this problem but suffer from the lack of panel data as well as from a lack of unit level data for the entire manufacturing sector.

One of our contributions lies in the construction of a combined unit-level dataset that encompasses the entire Indian manufacturing sector at one point in time (2015). The obstacles posed in integrating ASI data with NSSO data have been discussed in the paper and solutions have been provided. Of course, the use of cross-section rather than panel data only allows us to estimate partial productivity measures.

A second caveat is that the distinction between productivity and profitability is important to keep in mind. Registration imposes higher costs, including higher wages and benefits, compliance with standards, maintaining records, as well as potential bribes to inspectors. Thus higher productivity need not straightforwardly imply higher profitability. As has been recognised in the literature, this may make small firms reluctant to register (Amirapu and Gechter, 2020; Bhattacharjea, 2021; Raj SN and Sen, 2019). Though it is interesting to note that while organised units pay higher wages, controlling for unit size, capital endowment and other factors, the increase in wages is less than the increase in productivity. Though we have not examined profits in this study, this finding suggests that organised small units are not only more productive, they are also more profitable than unorganised units. In addition at the smaller end of our chosen firm size spectrum (less than ten workers), profits may not be large enough in absolute terms to enable meaningful investments.[13]

Thirdly, since firm characteristics are equally causes as consequences of a unit being organised or

[13]We are indebted to Aditya Bhattacharjea for these points.

unorganised and since units are not randomly assigned to a sector, the empirical approach employed here does not allow us to make causal claims regarding the impact of being organised. Rather, our purpose is to quantify the extent to which units in the relatively more dynamic part of the unorganised sector come close to productivity performance observed in organised units, when we match them on observed characteristics. This exercise is of importance because it improves our understanding of the dynamic part of the unorganised sector and allows us to estimate its potential in driving productivity growth as well as job creation.

The results presented here show that differences in observed characteristics or endowments such as the factor inputs as well as the state in which the unit is located or industry of operation account for the major part of the productivity difference between organised and unorganised units. However, it is also true that unobserved or uncontrolled for factors such as local geography (type of town or city, neighbourhood), managerial or entrepreneurial ability, quality of labour and discrimination in factor or product markets account for a substantial part of the productivity differential.

Since capital productivity is higher for unorganised units, one implication of reported findings is that policy support to unorganised firms that make makes it easier for them to deploy capital (e.g. easing credit constraints) can have significant effects even if unobserved factors still matter for performance. Conversely the fact that labour productivity is higher for organised units suggests possible labour constraints, but we focus here on the unorganised sector implications.

The potential for job creation among unorganised small firm is large as shown by the numbers in Table 2. Putting the size of the sector together with our findings on its performance suggests one possible policy path to large-scale job creation in the Indian manufacturing sector. Since the absolute numbers of unorganised manufacturing units are nine times more than organised units, even with comparable output elasticity of employment, the number of jobs created for a given amount of growth will be much larger in the unorganised sector.

Finally, it is important to distinguish policy that aims to foster growth of the modern segment of the unorganised sector from policy that aims to formalise informal firms. The former is about enabling scale-up and employment generation with formalisation being the by-product. For example, if absence of local public infrastructure such as electricity and water supply, security, and roads prevent unorganised firms from growing, supplying such infrastructure does not require unorganised

firms to formalise. Indeed, a lack of public infrastructure is an often-cited reason for failure of Indian manufacturing to expand (R. Nagaraj, 2017).

Further, as we saw with the analysis of the "degree of organisedness", there is considerable heterogeneity within the unorganised sector with units registered under some Government body tending to perform better. Though in many cases such registration is in name only and has no material consequences for the firm, it is possible to imagine support measures such as subsidised credit or other inputs being targeted via the information that such registration provides. This does not require the unorganised unit to become any more formal than it already is. In contrast, formalisation can simple be achieved by bringing nano and micro units into the regulatory net. But does this enable their scale-up and productivity growth is really the relevant question from a structural transformation perspective.

6 Conclusion

The manufacturing sector has played a key role in nearly every successful structural transformation in the past two hundred years. Even as the modern services sector plays a more substantive role in this process, manufacturing retains its relevance. Enabling firm size, productivity, and employment growth in this sector is a major policy priority for an economy such as India where the landscape is dominated by micro and small enterprises.

In this paper we have drawn attention to the relatively larger unorganised sector units in the Indian manufacturing sector as potential vehicles of transformation. Though proportionately tiny (around 5 percent of all unorganised units), they are far more numerous in absolute terms than similar sized units in the organised sector. We have shown that in terms of partial productivity measures, these unorganised units matched closely on observable characteristics to their organised counterparts, do bridge a large part of the productivity gap.

The principle policy conclusion is that, with the right kind of policy support, this sub-sector of dynamic or profit-oriented unorganised manufacturing units presents significant possibilities for employment generation and structural transformation.

References

Allen, J., Nataraj, S., & Schipper, T. C. (2018). Strict duality and overlapping productivity distributions between formal and informal firms. Journal of Development Economics, 135, 534–554.

Amaral, P. S., & Quintin, E. (2006). A competitive model of the informal sector. Journal of Monetary Economics, 53(7), 1541–1553.

