Input Usage and Productivity in Indian Manufacturing Plants

Ejaz Ghani

William R. Kerr

Stephen D. O'Connell

The World Bank
Poverty Reduction and Economic Management Network
Economic Policy and Debt Unit
October 2013

Abstract

This paper analyzes the scale and productivity consequences of varied input use in Indian manufacturing using detailed plant-level data. Counts of distinct material inputs are higher in urban settings than in rural locations, unconditionally and conditional on plant size, and they are also higher in the organized sector than in the unorganized sector. At the district level, higher input usage in the organized sector is generally observed in wealthier districts and those with greater literacy rates. If looking within states, the usage is more

closely associated with electricity access, population density, and closer spatial proximity to one of India's largest cities. Plants in the organized sector utilizing a greater variety of inputs display higher productivity, with the effects mostly concentrated among smaller plants with fewer than 50 employees. For the unorganized sector, there is little correlation of input counts and local conditions, for better or for worse, and a more modest link to productivity outcomes.

This paper is a product of the Economic Policy and Debt Unit, Poverty Reduction and Economic Management Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at Eghani@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Input Usage and Productivity in Indian Manufacturing Plants

Ejaz Ghani, William R. Kerr and Stephen D. O'Connell

Keywords: Inputs, productivity, development, manufacturing, urban, India.

JEL: D24, L23, L25, L60, O10, O14, O17, O18, O40, R00, R10, R11, R12, R34

Author institutions and contact details: Ghani: World Bank, Eghani@worldbank.org; Kerr: Harvard University, Bank of Finland, and NBER, wkerr@hbs.edu; O'Connell: World Bank and CUNY Graduate Center, soconnell@gc.cuny.edu.

Acknowledgments: Funding for this project was provided by the World Bank. The views expressed here are those of the authors and not of any institution they may be associated with.

1. Introduction

An old adage goes, “You are what you eat.” While the original saying was mainly intended to provoke a better diet, the spirit of this adage sits deep within how economists view the determinants of firm productivity. In particular, economists have long recognized that access to material inputs and their qualities can influence the performance and success of firms. Not surprisingly, the sourcing of inputs thus drives the location choices of many firms. Moreover, many discussions of the differences in productivity across countries highlight the weak input conditions in developing or emerging economies that limit firm performance.

This paper uses establishment-level manufacturing data for India to trace out one feature of this process. Over the last two decades, endogenous growth theory has emphasized a particular gain for differentiated input access over and above the basic qualities of available inputs to the firm. The source of this productivity advantage comes from the fact that differentiated inputs allow firms to better match the needs that they have with the tools and materials they utilize. For example, it is possible for a firm both to cut a board and open up a can with a saw, and it is likely also possible (with enough time) to do both tasks if the firm only possesses a can opener. Nevertheless, we would generally anticipate firms to be more productive if they have access to both tools. The essential gain that the literature emphasizes is that this gain is not due to the better quality of the saw or the can opener, but instead their joint use. By having a greater range of potential inputs, firms are best able to match needs with inputs and thus encounter less diminishing productivity in their materials usage.¹

We empirically quantify these patterns in the Indian data with a focus on raw material inputs into the firm. We consider the organized and unorganized manufacturing sectors in the year 2000, and we evaluate spatial dimensions and productivity connections to varied input use. A very attractive feature of Indian data is the direct collection of input records for all plants at the establishment level. This allows us to simultaneously compare inputs to the local or industry conditions of firms and to consider productivity implications.

¹ Seminal theoretical contributions include Romer (1987, 1990), Rivera-Batiz and Romer (1991), and Grossman and Helpman (1991). Evidence at the firm level from the trade literature on these links includes Muendler (2004), Halpern, Koren, and Szeidl (2006), Amiti and Konings (2007), and Kasahara and Rodrigue (2008). Goldberg et al. (2010a,b) provide a detailed review of the sector- and product-level studies also undertaken.

We first document substantial differences across India in terms of input usage. Counts of distinct material inputs are higher in urban settings than in rural locations—unconditionally and also after conditioning on plant size—and they are also higher in the organized sector than in the unorganized sector. At the district level, higher input usage in the organized sector is generally observed in wealthier districts and those with greater literacy rates. If looking within states, the usage is more closely associated with electricity access, population density, and closer spatial proximity to one of India’s largest cities. Plants in the organized sector utilizing a greater variety of inputs display a higher productivity, with the effects mostly concentrated among smaller plants with fewer than 50 employees. For the unorganized sector, there is very little correlation of input counts and local conditions, for better or for worse, and a modest link to productivity outcomes. These conditions identify for policy makers that input access is not an abstract theoretical concept, but finds direct expression in Indian outcomes, especially for small firms. Policies that can aid the development of input access (e.g., trade reforms, transportation development) may directly aid the productivity of these firms. As we discuss later, our work may also highlight that one of the costs of unorganized firms is less productive use of inputs and scaling.

Our study is motivated by the pivotal work of Goldberg et al. (2010a,b). Goldberg et al. (2010a,b) identify how trade liberalization for India increased the number of imported inputs used by organized sector firms included in the Prowess database. These authors further show a link of this development to a larger number of products that the firm produced, with this metric showing one aspect of firm productivity. In many respects, we step back from these more advanced studies: rather than focus on the causal change in inputs within the organized sector due to a particular reform or event, we want to first characterize the differences in input access that exist spatially and across different sectors of the economy. To the best of our knowledge, neither of these dimensions has been studied in the Indian context, yet establishing these facts and perspectives is an important stepping stone for understanding how events like trade reform or large infrastructure projects like highway improvements (e.g., Datta 2011, Ghani, Goswami, and Kerr 2012b) will influence the distribution of activity across the economy. It also allows us to understand better the degree to which the growth seen in the organized sector by Goldberg et

al. (2010a,b) may translate into unorganized sector, which accounts for the vast majority of Indian employment.

Our work is more broadly related to recent studies of Indian manufacturing productivity development (e.g., Kathuria et al. 2010, Trivedi et al. 2011, Bollard, Klenow, and Sharma 2013). In particular, many scholars today argue that India still contains extensive misallocation of activity across plants and regions (e.g., Hsieh and Klenow 2009, Desmet et al. 2011). The patterns identified in this paper highlight an important difference across firms in their input access or usage, which then correlates with lower productivity. These patterns help open up lines of potential inquiry about the reasons for these differences and their persistence as sources of misallocation. In this paper we observe a limited cross-sectional role for local industrial diversity, but more complex spatial dynamics may still exist.²

The remainder of this paper is as follows: Section 2 describes the data used for this paper and Section 3 describes the broad patterns of input usage across sectors of Indian manufacturing. Section 4 analyzes the traits of districts and industries that are associated with higher input usage. Section 5 estimates the link between higher counts of distinct inputs and manufacturing productivity at the establishment level. The final section concludes and provides some thoughts about future work.

2. Indian Manufacturing Data and Usage Information

This study employs cross-sectional surveys of manufacturing establishments carried out by the government of India in 2000-2001 (hereafter, we will only refer to the initial year of 2000 for simplicity). The organized and unorganized sectors of Indian manufacturing are surveyed separately. The organized sector comprises establishments with more than 10 workers if the establishment uses electricity. If the establishment does not use electricity, the threshold is 20 workers or more. These establishments are required to register under the India Factories Act of 1948. The unorganized sector is, by default, comprised of establishments which fall outside the scope of the Factories Act. The organized sector accounts for over 80% of India's manufacturing

² The literature surrounding the specialization and diversity of cities includes Jacobs (1969), Porter (1990), Glaeser et al. (1992), Henderson (1997), Feldman and Audretsch (1999), Duranton and Puga (2000, 2001), and Lall et al. (2003). Henderson (2010) provides a broader review of the role of cities in developing countries.

output, while the unorganized sector accounts for over 80% and 99% of Indian manufacturing employment and establishments, respectively (Ghani, Kerr, and O’Connell 2013a).

The organized sector is surveyed by the Central Statistical Organization through the Annual Survey of Industries (ASI). Our data for the unorganized sector come from the National Sample Survey (NSS) Organization’s periodic “Survey of Unorganized Manufactures”. These surveys are used for many published reports on the state of Indian businesses and government agency monitoring of the Indian economy. With respect to establishment counts surveyed, the Indian sampling frame is about three times larger than that used with the Annual Survey of Manufacturing conducted in the United States.

Establishments are surveyed with state and four-digit National Industry Classification (NIC) stratification. The surveys provide sample weights to construct population-level estimates. We utilize the micro-level data directly when considering productivity estimations, and these estimations are weighted by these survey weights. We also undertake analyses at the district or district-industry level. For these estimations, we use the sample weights to first construct population-level estimates that can then be analyzed. Districts are administrative subdivisions of Indian states or territories that provide meaningful local economic conditions. The average district size is around 5,500 square kilometers—roughly twice the size of a U.S. county—and there is substantial variability in district size (standard deviation of ~5,500 square kilometers). Indian districts can be effectively considered as self-contained labor markets and, to some degree, economic units.

Our core sample contains 415 districts that are located in 31 Indian states.³ We set two criteria with respect to district size for inclusion: that the district has a population of at least one million in the 2001 census and has 10 or more establishments sampled. Finally, we drop from the sample any observations reporting zero or missing data in major input or output categories (total output/sales, total persons engaged, or total spending on intermediate inputs), or observations reporting an industrial classification outside of manufacturing. The exclusions are minor in terms of economic activity, and the resulting data account for over 95% of employment in the manufacturing sector.

³ The states of Arunachal Pradesh, Meghalaya and Sikkim and the Union territory of Lakshadweep are not present in our sample due to the ASI sampling frame not extending to these areas.

We use the two-digit level of the NIC system for industry variation. This level of aggregation contains 22 manufacturing industries. Discussions of India’s industrial landscape and its long-run transformation often use the terms “traditional” and “modern”, although there are no established or precise definitions of these groups. Following Ghani, Kerr, and O’Connell (2013a), we classify an industry as being modern if its unorganized sector share is less than the unweighted average unorganized share across manufacturing industries. We report several traits below that utilize this separation, and the notes to Table 3a document the industries included in the two groupings explicitly.⁴

The NSS surveys the ownership type of each establishment. Establishments can be listed as male proprietary, female proprietary, other owned, cooperative, household partnership, multi-household partnership, private LLC, and unknown. In some of our analyses of the unorganized sector, we separately consider establishments listed as either male proprietary or female proprietary. These two groups constitute 98% of establishments in the informal manufacturing sector in 2000. Ghani, Kerr, and O’Connell (2012) in particular emphasize the partitioned business networks in India for female-owned businesses, and this work extends these earlier studies to understand the implications for input usage.

Finally, and most important, we collect raw material input counts information from the survey questionnaires (ASI ‘00 - Block H, NSS ‘00 (56th Round) - Block 3). In each survey, respondents are asked to list the first five major raw material inputs into their production, providing product codes, quantities and values. From this block of information we extract the total number of reported inputs (with a maximum of five as per the survey questionnaires). The minimum input count is zero and the median input count is two inputs.

