Stockholm School of Economics
Department of Economics
5350 Master's Thesis in Economics
Academic Year 2022/2023

Firm Dynamics, Productivity and Demand Shocks: A Study of Indian Manufacturing Plants

Syed Ravish Usman Rizvi (42145)

Abstract

Firm growth in the Indian manufacturing sector has puzzled economists for many years due to the persistence of many small and unproductive plants. I study the importance of producer-specific demand, along with productivity and prices, for firm dynamics in this setting. First, I use micro-level panel data from India to estimate these measures at the plant and individual product levels. I find high dispersion and low persistence in these variables relative to the US data. Moreover, the rate of persistence varies by the category of producers with larger plants observing higher rates of persistence in productivity, which is further associated with a significantly higher persistence in demand shocks. Finally, I study the evolution in these measures over the life cycle of plants and show that productivity and prices increase over time with younger plants being the least productive but observing higher demand, a potential factor in their survival. This is in contrast to the US and thus, establishes an important distinction in the markets of the two countries. Furthermore, the results highlight the role of the idiosyncratic demand component in firm dynamics within the Indian manufacturing sector.

JEL classification: C26, D22, D24, D43, E23, L11, L60

Keywords: Productivity, firm dynamics, demand, manufacturing

Supervisor: Sampreet Goraya

Date submitted: 15 May 2023

Date examined: 2 June 2023

Discussant: Gosta Lycke

Examiner: Kelly Ragan

Acknowledgments

I am deeply grateful to my supervisor, Sampreet Goraya, for his constant mentoring and guidance throughout this last year. Special thanks to Abhijeet Singh for nudging me in the right direction early on in the master's program as well as Mambuna Njie for valuable comments and discussions. Finally, I dedicate my work to my family, none of this would be possible without their unwavering support.

Table of Contents

1 Introduction

In developed countries with efficient markets, it is believed that markets reallocate demand from less productive to more productive firms, which constitutes a significant source of firm growth (see, e.g., Baily, Hulten, and Campbell (1992); Bartelsman and Doms (2000); Aw, Chen, and Roberts (2001); and Foster, Haltiwanger, and Krizan (2006)). Foster, Haltiwanger, and Syverson (2008) undertake a decomposition of aggregate productivity growth in the US and find that firms not only become more productive over time, but markets also reallocate share to the more productive firms contributing to the overall productivity growth. In contrast, Hsieh and Klenow (2009) document significant dispersion in the productivity of Indian firms along with high volatility and in later work, also show the persistence of a large number of small and unproductive firms, even in narrowly defined sectors. Studying the life cycle of manufacturing plants, they find that the US has a higher number of large firms that start small and grow much faster over time in size and productivity than India, where there is a persistence of a large number of small, unproductive firms.

While Hsieh and Klenow (2014) show the patterns in productivity over the life cycle of plants and explain about one-third of the difference between India and the US through tax-like wedges, I explore the role of the demand side, in addition to productivity, and analyze the firm dynamics in the Indian manufacturing sector, in order to compare the results with Foster, Haltiwanger, and Syverson (2008) and further our understanding of the reason behind the persistence of small and unproductive plants in the market. They distinguish between revenue and physical measures of productivity and show how the former confounds factor price effects and idiosyncratic demand components with efficiency differences. The latter can include long-standing buyer-seller relationships or connections with authorities that help businesses clear bureaucratic hurdles. Thus, if Indian establishments are able to exploit such avenues of demand that are specific to them and if this component is quite persistent such that they do not lose their market share to bigger and more productive firms, it would provide a further explanation for the persistence of less productive firms over time.

In the following analyses, I implement a select set of exercises following Foster, Haltiwanger, and Syverson (2008) but using the ASI (Annual Survey of Industries) panel data for the period 1998-2016. These revolve around estimating the production function and idiosyncratic demand levels for the plants as well as studying the persistence and evolution of various measures of productivity, prices, and demand. While they only analyze certain products from the sample and only one product per firm, given the rich micro-level panel data at my disposal which contains detailed information on all products manufactured by registered factories in India, I execute a graded analysis, exploring both aggregate plant-level dynamics and individual product-level dynamics for all plants. This distinction is important as a lot of earlier research has been conducted at the plant level but our variables of interest are likely to vary even within plants over the different products. Additionally, I differentiate between multi-product firms to see if this aspect might be playing a role in how firms exploit the demand component for their growth.

Some of the results validate past research and instill confidence in our analysis, such as the high dispersion in productivity measures but low growth over the life cycle of plants, high persistence in the producer-specific demand measure, and the least productive plants being the ones to exit the market. However, most results are in contrast to Foster, Haltiwanger, and Syverson (2008). I find both physical and revenue productivity to be slightly more volatile than in the US. Moreover, results from the evolution of productivity are completely opposite to the US, with younger firms being 3-4% less productive relative to old incumbents in India. These findings are irrespective of whether we consider the plant-level or individual product-level data.

The most interesting results come in the form of the producer-specific demand and also highlight the importance of the graded analysis. In the analysis of persistence in productivity and prices, I find that the rate of persistence is higher in the product-level sample compared to the plant level. This divergence was intriguing and to investigate what could cause this, I studied how these estimates vary by categories of multi-product plants using the product-level data. This revealed that the largest plants (those manufacturing the most number of products) observe significantly higher persistence in revenue productivity and this is not driven by physical efficiency or prices, but by the idiosyncratic demand component which is found to be as high as 0.96. The main discovery relates to the evolution of this measure. Analyzing how our variables of interest change over the life cycle of plants, I find that young producers are able to attract higher demand relative to old incumbents which seems to be the key to their survival given that they are less productive and also observe lower prices.

There are three main takeaways from this paper. First, I emphasize the importance of differentiating between plant-level and individual product-level analysis. Second, I document how the Indian market differs from its counterpart in the US in terms of the persistence and evolution in productivity, prices and idiosyncratic demand measures. Finally, I highlight the role of producer-specific demand in firm dynamics within the Indian manufacturing sector. These findings open up many avenues for future research, such as estimating the causal impact of this demand component on firm and aggregate growth as well as modeling such economic behavior whereby these elements can hinder the reallocation of market share and efficient firm dynamics.

The rest of the paper proceeds as follows. The next section presents an overview of the literature and where this study fits in. Section 3 provides a description of the data and how the different datasets were prepared for the analysis. Section 4 describes the empirical strategy for estimating the production function and demand shocks. Sections 5 and 6 present all the results relating to the estimation and firm dynamics. Section 7 includes some robustness checks and finally, section 8 concludes with a brief discussion of the results and future research avenues.

2 Related Literature

The issue of the Indian manufacturing sector has attracted considerable attention in the last couple of decades, with panel datasets becoming available. On the one hand, researchers have tackled the issue from the perspective of the misallocation of resources. As mentioned earlier, Hsieh and Klenow (2014) develop a general equilibrium model in which they incorporate tax-like wedges which allow them to explain about one-third of the difference in the life-cycle patterns between India and the US. Quantitatively, this can mean a 25% drop in aggregate TFP. Their reasoning is that returns to investment in India are lower due to high taxes or labor costs which hinder plant investment and growth. In related work, Alfaro and C. (2014) show continued dominance of state-owned enterprises and older manufacturing enterprises, despite significant entry of new firms.