Amin, M., & Islam, A. (2015). Are large informal firms more productive than the small informal firms? evidence from firm-level surveys in africa. World Development, 74, 374–385.

Amirapu, A., & Gechter, M. (2020). Labor regulations and the cost of corruption: Evidence from the indian firm size distribution. Review of Economics and Statistics, 102(1), 34–48.

Amirapu, A., & Subramanian, A. (2015). Manufacturing or services? an indian illustration of a development dilemma. Center for Global Development Working Paper 408.

Báez-Morales, A. (2015). Differences in efficiency between formal and informal micro firms in mexico. AQR Working Papers 15/10.

Baldwin, J. R., Jarmin, R. S., & Tang, J. (2002). The trend to smaller producers in manufacturing: A canada/us comparison. Statistics Canada, Economic Analysis Series 1F0027MIE, (003).

Bartelsman, E., Haltiwanger, J., & Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. American Economic Review, 103(1), 305–334.

Basole, A., Abraham, R., Shrivastava, A., & Jayadev, A. (2018). State of working india (tech. rep.). Centre for Sustainable Employment. Azim Premji University.

Basole, A., Basu, D., & Bhattacharya, R. (2015). Determinants and impacts of subcontracting: Evidence from india's unorganized manufacturing sector. International Review of Applied Economics, 29(3), 374–402.

Bedi, J. S., & Banerjee, P. K. (2007). Discrepancies and validation of indian manufacturing data. Economic and Political Weekly, 883–891.

Benjamin, N. C., & Mbaye, A. A. (2012). The informal sector, productivity, and enforcement in w est a frica: A firm-level analysis. Review of Development Economics, 16(4), 664–680.

Bhattacharjea, A. (2021). Labour market flexibility in indian manufacturing: A critical survey of the literature. International Labour Review, 160(2), 197–217.

Bhattacharya, R., Bhattacharya, S., & Sanyal, K. K. (2013). Dualism in the informal economy:

Exploring the indian informal manufacturing sector. Development and Sustainability, 339–362.

Dabla-Norris, E., Gradstein, M., & Inchauste, G. (2008). What causes firms to hide output? the determinants of informality. Journal of Development Economics, 85(1-2), 1–27.

De Soto, H., et al. (1989). The other path. Harper & Row New York.

Dimova, R. D., Nordman, C. J., & Roubaud, F. (2008). Allocation of labour in urban west africa: Implication for development policies. IZA Discussion Paper.

Djankov, S., Glaeser, E., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2003). The new comparative economics. Journal of Comparative Economics, 31(4), 595–619.

Fajnzylber, P., Maloney, W. F., & Montes-Rojas, G. V. (2009). Releasing constraints to growth or pushing on a string? policies and performance of mexican micro-firms. Journal of Development Studies, 45(7), 1027–1047.

Farrell, D. (2004). Hidden dangers. The McKinsey Quarterly, 3, 26.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In Handbook of labor economics (pp. 1–102). Elsevier.

Gelb, A., Mengistae, T., Ramachandran, V., & Shah, M. K. (2009). To formalize or not to formalize? comparisons of microenterprise data from southern and east africa. CGDev Working Paper 175.

Ghani, E., & Kharas, H. (2010). The service revolution (tech. rep.). The World Bank. World Bank, Washington, DC.

Ghose, A. K. (2020). Structural transformation of india's economy. IHD Working Paper 01-2020.

Griffin, K., Khan, A. R., & Ickowitz, A. (2002). Poverty and the distribution of land. Journal of Agrarian Change, 2(3), 279–330.

Gupta, G., & Basole, A. (2020). India's information technology industry: Prospects for growth and role in structural transformation. Decision, 47(4), 341–361.

Hasan, R., & Jandoc, K. R. (2010). The distribution of firm size in india: What can survey data tell us? Asian Development Bank Economics Working Paper Series, (213).

Jann, B., et al. (2008). A stata implementation of the blinder-oaxaca decomposition. Stata Journal, 8(4), 453–479.

Kannan, K., & Raveendran, G. (2009). Growth sans employment: A quarter century of jobless growth in india's organised manufacturing. Economic and Political Weekly, 80–91.

Kapoor, R. (2022). Stylized facts on the evolution of the enterprise size: Distribution in india's manufacturing sector. ICRIER Working Paper 409.

Kathuria, V., RAJ SN, R., & Sen, K. (2010). Organised versus unorganised manufacturing performance in the post-reform period. Economic and Political Weekly, 55–64.

Kesar, S., & Bhattacharya, S. (2020). Dualism and structural transformation: The informal manufacturing sector in india. The European Journal of Development Research, 32, 560–586.

Krishna, K., Goldar, B., Aggarwal, S. C., Das, D. K., Erumban, A. A., Das, P. C., et al. (2018). Productivity growth and levels in formal and informal manufacturing in india. Centre for Development Economics Working Paper No. 291.

Kumari, R. (2022). Differences in efficiency in the formal–informal dichotomy—a study of indian firms. The Indian Economic Journal, 70(2), 289–312.

La Porta, R., & Shleifer, A. (2008). The unofficial economy and economic development. NBER Working Paper No. w14520.