We focus on the raw count of distinct materials inputs in this paper. As we describe at many points below, we typically consider deviations of these input counts from the average pattern of plants in each industry to control for differences in typical production techniques across industries. We will also control directly for the intensity of material use as input for firms. Thus, the input count measure in our strictest estimations is modeling the variety of raw material

⁴ For clarity, we use term “sector” when discussing organized and unorganized sectors, “industries” when discussing the 22 manufacturing industries within manufacturing, and “groups” when describing “traditional” and “modern” industry groups.

inputs employed by the firm conditional on industry and the overall amount of material inputs consumed. As noted in the introduction, theory would suggest that productivity can be enhanced through lower diminishing returns through better suited inputs. By contrast, we do not model directly how much of each input is used or their relative concentrations (e.g., a Herfindahl-Hirschman Index) as we do not know the optimal mix of inputs for each industry. That is, it may be quite beneficial to an industry to have a small amount of a specific input, but it would never be the case that they would want equal proportions of that input to the others utilized.

3. Patterns of Input Usage in the Indian Manufacturing Sector

This section provides a broad descriptive foundation for input usage in Indian manufacturing. Panel A of Table 1a starts by documenting the overall level of input usage by sector of the economy. The first column provides descriptive statistics for the organized sector, while the second through fourth columns consider the unorganized sector. These descriptive statistics are unweighted representations of the raw survey data.

The mean level of input usage for plants in the organized sector is 2.60 inputs, with a standard deviation of 1.79. This level is higher than in the unorganized sector, where the average is 2.35 inputs, and this difference is statistically significant at a 10% level. In both sectors, the median value is 2 inputs. Within the unorganized sector, female-owned businesses have an even lower input usage at 1.91 inputs, compared to the 2.44 average for male-owned businesses. This partially reflects the smaller and often household-based nature of these businesses.

Panels B and C repeat Panel A and separate establishments into those located in urban versus rural areas.⁵ There is a clear pattern of higher input usage in urban areas compared to rural areas for all types of establishments. For the organized sector, plants in urban areas utilize 2.78 inputs compared to 2.32 inputs in rural areas. The raw gap is even larger in the unorganized

⁵ Whether or not a plant is located in an urban area is collected directly from the survey and is not part of the sample stratification. The definition of an urban setting for our surveys is (a) All statutory places with a municipality, corporation, cantonment board or notified town area committee, etc., or (b) A place satisfying the following three criteria simultaneously: i) a minimum population of 5,000; ii) at least 75% of male working population engaged in non-agricultural pursuits; and iii) a density of population of at least 400 per sq. km. (1,000 per sq. mile). The 2000 surveys use classifications based upon the 1991 Census for these demarcations.

sector, at 2.53 inputs in urban areas versus 2.09 inputs in rural areas, and it holds along the gender dimension, too. These differences are also statistically significant at a 10% level.

Within the organized sector, we will later show in Table 10 that the average input usage for small organized sector plants, defined to be fewer than 50 employees, at 2.42 inputs is very comparable to the unorganized sector's 2.44 inputs. By contrast, plants with more than 100 employees have on average 2.99 inputs. Thus, the unorganized nature of businesses does not necessarily reduce their input usage, and the sizes of firms and industry choices could play the governing roles.

Table 1b analyzes this issue by repeating Table 1a for a normalized input count. We calculate the normalized input count by pooling the two sectors together and running a regression of input counts on industry fixed effects and establishment size fixed effects. For establishment size effects, we utilize the bins of 0-2, 3-5, 6-10, 11-20, 21-50, 51-100, and 101+ employees. We take the residuals from this regression as our normalized count, which effectively removes from each plant's input count the average for the plant's industry and size category. While the differences remain statistically significant, this normalization closes most of the gap between organized and unorganized sector establishments. On the other hand, the differences between urban and rural areas persist.

Figures 1-6 provide visual depictions of the spatial variation in input usage. Figures 1-2 combine the organized and unorganized sectors, while Figures 3-4 and 5-6 document patterns for the organized and unorganized sectors, respectively. Within each pair, the first figure utilizes the raw count of inputs for all establishments in the district. As input usage naturally varies across industries, the second figure in each pair considers a normalized version where input usage for each plant is measured as the deviation from the national average for the plant's industry (for this purpose, we retain the establishment size differences).

Table 2 documents some of the larger values observed at the district level in our survey, and Tables 3a and 3b provide detailed counts of input levels by two-digit industries. For reference, Appendix Tables 2a and 2b provide a comparable documentation of the sizes of industries. Table 3a compares input usage across industries for the organized and unorganized sector. The overall pattern reflected in Table 1 of input usage being higher in the organized sector is generally observed at the industry level. In 17 of the 22 industries, the average count of

inputs used in the organized sector exceeds the unorganized sector. Of the five cases where the unorganized sector is higher, three of those cases are due to very small cell sizes related to advanced industries in the unorganized sector producing noisy estimates (NIC industries 23, 30, and 32, see Appendix Table 2a). A second approach to quantifying this deviation is to divide the mean input count in the organized sector by the mean input count in the unorganized sector. The unweighted average of this ratio across industries is 1.11, with a standard deviation of 0.18.

Despite this overall pattern, the bottom of Table 3a notes the interesting fact that input usage in the unorganized sector is overall higher in traditional industries than the organized sector, while the opposite is true for modern industries based on the average across all plants within each group. If one excludes the “Office, accounting and computing machinery” industry (NIC 30), which has a very limited unorganized sector presence, there is a 0.07 correlation between the share of an industry’s establishments in the organized sector and the ratio of input counts between the organized and unorganized sectors. In other words, industries where the organized sector is a greater share of activity tend to have larger deviations in input usage, but this is very weak.

Table 3b continues with the unorganized sector and compares the input usage at the industry level by gender of owner. In 16 of the 22 industries, male-owned establishments show a higher input usage than female-owned establishments within the unorganized sector. Some of the deviations (e.g., NIC 23 and 35) again appear to be due to small cell sizes, as shown in Appendix Table 2b. Creating a ratio of average input counts for male-owned establishments compared to average input counts for female-owned establishments, the unweighted average across the industries is 1.17, with a standard deviation of 0.23. The male-to-female ratio exceeds one in both the traditional and modern sectors.

Finally, Table 4 describes the correlation of input usage across sectors within districts. This analysis provides a sense of whether districts with high average input counts in the organized sector also display high average input counts in the unorganized sector. There is no clear prediction about whether this correlation should exist, especially in the Indian manufacturing context. On one hand, input usage could be pushed up in both sectors by shared infrastructure, common geographic location, uniform local legal environment, and similar attributes. On the other hand, the long-term migration patterns of the two sectors are quite

different, with the organized sector moving out of urban areas and the reverse being true for the unorganized sector (Ghani, Goswami, and Kerr 2012a), and tight interconnections have yet to be established empirically between the sectors. These factors would suggest that the two input conditions may operate independently.

Panel A of Table 4 shows a very limited correlation between raw average input counts for the organized and unorganized sectors in India at the district level. While the 0.11 correlation is statistically significant, its economic magnitude is quite small. Panel B reaches a similar conclusion after first normalizing input usage for each establishment by the average of its industry and sector (i.e., the residual from a regression of input counts on industry and sector fixed effects) and then aggregating to district-level values. On the other hand, there is a high correlation among male- and female-owned establishments in the unorganized sector. In the next section we consider traits of districts that are associated with greater input usage. These base correlations suggest we may observe quite different determinants for the organized and unorganized sectors, with more consistent results likely emerging within the unorganized sector across the gender of business owners.

Appendix Table 3 provides an extended set of correlations that separate urban versus rural areas of districts. The most important observation from these correlations is that input usage in districts tends to be highly correlated across sectors between urban and rural locations. For example, the average input count of the organized sector in urban areas displays a 0.47 correlation with the average input count of organized sector in rural areas in the same district. Similar correlations are found in the unorganized sector. Thus, higher correlations exist by sectors across urban and rural locations within districts than exist across sectors—organized versus unorganized—within each district. As a consequence, we devote most of the remainder of this study to describing differences between organized and unorganized sectors, paying less attention to the urbanization of locations.

4. Correlations of Input Usage to District and Industry Traits

Table 5 presents an investigation of the degree to which district and industry traits are associated with higher input usage with a set of univariate correlations between district conditions and the average input usage observed among the plants in each district. The traits in

Panel A are taken from the 2001 Population Census for India, while those in Panel B are derived directly from the manufacturing surveys. These correlations are unweighted and treat each industry independently. We select the specific attributes listed due to their frequent use in studies of Indian manufacturing and its spatial variations.⁶

Within Panel A, several factors display strong univariate correlations across all four columns: literacy rate, strength of household banking sector, urbanization rate, and consumption per capita levels. Population density also comes across very strongly for the organized sector, while infrastructure access in the form of paved roads has a univariate link for the unorganized sector. Of course, these factors are likely to be highly correlated among themselves, and so we turn shortly to a multivariate analysis.

Within Panel B, the first three rows consider the organized sector's share of local manufacturing activity. When using employments or output, there is strong correlation of a larger organized sector share in the district to greater average input counts by local plants in both sectors. On the other hand, we observe in the fourth through sixth rows of Panel B a weaker connection to the degree to which the local unorganized sector contains a high share of female-owned businesses relative to male-owned businesses (i.e., the relative composition of the local unorganized sector). The negative connections observed in the fourth column are likely a consequence of greater local female ownership encouraging more entry on the margin by smaller and frequently household-based female entrepreneurs (Ghani, Kerr, and O'Connell 2013b). These marginal entrants tend to have smaller businesses and likely lower input counts.

The final three rows of Panel B provide perhaps the most interesting result, which is a null finding. One might have expected that substantial local industrial diversity would have been associated with greater counts of distinct inputs used by firms, an argument that dates back to Jacobs (1969) and others. To assess, we calculate a Herfindahl-Hirschman Index (HHI) of the

⁶ Appendix Table 1 provides descriptive statistics on district-level traits. The construction and definition of most traits are fairly intuitive. Infrastructure variables are calculated as the percentage of villages that report access to that form of infrastructure (e.g., electricity, paved roads). Travel time to one of India's ten largest cities is calculated from Lall, Wang, and Deichmann (2011) based upon the driving time from the central node of a district. The ten cities are Ahmedabad, Bangalore, Bhubaneswar, Chennai, Delhi, Guwahati, Hyderabad, Kolkata, Mumbai, and Patna. The strength of household banking variable is calculated the share of households that report a banking relationship.

districts manufacturing base, aggregating together the organized and unorganized sectors. A higher value of this index indicates a more concentrated local industrial environment in terms of the manufacturing industries represented. In general, there seems to be no connection between local industrial diversity and greater average counts of inputs used in Indian manufacturing. This even holds with our measure considering input usage in a binary form, compared to an intensity of use formulation. The one exception is a small weak correlation in Column 1 for more concentrated local environments based upon establishment distributions having reduced average input counts for the organized sector. This overall weakness for the city may reflect the longer spatial horizons for material inputs between customer and supplier firms.⁷

Table 6a considers multivariate analyses of these local traits and continues to analyze the organized and unorganized sectors separately. We calculate the average raw input use among plants in each district d and industry j , weighting plants by their sample weights in the survey. We regress this average input usage for each district on a vector Z_d of district traits and a vector γ_j of industry fixed effects using the specification:

$$Inputs_{d,j} = \beta \cdot Z_d + \gamma_j + \varepsilon_{d,j}.$$

District-level traits in the vector Z_d are taken from the 2001 Census and follow Panel A of Table 5. Non-logarithm variables are transformed to have unit standard deviation for interpretation. The industry effects directly control for broad differences across industries in their average input counts. We cluster standard errors at the district level to reflect the repeated mapping of district traits to the district-industry observations.