Bau and Matray (2023) study another strand of the misallocation literature through financial frictions. They exploit the staggered roll-out of foreign capital liberalization to estimate its impact on the misallocation of capital across firms. They estimate the marginal product of firms and find that the firms with the highest MRPK increased not only their capital but also revenues and wage bills. Another type of friction that has been recently explored by Boehm and Oberfield (2020) is the one arising from imperfect contract enforcement in the market for inputs. They also use the ASI data to document how areas with poorer contract enforceability observe more distortion in input sourcing and production decisions, with industries that rely on non-standardized input decreasing their share of intermediate inputs altogether and tilting it toward homogenous inputs. They also develop a multi-industry general equilibrium model and find substantial gains from reducing court congestion.

The other side of the literature has focused on the impact of aggregate-level shocks on firm outcomes in India. These include large-scale infrastructure projects which improve market access along with general equilibrium effects and weather-related shocks which hamper incomes and demand in the rural economy. Shamdasani (2021) studies the impact of a national road-building program that improved rural infrastructure and access to markets. This allowed for the use of better inputs and practices, and an increase in labor hiring leading to an overall rise in rural incomes. Santangelo (2019) documents how weather shocks can be transmitted to firms in the rural economy through lower agricultural productivity, incomes and demand. This negatively impacts firm output and employment.

Earlier work has also focused on various policy reforms that have been implemented in India over the last few decades. Goldberg et al. (2010a) are the first authors to use product-level data for India. They find that falling input tariffs post the 1990s trade reform period account for more than 30 percent of new product introductions. On the other hand, Goldberg et al. (2010b) uncover little evidence for "creative destruction" during this period with no link between declines in tariffs on final goods induced by India's 1991 trade reform and product dropping. Prior to this era, Besley and Burgess (2004) show how excessive labor regulation blunted investment incentives giving India an unfavorable business climate and resulting in a movement towards informal sector enterprises. In a related paper, Aghion et al. (2008) find that the effects of dismantling the License Raj, a system of central controls regulating entry and production activity in the manufacturing sector, differed among states based on labor market regulations with industries in states with pro-market regulations growing faster than others.

In recent research, Martin, Nataraj, and Harrison (2017) study the de-reservation of certain industries which had been geared towards small establishments and find that output, employment and investment grew post the reform. Muralidharan, Niehaus, and Sukhtankar (2017) study the general equilibrium effects of a large-scale public employment program and discover a rapid growth in non-agricultural enterprise counts and employment, consistent with a role for local demand in structural transformation. Boehm, Dhingra, and Morrow (2022) utilize the trade reforms of the 1990s which brought down tariffs on intermediate inputs to show that the similarity of a firm's and an industry's input mix determines firm choices in terms of product diversification.

This paper relates more closely to the previous strands of the literature in that I try to relate the productivity aspect of firm growth with the demand component, the basic premise being that firms can either grow by becoming more productive or by adding more demand. The contributions include differentiating between the revenue and physical aspects of productivity in the context of Indian firms, using the detailed nature of the ASI data to conduct both plant-level and product-level analyses for comparison, establishing basic facts about the multi-product nature of plants in the sample and finally, studying productivity, prices and demand together in order to analyze firm dynamics in the Indian manufacturing sector.

3 Data

I began by obtaining ASI panel data from the Data Informatics and Innovation Division (DIID) in the Ministry of Statistics and Programme Implementation (MOSPI) of the Government of India. The ASI is a census of manufacturing establishments with more than 100 employees and a random sample of formally registered establishments with fewer than 100 employees. It provides comprehensive information relating to inputs, output, employment, assets, etc. of the registered factories. According to India's Factories Act of 1948, establishments with more than 20 workers (the threshold is 10 or more workers if the establishment uses electricity) are required to be formally registered.

The data corresponds to the period 1998-2018 and comes with unique factory identifiers which act as the panel variable in the analyses. Moreover, plants are required to report detailed information at the product level, both the ones used in production as inputs and the ones manufactured as output by the plant. Hence, factories report information such as quantity consumed and purchase value of input items and quantity manufactured, quantity sold, gross sale value, taxes, and ex-factory value of the various products manufactured. Therefore, the ASI is a very rich dataset but also required extensive and careful data work as I attempted to exploit both aggregate plant-level and individual plant-product-level information in my analysis. For this purpose, I created two separate datasets, one with unique plant-level observations for every year where I aggregated all variables for each plant, and the other with individual product-level observations for each plant in every year, where I apportioned the share of all the cost variables (capital, labor and input materials) to each plant-product combination using its revenue share (as done in Foster, Haltiwanger, and Syverson (2008)).

Another set of information from the dataset which is extremely vital for my analysis relates to the industry and product classifications. These are used to compare producers of homogeneous goods, that is, plants operating in the same industry and producers of the same products. This aspect also required careful attention as both of these classifications underwent multiple revisions during the period of analysis and hence, require matching so that they are comparable over time. For the purpose of this paper, I use the 4-digit National Industrial Classification (NIC-1998) codes and 7-digit National Product Classification for Manufacturing Sector (NPCMS-2011). While I could take the help of the replication package provided by Bau and Matray (2023) in dealing with the industry codes, very few researchers have used the ASI product-level information for this period and so, replication packages for ASI product-level data are not available like those for plant-level data. Thus, I had to develop concordances for these myself using documentation provided by the Central Statistical Organization (CSO) of India. For a detailed description of the process behind putting together the data as well as the cleaning of industry and product codes, refer to Appendix A.

The main ASI variables I use in the analysis are factory identifiers, state codes, the initial year of production, quantity manufactured and sold, gross sale value, labor compensation, materials consumed, fixed capital, prices (plant-product level and average plant level), industry classification, product codes, and sampling weights (these are used in all regressions). Labor compensation is obtained as the sum of wages, bonuses, fund contributions, and welfare expenses. Materials include both imports and indigenous inputs used in production. Fixed capital is the average value of fixed assets in a year net of depreciation. Prices are computed as sales divided by the quantity sold. The initial year of production is used to compute the factory age. For the plant-level analysis, I use factory identifiers as the panel variables while for the product-level exercises, I use the combination of the factory identifier and product code as a unique ID to track individual plant products over time. For the analyses that follow, I refer to this as my product-level data and the former as my plant-level data.

The production-related variables are in revenue terms and need to be deflated. As mentioned before, I obtain deflators from the Bau and Matray (2023) replication package which contains separate input, output, capital and GDP deflators. These deflators are till 2015 and I restrict my analysis to this point in time as after this, India underwent some major economic changes in the form of Demonetization in 2016 and the introduction of the Goods and Services Tax in 2017, both of which would have directly impacted the firm characteristics I study. I also restrict my analysis to manufacturing industries which involved only keeping data corresponding to 15-36 NIC codes at the 2-digit level. Furthermore, in order to implement the cost share method as in Foster, Haltiwanger, and Syverson (2008) (see also Foster, Haltiwanger, and Krizan (2001)), we require the user cost of capital. I compute this using the real interest rates for India made available by the World Bank and assuming the depreciation rate as in De Loecker, Eeckhout, and Unger (2020). This results in the final data set which is used for all analyses.

Table 1: Sampling Restrictions (Product-level Data)

		Observations	Firms	Age	Size
0.	Full sample	1,626,096	332,038	14	52
1.	Excl. industries other than manufacturing	1,543,620	308,943	14	55
2.	Excl. years post 2015	981,081	173,850	14	59
3.	Excl. firms with missing production variables	973,984	172,451	14	59
4.	Excl. firms with non-continuous data	246,417	47,459	14	130

Notes: The table presents sampling restrictions applied to the product-level data from ASI. Age and Size refer to the median age and number of employees of the sample respectively.