Levy, S. (2010). Good intentions, bad outcomes: Social policy, informality, and growth in mexico. Brookings Institution Press.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. The Manchester School, 22, 139–191.

Marcouiller, D., & Young, L. (1995). The black hole of graft: The predatory state and the informal economy. The American Economic Review, 85(3), 630–646.

McKenzie, D., & Sakho, Y. S. (2010). Does it pay firms to register for taxes? the impact of formality on firm profitability. Journal of Development Economics, 91(1), 15–24.

Mehrotra, S., & Giri, T. (2019). The size structure of india's enterprises: Not just the middle is missing. Centre for Sustainable Employment Working Paper 25.

Nagaraj, D. R. (2018). Of "missing middle" and size-based regulaton: A new fronter in the labour market flexibility. Centre for Sustainable Employment Working Paper 7.

Nagaraj, R. (2017). Economic reforms and manufacturing sector growth: Need for reconfiguring the industrialisation model. Economic and Political Weekly, 52(2), 61–68.

Nagaraj, R. (Ed.). (2021). Industrialisation for employment and growth in india. Cambridge University Press.

Nelson, E. G., & De Bruijn, E. J. (2005). The voluntary formalization of enterprises in a developing

economy—the case of tanzania. Journal of International Development, 17(4), 575–593.

Perry, G. (2007). Informality: Exit and exclusion. World Bank Publications.

Rabbani, G., & Raj SN, R. (2023). Small firms in indian manufacturing: Some stylized facts. Journal of the International Council for Small Business, 1–13.

Raj SN, R., & Sen, K. (2019). What do we know about firms in the informal manufacturing sector in india? Globalization, Labour Market Institutions, Processes and Policies in India, 47–86.

Rand, J., & Torm, N. (2012). The benefits of formalization: Evidence from vietnamese manufacturing smes. World development, 40(5), 983–998.

Ranis, G., & Stewart, F. (1999). V-goods and the role of the urban informal sector in development. Economic Development and Cultural Change, 47(2), 259–288.

Schmitz, H., & Nadvi, K. (1999). Clustering and industrialization: Introduction.

Shekar, K. C., & Nataraj, M. (2022). Subcontracting and enterprise development in india's informal manufacturing sector. Journal of Small Business and Enterprise Development, (ahead-of-print).

Straub, S. (2005). Informal sector: The credit market channel. Journal of Development Economics, 78(2), 299–321.

Taymaz, E., et al. (2009). Informality and productivity: Productivity differentials between formal and informal firms in turkey. Country Economic Memorandum (CEM).

Thomas, J. J. (2013). Explaining the 'jobless' growth in indian manufacturing. Journal of the Asia Pacific Economy, 18(4), 673–692.

Ulyssea, G. (2010). Regulation of entry, labor market institutions and the informal sector. Journal of Development Economics, 91(1), 87–99.

Van Ark, B., & Monnikhof, E. (1996). Size distribution of output and employment: A data set for manufacturing industries in five oecd countries, 1960s-1990. OECD Economics Department Working Papers, (166).

Van Biesebroeck, J. (2005). Firm size matters: Growth and productivity growth in african manufacturing. Economic Development and Cultural Change, 53(3), 545–583.

Tables

Table 1: A comparison of ASI and NSSO survey procedure

Variable	ASI	NSSO
Ref period	Accounting year	Last 30 days or past year.
Survey unit	Factory	Enterprise
Geography	Rural/urban, state	Same
Ownership	Prop/part/trust/private	Similar
Industry	NIC 2008	Same
Labour	Positions not workers	Same
Capital	Book value	Market value
Value-added	Expenses minus receipts	Same
Wages	Payments to hired workers	Same

Source: ASI and NSSO Instruction Manuals

Table 2: A comparison of ASI and NSSO unit characteristics

Characteristic (2015)	ASI	NSSO
Full dataset		
Total workers (millions)	14	35
Median workers	21	1
Median GVA per worker (INR/year)	263,188	39,804
% small firms (5-49 workers)	66	5
Small firms dataset		
Estimated number of small firms	102,078	914,376
Total workers (millions)	1.8	7.43
Hired workers (millions)	1.8	5.69
Median workers	15	6
Median GVA per worker (INR/year)	250,344	108,077
Share of total workers (%)	20	80
Share of hired workers (%)	24	76
Share of GVA (%)	46	54
Share of capital (%)	54	46

Source: Authors' estimates based on ASI 2015-16 and NSSO 2015-16

Table 3: Descriptive Statistics for Organised and Unorganised Units

	Organised	Unorganised
% urban	61.9	72.0
% home based	0	22.68
Number of workers	18.4	8.1
	(11.4)	(4.8)
Hired workers	17.9	6.2
	(11.6)	(5.0)
Female share	0.12	0.17
	(0.23)	(0.26)
Capital stock (INR lakhs)	88.61	8.06
	(419.78)	(19.98)
Annual GVA per unit (INR lakhs)	85.36	11.02
	(721.80)	(13.32)
Annual GVA per worker (INR lakhs)	4.22	1.30
	(36.76)	(1.06)
Annual wage rate (INR lakhs)	1.55	0.86
	(4.38)	(0.43)

Note: Standard deviation in parenthesis.