Given the small cells that some district and industry combinations form, we weight observations by an interaction of log district size and log industry size. This interaction approach places greater emphasis on the observations expected to be substantial in size due to the associated districts and industries being large. The advantage of this approach compared to directly weighting by observed district-industry size is that it places less emphasis on very abnormal spatial concentrations of activity, where reverse causality of particular industry agglomerations can overly shape local traits (e.g., Glaeser and Kerr 2009). The weights

⁷ For example, Rosenthal and Strange (2001, 2004) and Ellison, Glaeser, and Kerr (2010).

aggregate across employment in the organized and unorganized sectors so that a single weight can be applied in all specifications.

Columns 1 and 2 of Table 6a report results for the organized sector. In the first column, the strongest predictors for high input counts are the log per capita consumption of the district, followed by its literacy rate and the strength of the local banking environment. A district age profile that favors the working age ranges has a negative partial correlation. The per capita consumption result is the easiest to understand, given that it is a close correlate of district per capita GDP. Simply put, wealthier districts also display higher typical input counts in local plants. A 10% increase in per capita consumption is associated with 0.03 higher inputs on average. This is substantial in economic magnitude—for example, the entire urban-rural gap in Table 1a for the organized sector is just 0.46 inputs.

The connections of the literacy rate and local household banking strength to input usage are harder to make. The banking result will prove to be fleeting once we adjust the specification, so we do not dwell on it. By contrast, the connection of higher literacy rates to greater local input usage is very robust. One interpretation of this pattern is that greater local education at a general level (i.e., compared to the graduate education share that we also model as a covariate) allows for more complex production techniques that includes more input usage. In other words, there is complementarity between worker human capital and the variety of inputs that a plant utilizes. We are extremely cautious about this conclusion, however, given that measurement error in other wealth variables like log per capita consumption may be loading onto the literacy rate. We hope to further refine this finding in future work.

In Column 2, we continue with the organized sector patterns and introduce a vector of state fixed effects into the estimation. These fixed effects restrict the identification to variations across districts within states. This adjustment can be both helpful and limiting. On one hand, these fixed effects pick up many additional factors that we did not model (e.g., state legal environments), and we will thus place more faith on results that survive this more stringent specification. On the other hand, the earlier figures show some clear differences in input usage at a regional level, and some of our explanatory variables are also regional in scope. In these cases, state fixed effects might remove some of the most interesting variations.

After introducing state fixed effects, the log per capita consumption variable remains the strongest effect in terms of economic magnitudes. It is no longer statistically significant at a 10% level, but it remains reasonably well estimated. Thus, the connection of input usage to wealthier or more productive districts within states continues to hold, matching the theory outlined in the introduction. Literacy rates also continue to display a strong association. Moreover, we also now find evidence of population density and electricity access being among the more important associations. Proximity to one of India's ten largest cities also raises average input counts. Interestingly, within states, a higher urbanization rate at the district level is associated with lower input counts for organized sector plants once controlling for the other covariates.

Columns 3 and 4 report results from similar estimations for the unorganized sector. As foreshadowed by the limited correlations in Table 4 in average input counts across sectors at the district level, the patterns in Columns 3 and 4 are quite distinct. Without conditioning on state fixed effects, the literacy rates of districts, the infrastructure measures related to electricity and paved roads, and the urbanization rate of the district have the strongest connection to higher local input counts, while population density is associated with lower usage. Once controlling for state fixed effects, very few district traits other than the local literacy rate retain their association to input counts.

Table 6b finds very similar results when we use normalized version of input counts examined in Table 1b (i.e., removing industry and plant size averages). Appendix Tables 4a and 4b provide additional variants on these specifications. We include in our baseline estimation three explanatory variables that are related to manufacturing specifically (the last three rows in Table 6). We first control for the overall log employment in the local district-industry in the organized and unorganized sectors. These measures provide a check against district-industry size introducing a selection effect (e.g., large district-industry agglomerations inducing the entry of smaller firms that lower average input counts measured for the district-industry) that biases our estimates for local conditions (e.g., Figueiredo, Guimaraes, and Woodward 2009). We likewise model in the baseline estimation the local diversity of manufacturing activity. We find quite similar patterns for the other explanatory variables in the appendix when excluding these three metrics. We also find similar results when dropping the estimation weights that interact district and industry size. Finally, we observe that the patterns for the male- and female-owned

establishments are reasonably similar within the unorganized sector, which is not surprising given the high correlation we observed across owner gender in Table 4.

On a whole, we make the following conclusions from these estimations. For the organized sector, higher input usage is generally observed in wealthier districts and those with greater literacy rates. If looking within states, the usage is more closely associated with electricity access and population density. Proximity to one of India's largest cities is also connected to higher average input counts. By contrast, input usage in the unorganized sector has much more erratic variation across districts. The literacy rate is again a robust finding, but there is otherwise too much noise to draw sharp conclusions.

Finally, Table 7 closes this analysis by quickly considering industry traits associated with greater input usage. We are intrinsically less interested in these industry-level traits compared to the spatial traits, as we typically try to control directly for industry production techniques using industry fixed effects, as in Table 6a. Nonetheless, we can establish through these correlations some insights for our work. First, the top of Panel A suggests that our input count measure is quite distinct from other metrics one might measure for industries. Greater counts of distinct inputs are not simply reflecting greater dependency on material inputs or a similarly defined trait. This provides an empirical rationale for a more careful consideration of them.

Second, and quite interesting, the bottom of Table 7 looks a lot like Panel B of Table 5, despite the fact that the two tables exploit different variations. Specifically, Table 5 collapses over industries to consider correlations at the district level, while the opposite is true in Table 7 where we have aggregated up to the national level to consider broad variations that exist across industries. Despite this change of perspective, we again find a strong correlation of a larger organized sector share to greater average input counts by organized sector plants. This is likely indicating that situations where more distinct inputs are required for production are also pushing towards a production structure that favors more organized sector firms. On the other hand, a high share of female-owned businesses for an industry is associated with lower input usage for female-owned businesses in that industry.

5. Input Usage in Manufacturing Production Function Estimations

We now turn to the link between input counts and the productivity of manufacturing establishments. Table 8 provides a simple analysis of whether greater input counts in plants is associated with stronger productivity in the organized sector. We estimate a simple production function with log output Y of each establishment i in a district d and industry j as the dependent variable. The specification takes the form:

$$Y_{i,d,j} = \beta \cdot Inputs_i + \mu \cdot X_i + \phi \cdot Z_d + \gamma_j + \varepsilon_{i,d,j}.$$

Our core regressor is an indicator variable for the raw input usage for each plant being greater than its industry's median value. We later test variants on this metric design. We include a vector X_i of plant inputs into the production function: log employees, log book values of capital, and log costs of materials. We exclude plants with missing values for these metrics. As noted above, the input counts measure that we are studying is quite distinct from the intensity of material inputs into a production function. We thus control for this latter intensity directly. Because we rely on revenue data to calculate productivity, we face a common limitation in the literature that we cannot separate the efficiency or productivity of plants in terms of real inputs and outputs from other factors like their mark-ups or quality (e.g., Foster, Haltiwanger, and Syverson 2008, De Loecker 2011).

Regressions include three-digit industry fixed effects γ_j to capture regular differences in production techniques and spatial locations across industries. We also control for other district traits described below with a vector of district-level controls Z_d . We use establishment weights from the surveys to weight plants, and we report robust standard errors.

Column 1 provides a baseline estimation of the plant-level production function before district-level conditions are incorporated. These underlying parameters for the production function, emphasizing employees and materials, are very stable across estimations. Column 2 introduces the input count measure and finds that establishments with above median input counts for their industry have 0.039 higher log output than those with input counts below the median, holding everything else constant. We later show in Table 10 when using a linear metric design that an increase in one input is associated with a 0.007 increase in log output, indicative of some non-linearity in this effect. To provide additional perspective, the coefficient in the first row

suggests a 10% increase in employment is associated with about a 0.012 output increase. The second row would suggest a 10% increase in capital is associated with about a 0.001 output increase. The linear role for counts of distinct material inputs into the establishment is thus bracketed by these other inputs into the firm production.⁸

Column 3 introduces the district average value of this indicator variable for above median inputs, finding a positive effect that is almost statistically significant. This coefficient suggests that moving 10% of local plants from below median values to above median values is associated with a 0.013 gain in log output for the plant. Column 4 introduces both measures and finds that the coefficients are comparable to their individual estimations.

Column 5 shows similar results when controlling for log manufacturing employment in district per square kilometer, log manufacturing employment in a plant's district-industry per square kilometer, log share of local manufacturing employment in the unorganized sector, and log share of district-industry manufacturing employment in the unorganized sector. The first two of these controls are often associated with the urbanization and agglomeration premiums of dense locations, and we use a per square kilometer normalization as Indian districts vary in spatial size.⁹

Columns 6 and 7 split the sample by traditional versus modern groups. The plant-specific input count effect is present in both groups, with some emphasis for traditional industries, while the local area affect is more closely associated with the traditional industries.

Table 9 repeats the analysis in Table 8 for establishments in the unorganized sector. In this sector, we generally find a more modest connection between higher input counts and plant output levels after conditioning on the base inputs. Compared to the organized sector, the

⁸ Since Olley and Pakes (1996), there has been a substantial amount of work attempting to address the potential simultaneity bias between input usage (broadly defined to include labor, materials, and so on) and plant productivity. Our current analysis cannot estimate these more complicated models given that it considers a cross-section of data. Looking forward, there is some scope for these models despite the fact that the general Indian data are repeated cross-sections rather than panel data for firms (e.g., Sivadasan 2009). The first step is to prepare consistent measures of input counts across Indian data surveys for many years. We are currently working on this task to provide a longitudinal dimension to this project.

⁹ See, for example, Ciccone and Hall (1996), Duranton and Puga (2004), and Rosenthal and Strange (2004). Indian studies include Lall, Shalizi, and Deichmann (2004), Lall and Mengistae (2005), Deichmann et al. (2008), and Fernandes and Sharma (2011).

economic magnitude is reduced by about a quarter. While statistically significant, the effect is also less precisely estimated despite having a sample size that is five-fold larger. We do not see as much evidence of a beneficial output effect from the local area as a whole having higher input usage, especially once controlling for the other urban traits in Column 5. Appendix Table 5 also finds very little connection to female-owned establishments specifically.

Table 10 provides a variety of extensions and robustness checks on these patterns, continuing with the organized sector. We repeat in Column 1 the base specification for convenience. Column 2 introduces separate indicator variables for plants being in the third or highest quartiles of input usage for their industries, finding very similar patterns to Column 1. There is some evidence, moreover, of a non-linear benefit to achieving moderately high input usage compared to moderately low input usage, with the results not being driven by extreme values. Column 3 reports the linear input count measure (taking values between zero inputs and five inputs) discussed earlier. This coefficient is lower than those using the indicator variable approach, as it does not measure the non-linear effects, but it remains economically and statistically important. Column 4 shows that the impact is weakened when we instead benchmark plants within each district-industry, which relates to the positive role for local input counts observed in Table 8.