Table 1 presents some basic facts about the data. The main dataset covering 18 years and 973,984 observations includes 172,451 unique manufacturing plants with a median age of 14 and a median size of 59 employees. These are divided into 125 unique industries and incorporate 6223 unique products.

4 Empirical Strategy

4.1 Production Function and Productivity

I use the production function employed by Bau and Matray (2023) which is of the following Cobb-Douglas form:

$$Y_{ijt} = A_{ijt} K_{ijt}^{\alpha_j^k} L_{ijt}^{\alpha_j^l} M_{ijt}^{\alpha_j^m} \tag{1}$$

where i denotes producer (plant or individual product), j denotes 4-digit industry or 5-digit product code and t denotes the year. Y, K, L, and M are deflated measures of sales, fixed capital, wage bill, and materials respectively.

Since the production function is in revenue terms, A is the producer-specific revenue productivity (TFPR) and our production function estimates are also in revenue terms. It is important to distinguish this from the physical measure of productivity (TFPQ) which can be extracted from TFPR by removing the revenue component, giving us the true measure of productive efficiency. Thus, while these TFP measures are positively correlated, TFPR confounds factor price effects and idiosyncratic demand components with efficiency differences. This is an important distinction that was made by Foster, Haltiwanger, and Syverson (2008) and has been extensively utilized ever since as earlier empirical work only focused on TFPR. This can be misleading as producers can have high TFPR not only because they are efficient but also if they observe high producer-specific demand.

For estimating the output elasticities with respect to each input ($\alpha_j' s$), I employ the simple cost shares approach from Foster, Haltiwanger, and Syverson (2008). This involves calculating the median cost share of each input at the 4-digit industry level, that is -

$$\alpha_j = median\{\frac{Inputcost}{rK + L + M}\} \tag{2}$$

Labor and material cost shares are calculated using the reported expenditures in the ASI data while the capital cost is calculated as the rental cost of capital and its share is computed as $\alpha_j^k = 1 - \alpha_j^l - \alpha_j^m$. Hence, this method requires some strong assumptions in that all inputs need to be variable and plants face constant returns to scale.

More advanced ways of production function estimation exist, such as the control function approach developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003), further enhanced through the Generalized Method of Moments (GMM) approach by Wooldridge (2009) and Ackerberg, Caves, and Frazer (2015). These proxy methods rely on a variable input that is observable and contains information on firms' productivity such that it can be used as a proxy. However, this requires a one-to-one mapping between the two as the function needs to be invertible. This mapping breaks down in the presence of other unobservable producer-level factors that might also drive changes in the observable proxy, such as demand shocks, which is exactly what has been shown by Foster, Haltiwanger, and Syverson (2008) and again here.

Once we have the parameters of the production function, we can obtain our measures of revenue productivity and physical productivity as follows:

$$TFPR_{ijt} = A_{ijt} = \frac{Y_{ijt}}{K_{ijt}^{\alpha_j^k} L_{ijt}^{\alpha_j^l} M_{ijt}^{\alpha_j^m}} \quad ; \quad TFPQ = \frac{TFPR_{ijt}}{P_{ijt}} \tag{3}$$

4.2 Idiosyncratic Demand

Next, we want to identify other idiosyncratic factors that contribute to firm survival and growth. This could include demand idiosyncrasies across local markets (some markets might be facing higher demand which would allow the local firms to set higher prices) or buyer-seller relationships. It is important to identify this component as high-demand producers could be more apt to survive even if they are less physically efficient than low-demand groups. I use this producer-specific demand measure, along with productivity and prices, extensively in the analysis to study such firm dynamics in India.

To get to such shocks, I follow Foster, Haltiwanger, and Syverson (2008) in estimating the following demand system at the 4-digit industry level and 3-digit product level:

$$ln(q_{it}) = \beta_0 + \beta_1 ln(p_{it}) + \beta_2 ln(Income_{mt}) + \sum_t \beta_t Year_t + \eta_{it} \tag{4}$$

where i, t and m denote producer (plant or individual products), year and market respectively, q is physical output, p is price, and η is a disturbance term that contains the demand shocks. We also need to control for a set of demand shifters, including a set of year dummies and the average plant income in the local market (identified by state in our case). This is a standard equation derived in the IO literature (for a detailed discussion on demand estimation, see Berry and Haile (2021)).

Clearly, Ordinary Least Squares (OLS) estimation could lead to (positively) biased estimates of price elasticity (α_1) as p_{it} is likely to be positively correlated with elements in the error term (η_{it}), such as demand shocks that are likely to impact producer price and output decisions directly. Thus OLS is likely to suffer from issues like Omitted Variable Bias (OVB) and simultaneity bias. A popular solution is to use an instrumental variable, with possible candidates ranging from cost-side factors to characteristics of competitors and variables from other markets. Luckily, we are in possession of an ideal instrument to resolve this issue, which is physical productivity (TFPQ). A producer's idiosyncratic technology is likely to impact its price decisions (inverse relationship) but is unlikely to be affected by any short-run demand shocks. Hence, I instrument price with TFPQ and the residual from this estimation, along with the estimated contribution of local income added back in, is used as the producer-specific demand measure in the analyses that follow.

This idiosyncratic demand component is particularly relevant in the context of India which constitutes a large geographical expanse covering heterogeneous populations and markets. Thus, local market idiosyncracies are likely to play a crucial role in how firms operate. Moreover, ethnic biases and bureaucratic hurdles might result in buyer-seller relationships and connections with political authorities hindering efficient market dynamics. Another way of thinking about this measure is that it is the output variation across plants due to shifts in the demand curve, rather than movements along the demand curve.

5 Estimation

5.1 Properties of the Data

In this section, I give a broad description of our variables of interest. In Appendix B.1, I present detailed summary statistics for the estimated output elasticities. The results are largely consistent with previous findings (see e.g., De Loecker, Goldberg, et al. (2016)). Table 2 shows correlations between our core variables (in logs) at the plant level after removing industry-year fixed effects, so that cross-industry heterogeneity does not drive our results. Table 3 replicates the same exercise at the product level after removing product-year fixed effects. Tables 4 and 5 present the distribution of the same variables in both samples.

Table 2: Plant-level Correlations

	Revenue Output	Physical Output	Price	Revenue Productivity	Physical Productivity
Revenue Output	1				
Physical Output	0.592	1			
Price	0.106	-0.694	1		
Revenue Productivity	0.323	0.191	0.0229	1	
Physical Productivity	-0.0452	0.721	-0.981	0.158	1

Notes: This table displays correlations between plant-level output, price and productivity measures (N = 560,845 plant-year observations). The variables are in logs, winsorized at the 1st and the 99th percentile and exclude industry-year fixed effects.