Table 4: Descriptive Statistics for Organised Degree

	Degree			
	1	2	3	4
% urban	57.8	72.7	83.3	61.9
% home based	38.2	24.0	8.4	0
Number of workers	6.9	7.5	9.6	18.4
	(3.1)	(4.1)	(5.9)	(11.4)
Hired workers	4.6	5.6	8.1	17.9
	(3.7)	(4.3)	(5.9)	(11.6)
Female share	0.23	0.17	0.12	0.12
	(0.28)	(0.27)	(0.22)	(0.23)
Capital stock	1.43	3.47	17.48	88.61
(INR lakhs)	(3.23)	(6.03)	(29.77)	(419.78)
Annual GVA per unit	6.34	8.59	17.01	85.36
(INR lakhs)	(8.33)	(8.03)	(17.42)	(721.80)
Annual GVA per worker	0.90	1.16	1.76	4.23
(INR lakhs)	(0.72)	(0.82)	(1.29)	(36.77)
Annual wage rate	0.72	0.85	0.97	1.55
(INR lakhs)	(0.39)	(0.41)	(0.44)	(4.38)

Note: Standard deviation in parenthesis.

Table 5: OLS Estimates of Organised Premium for 2015-16)

	Dependent variable:								
	Log GVA per unit								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Organised	1.69***	1.73***	0.92***	0.49***	0.47***	0.46***	0.48***	0.50***	0.50***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Urban		0.43***	0.15***	0.26***	0.26***	0.18***	0.18***	0.19***	0.17***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log capital stock			0.34***	0.26***	0.23***	0.21***	0.21***	0.23***	0.21***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Log workers				0.87***	0.87***	0.92***	0.93***	0.91***	0.92***
				(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Home based					−0.35***	−0.32***	−0.33***	−0.28***	−0.17***
					(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female share						−0.68***	−0.65***	−0.63***	−0.76***
						(0.01)	(0.01)	(0.02)	(0.02)
Type of ownership									
Partnership							−0.15***	−0.14***	−0.12***
							(0.01)	(0.01)	(0.01)
Others							−0.99***	−0.94***	−0.93***
							(0.06)	(0.06)	(0.06)
Private							0.13***	0.09***	0.14***
							(0.03)	(0.03)	(0.03)
Constant	13.51***	13.20***	9.31***	8.50***	8.88***	9.19***	9.21***	9.23***	9.38***
	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.08)
Industry (NIC 3)	No	No	No	No	No	No	No	Yes	Yes
State	No	No	No	No	No	No	No	No	Yes
Observations	30,217	30,217	29,965	29,965	29,965	28,414	28,414	28,295	28,295
R^2	0.21	0.24	0.55	0.66	0.67	0.69	0.70	0.72	0.74
Adjusted R^2	0.21	0.24	0.55	0.66	0.67	0.69	0.70	0.72	0.74

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Organised premium compared with degree of organised-ness

	Dependent variable:	
	Log GVA per unit	
	(1)	(2)
Organised	0.50***	
	(0.02)	
Organised degree 2		0.06***
		(0.01)
Organised degree 3		0.12***
		(0.01)
Organised degree 4		0.60***
		(0.02)
Urban	0.17***	0.15***
	(0.01)	(0.01)
Log capital stock	0.21***	0.20***
	(0.003)	(0.003)
Log workers	0.92***	0.91***
	(0.01)	(0.01)
Home based	−0.17***	−0.16***
	(0.01)	(0.01)
Female share	−0.76***	−0.75***
	(0.02)	(0.02)
Type of Ownership		
Partnership	−0.12***	−0.13***
	(0.01)	(0.01)
Others	−0.93***	−0.96***
	(0.06)	(0.06)
Private	0.14***	0.15***
	(0.03)	(0.03)
Constant	9.38***	9.45***
	(0.08)	(0.08)
Industry (NIC 3)	Yes	Yes
State	Yes	Yes
Observations	28,295	28,295
R^2	0.74	0.74
Adjusted R^2	0.74	0.74
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 7: Results of the Blinder-Oaxaca Decomposition

	Value	Percentage
Difference	1.72	100.0
Endowments	1.08	63.1
Coefficients	0.40	23.4
Interaction	0.23	13.5
Endowments		
Log workers	0.62	56.9
Log capital stock	0.49	45.4
Urban	-0.01	-1.1
Unpaid share	0.06	5.3
Female share	0.03	3.1
Proprietorship	-0.23	-20.9
Partnership	0.05	4.9
Others	-0.02	-1.5
Private	0.11	9.8
Coefficients		
Log workers	0.15	37.9
Log capital stock	-0.35	-88.2
Urban	-0.04	-9.5
Unpaid share	-0.01	-3.3
Female share	0.04	10.7
Proprietorship	-0.42	-105.4
Partnership	-0.02	-3.9
Others	0.00	0.4
Private	0.00	0.0
Interaction		
Log workers	0.06	23.8

Log capital stock	-0.07	-30.0
Urban	0.00	1.8
Unpaid share	0.01	3.4
Female share	-0.01	-5.7
Proprietorship	0.27	115.1
Partnership	-0.06	-23.9
Others	0.02	6.6
Private	-0.06	-27.4

Note: See text for details of decomposition.