The last three columns of Table 10 document an intriguing difference in the role of input counts for productivity gains across the establishment size distribution of the organized sector. Column 5 considers plants with fewer than 50 employees (~54% of the sample), Column 6 considers plants with 50-100 employees (20%), and Column 7 considers plants with over 100 employees (25%). The positive association of input counts to productivity is concentrated in the organized sector establishments with fewer than 50 employees. Near the bottom of the table, we document the mean input count for each group. As noted in the second section, the mean count for the organized firms with fewer than 50 employees is very similar to that observed in the unorganized sector and substantially weaker than the counts in large establishments in the organized sector.

The difference in the small-scale establishment results across the two sectors in Tables 9 and 10 is intriguing. Scholars in urban economics since Chinitz (1961) and Jacobs (1969) have considered how small firms are especially reliant on their local economy for inputs into the firm,

and Chinitz (1961) and the literature following him have emphasized how varied these local conditions are in terms of their appropriateness for small and young businesses compared to large establishments. The components of our study have identified that small firms in the organized sector have access to similar average input levels as unorganized sector firms. While we observe productivity consequences in both sectors, the impact is steepest among small organized sector firms. This may suggest that weaker input conditions for small organized sector firms limit their ability to outperform unorganized sector firms (or, said differently, that one cost of unorganized sector status is an inability to gain and exploit these distinct material inputs). This divergence is perhaps showing a variation of the Chinitz (1961) theme within India's manufacturing that depends upon sector. To provide greater traction for future study, we intend to develop a district-industry panel over time for the two sectors that can better separate how these input conditions matter.

6. Conclusions

The patterns that we document in this study are quite intriguing. Counts of distinct material inputs are clearly far from uniform across Indian manufacturing establishments. They are stronger in urban settings than in rural locations, and they are stronger in the organized sector than in the unorganized sector. Looking across districts, higher input counts are most closely associated with wealthier areas and those with higher literacy rates. Local infrastructure and population density are also important in some specifications, but these patterns are less robust. Plants in the organized sector utilizing a greater variety of inputs display a higher productivity, with the effects mostly concentrated among smaller plants with fewer than 50 employees. There is also a link between input usage and productivity in the unorganized sector, but it is more modest in size compared to the organized sector. It is likewise quite difficult to predict higher input counts in the unorganized sector using district-level conditions.

There are several directions that we hope to take with this study going forward. A first effort would continue on the current themes with the goal of expanding the range of inputs considered to include energy inputs, which are also separately tabulated on the surveys. This would be attractive for two reasons. First, energy scarcity and instability is an acute problem in India, leading to lost productivity during brown or black outs. As a consequence, many firms

self-provision at least part of their energy needs. This self-production of energy may be quite inefficient for many firms. We need a better understanding of how these various factors add up into productivity, and such a study would be substantially more challenging to model than our current focus on unique material inputs.

Second, and related, our current focus on material inputs has been almost exclusively from the perspective of individual firms, without worrying too much about spillovers or externalities outside of the firm. We believe this is reasonable for a study of material input counts, especially since most of the results indicate a positive role for these inputs into the productivity of the plants holding fixed other inputs. Moreover, where we observe an independent role for higher input usage in the local area, the spillover effect appears positive (although unidentified econometrically). This approach, however, would be inadequate for considering the self-production of energy by plants, given the potential pollution costs and related externalities across plants and other parts of society. Self-production may be associated with greater pollution through plants choosing dirtier but cheaper techniques, being less efficient at energy production than a dedicated facility, or a combination of these two.

This relates to another important area for future consideration. Ghani, Goswami, and Kerr (2012a) describe how the long-run trend for Indian manufacturing is that organized sector firms are moving towards rural locations, while the opposite is true for unorganized sector firms. Our current study considers a cross-section from 2000, and future work should marry these two studies by studying cohorts over time. While data limitations will prevent a perfect analysis, this study would help to answer some basic questions: Is the movement of firms in the organized sector from urban to rural areas associated with productivity declines due to poorer input conditions (e.g., land-use regulations pushing firms towards non-optimal conditions)? If input counts remain high, do we see instead positive uptake by other firms in the area? The latter might occur through diffusion of managerial practice (e.g., Bloom et al. 2013) or positive spillovers in local activity (e.g., unorganized sector firms benefit from better supplier conditions that result from organized sector firms being located in the district).¹⁰

¹⁰ Greenstone, Hornbeck, and Morretti (2010) find substantial productivity gains to existing rural firms in locations that win bids for “million dollar plants” in the United States, while Falck et al. (2013) find adverse effects for local West German incumbents from relocating East German firms. The present Indian case may fall in between,

A final area of interest also concerns itself with the trends in input usage over time and their spatial diffusion. We hope to consider in future work how major infrastructure projects within India and India's international trade reforms shaped the spatial access of firms to production inputs. One prominent example of an infrastructure project is the Golden Quadrangle highway project that Datta (2011) finds quickly impacted input sourcing decisions and inventory management of organized sector firms. Ghani, Goswami, and Kerr (2012b) further link the Golden Quadrangle highway project to shifts in the spatial organization of India's organized sector. We hope to consider how changes in input access link to these two features and whether this improved the overall efficiency of India's manufacturing sector. We would also like to complement the prior work on trade reforms (e.g., Goldberg et al. 2010a,b, Nataraj 2011) to observe how trade impacts the spatial development of Indian productivity given the substantial heterogeneity in input usage across regions of India. Both of these research strands, in addition to being worthwhile studies independently, can help shed light on the distribution impacts across India from its economic development and growth.

given that the migration may be less optimal for the local incumbents than cases where plant bidding occurred, but more optimal than the random wartime dispersal that occurred in Germany.

References

- Amiti, Mary, and Jozef Konings, "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia", *American Economic Review* 97:5 (2007), 1611-1638.
- Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence from India", *Quarterly Journal of Economics* 128:1 (2013), 1-51.
- Bollard, Albert, Peter Klenow, and Gunjan Sharma, "India's Mysterious Manufacturing Miracle", *Review of Economic Dynamics* 16:1 (2013), 59-85.
- Ciccone, Antonio, and Robert Hall, "Productivity and the Density of Economic Activity", *American Economic Review* 86:1 (1996), 54-70.
- Chinitz, Benjamin, "Contrasts in Agglomeration: New York and Pittsburgh", *American Economic Review* 51:2 (1961), 279-289.
- Datta, Saugato, "The Impact of Improved Highways on Indian Firms", *Journal of Development Economics*, 99:1 (2011), 46-57.
- De Loecker, Jan, "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity", *Econometrica* 79:5 (2011), 1407-1451.
- Deichmann, Uwe, Somik Lall, Stephen Redding, and Anthony Venables, "Industrial Location in Developing Countries", *World Bank Research Observer* 23:2 (2008), 219-246.
- Desmet, Klaus, Ejaz Ghani, Stephen O'Connell, and Esteban Rossi-Hansberg, "The Spatial Development of India", *World Bank Policy Research Paper* 6060 (2012).
- Duranton, Gilles, and Diego Puga, "Diversity and Specialization in Cities: Why, Where and When Does It Matter?", *Urban Studies* 37:3 (2000), 533-555.
- Duranton, Gilles, and Diego Puga, "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products", *American Economic Review* 91 (2001), 1454-1477.
- Duranton, Gilles, and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies", in Vernon Henderson and Jacques François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4 (Amsterdam: North-Holland, 2004), 2063-2117.
- Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", *American Economic Review* 100 (2010), 1195-1213.
- Falck, Oliver, Christina Guenther, Stephan Heblich, and William Kerr, "From Russia with Love: The Impact of Relocated Firms on Incumbent Survival", *Journal of Economic Geography* 13:3 (2013), 419-449.
- Feldman, Maryann, and David Audretsch, "Innovation in Cities: Science-based Diversity, Specialization and Localized Competition", *European Economic Review* 43:2 (1999), 409-429.
- Fernandes, Ana, and Gunjan Sharma, "Together We Stand? Agglomeration in Indian Manufacturing", *World Bank Policy Working Paper* 6062 (2012).

- Figueiredo, Octávio, Paulo Guimaraes, and Douglas Woodward, “Localization Economies and Establishment Size: Was Marshall Right after All?”, *Journal of Economic Geography* 9 (2009), 853-868.
- Foster, Lucia, John Haltiwanger, and Chad Syverson, “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?”, *American Economic Review* 98:1 (2008), 394-425.
- Ghani, Ejaz, William Kerr, and Stephen O’Connell, “Spatial Determinants of Entrepreneurship in India”, National Bureau of Economic Research Working Paper 17514 (2011). Forthcoming in *Regional Studies*.
- Ghani, Ejaz, William Kerr, and Stephen O’Connell, “Local Industrial Structures and Female Entrepreneurship in India”, National Bureau of Economic Research Working Paper 17596. Forthcoming in *Journal of Economic Geography*.
- Ghani, Ejaz, William Kerr, and Stephen O’Connell, “The Exceptional Persistence of India’s Unorganized Sector”, World Bank Policy Working Paper 6454 (2013a).
- Ghani, Ejaz, William Kerr, and Stephen O’Connell, “Female Business Ownership and Informal Sector Persistence”, World Bank Policy Working Paper 6612 (2013b).
- Ghani, Ejaz, Arti Goswami, and William Kerr, “Is India's Manufacturing Sector Moving Away From Cities?”, National Bureau of Economic Research Working Paper 17992 (2012a).
- Ghani, Ejaz, Arti Goswami, and William Kerr, “Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing”, National Bureau of Economic Research Working Paper 18524 (2012b).
- Glaeser, Edward, Heidi Kallal, José Scheinkman, and Andrei Shleifer. “Growth in Cities”, *Journal of Political Economy* 100:6 (1992), 1126-1152.
- Glaeser, Edward, and William Kerr, “Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?”, *Journal of Economics and Management Strategy* 18:3 (2009), 623-663.
- Goldberg, Pinelopi, Amit Khandelwal, Nina Pavcnik, and Petia Topalova, “Imported Intermediate Inputs and Domestic Product Growth: Evidence from India”, *Quarterly Journal of Economics* 125:4 (2010a), 1727-1767.
- Goldberg, Pinelopi, Amit Khandelwal, Nina Pavcnik, and Petia Topalova, “Multi-product Firms and Product Turnover in the Developing World: Evidence from India,” *Review of Economics and Statistics* 92:4 (2010b), 1042-1049.
- Greenstone, Michael, Richard Hornbeck and Enrico Moretti, “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings”, *Journal of Political Economy* 118 (2010), 536-598.
- Grossman, Gene, and Elhanan Helpman, *Innovation and Growth in the Global Economy* (Cambridge: MIT Press, 1991).
- Halpen, Lazlo, Miklos Koren, and Adam Szeidl, “Imports and Productivity”, Working Paper (2006).