Table 3: Product-level Correlations

	Revenue Output	Physical Output	Price	Revenue Productivity	Physical Productivity
Revenue Output	1				
Physical Output	0.689	1			
Price	0.141	-0.557	1		
Revenue Productivity	0.214	0.133	0.0272	1	
Physical Productivity	-0.0955	0.575	-0.977	0.172	1

Notes: This table displays correlations for product-level output, price and productivity measures (N = 973,984 plant-product-year observation). The variables are in logs, winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

The first observation that jumps out is that the correlations and distributions for all variables are quite similar between both samples of data. All the correlations are in the same direction and the variables are highly dispersed. The latter finding is in contrast to Foster, Haltiwanger, and Syverson (2008), with the standard deviations of our variables being more than double than that of the US data. This is expected as we have a much larger dataset (they only consider a few products) and it has been shown in previous work, such as Hsieh and Klenow (2014), how the manufacturing sector in India has huge dispersion in terms of the size of existing plants. Consequently, our output and productivity measures are also much less correlated than the US. However, there are some correlations where the magnitude is larger as well, particularly, the ones relating to prices. This could be an indication of high price sensitivity in the Indian markets.

Table 4: Plant-level Summary Statistics

	Obs.	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Output	560715	2.038	-4.918	-2.620	-1.355	0.024	1.389	2.619	4.546
Physical Output	542752	3.041	-7.569	-4.032	-1.880	0.104	1.992	3.771	6.878
Price	542673	2.485	-5.917	-3.040	-1.470	-0.029	1.257	3.383	6.209
Revenue Productivity	559836	0.396	-1.748	-0.313	-0.101	0.027	0.163	0.348	0.974
Physical Productivity	541817	2.523	-6.298	-3.442	-1.297	0.050	1.513	3.080	5.973

Notes: This table displays summary statistics for plant-level output, price and productivity measures. The variables are in logs, winsorized at the 1st and the 99th percentile and exclude industry-year fixed effects. It can be seen that there is a large dispersion in the variables (except in TFPR).

Table 5: Product-level Summary Statistics

	Obs.	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Output	972642	2.285	-5.955	-3.003	-1.405	0.114	1.532	2.846	4.902
Physical Output	927095	2.874	-7.409	-3.664	-1.727	0.094	1.779	3.517	6.866
Price	926225	2.137	-5.824	-2.626	-0.874	0.036	0.884	2.479	6.027
Revenue Productivity	971634	0.380	-1.544	-0.354	-0.132	0.018	0.177	0.378	0.949
Physical Productivity	925250	2.176	-6.051	-2.545	-0.938	-0.031	0.922	2.673	5.893

Notes: This table displays summary statistics for product-level output, price and productivity measures. The variables are in logs, winsorized at the 1st and the 99th percentile and exclude product-year fixed effects. It can be seen that the numbers are similar to the plant-level data.

The inverse correlation between physical productivity (TFPQ) and price is by construction and in line with more efficient plants having lower marginal costs and, in turn, charging lower prices. Other results, such as the positive correlation between revenue productivity (TFPR) and sales (Revenue Output), as well as TFPQ and quantity manufactured (Physical Output), are along expectations and reassuring for our analysis since these variables are directly related (being more productive allows the plants to produce and sell more). Similarly, the dispersions give confidence as well as they are in line with Hsieh and Klenow (2009) who cross-sectionally show a much higher dispersion in productivity measures for India in the 1990s, compared to the US, and lower dispersion in TFPR relative to TFPQ. This happens because there is also substantial price dispersion across producers in the same industry and the strong inverse correlation between prices and TFPQ results in a less dispersed TFPR.

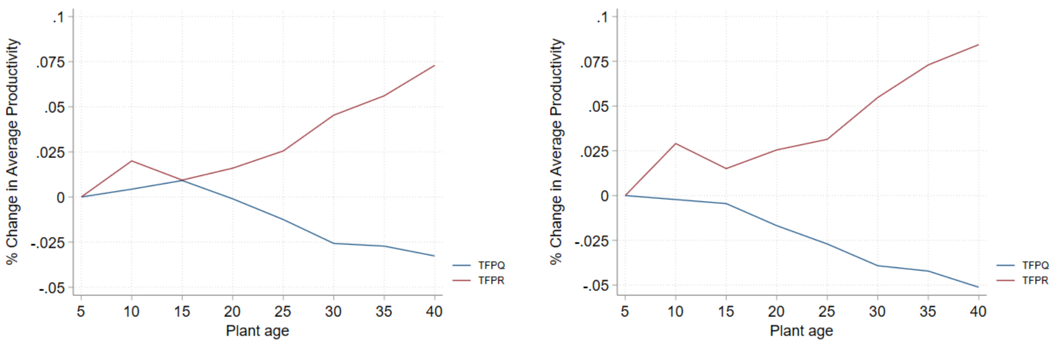
There are two instances where the results differ between the Indian and US data. These are the TFPR-price correlation going from positive to negative and vice versa for the TFPQ-sales correlations. The former is consistent with theory as the variables are directly proportional and thus, lends further credibility to our strategy. The latter seems puzzling at first, however, notice again the highly negative correlation between price and TFPQ. Hence, being more efficient does not necessarily result in higher sales, perhaps because this combination of productive efficiency and prices are not enough to attract more demand and there are certain idiosyncrasies at play.

I explore these dynamics surrounding productivity and prices further in subsequent sections, also incorporating the demand component to draw a more holistic picture of the Indian manufacturing sector. In all the empirical exercises that follow, I present results both at the plant and product levels to highlight the importance of differentiating the analysis as such, something which has not always been achieved in previous literature. In all regression, I employ sampling weights and control for product, industry and year-fixed effects in order to identify true variations between plants and products.

5.2 Productivity over the life cycle

Next, I replicate an exercise from Hsieh and Klenow (2014) who study TFPQ over the life cycle of plants, except that I examine both TFP measures and for both plant-level and product-level samples. Figure 1 plots the average productivity over the life cycle of plants in India using the data at the plant level while Figure **??** does the same for the product level.

Figure 1: Productivity over Plant Age



(a) Plant-level TFP

(b) Product-level TFP

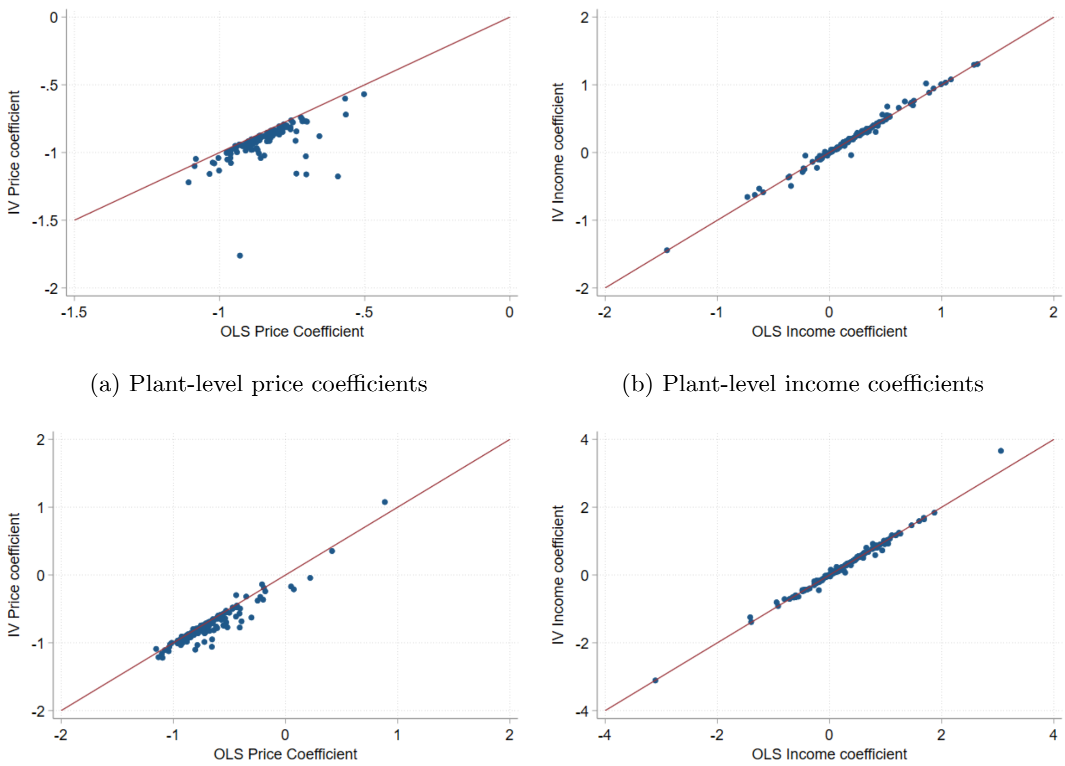
Notes: These figures plot the percentage change in the average productivity (relative to the youngest age group) for each age group in the plant- and product-level samples. The red line corresponds to TFPR while the blue line represents TFPQ. Both variables are in logs and winsorized at the 1st and the 99th percentile.