Figures

Figure 1: Full distribution of GVA per unit

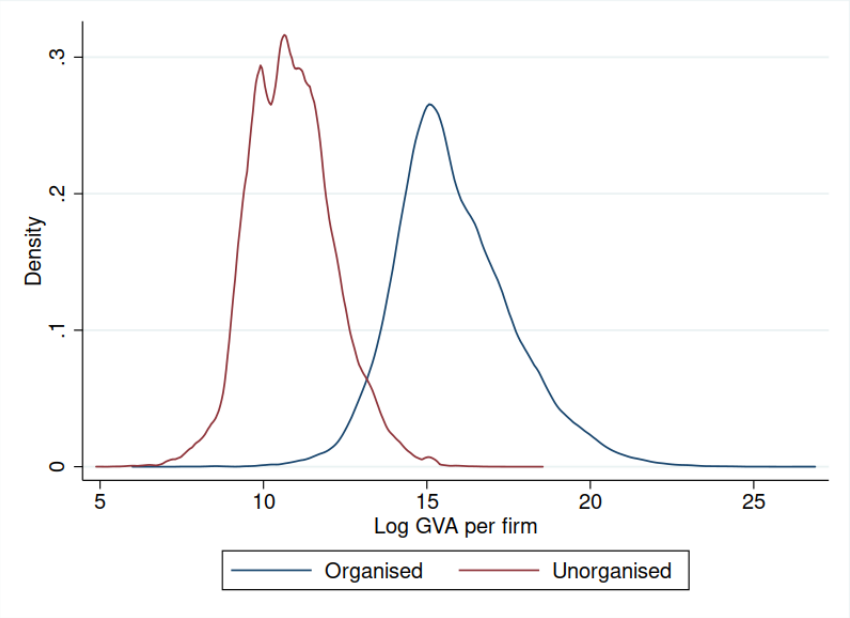


Figure 2: Distribution of GVA per firm among small firms

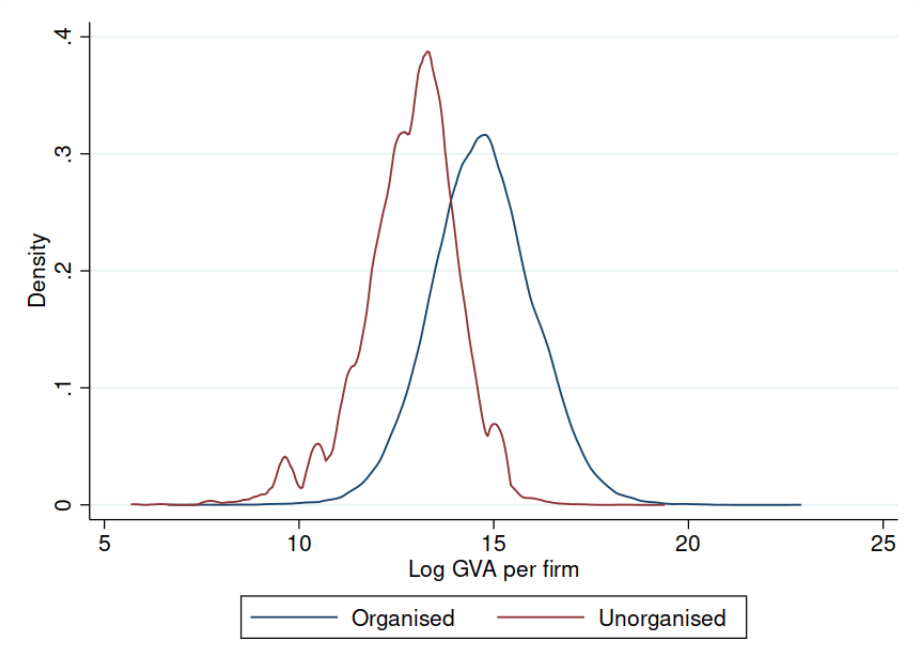


Figure 3: Full distribution of GVA per worker

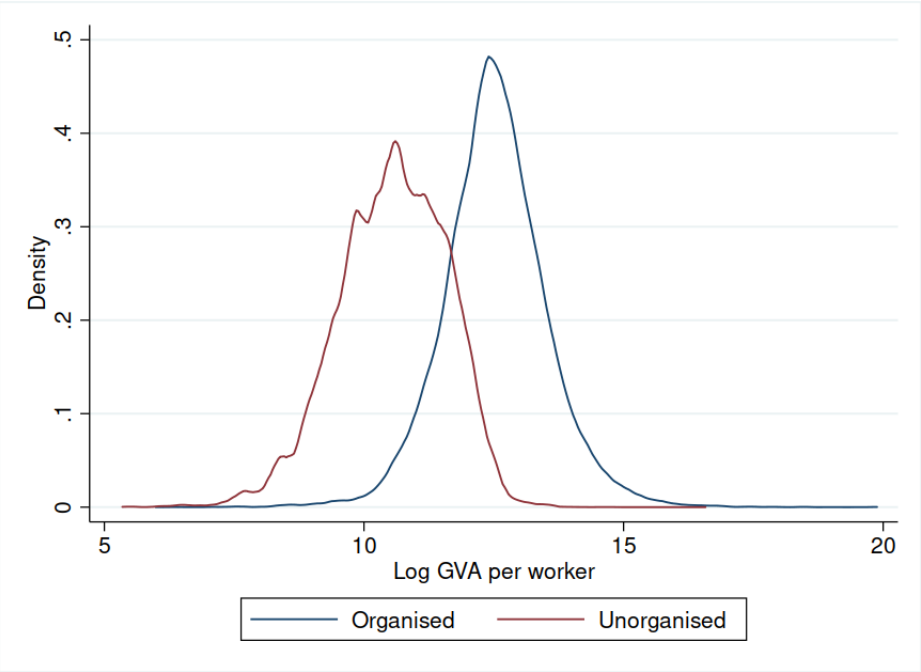