- Henderson, J. Vernon, "Externalities and Industrial Development, *Journal of Urban Economics* 42:3 (1997), 449-470.
- Henderson, Vernon, "Cities and Development", *Journal of Regional Science* 50 (2010), 515-540.
- Hsieh, Chang-Tai, and Peter Klenow, "Misallocation and Manufacturing TFP in China and India", *Quarterly Journal of Economics* 124 (2009), 1403-1448.
- Jacobs, Jane, *The Economy of Cities* (New York: Random House, 1969).
- Kasahara, Hiroyuki and Joel Rodrigue, "Does the Use of Imported Intermediates Increase Productivity?", *Journal of Development Economics* 87:1 (2008), 106-118.
- Kathuria, Vinish, Seethamma Natarajan, Rajesh Raj, and Kunal Sen, "Organized versus Unorganized Manufacturing Performance in India in the Post-Reform Period", MPRA Working Paper No. 20317 (2010).
- Lall, Somik, Jun Koo, and Sanjoy Chakravorty, "Diversity Matters: The Economic Geography of Industry Location in India", *World Bank Publications* 3072 (2003).
- Lall, Somik, and Taye Mengistae, "The Impact of Business Environment and Economic Geography on Plant Level Productivity: An Analysis of Indian Industry", Working Paper (2005).
- Lall, Somik, Zmarak Shalizi, and Uwe Deichmann, "Agglomeration Economies and Productivity in Indian Industry", *Journal of Development Economics* 73:2 (2004), 643-673.
- Lall, Somik, Hyoung Wang, and Uwe Deichmann, "Infrastructure and City Competitiveness in India", Working Paper (2011).
- Muendler, Marc-Andreas, "Trade, Technology, and Productivity: A Study of Brazilian Manufacturers, 1986-1998", Working Paper (2004).
- Nataraj, Shanthi, "The Impact of Trade Liberalization on Productivity: Evidence from India's Formal and Informal Manufacturing Sectors", *Journal of International Economics* 85 (2011), 292-301.
- Olley, Steven, and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry", *Econometrica* 64:6 (1996), 1263-1297.
- Porter, Michael, *The Competitive Advantage of Nations* (New York, NY: The Free Press, 1990).
- Rivera-Batiz, Luis, and Paul Romer, "Economic Integration and Endogenous Growth", *Quarterly Journal of Economics* 106:2 (1991), 531-555.
- Romer, Paul, "Growth Based on Increasing Returns Due to Specialization", *American Economic Review*, 77:2 (1987), 56-62.
- Romer, Paul, "Endogenous Technological Change", *Journal of Political Economy* 98:5 (1990), 71-102.
- Rosenthal, Stuart, and William Strange, "The Determinants of Agglomeration", *Journal of Urban Economics* 50 (2001), 191-229.

- Rosenthal, Stuart, and William Strange, “Evidence on the Nature and Sources of Agglomeration Economies”, in Vernon Henderson and Jacques François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4 (Amsterdam: North-Holland, 2004), 2119-2171.
- Sivadasan, Jagadeesh, “Barriers to Competition and Productivity: Evidence from India”, *The B.E. Journal of Economic Analysis & Policy* 9:1 (2009), Article 42.
- Trivedi, Pushpa, L. Lakshmanan, Rajeev Jain, and Yogesh Gupta, “Productivity, Efficiency, and Competitiveness of the Indian Manufacturing Sector”, Reserve Bank of India Paper (2011).
- World Bank, “Planning, Connecting, and Financing Cities—Now”, *Urbanization Review Flagship Report* (2012).

Figure 1: Average Input Count, All Plants

Figure 2: Normalized Input Count, All Plants

Figure 3: Average Input Count, Organized Plants

Figure 4: Normalized Input Count, Organized Plants

Figure 5: Average Input Count, Unorganized Plants

Figure 6: Normalized Input Count, Unorganized Plants

Table 1a: Input counts in manufacturing plants, 2000

	Organized sector	Unorganized sector		
		Total	Male owned	Female owned
A. Input counts among all plants				
Mean value	2.60†	2.35	2.44	1.91
Standard deviation	1.79	1.48	1.52	1.18
Median value	2.00	2.00	2.00	2.00
B. Input counts among plants in urban areas				
Mean value	2.78†	2.53	2.64	2.00
Standard deviation	1.81	1.53	1.56	1.25
Median value	3.00	2.00	2.00	2.00
C. Input counts among plants in rural areas				
Mean value	2.32†*	2.09*	2.15*	1.77*
Standard deviation	1.74	1.35	1.40	1.06
Median value	2.00	2.00	2.00	2.00

Notes: Descriptive statistics taken from Annual Survey of Industries (ASI) and National Sample Statistics (NSS). Input counts range from zero inputs to a maximum of five inputs. An asterisk * indicates rural-area mean is different from urban-area sample at 10% level of significance. The symbol † denotes organized sector mean is different from unorganized sample mean at 10% level of significance. Appendix Table 1 provides additional descriptive statistics for district-level conditions.

Table 1b: Table 1a using normalized input counts

	Organized sector	Unorganized sector		
		Total	Male owned	Female owned
A. Input counts among all plants				
Mean value	-0.05†	0.01	0.10	-0.38
Standard deviation	1.67	1.32	1.35	1.11
Median value	-0.17	-0.21	-0.07	-0.49
B. Input counts among plants in urban areas				
Mean value	0.10†	0.09	0.19	-0.37
Standard deviation	1.64	1.38	1.41	1.17
Median value	0.13	-0.06	0.09	-0.49
C. Input counts among plants in rural areas				
Mean value	-0.30†*	-0.10*	-0.04*	-0.40*
Standard deviation	1.68	1.21	1.24	1.01
Median value	-0.61	-0.39	-0.34	-0.49

Notes: See Table 1a. Normalized input counts are residuals from a regression of input counts on industry fixed effects and plant size dummies.

Table 2: Detailed input count levels across district urban areas

District	State	Mean input counts	Plant count
A. Districts with highest average values in urban areas for organized sector			
Bharuch	GUJARAT	3.94	96
Raisen	MADHYA PRADESH	3.90	29
North Goa	GOA	3.79	33
Kolhapur	MAHARASHTRA	3.64	45
Ernakulam	KERALA	3.58	139
AllahabadBanda	UTTAR PRADESH	3.54	37
Bhopal	MADHYA PRADESH	3.52	60
Thane	MAHARASHTRA	3.48	419
MeerutMuzaffarnagar	UTTAR PRADESH	3.48	83
Ambala	HARYANA	3.47	34
Vadodara	GUJARAT	3.44	196
Dehradun	UTTARANCHAL	3.42	33
<i>Average across urban areas for organized sector</i>		2.78	64
B. Districts with highest average values in urban areas for unorganized sector			
Mainpuri	UTTAR PRADESH	4.15	41
Hamirpur	HIMACHAL PRADESH	4.09	144
South Goa	GOA	4.04	159
Godda	JHARKHAND	4.04	28
Bilaspur	HIMACHAL PRADESH	4.01	71
Udhampur	JAMMU & KASHMIR	3.98	123
Pithoragarh	UTTARANCHAL	3.94	31
Churachandpur	MANIPUR	3.89	57
Kurukshetra	HARYANA	3.88	106
Khandhamal	ORISSA	3.86	28
Fatehpur	UTTAR PRADESH	3.80	30
Una	HIMACHAL PRADESH	3.78	253
<i>Average across urban areas for unorganized sector</i>		2.53	178

Notes: See Table 1. Sample is restricted to urban and rural areas with at least 25 reporting plants.

Table 3a: Detailed input counts for industries, 2000

NIC	Industry Description	Organized sector			Unorganized sector		
		Mean	SD	Median	Mean	SD	Median
15	Food products and beverages	1.76	1.55	1.00	2.58	1.58	2.00
16	Tobacco products	2.09	1.39	2.00	1.51	0.86	1.00
17	Textiles	1.88	1.50	1.00	1.61	0.96	1.00
18	Wearing apparel; dressing and dyeing of fur	3.11	2.01	4.00	2.87	1.62	3.00
19	Leather; luggage, handbags, saddlery, harness and footwear	2.87	1.94	3.00	3.41	1.40	4.00
20	Wood and wood products, except furniture; straw and plating	1.86	1.51	1.00	1.52	0.93	1.00
21	Paper and paper products	2.80	1.55	3.00	2.31	1.24	2.00
22	Publishing, printing and reproduction of recorded media	2.88	1.61	3.00	2.54	1.36	2.00
23	Coke, refined petroleum and nuclear fuel	1.62	1.27	1.00	1.75	0.91	1.00
24	Chemicals and chemical products	3.88	1.65	5.00	3.37	1.42	3.00
25	Rubber and plastic products	2.86	1.68	3.00	2.09	1.28	2.00
26	Other non-metallic mineral products	2.54	1.59	2.00	2.07	1.14	2.00
27	Basic metals	2.23	1.57	2.00	2.01	1.29	2.00
28	Fabricated metal products, except machinery and equipments	2.23	1.63	2.00	2.09	1.29	2.00
29	Machinery and equipment, n.e.c.	3.56	1.69	4.00	2.72	1.50	3.00
30	Office, accounting and computing machinery	3.12	2.04	4.00	3.88	1.58	5.00
31	Electrical machinery and apparatus, n.e.c.	3.83	1.56	5.00	3.33	1.50	3.00
32	Radio, television, and communication equipment and apparatus	3.43	1.85	4.00	3.81	1.32	4.00
33	Medical, precision and optical instruments, watches and clocks	3.50	1.68	4.00	2.74	1.24	2.00
34	Motor vehicles, trailers and semi-trailers	2.97	1.81	3.00	2.58	1.50	2.00
35	Other transport equipment	2.74	1.86	3.00	2.33	1.46	2.00
36	Furniture, manufacturing n.e.c.	2.73	1.92	3.00	2.44	1.43	2.00
	Traditional	2.19	1.67	2.00	2.33	1.47	2.00
	Modern	3.26	1.78	4.00	2.68	1.50	2.00

Notes: See Table 1. "n.e.c." stands for Not Elsewhere Classified. Appendix Tables 2a and 2b provide details on industry sizes. "Modern" industries are comprised of the following (NIC 98 2-digit): 23, 24, 25, 27, 29, 30, 31, 32, 33, 34, 35.

Table 3b: Detailed input counts for industries by gender in unorganized sector, 2000

NIC	Industry Description	Unorganized sector, male			Unorganized sector, female		
		Mean	SD	Median	Mean	SD	Median
15	Food products and beverages	2.58	1.60	2.00	2.53	1.41	2.00
16	Tobacco products	1.67	0.85	1.00	1.20	0.77	1.00
17	Textiles	1.74	1.03	1.00	1.30	0.70	1.00
18	Wearing apparel; dressing and dyeing of fur	3.24	1.68	4.00	2.13	1.19	2.00
19	Leather; luggage, handbags, saddlery, harness and footwear	3.43	1.39	4.00	2.77	1.49	2.00
20	Wood and wood products, except furniture; straw and plating	1.56	0.98	1.00	1.29	0.62	1.00
21	Paper and paper products	2.48	1.28	2.00	1.75	0.92	2.00
22	Publishing, printing and reproduction of recorded media	2.50	1.36	2.00	2.71	1.45	2.00
23	Coke, refined petroleum and nuclear fuel	1.64	0.88	1.00	2.00	0.00	2.00
24	Chemicals and chemical products	3.31	1.42	3.00	3.47	1.36	3.00
25	Rubber and plastic products	2.07	1.28	2.00	1.77	1.03	1.00
26	Other non-metallic mineral products	2.07	1.14	2.00	2.01	1.16	2.00
27	Basic metals	2.04	1.34	2.00	1.67	0.80	1.00
28	Fabricated metal products, except machinery and equipments	2.06	1.27	2.00	2.22	1.42	2.00
29	Machinery and equipment, n.e.c.	2.58	1.46	2.00	2.77	1.58	3.00
30	Office, accounting and computing machinery	3.30	1.83	4.00			
31	Electrical machinery and apparatus, n.e.c.	3.30	1.49	3.00	2.80	1.57	3.00
32	Radio, television, and communication equipment and apparatus	3.74	1.40	4.00	2.43	0.53	2.00
33	Medical, precision and optical instruments, watches and clocks	2.62	1.21	2.00	1.67	0.52	2.00
34	Motor vehicles, trailers and semi-trailers	2.63	1.50	2.00	2.40	1.76	2.00
35	Other transport equipment	2.31	1.48	2.00	3.00	1.47	3.00
36	Furniture, manufacturing n.e.c.	2.46	1.45	2.00	1.96	1.09	2.00
	Traditional	2.43	1.52	2.00	1.89	1.17	2.00
	Modern	2.63	1.49	2.00	2.49	1.45	2.00

Notes: See Table 3a.