In general, the earlier finding that manufacturing plants in India see very little growth in productivity over their life is reinstated. At the plant level, I find that although TFPR rises almost throughout, it only sees approximately 7 percent growth by age 40 compared to infants (less than 5 years old). On the other hand, TFPQ sees some growth in the first decade of a plant's life but then falls, although by very little. This finding is reaffirmed at the product level with a strikingly similar graph, except that now TFPQ does not see any growth at all and is, in fact, 5 percent lower by age 40 compared to young plants (up to 5 years old). Although the variables differ from Hsieh and Klenow (2014) in terms of computation, they exhibit a similar pattern in productivity over the life cycle of Indian plants. Of course, this analysis hides industry and year-specific effects and I return to this point in a later section on the evolution of these variables.

5.3 Demand Estimation

Figure 2 presents the estimates obtained from the demand system with IV and OLS techniques using the plant and product level data. Firstly, both sets of coefficients are quite similar for the two datasets, with the price elasticities being around -1 and the income estimates around 0. Secondly, almost all the price elasticities are negative and many exceed one in absolute value, especially for the product-level estimates. These estimates are reassuring since price-setting producers should be operating in the elastic portion of their demand curves. Additionally reassuring for our demand estimation strategy is that using the plant's physical productivity as an instrument for price delivers lower (more elastic) estimates than OLS which is what we would expect due to the positive bias present in them, as discussed before, as well as the ability of TFPQ to instrument for endogenous prices.

Figure 2: Demand Estimates



(a) Plant-level price coefficients

(b) Plant-level income coefficients

(c) Product-level price coefficients

(d) Product-level income coefficients

Notes: These figures plot the coefficients from the demand system estimated using TFPQ as IV along with OLS for comparison and a 45-degree line (in red). The plant-level estimates are by 4-digit NIC codes while the product-level estimates are by 3-digit NPCMS codes. All regressions include year-fixed effects and sampling weights. Standard errors, clustered by producer, are in parentheses.

The residuals from the IV estimation, along with the estimated contribution of local income added back in, are used to compute the producer-specific measure of demand. This measure is virtually uncorrelated with all of our measures of interest, that is, TFPR, TFPQ and Prices (Plant-level: 0.06, -0.08, 0.07; Product-level: -0.02, -0.04, 0.04 respectively). This is again reassuring for our estimation strategy as this demand measure captures the variation in output after taking into account the variations in productivity and movements along the demand curve. For reference, Foster, Haltiwanger, and Syverson (2008) get price elasticities between -5.93 and -0.52 and income coefficients between -0.23 and 0.76 from the IV estimation for US data. Their coefficients on local income are also similar between the two techniques of estimation. Finally, their producer-specific demand is also uncorrelated with TFPQ (0.01) but positively correlated with TFPR (0.28) and prices (0.34). For a detailed look at our demand measure, refer to Appendix B.2.

6 Firm Dynamics

6.1 Persistence

Earlier work, including Foster, Haltiwanger, and Syverson (2008), has found that conditional on survival, there is substantial persistence in productivity and prices. I attempt to see if this finding is reinstated with Indian data and how it differs between the plant and the product level data, through simple autoregressive regressions of each measure on its own lag. The results are reported in Table 6. Standard errors are clustered by producer and I control for industry-year and product-year fixed effects respectively.

I find that plant-level productivity and prices are persistent to a degree of only about 0.55-0.6 however, the rate increases to 0.6-0.7 for the product-level sample. In comparison, Foster, Haltiwanger, and Syverson (2008) find the magnitude of persistence to be around 0.75-0.8 in their data. Hence, the productivity of Indian manufacturing plants is slightly more volatile than that of their counterparts in the US and moreover, this differs going from the aggregate plant-level data to its individual products.

Table 6: Persistence of Productivity, Prices and Demand Shock

Variables	Plant-level	Product-level
TFPR	0.559***	0.601***
	(0.00565)	(0.00450)
TFPQ	0.609***	0.708***
	(0.00343)	(0.00305)
Prices	0.611***	0.716***
	(0.00346)	(0.00309)
Demand shock	0.880***	0.868***
	(0.00247)	(0.00217)

Notes: This table reports the results of regressing TFP, price and idiosyncratic demand on its own lag (shown by row) at the plant and product level (shown by column). Thus, the coefficients depict the persistence of these measures. The variables are in logs (except Demand shock) and standard errors, clustered at the producer level, are shown in parentheses. I control for industry-year and product-year fixed effects respectively, and also employ sampling weights. *** p<0.01, ** p<0.05, * p<0.1

While this is already an interesting result, it is perhaps one that is expected given the low productivity growth observed in Indian plants. Another intriguing aspect is that the producer-specific demand is highly persistent and this is in fact in line with the US where the value is found to be as high as 0.9. This is a crucial finding as it hints at the importance of this idiosyncratic demand channel in the growth of Indian manufacturing plants and market selection. A question that arises from this graded analysis between plants and their individual products is: In what way does being a multi-product plant impact productivity? Are they better off than single-product plants? This is what I explore next.

The ASI is a large dataset representing the complete set of manufacturing establishments in India. To further understand how our variables differ between different types of producers, I use the product-level data to categorize plants into single-product plants (Category 1) and multi-product plants. The latter are further categorized into Category 2 (plants producing 2-3 products), Category 3 (plants producing 4-10 products) and the rest are plants producing more than 10 products (Category 4). For a detailed look at how our variables of interest differ between these categories, refer to Appendix C.

From the summary statistics of each category, it can be seen that products belonging to Category 4 plants are the most productive in terms of TFPR on average, however, this is not driven by productive efficiency or producer-specific demand, but by higher prices. Thus, products in larger plants enjoy the highest prices on average but seem to lose out on individual efficiency. This is the opposite for single-product plants which are efficient and are able to exploit the highest idiosyncratic demand but rank the lowest in terms of prices and revenue productivity.