Figure 4: Distribution of GVA per worker among small firms

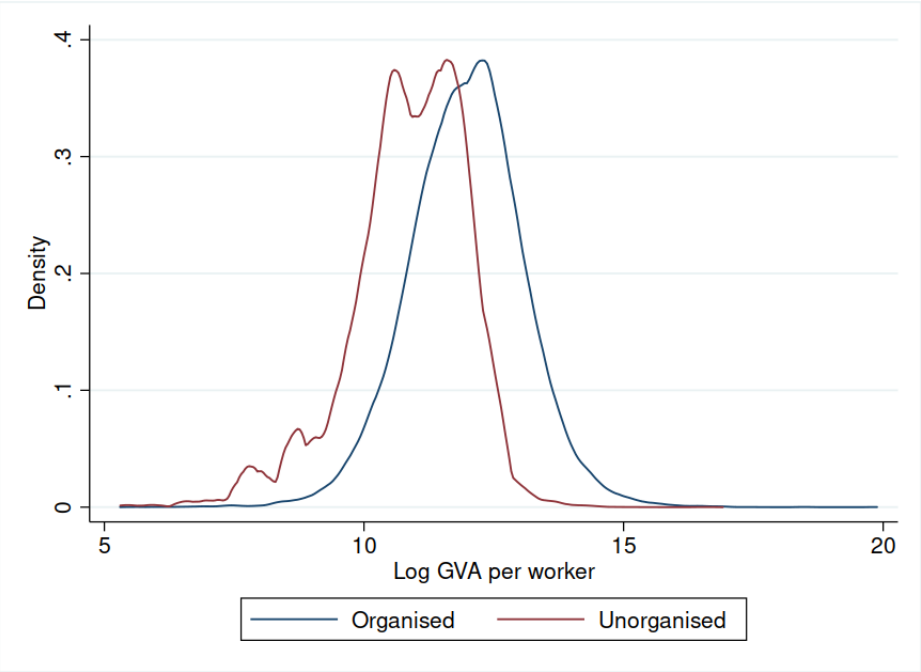


Figure 5: Distribution of GVA per worker over organised degree

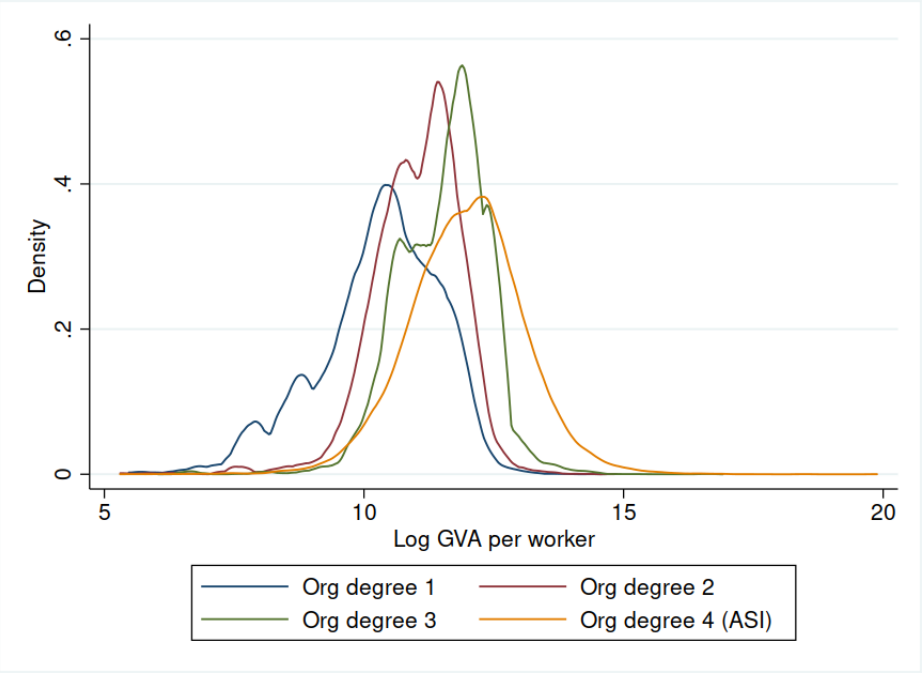


Figure 6: Distribution of log capital per unit before and after matching