Table 4: Correlation of input counts across sectors within district, 2000

	Organized sector	Unorganized sector		
		Total	Male owned	Female owned
	(1)	(2)	(3)	(4)
A. Total district				
Organized sector	1.000			
Unorganized sector, total	0.108*	1.000		
Unorganized sector, male owned	0.069	0.668*	1.000	
Unorganized sector, female owned	0.099	0.950*	0.542*	1.000
B. Total district, after first normalizing input usage for plant by industry-sector average nationally				
Organized sector	1.000			
Unorganized sector, total	0.083	1.000		
Unorganized sector, male owned	0.041	0.639*	1.000	
Unorganized sector, female owned	0.055	0.931*	0.500*	1.000

Notes: Table documents correlations between input counts across sectors within districts. An asterisk denotes a correlation is statistically significant at the 10% level. Appendix Table 3 provides correlations across urban and rural areas within districts.

Table 5: Correlation between mean input counts and district traits, 2000

	Organized sector	Unorganized sector		
		Total	Male owned	Female owned
	(1)	(2)	(3)	(4)
A. District traits from 2001 Population Census				
Log population	0.072	-0.068	-0.112*	-0.049
Log population density	0.258*	-0.033	-0.044	-0.019
Age profile (demographic dividend)	0.082	0.190*	0.226*	0.126*
Share of population in scheduled caste	-0.091	0.121*	0.080*	0.094*
Share of population in scheduled tribe	0.029	-0.080	-0.046	-0.040
Educated worker share (% pop graduate)	0.315*	0.093*	0.112*	0.076
Literacy rate	0.196*	0.327*	0.370*	0.168*
Infrastructure: electricity access	0.063	-0.044	-0.016	-0.057
Infrastructure: paved roads	0.059	0.256*	0.282*	0.147*
Travel time to nearest of India's ten largest cities	-0.053	-0.003	0.025	-0.088*
Strength of household banking sector	0.303*	0.313*	0.274*	0.204*
Urbanization rate (% urban)	0.262*	0.216*	0.231*	0.156*
Log consumption per capita	0.277*	0.273*	0.302*	0.151*
B. District traits for manufacturing sector from ASI/NSS in 2000				
Organized sector share of establishments	-0.094	0.026	0.025	0.053
Organized sector share of employment	0.133*	0.160*	0.160*	0.099*
Organized sector share of output	0.168*	0.170*	0.168*	0.093*
Female share of unorg. sector establishments	-0.005	0.034	0.135*	-0.038
Female share of unorg. sector employment	-0.065	-0.053	0.071	-0.107*
Female share of unorg. sector output	-0.124*	-0.069	0.024	-0.085*
HHI index of establishments across sectors	-0.106*	0.020	0.019	0.054
HHI index of employment across sectors	-0.018	0.018	0.017	-0.007
HHI index of output across sectors	0.020	0.038	0.034	-0.007

Notes: Table documents correlations between input counts and district conditions. District traits in Panel A are from the 2001 Population Census. District traits are expressed in log values or percentage point values as indicated. District traits in Panel B are calculated from the ASI/NSS data. An asterisk denotes a correlation is statistically significant at the 10% level. Consumption per capita taken from NSS Round 55 (household survey).

Table 6a: Analysis of mean input counts by district-industry

	Organized sector		Unorganized sector	
	(1)	(2)	(3)	(4)
Log of district population	-0.003 (0.024)	0.094+ (0.053)	-0.043+ (0.024)	0.018 (0.033)
Log of district population density	0.046 (0.030)	0.103++ (0.045)	-0.116+++ (0.020)	-0.037 (0.030)
Age profile (demographic dividend)	-0.107+++ (0.037)	-0.080 (0.062)	0.026 (0.025)	-0.067+ (0.039)
Educated worker share (% pop graduate)	0.017 (0.040)	0.078 (0.057)	0.019 (0.027)	0.064+ (0.034)
Literacy rate	0.100+++ (0.039)	0.133++ (0.052)	0.061++ (0.025)	0.092+++ (0.033)
Infrastructure: electricity access	0.038 (0.025)	0.109+++ (0.038)	0.049+++ (0.019)	0.044 (0.031)
Infrastructure: paved roads	-0.037 (0.028)	0.000 (0.047)	0.080+++ (0.020)	-0.025 (0.032)
Log travel time to closest large city	0.030 (0.019)	0.069+++ (0.025)	-0.001 (0.013)	0.003 (0.017)
Strength of household banking environment	0.079+++ (0.026)	-0.056 (0.047)	0.002 (0.019)	-0.013 (0.030)
Urbanization rate	-0.009 (0.047)	-0.140++ (0.064)	0.087+++ (0.029)	-0.006 (0.040)
Log per capita consumption	0.218+ (0.124)	0.271 (0.170)	0.013 (0.075)	0.138 (0.098)
Log organized employment for district-industry	0.013 (0.012)	0.012 (0.012)	-0.005 (0.012)	-0.008 (0.012)
Log unorganized employment for district-industry	0.036+++ (0.013)	0.033++ (0.013)	-0.011++ (0.005)	-0.018+++ (0.005)
HHI of local employment across mfg industries	-0.026 (0.019)	-0.031 (0.022)	0.039 (0.025)	0.053++ (0.026)
Industry fixed effects	Yes	Yes	Yes	Yes
State fixed effects		Yes		Yes
Observations	3,589	3,589	10,674	10,674
Adjusted R-squared	0.307	0.330	0.270	0.289

Notes: Estimations quantify the relationship between average district-industry input counts and local conditions. District-level traits are taken from the 2001 Census. Estimations weight observations by an interaction of district size and industry size, include industry fixed effects, and cluster standard errors by district. Non-logarithm variables are transformed to have unit standard deviation for interpretation. Appendix Tables 4a and 4b provide additional specifications.

Table 6b: Table 6a using normalized input counts

	Organized sector		Unorganized sector	
	(1)	(2)	(3)	(4)
Log of district population	-0.001 (0.023)	0.100+ (0.052)	-0.044+ (0.024)	0.024 (0.033)
Log of district population density	0.048 (0.030)	0.102++ (0.044)	-0.114+++ (0.020)	-0.034 (0.030)
Age profile (demographic dividend)	-0.105+++ (0.036)	-0.068 (0.061)	0.027 (0.025)	-0.076+ (0.039)
Educated worker share (% pop graduate)	0.016 (0.040)	0.074 (0.056)	0.022 (0.027)	0.059+ (0.034)
Literacy rate	0.099++ (0.038)	0.132++ (0.051)	0.062++ (0.025)	0.095+++ (0.033)
Infrastructure: electricity access	0.038 (0.025)	0.106+++ (0.038)	0.047++ (0.019)	0.037 (0.031)
Infrastructure: paved roads	-0.039 (0.027)	-0.001 (0.046)	0.074+++ (0.019)	-0.026 (0.032)
Log travel time to closest large city	0.031+ (0.018)	0.070+++ (0.025)	-0.003 (0.013)	0.001 (0.017)
Strength of household banking environment	0.080+++ (0.026)	-0.053 (0.046)	-0.004 (0.019)	-0.016 (0.030)
Urbanization rate	-0.003 (0.046)	-0.131++ (0.063)	0.079+++ (0.029)	0.000 (0.040)
Log per capita consumption	0.212+ (0.123)	0.249 (0.168)	0.007 (0.075)	0.132 (0.097)
Log organized employment for district-industry	0.018 (0.012)	0.017 (0.012)	-0.020+ (0.012)	-0.025++ (0.012)
Log unorganized employment for district-industry	0.017 (0.013)	0.014 (0.013)	-0.009++ (0.004)	-0.015+++ (0.005)
HHI of local employment across mfg industries	-0.034+ (0.019)	-0.039+ (0.022)	0.028 (0.025)	0.044+ (0.026)
Industry fixed effects	Yes	Yes	Yes	Yes
State fixed effects		Yes		Yes
Observations	3589	3589	10674	10674
Adjusted R-squared	0.119	0.148	0.058	0.083

Notes: See Table 6a. Normalized input counts are residuals from a regression of input counts on industry fixed effects and plant size dummies.

Table 7: Correlation between industry traits and mean input counts, 2000

	Organized sector	Unorganized sector		
		Total	Male owned	Female owned
	(1)	(2)	(3)	(4)
Log labor intensity	-0.189	-0.246*	-0.169	-0.246*
Log capital intensity	0.173	0.039	-0.036	-0.154
Log materials intensity	0.131	0.080	0.053	0.103
Log average wage	-0.012	0.003	0.045	0.262*
Log financial dependency	0.134	-0.064	-0.045	0.053
Log import dependency	0.034	0.142	0.044	0.140
Organized sector share of establishments	0.128	-0.032	0.019	0.045
Organized sector share of employment	0.470*	0.211	0.245*	0.215
Organized sector share of output	0.365*	0.213	0.223*	0.300*
Female share of unorg. sector establishments	-0.036	0.006	0.152	-0.153
Female share of unorg. sector employment	-0.160	-0.200	0.016	-0.258*
Female share of unorg. sector output	-0.408*	-0.255*	0.022	-0.253*

Notes: See Table 5. Table documents correlations between input counts and industry traits.

Table 8: Estimations of manufacturing production functions, organized sector

	DV: Log output in manufacturing establishment, 2000 sample						
	Full	Full	Full	Full	Full	Traditional	Modern
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log employment	0.121+++ (0.006)	0.121+++ (0.006)	0.123+++ (0.006)	0.122+++ (0.006)	0.125+++ (0.007)	0.106+++ (0.008)	0.147+++ (0.010)
Log capital	0.011+++ (0.004)	0.011+++ (0.004)	0.010+++ (0.004)	0.010+++ (0.004)	0.008++ (0.004)	-0.001 (0.005)	0.028+++ (0.006)
Log materials	0.931+++ (0.006)	0.930+++ (0.006)	0.931+++ (0.006)	0.930+++ (0.006)	0.933+++ (0.006)	0.943+++ (0.008)	0.907+++ (0.010)
(0,1) Input counts above industry median		0.039+++ (0.009)		0.031+++ (0.010)	0.034+++ (0.010)	0.035+++ (0.013)	0.028+ (0.015)
District average of inputs counts metric			0.129+++ (0.031)	0.102+++ (0.033)	0.122+++ (0.033)	0.117+++ (0.039)	0.053 (0.059)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional urban traits					Yes		
Observations	25,106	25,106	25,106	25,106	24,139	14,749	10,357

Notes: Estimations consider simple production functions for manufacturing establishments in India. Indian data are taken from ASI and NSS. Additional urban traits include log manufacturing employment in district, log manufacturing employment in plant's district-industry, log share of local manufacturing employment in the unorganized sector, and log share of district-industry manufacturing employment in the unorganized sector. Estimations report robust standard errors, include industry fixed effects, and weight observations by sample weights. + significant at 10% level; ++ significant at 5% level; +++ significant at 1% level.