Table 7: Persistence by Producer Category

Variables	Category 1 (1 product)	Category 2 (2-3 products)	Category 3 (4-10 products)	Category 4 (>10 products)
TFPR	0.663***	0.558***	0.532***	0.837***
	(0.00737)	(0.00729)	(0.00900)	(0.0240)
TFPQ	0.698***	0.703***	0.707***	0.636***
	(0.00607)	(0.00488)	(0.00503)	(0.0577)
Prices	0.705***	0.714***	0.715***	0.639***
	(0.00615)	(0.00492)	(0.00514)	(0.0600)
Demand shock	0.899***	0.856***	0.834***	0.966***
	(0.00368)	(0.00367)	(0.00407)	(0.0193)

Notes: This table reports the results of regressing TFP, price and idiosyncratic demand on its own lag (shown by row) for each category of producers (shown by column) using the product-level data. Thus, the coefficients depict the persistence of these measures for each category of producers. The variables are in logs (except Demand shock) and standard errors, clustered at the producer level, are shown in parentheses. I control for product-year fixed effects and also employ sampling weights. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Next, I rerun the persistence exercise for each category of producer separately and the results from the regressions are presented in Table 7. The first thing that stands out is that all variables are quite persistent with respect to single-product plants compared to other categories. Hence, even if on average they are less productive and observe lower prices, these measures are not volatile. More importantly, the rate of persistence in revenue productivity is significantly higher for products belonging to Category 4 plants and this does not seem to be driven by productive efficiency or prices but by an extremely high persistence in the idiosyncratic demand measure (almost equal to 1). This finding again highlights the role of this demand component in the productivity of Indian manufacturing plants. Additionally, more volatile productive efficiency in higher category plants complements the finding from the summary statistics to indicate a tradeoff between expanding the pool of products a unit manufactures and maintaining efficiency in individual products.

6.2 Evolution

In an earlier section, we saw the patterns in productivity over age for plants. Now, I refine that analysis and study the dynamics of our variables of interest for young producers compared to those of more mature plants. Young plants are categorized as being less than five years old, medium as being between 5-15 years of age and the rest of the incumbents are classified as old. I also include a dummy for exiting plants and estimate the following specification:

$$x_{it} = \beta_0 + \beta_1 Young_{it} + \beta_2 Medium_{it} + \beta_3 Exit_{it} + \sum_{IT} \beta_{IT} IndYear_{IT} + \mu_{it} \tag{5}$$

where i and t denote plant and year respectively, x is the measure (TFP, price, or demand) and we also include Industry-Year fixed effects (IndYear). The coefficients on the indicator variables measure the percentage difference between the productivity and price of the respective age category and the excluded group (old incumbents). Since the demand measure is not in logs, the coefficients there simply reflect the average difference between the age groups. For reference, I provide detailed summary statistics of our variables of interest by age categories for both samples in Appendix D.

Table 8: Evolution of Productivity, Prices and Demand Shock (plant level)

Variables	TFPR	TFPQ	Prices	Demand shock
Young (<5 years)	-0.0321***	-0.125***	0.0883***	0.0944***
	(0.00282)	(0.0184)	(0.0181)	(0.0161)
Medium (5-20 years)	-0.0178***	-0.0830***	0.0597***	0.120***
	(0.00243)	(0.0165)	(0.0163)	(0.0146)
Exit	-0.0221***	-0.0188	-0.00683	-0.453***
	(0.00263)	(0.0174)	(0.0171)	(0.0147)
Constant	0.635***	-6.360***	6.984***	5.515***
	(0.00202)	(0.0138)	(0.0136)	(0.0129)

Notes: This table reports the results of regressing plant-level TFP, price and idiosyncratic demand (shown by column) on indicators for young, medium and exiting establishments (shown by row). The excluded group includes continuing plants older than 15 years. The variables are in logs and standard errors, clustered at the plant level, are shown in parentheses. I control for industry-year fixed effects and also employ sampling weights. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The results of this estimation are reported in Tables 8 and 9. The first major finding is that early-stage plants are not more productive than old incumbents on average, irrespective of the level of analysis (plant-level or product-level) and the measure of productivity (revenue or physical). In fact, in terms of revenue productivity, they are 3-4% less productive. Additionally, prices seem to fall at the plant level which might seem puzzling. However, this is only an average price measure for the plant while the price at the individual product level is more accurate. Here, younger plants observe around 4% lower prices relative to the old incumbents. These results are in contrast to Foster, Haltiwanger, and Syverson (2008) who find that young firms have higher TFP than incumbents in the US but no significant difference in terms of prices. Our results could be due to potentially offsetting factors such as learning-by-doing, startup costs, and misallocation (Hsieh and Klenow (2009)), which would keep young manufacturing establishments in India from immediately reaching their production frontier.

Table 9: Evolution of Productivity, Prices and Demand Shock (product level)

Variables	TFPR	TFPQ	Prices	Demand shock
Young (<5 years)	-0.0438***	-0.00229	-0.0436***	0.196***
	(0.00193)	(0.0109)	(0.0107)	(0.0115)
Medium (5-20 years)	-0.0236***	0.0133	-0.0385***	0.174***
	(0.00156)	(0.00936)	(0.00919)	(0.00999)
Exit	-0.0326***	0.0593***	-0.0921***	-0.350***
	(0.00183)	(0.0104)	(0.0102)	(0.0109)
Constant	0.607***	-6.777***	7.372***	4.049***
	(0.00124)	(0.00743)	(0.00729)	(0.00820)

Notes: This table reports the results of regressing product-level TFP, price and idiosyncratic demand (shown by column) on indicators for young, medium and exiting establishments (shown by row). The excluded group includes continuing plants older than 15 years. The variables are in logs and standard errors, clustered at the plant-product level, are shown in parentheses. I control for product-year fixed effects and also employ sampling weights. *** p<0.01, ** p<0.05, * p<0.1

The question naturally arises: how do these less productive plants survive while also observing lower prices? The answer, again, seems to lie in the producer-specific demand measure. It can be seen that these younger producers face higher idiosyncratic demand than older plants and the result holds both at the plant level and individual product level. This is also in contrast to the US where the demand measure is found to be lower for each category relative to old incumbents. Perhaps the plants in our data are able to exploit new markets and compete on prices or take advantage of certain local connections and relationships with buyers which allows them to survive in the industry. Regardless, the finding asserts the importance of this producer-specific demand measure for firm dynamics in the Indian manufacturing industry, as has also been highlighted in previous sections. Finally, exiting plants are among the least productive out of all categories, in terms of both TFPR and TFPQ, and also face the lowest idiosyncratic demand, which probably explains their non-survival and is reassuring for the validity of this analysis.

7 Robustness

7.1 Continuous Data

A peculiar feature of the ASI is that it is a census of manufacturing establishments with more than 100 employees and a random sample of formally registered establishments with fewer than 100 employees. As a result, there are many plants that feature in the data in a non-regular fashion which makes their information discontinuous and might also raise concerns about data quality and validity of results. To check this, I restrict the dataset to only those firms which include a continuous stream of data from the first year of their appearance till the last and rerun the main exercises.

Clearly, all the estimates seem to be robust. Appendix E shows similar patterns in productivity over the life cycle of plants as observed before. Table 11 displays the same heterogeneity between producer categories that was observed in the main sample with smaller plants displaying more volatile TFPR and the results being driven by the idiosyncratic demand component. Finally, the results from the persistence (10 and evolution (12) of all variables are also along previous lines, although the plant-level evolution results are less precisely estimated now due to the lower number of observations after the restriction imposed on the dataset. More importantly, the overall finding concerning the role of the producer-specific demand measure in firm dynamics stays intact.