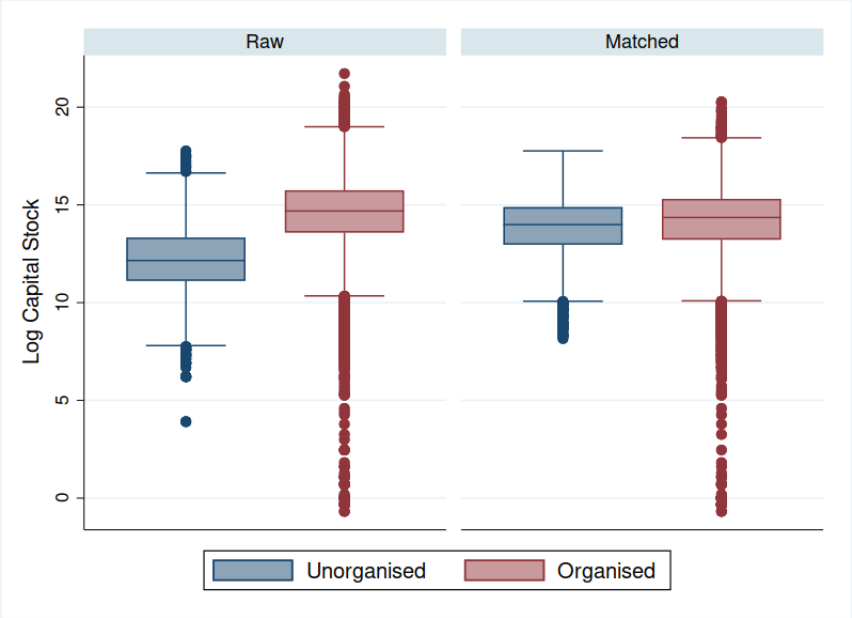
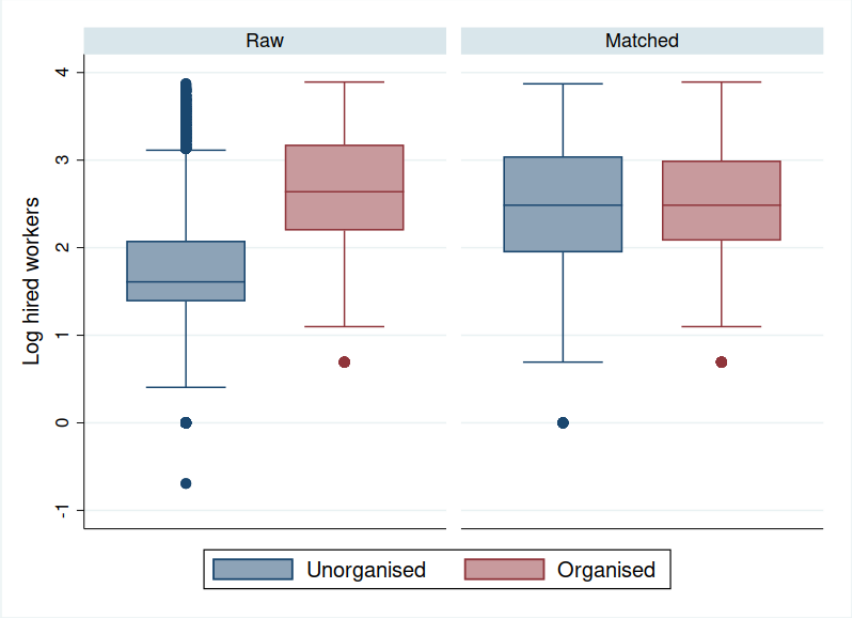


Figure 7: Distribution of log hired workers per unit before and after matching



Appendix Tables and Figures

Figure 1: Distribution of market value to book value ratio at NIC 3 digit level from Prowess IQ