Table 9: Estimations of manufacturing production functions, unorganized sector

	DV: Log output in manufacturing establishment, 2000 sample						
	Full	Full	Full	Full	Full	Traditional	Modern
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log employment	0.362+++ (0.012)	0.361+++ (0.012)	0.362+++ (0.012)	0.361+++ (0.012)	0.355+++ (0.013)	0.359+++ (0.013)	0.408+++ (0.034)
Log capital	0.089+++ (0.004)	0.089+++ (0.004)	0.088+++ (0.004)	0.089+++ (0.004)	0.089+++ (0.004)	0.089+++ (0.004)	0.085+++ (0.011)
Log materials	0.628+++ (0.005)	0.626+++ (0.005)	0.627+++ (0.005)	0.626+++ (0.005)	0.626+++ (0.005)	0.626+++ (0.005)	0.640+++ (0.017)
(0,1) Input counts above industry median		0.029++ (0.012)		0.024+ (0.012)	0.025++ (0.012)	0.028++ (0.013)	-0.093+++ (0.033)
District average of inputs counts metric			0.068+ (0.037)	0.046 (0.040)	0.025 (0.040)	0.042 (0.041)	0.139 (0.108)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional urban traits					Yes		
Observations	120,599	120,599	120,599	120,599	120,599	113,003	7,596

Notes: See Table 8.

Table 10: Extended estimations of manufacturing production functions, organized sector

	DV: Log output in manufacturing establishment, 2000 sample						
	Full	Full	Full	Full	<50 empl	50-100 empl	>100 empl
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log employment	0.121+++ (0.006)	0.120+++ (0.006)	0.121+++ (0.006)	0.121+++ (0.006)	0.123+++ (0.011)	0.142+++ (0.035)	0.116+++ (0.019)
Log capital	0.011+++ (0.004)	0.011+++ (0.004)	0.011+++ (0.004)	0.011+++ (0.004)	0.003 (0.004)	0.025+++ (0.009)	0.064+++ (0.011)
Log materials	0.930+++ (0.006)	0.930+++ (0.006)	0.931+++ (0.006)	0.931+++ (0.006)	0.933+++ (0.007)	0.914+++ (0.016)	0.907+++ (0.019)
(0,1) Input counts above industry median	0.039+++ (0.009)				0.046+++ (0.011)	0.016 (0.025)	0.029 (0.033)
(0,1) Input counts in 50th-75th percentile		0.031++ (0.015)					
(0,1) Input counts above 75th percentile		0.050+++ (0.012)					
Raw input counts on 0-5 scale			0.007+ (0.004)				
(0,1) Input counts above district-industry average				0.014 (0.009)			
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean input count	2.60	2.60	2.60	2.60	2.42	2.64	2.99
Observations	25,106	25,106	25,106	25,106	13,635	5,071	6,400

Notes: See Table 8.

Appendix Table 1: Descriptive statistics for districts

	Mean	St. dev.	Median	Min	Max
District population	3,606,737	2,684,933	3,059,423	33,224	13,900,000
District population density (persons per sq. km.)	1,793	4,743	484	2	24,963
Age profile (demographic dividend)	1.47	0.28	1.46	0.92	2.59
Educated worker share (% pop graduate)	7.3%	3.8%	6.1%	1.1%	21.5%
Literacy rate	59.1%	11.7%	59.2%	26.7%	85.4%
Infrastructure: electricity access	0.32	0.31	0.25	0.00	1.00
Infrastructure: paved roads	0.73	0.24	0.80	0.13	1.00
Proximity to India's ten largest cities (min driving)	400	256	404	0	1,244
Strength of household banking environment	0.38	0.15	0.36	0.03	0.76
Urbanization rate	0.35	0.24	0.28	0.00	1.00
Consumption per capita (1999 INR per month)	604	187	569	277	1,331

Notes: See Table 1.

Appendix Table 2a: Sizes of industries, 2000

NIC	Industry Description	Organized sector			Unorganized sector		
		Plants	Empl	Output	Plants	Empl	Output
15	Food products and beverages	21.4	1,261.4	1,234,459	1,743.6	4,622.3	392,886
16	Tobacco products	2.2	475.2	103,597	169.7	485.8	16,820
17	Textiles	12.3	1,245.9	773,018	724.5	2,034.3	149,788
18	Wearing apparel; dressing and dyeing of fur	2.8	329.2	141,507	2,018.0	3,257.1	93,687
19	Leather; luggage, handbags, saddlery, harness and footwear	2.2	135.7	88,793	130.5	287.3	22,159
20	Wood and wood products, except furniture; straw and plating	2.7	45.6	18,731	1,226.7	2,646.0	76,124
21	Paper and paper products	3.2	176.1	178,617	59.6	189.9	17,998
22	Publishing, printing and reproduction of recorded media	3.0	116.7	53,516	109.4	382.3	36,042
23	Coke, refined petroleum and nuclear fuel	0.8	66.7	771,868	5.6	19.0	2,807
24	Chemicals and chemical products	9.9	779.8	1,439,134	38.8	198.9	40,877
25	Rubber and plastic products	6.4	251.1	258,618	50.5	216.3	42,320
26	Other non-metallic mineral products	10.5	428.3	288,191	623.9	2,562.2	116,946
27	Basic metals	6.5	551.0	820,940	15.5	73.1	28,293
28	Fabricated metal products, except machinery and equipments	7.9	292.6	181,534	282.1	797.0	80,883
29	Machinery and equipment, n.e.c.	8.9	422.9	373,526	52.5	224.1	43,967
30	Office, accounting and computing machinery	0.2	17.1	36,757	0.2	0.9	309
31	Electrical machinery and apparatus, n.e.c.	3.7	229.0	237,760	35.8	183.6	283,943
32	Radio, television, and communication equipment and apparatus	1.0	109.1	166,577	4.0	25.7	2,995
33	Medical, precision and optical instruments, watches and clocks	0.9	58.8	42,137	4.7	19.3	3,511
34	Motor vehicles, trailers and semi-trailers	2.5	257.3	389,512	9.3	45.9	9,931
35	Other transport equipment	1.8	182.7	200,111	8.3	38.2	7,273
36	Furniture, manufacturing n.e.c.	2.1	119.3	105,705	435.0	1,078.1	90,704
	Traditional	70.3	4,626.0	3,167,666	7,523.1	18,342.3	1,094,037
	Modern	42.8	2,925.5	4,736,939	225.1	1,045.0	466,225

Notes: See Table 3a. Plants and employments are expressed in thousands. Output is expressed in millions of rupees. "n.e.c." stands for Not Elsewhere Classified.

Appendix Table 2b: Sizes of industries by gender in unorganized sector, 2000

NIC	Industry Description	Unorganized sector, male			Unorganized sector, female		
		Plants	Empl	Output	Plants	Empl	Output
15	Food products and beverages	1,508.4	4,046.5	311,006	205.0	383.1	14,901
16	Tobacco products	104.0	391.7	12,408	65.1	88.0	727
17	Textiles	481.3	1,531.8	110,196	227.2	373.6	13,277
18	Wearing apparel; dressing and dyeing of fur	1,299.7	2,330.6	73,041	702.7	833.2	11,605
19	Leather; luggage, handbags, saddlery, harness and footwear	126.4	269.6	18,685	1.7	4.2	501
20	Wood and wood products, except furniture; straw and plating	992.8	2,229.6	61,395	216.1	347.1	3,886
21	Paper and paper products	37.1	132.5	11,436	20.5	39.3	1,274
22	Publishing, printing and reproduction of recorded media	95.2	306.0	23,234	5.6	22.3	1,033
23	Coke, refined petroleum and nuclear fuel	5.4	17.6	2,468	0.1	0.3	11
24	Chemicals and chemical products	27.2	127.9	16,866	5.1	13.8	1,058
25	Rubber and plastic products	33.2	133.1	18,986	9.1	19.3	2,640
26	Other non-metallic mineral products	581.0	2,107.4	77,685	25.3	73.3	4,310
27	Basic metals	13.3	51.9	8,266	0.3	1.7	347
28	Fabricated metal products, except machinery and equipments	261.6	690.7	60,353	5.4	24.1	2,747
29	Machinery and equipment, n.e.c.	45.3	176.8	29,657	1.3	6.4	1,118
30	Office, accounting and computing machinery	0.1	0.6	115	-	-	-
31	Electrical machinery and apparatus, n.e.c.	30.9	142.9	276,843	0.9	4.3	401
32	Radio, television, and communication equipment and apparatus	2.8	14.2	2,075	0.1	1.0	155
33	Medical, precision and optical instruments, watches and clocks	3.7	12.4	1,959	0.1	0.4	47
34	Motor vehicles, trailers and semi-trailers	7.3	28.7	5,156	0.2	1.0	137
35	Other transport equipment	7.4	31.4	5,801	0.1	0.3	91
36	Furniture, manufacturing n.e.c.	371.6	937.3	76,912	55.3	92.2	3,724
	Traditional	1,529.7	2,280.4	57,984	5,859.1	14,973.7	836,350
	Modern	17.3	48.5	6,006	176.6	737.5	368,193

Notes: See Appendix Table 2a.

App. Table 3: Correlation of input counts across urban and rural sectors within district, 2000

	Urban areas				Rural areas			
	Organized sector	Unorganized sector			Organized sector	Unorganized sector		
		Total	Male owned	Female owned		Total	Male owned	Female owned
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urban - Organized sector	1.000							
Urban - Unorg. sector, total	0.002	1.000						
Urban - Unorg. sector, male	-0.018	0.660*	1.000					
Urban - Unorg. sector, female	-0.010	0.950*	0.519*	1.000				
Rural - Organized sector	0.470*	0.025	0.038	0.016	1.000			
Rural - Unorg. sector, total	0.026	0.452*	0.389*	0.437*	0.066	1.000		
Rural - Unorg. sector, male	-0.059	0.364*	0.348*	0.346*	0.040	0.614*	1.000	
Rural - Unorg. sector, female	0.022	0.437*	0.374*	0.445*	0.063	0.962*	0.496*	1.000

Notes: See Panel A of Table 4.