Table 10: Persistence of Productivity, Prices and Demand Shocks (Continuous Data)

Variables	Plant-level	Product-level
TFPR	0.558***	0.617***
	(0.00756)	(0.00565)
TFPQ	0.608***	0.708***
	(0.00459)	(0.00402)
Prices	0.610***	0.715***
	(0.00467)	(0.00408)
Demand shock	0.877***	0.876***
	(0.00317)	(0.00273)

Notes: This table reports the results of regressing TFP, price and idiosyncratic demand on its own lag (shown by row) for each producer level (shown by column) using only continuous data. Hence, the coefficients depict the persistence of these measures. The variables are in logs (except Demand shock) and standard errors, clustered at the producer level, are shown in parentheses. I control for industry-year and product-year fixed effects respectively and also employ sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Persistence by Producer Category

Variables	Category 1 (1 product)	Category 2 (2-3 products)	Category 3 (4-10 products)	Category 4 (>10 products)
TFPR	0.672***	0.566***	0.556***	0.860***
	(0.00893)	(0.00953)	(0.0121)	(0.0261)
TFPQ	0.703***	0.701***	0.702***	0.645***
	(0.00780)	(0.00669)	(0.00706)	(0.0668)
Prices	0.708***	0.711***	0.711***	0.643***
	(0.00790)	(0.00676)	(0.00734)	(0.0697)
Demand shock	0.904***	0.865***	0.842***	0.985***
	(0.00488)	(0.00438)	(0.00556)	(0.0202)

Notes: This table reports the results of regressing TFP, price and idiosyncratic demand on its own lag (shown by row) for each category of producers (shown by column) using only continuous data. Thus, the coefficients depict the persistence of these measures for each category. The variables are in logs (except Demand shock) and standard errors, clustered at the producer level, are shown in parentheses. I control for product-year fixed effects and also employ sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Table 12: Evolution of Productivity, Prices and Demand Shocks (plant level)

Variables	TFPR	TFPQ	Prices	Demand shock
Young (<5 years)	-0.0313***	-0.0113	-0.0231	-0.574***
	(0.00559)	(0.0349)	(0.0342)	(0.0352)
Medium (5-20 years)	-0.00141	0.0508	-0.0569*	-0.243***
	(0.00511)	(0.0326)	(0.0319)	(0.0338)
Exit	-0.0236***	0.167***	-0.193***	-1.392***
	(0.00686)	(0.0424)	(0.0415)	(0.0384)
Constant	0.674***	-6.401***	7.070***	6.907***
	(0.00434)	(0.0272)	(0.0266)	(0.0296)

Notes: This table reports the results of regressing plant-level TFP, price and idiosyncratic demand (shown by column) on indicators for young, medium and exiting plants (shown by row) using only continuous data. The excluded group includes incumbents older than 15 years. The variables are in logs (except Demand shock) and standard errors, clustered at the plant level, are shown in parentheses. I control for industry-year fixed effects and employ sampling weights. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 13: Evolution of Productivity, Prices and Demand Shocks (product level)

Variables	TFPR	TFPQ	Prices	Demand shock
Young (<5 years)	-0.0438***	0.0187	-0.0640***	0.111***
	(0.00420)	(0.0237)	(0.0232)	(0.0263)
Medium (5-20 years)	-0.0209***	0.0803***	-0.104***	0.126***
	(0.00345)	(0.0207)	(0.0204)	(0.0234)
Exit	-0.00941*	0.139***	-0.142***	-0.620***
	(0.00564)	(0.0293)	(0.0286)	(0.0320)
Constant	0.648***	-6.796***	7.439***	4.204***
	(0.00270)	(0.0162)	(0.0159)	(0.0190)

Notes: This table reports the results of regressing product-level TFP, price and idiosyncratic demand (shown by column) on indicators for young, medium and exiting plants (shown by row) using only continuous data. The excluded group includes incumbents older than 15 years. The variables are in logs (except Demand shock) and standard errors, clustered at the plant-product level, are shown in parentheses. I control for product-year fixed effects and also employ sampling weights. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

7.2 Alternate Estimation Strategy

Until now, for estimating the production function using product-level data, we assumed that the inputs used in the production of various products are proportional to their share in total plant sales. As pointed out by De Loecker, Goldberg, et al. (2016), the allocation of inputs in multi-product firms is unobserved and can lead to bias in the estimates of the output elasticities. They propose an alternate strategy to address this bias which relaxes the above assumption by exploiting the data on single-product plants only to obtain estimates of the production function. This approach assumes that the physical relationship between inputs and outputs is the same for single- and multi-product firms that manufacture the same product. While this assumption may appear strong, it is already implicitly employed in all previous work that pools data across single- and multi-product firms (e.g., Olley and Pakes (1996) or Levinsohn and Petrin (2003)).

Table 14: Persistence of Productivity, Prices and Demand Shock (Alternate Estimation)

Variables	TFPR	TFPQ	Demand shock
Lag TFPR,	0.600***		
	(0.00450)		
Lag TFPQ,		0.708***	
		(0.00305)	
Lag Demand shock,			0.868***
			(0.00217)

Notes: This table reports the results of regressing TFP and idiosyncratic demand (shown by column) on its own lag (shown by row) using the production function estimates obtained from single-product plants. Hence, the coefficients on the lag depict the persistence in these measures. The TFP measures are in logs and standard errors, clustered at the plant-product level, are shown in parentheses. I control for product-year fixed effects and also employ sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Another concern that arises from the approach of using single-product plants to estimate the production function for all plants is that the estimates may suffer from a selection bias. The selection bias arises if plants' choice to add a second product and become multi-product depends on its unobserved productivity and/or input use. This potential selection bias is addressed by our use of an unbalanced panel that consists of plants that are single-product at a given point in time. At time t, the unbalanced panel includes both plants that always remain single-product producers and those that add products later. This helps address the selection concern arising from the nonrandom event that a plant becomes a multi-product producer based on unobserved productivity.

Table 15: Persistence by Producer Category (Alternate Estimation)

Variables	Category 1 (1 product)	Category 2 (2-3 products)	Category 3 (4-10 products)	Category 4 (>10 products)
TFPR	0.663***	0.557***	0.527***	0.838***
	(0.00733)	(0.00731)	(0.00909)	(0.0240)
TFPQ	0.698***	0.703***	0.707***	0.636***
	(0.00607)	(0.00488)	(0.00503)	(0.0577)
Demand shock	0.899***	0.856***	0.834***	0.966***
	(0.00368)	(0.00366)	(0.00406)	(0.0193)

Notes: This table reports the results of regressing TFP, price and idiosyncratic demand on its own lag (shown by row) for each category of producers (shown by column) using the production function estimates obtained from single-product plants. Thus, the coefficients depict the persistence of these measures for each category of producers. The variables are in logs (except Demand shock) and standard errors, clustered at the producer level, are shown in parentheses. I control for product-year fixed effects and also employ sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Table 16: Evolution of Productivity, Prices and Demand Shock (Alternate Estimation)

Variables	TFPR	TFPQ	Demand shock
Young (<5 years)	-0.0449***	-0.00411	0.198***
	(0.00195)	(0.0110)	(0.0115)
Medium (5-20 years)	-0.0249***	0.0112	0.176***
	(0.00157)	(0.00942)	(0.0100)
Exit	-0.0326***	0.0579***	-0.346***
	(0.00185)	(0.0105)	(0.0110)
Constant	0.617***	-6.766***	4.060***
	(0.00126)	(0.00751)	(0.00826)

Notes: This table reports the results of regressing product-level TFP and idiosyncratic demand (shown by column) on indicators for young, medium and exiting establishments (shown by row) using the production function estimates obtained from single-product plants. The excluded group includes continuing plants older than 15 years. The variables are in logs (except Demand shock) and standard errors, clustered at the plant-product level, are shown in parentheses. I control for product-year fixed effects and also employ sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Additionally, I also exploit the considerable length of the dataset and allow the estimates to vary over time. With 18 years of effective data (1998-2016), I divide the panel into 3 periods of 6 years each and then calculate the average cost share pertaining to each input by industry and time period. Finally, I rerun all the main exercises for these revised estimates from the product-level dataset and present the result below. Again, I find the results to be remarkably robust. The variables (Appendix F), persistence rates, both overall (Table 14) and by producer category (Table 15), and evolution estimates (16) are almost identical to the main sample (I exclude the price variable as it does not undergo any change in this alternate estimation).