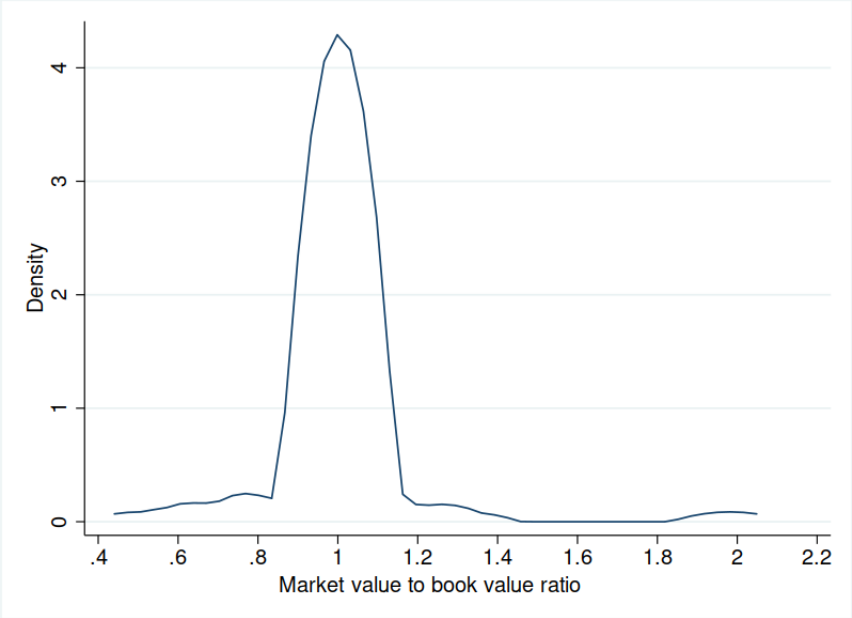


Table 1: GVA: Comparing old and new capital variable

	Dependent variable:			
	ln_gva			
	Old	New	Old	New
organised	0.50***	0.50***		
	(0.02)	(0.02)		
factor(formal_deg)2			0.06***	0.06***
			(0.01)	(0.01)
factor(formal_deg)3			0.12***	0.12***
			(0.01)	(0.01)
factor(formal_deg)4			0.60***	0.60***
			(0.02)	(0.02)
urban	0.17***	0.17***	0.15***	0.15***
	(0.01)	(0.01)	(0.01)	(0.01)
ln_capital_notlb	0.21***		0.20***	
	(0.003)		(0.003)	
ln_adj_K_3		0.21***		0.20***
		(0.003)		(0.003)
ln_workers	0.91***	0.92***	0.91***	0.91***
	(0.01)	(0.01)	(0.01)	(0.01)
home_based	−0.16***	−0.17***	−0.16***	−0.16***
	(0.01)	(0.01)	(0.01)	(0.01)
Femshare	−0.76***	−0.76***	−0.75***	−0.75***
	(0.02)	(0.02)	(0.02)	(0.02)
factor(Type_ownership)2	−0.12***	−0.12***	−0.13***	−0.13***
	(0.01)	(0.01)	(0.01)	(0.01)
factor(Type_ownership)3	−0.93***	−0.93***	−0.95***	−0.96***
	(0.06)	(0.06)	(0.06)	(0.06)
factor(Type_ownership)4	0.14***	0.14***	0.15***	0.15***
	(0.03)	(0.03)	(0.03)	(0.03)
Constant	9.37***	9.38***	9.43***	9.45***
	(0.08)	(0.08)	(0.08)	(0.08)
Industry Control	Yes	Yes	Yes	Yes
State Control	Yes	Yes	Yes	Yes
Observations	28,295	28,295	28,295	28,295
R^2	0.74	0.74	0.74	0.74
Adjusted R^2	0.74	0.74	0.74	0.74
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression Analysis on Young Firms

	Dependent variable:	
	ln_gva	
	(1)	(2)
organised	0.56***	
	(0.05)	
factor(formal_deg)2		0.09***
		(0.02)
factor(formal_deg)3		0.17***
		(0.02)
factor(formal_deg)4		0.69***
		(0.05)
urban	0.26***	0.24***
	(0.02)	(0.02)
ln_adj_K_3	0.16***	0.14***
	(0.01)	(0.01)
ln_workers	0.91***	0.89***
	(0.02)	(0.02)
home_based	−0.22***	−0.21***
	(0.02)	(0.02)
Femshare	−0.73***	−0.73***
	(0.03)	(0.03)
factor(Type_ownership)2	0.02	0.02
	(0.03)	(0.03)
factor(Type_ownership)3	−3.33***	−3.35***
	(0.15)	(0.15)
factor(Type_ownership)4	0.25***	0.28***
	(0.08)	(0.08)
Constant	10.41***	10.51***
	(0.27)	(0.27)
Industry control	*Yes*	*Yes*
State control	*Yes*	*Yes*
Observations	4,967	4,967
R^2	0.72	0.72
Adjusted R^2	0.71	0.71
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 3: OLS Estimates of Organised Premium (Labour Productivity)

	Dependent variable:						
	ln_gva_per_worker						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
organised	0.95***	1.00***	0.42***	0.49***	0.47***	0.48***	0.51***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
urban		0.48***	0.28***	0.26***	0.26***	0.24***	0.21***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ln_adj_K_3			0.24***	0.26***	0.23***	0.25***	0.24***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ln_workers				−0.13***	−0.13***	−0.15***	−0.15***
				(0.01)	(0.01)	(0.01)	(0.01)
home_based					−0.35***	−0.30***	−0.21***
					(0.01)	(0.01)	(0.01)
Constant	11.52***	11.18***	8.38***	8.50***	8.88***	8.98***	9.19***
	(0.005)	(0.01)	(0.02)	(0.03)	(0.03)	(0.05)	(0.08)
Industry control	*No*	*No*	*No*	*No*	*No*	*Yes*	*Yes*
State control	*No*	*No*	*No*	*No*	*No*	*No*	*Yes*
Observations	30,217	30,217	29,965	29,965	29,965	29,846	29,846
R^2	0.11	0.17	0.44	0.44	0.46	0.50	0.53
Adjusted R^2	0.11	0.17	0.44	0.44	0.46	0.50	0.53
Note:							*p<0.1; **p<0.05; ***p<0.01