Appendix Table 4a: Analysis of mean input counts by district-industry, weighted

	Organized sector			Unorganized sector			Unorganized sector, male			Unorganized sector, female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log of district population	0.019 (0.022)	-0.003 (0.024)	0.094+ (0.053)	-0.072+++ (0.020)	-0.043+ (0.024)	0.018 (0.033)	-0.085+++ (0.024)	-0.068++ (0.030)	-0.015 (0.040)	-0.086+ (0.045)	-0.041 (0.053)	0.044 (0.076)
Log of district population density	0.048 (0.030)	0.046 (0.030)	0.103++ (0.045)	-0.112+++ (0.020)	-0.116+++ (0.020)	-0.037 (0.030)	-0.113+++ (0.024)	-0.114+++ (0.024)	-0.032 (0.036)	-0.075+ (0.045)	-0.072 (0.044)	-0.043 (0.068)
Age profile (demographic dividend)	-0.098+++ (0.036)	-0.107+++ (0.037)	-0.080 (0.062)	0.018 (0.025)	0.026 (0.025)	-0.067+ (0.039)	0.003 (0.030)	0.008 (0.030)	-0.045 (0.047)	-0.026 (0.056)	-0.019 (0.056)	-0.073 (0.089)
Educated worker share (% pop graduate)	-0.002 (0.039)	0.017 (0.040)	0.078 (0.057)	0.018 (0.027)	0.019 (0.027)	0.064+ (0.034)	0.051 (0.033)	0.052 (0.033)	0.056 (0.040)	0.024 (0.055)	0.024 (0.055)	0.132+ (0.076)
Literacy rate	0.091++ (0.038)	0.100+++ (0.039)	0.133++ (0.052)	0.056++ (0.025)	0.061++ (0.025)	0.092+++ (0.033)	0.034 (0.031)	0.036 (0.031)	0.079+ (0.041)	0.040 (0.053)	0.037 (0.053)	0.102 (0.068)
Infrastructure: electricity access	0.039 (0.025)	0.038 (0.025)	0.109+++ (0.038)	0.050+++ (0.019)	0.049+++ (0.019)	0.044 (0.031)	0.060+++ (0.022)	0.060+++ (0.022)	0.027 (0.037)	0.028 (0.041)	0.024 (0.041)	0.075 (0.068)
Infrastructure: paved roads	-0.028 (0.027)	-0.037 (0.028)	0.000 (0.047)	0.077+++ (0.020)	0.080+++ (0.020)	-0.025 (0.032)	0.096+++ (0.024)	0.098+++ (0.024)	-0.030 (0.038)	0.082+ (0.042)	0.086++ (0.042)	-0.064 (0.072)
Log travel time to closest large city	0.030+ (0.018)	0.030 (0.019)	0.069+++ (0.025)	-0.002 (0.013)	-0.001 (0.013)	0.003 (0.017)	0.022 (0.016)	0.023 (0.016)	0.014 (0.020)	0.018 (0.030)	0.023 (0.030)	0.036 (0.038)
Strength of household banking environment	0.084+++ (0.026)	0.079+++ (0.026)	-0.056 (0.047)	0.003 (0.019)	0.002 (0.019)	-0.013 (0.030)	-0.018 (0.023)	-0.019 (0.023)	-0.016 (0.035)	-0.040 (0.040)	-0.035 (0.040)	-0.043 (0.071)
Urbanization rate	0.012 (0.046)	-0.009 (0.047)	-0.140++ (0.064)	0.085+++ (0.029)	0.087+++ (0.029)	-0.006 (0.040)	0.073++ (0.035)	0.075++ (0.035)	-0.004 (0.047)	0.135++ (0.064)	0.145++ (0.064)	0.023 (0.092)
Log per capita consumption	0.268++ (0.120)	0.218+ (0.124)	0.271 (0.170)	-0.003 (0.074)	0.013 (0.075)	0.138 (0.098)	-0.002 (0.091)	0.013 (0.093)	0.098 (0.119)	0.134 (0.156)	0.147 (0.161)	0.113 (0.214)
Log organized employment for district-industry		0.013 (0.012)	0.012 (0.012)		-0.005 (0.012)	-0.008 (0.012)		-0.012 (0.015)	-0.015 (0.015)		-0.022 (0.027)	-0.020 (0.027)
Log unorganized employment for district-industry		0.036+++ (0.013)	0.033++ (0.013)		-0.011++ (0.005)	-0.018+++ (0.005)		-0.007 (0.006)	-0.013++ (0.006)		-0.011 (0.009)	-0.015 (0.009)
HHI of local employment across mfg industries		-0.026 (0.019)	-0.031 (0.022)		0.039 (0.025)	0.053++ (0.026)		0.012 (0.033)	0.036 (0.035)		0.127 (0.102)	0.172 (0.123)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects			Yes			Yes			Yes			Yes
Observations	3,589	3,589	3,589	10,674	10,674	10,674	6,671	6,671	6,671	2,081	2,081	2,081
Adjusted R-squared	0.306	0.307	0.330	0.269	0.270	0.289	0.304	0.304	0.327	0.258	0.259	0.269

Notes: See Table 6a.

Appendix Table 4b: Analysis of mean input counts by district-industry, unweighted

	Organized sector			Unorganized sector			Unorganized sector, male			Unorganized sector, female		
	(1)	(1)	(2)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log of district population	0.003 (0.022)	-0.022 (0.023)	0.083 (0.050)	-0.055+++ (0.017)	-0.025 (0.020)	0.052+ (0.028)	-0.061+++ (0.021)	-0.044+ (0.025)	0.027 (0.035)	-0.073+ (0.039)	-0.048 (0.046)	0.035 (0.063)
Log of district population density	0.055+ (0.029)	0.052+ (0.029)	0.105++ (0.043)	-0.100+++ (0.017)	-0.102+++ (0.017)	-0.024 (0.025)	-0.113+++ (0.020)	-0.113+++ (0.020)	-0.032 (0.029)	-0.053 (0.038)	-0.047 (0.039)	0.014 (0.057)
Age profile (demographic dividend)	-0.119+++ (0.035)	-0.132+++ (0.036)	-0.123++ (0.059)	0.002 (0.020)	0.009 (0.020)	-0.049 (0.033)	-0.017 (0.025)	-0.013 (0.025)	-0.026 (0.041)	-0.053 (0.043)	-0.052 (0.043)	-0.093 (0.072)
Educated worker share (% pop graduate)	0.010 (0.038)	0.037 (0.038)	0.090+ (0.054)	0.000 (0.023)	0.005 (0.023)	0.038 (0.029)	0.031 (0.028)	0.034 (0.028)	0.019 (0.035)	-0.028 (0.049)	-0.028 (0.049)	0.081 (0.066)
Literacy rate	0.108+++ (0.036)	0.123+++ (0.037)	0.162+++ (0.049)	0.055+++ (0.021)	0.060+++ (0.021)	0.084+++ (0.027)	0.052++ (0.025)	0.054++ (0.025)	0.083++ (0.033)	0.052 (0.043)	0.051 (0.043)	0.104+ (0.055)
Infrastructure: electricity access	0.036 (0.024)	0.035 (0.024)	0.117+++ (0.037)	0.031++ (0.015)	0.028+ (0.015)	0.007 (0.025)	0.049+++ (0.018)	0.048+++ (0.018)	0.009 (0.030)	-0.016 (0.033)	-0.023 (0.034)	-0.017 (0.054)
Infrastructure: paved roads	-0.010 (0.026)	-0.022 (0.027)	0.023 (0.045)	0.055+++ (0.016)	0.056+++ (0.016)	-0.054++ (0.026)	0.073+++ (0.019)	0.074+++ (0.019)	-0.045 (0.031)	0.040 (0.035)	0.040 (0.035)	-0.120++ (0.056)
Log travel time to closest large city	0.030+ (0.017)	0.030+ (0.018)	0.071+++ (0.024)	0.006 (0.011)	0.006 (0.011)	-0.003 (0.014)	0.024+ (0.013)	0.025+ (0.013)	0.006 (0.016)	0.006 (0.026)	0.008 (0.026)	0.017 (0.033)
Strength of household banking environment	0.084+++ (0.025)	0.077+++ (0.025)	-0.073+ (0.044)	0.036++ (0.016)	0.033++ (0.016)	0.031 (0.025)	0.025 (0.019)	0.023 (0.019)	0.040 (0.030)	0.004 (0.035)	0.010 (0.035)	0.001 (0.058)
Urbanization rate	0.001 (0.044)	-0.028 (0.045)	-0.152++ (0.061)	0.096+++ (0.025)	0.093+++ (0.025)	-0.000 (0.034)	0.092+++ (0.030)	0.088+++ (0.031)	0.010 (0.041)	0.152+++ (0.054)	0.155+++ (0.054)	0.018 (0.076)
Log per capita consumption	0.246++ (0.116)	0.174 (0.119)	0.287+ (0.161)	0.078 (0.061)	0.056 (0.062)	0.129 (0.081)	0.037 (0.073)	0.021 (0.075)	0.074 (0.098)	0.171 (0.129)	0.152 (0.134)	0.099 (0.175)
Log organized employment for district-industry		0.017 (0.012)	0.015 (0.012)		0.013 (0.010)	0.011 (0.010)		0.007 (0.012)	0.004 (0.012)		-0.010 (0.022)	-0.006 (0.023)
Log unorganized employment for district-industry		0.040+++ (0.013)	0.036+++ (0.013)		-0.005 (0.004)	-0.013+++ (0.004)		-0.001 (0.005)	-0.009+ (0.005)		0.000 (0.009)	-0.007 (0.009)
HHI of local employment across mfg industries		-0.041++ (0.018)	-0.047++ (0.021)		0.071+++ (0.021)	0.077+++ (0.022)		0.038 (0.027)	0.045 (0.029)		0.095 (0.078)	0.142 (0.098)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects			Yes			Yes			Yes			Yes
Observations	3,589	3,589	3,589	10,674	10,674	10,674	6,671	6,671	6,671	2,081	2,081	2,081
Adjusted R-squared	0.313	0.316	0.339	0.258	0.259	0.277	0.284	0.284	0.305	0.237	0.237	0.250

Notes: See Table 6a and Appendix Table 4a. Estimations are unweighted.

App. Table 5: Extensions of Table 9 for female-owned businesses

	DV: Log output in manufacturing establishment, 2000 sample						
	Full	Full	Full	Full	Full	Full	Full
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log employment	0.381+++ (0.037)	0.381+++ (0.036)	0.381+++ (0.037)	0.380+++ (0.037)	0.383+++ (0.037)	0.371+++ (0.037)	0.377+++ (0.038)
Log capital	0.062+++ (0.008)	0.063+++ (0.008)	0.063+++ (0.008)	0.063+++ (0.008)	0.063+++ (0.008)	0.062+++ (0.008)	0.060+++ (0.008)
Log materials	0.662+++ (0.014)	0.658+++ (0.015)	0.661+++ (0.014)	0.660+++ (0.016)	0.660+++ (0.016)	0.657+++ (0.016)	0.654+++ (0.016)
(0,1) Input counts above industry median		0.050 (0.031)		0.023 (0.037)	0.022 (0.038)	0.027 (0.039)	0.024 (0.037)
District average of inputs counts metric			0.257+++ (0.082)	0.234++ (0.100)	0.234++ (0.101)	0.198+ (0.107)	0.201+ (0.104)
Log female-owned firms in district-industry					-0.007 (0.014)	-0.019 (0.015)	-0.055+++ (0.020)
Log male-owned firms in district-industry						0.029++ (0.012)	-0.009 (0.017)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional urban traits							Yes
Observations	21,516	21,516	21,516	21,516	21,516	20,472	20,472

Notes: See Table 8.