7.3 Including Region Fixed Effects

The ASI data covers the whole geography of India, divided into 36 different regions (28 states and 8 union territories). All of these regions vary greatly in terms of markets, culture, topography, etc. Hence, it is possible that the previous results might be driven by region-specific effects. To check this, I also control for region-year fixed effects using the state information pertaining to each plant available in the panel dataset which was also used in the demand estimation. The results from all regression exercises are presented below and again, they seem quite robust.

Tables 17 and 18 show the estimates from the persistence regressions and for both, they resemble the earlier results with the overall rate ranging from 0.55 to 0.7 and larger plants (in terms of the number of products manufactured) displaying significantly higher persistence in TFPR, driven by the producer-specific demand component. Similarly, Tables 19 and 20 depict how these measures evolve over the life of plants and reiterate that younger plants are relatively unproductive but are able to survive potentially due to higher idiosyncratic demand.

Table 17: Persistence of Productivity, Prices and Demand Shock

Variables	Plant-level	Product-level
TFPR	0.556***	0.597***
	(0.00565)	(0.00448)
TFPQ	0.605***	0.708***
	(0.00342)	((0.00303))
Prices	0.608***	0.715***
	(0.00345)	(0.00307)
Demand shock	0.876***	0.864***
	(0.00251)	(0.00222)

Notes: This table reports the results of regressing TFP, price and idiosyncratic demand on its own lag (shown by row) for each producer level (shown by column). The variables are in logs (except Demand shock) and standard errors, clustered at the producer level, are shown in parentheses. In addition to industry-year fixed effects, I also control for region-year fixed effects, while employing sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Table 18: Persistence by Producer Category

Variables	Category 1 (1 product)	Category 2 (2-3 products)	Category 3 (4-10 products)	Category 4 (>10 products)
TFPR	0.660***	0.555***	0.522***	0.870***
	(0.00718)	(0.00708)	(0.00863)	(0.0393)
TFPQ	0.697***	0.703***	0.706***	0.624***
	(0.00602)	(0.00482)	(0.00493)	(0.0596)
Prices	0.703***	0.713***	0.714***	0.623***
	(0.00611)	(0.00486)	(0.00504)	(0.0624)
Demand shock	0.895***	0.850***	0.828***	0.975***
	(0.00377)	(0.00373)	(0.00410)	(0.0228)

Notes: This table reports the results of regressing TFP, price and idiosyncratic demand on its own lag (shown by row) for each category of producers (shown by column) using the product-level data. Thus, the coefficients depict the persistence of these measures for each category of producers. The variables are in logs (except Demand shock) and standard errors, clustered at the producer level, are shown in parentheses. In addition to industry-year fixed effects, I also control for region-year fixed effects now, while employing sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Table 19: Evolution of Productivity, Prices and Demand Shock (plant level)

Variables	TFPR	TFPQ	Prices	Demand shock
Young (<5 years)	-0.0414***	-0.117***	0.0702***	0.0433***
	(0.00285)	(0.0187)	(0.0184)	(0.0162)
Medium (5-20 years)	-0.0188***	-0.0640***	0.0400**	0.0875***
	(0.00244)	(0.0166)	(0.0164)	(0.0146)
Exit	-0.0238***	0.0111	-0.0387**	-0.471***
	(0.00266)	(0.0176)	(0.0173)	(0.0148)
Constant	0.637***	-6.377***	7.004***	5.541***
	(0.00205)	(0.0139)	(0.0137)	(0.0129)

Notes: This table reports the results of regressing plant-level TFP, price and idiosyncratic demand (shown by column) on indicators for young, medium and exiting establishments (shown by row). The excluded group includes continuing plants older than 15 years. The variables are in logs (except Demand shock) and standard errors, clustered at the plant level, are shown in parentheses. In addition to industry-year fixed effects, I also control for region-year fixed effects now, while employing sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Table 20: Evolution of Productivity, Prices and Demand Shock (product level)

Variables	TFPR	TFPQ	Prices	Demand shock
Young (<5 years)	-0.0510***	-0.0227**	-0.0305***	0.154***
	(0.00194)	(0.0111)	(0.0109)	(0.0116)
Medium (5-20 years)	-0.0254***	0.00785	-0.0350***	0.145***
	(0.00156)	(0.00944)	(0.00926)	(0.0100)
Exit	-0.0363***	0.0591***	-0.0961***	-0.374***
	(0.00185)	(0.0106)	(0.0104)	(0.0110)
Constant	0.610***	-6.772***	7.369***	4.075***
	(0.00125)	(0.00754)	(0.00739)	(0.00825)

Notes: This table reports the results of regressing product-level TFP, price and idiosyncratic demand (shown by column) on indicators for young, medium and exiting establishments (shown by row). The excluded group includes continuing plants older than 15 years. The variables are in logs (except Demand shock) and standard errors, clustered at the plant-product level, are shown in parentheses. In addition to industry-year fixed effects, I also control for region-year fixed effects now, while employing sampling weights. *** p<0.01, ** p<0.05, * p<0.1

8 Discussion and Conclusion

In this paper, I present evidence for the crucial role that the producer-specific measure of demand is playing in the state of the Indian manufacturing sector, which is historically characterized by high dispersion in productivity and the persistence of small and unproductive plants. I exploited the detailed ASI panel data to derive the various components of revenue productivity, including physical efficiency, prices and an idiosyncratic demand measure. Then I conducted a detailed analysis of the role these measures play in firm dynamics with respect to the Indian data.

Firstly, comparing the results with Foster, Haltiwanger, and Syverson (2008), I find that the Indian market differs in a few aspects from the US. This includes a lower persistence in productivity and prices as well as variations in their evolution over the life cycle of plants, with younger establishments being less productive compared to older incumbents but facing higher idiosyncratic demand. Moreover, the rich ASI dataset allowed me to study various categories of producers based on the number of products that they manufacture. In this respect, I find that products belonging to larger plants observe significantly higher persistence in revenue productivity, which is driven not by productive efficiency or prices but by an even higher persistence in the demand component. These results are found to be robust along a number of metrics.

The finding that idiosyncratic producer demand plays a crucial role in firm dynamics in India is particularly interesting and opens up avenues for further research. Note that our results imply that these manufacturers have some degree of market power. Using our current dataset, we can study this by estimating markups and establishing a relationship between productivity and markups (do more productive firms have lower costs or charge higher prices?). This would be based on the research done by De Loecker, Eeckhout, and Unger (2020) relating markups and market power and might help us in understanding further why firms are unable to add more demand and grow. However, this would require the control function approach to production function estimation, and as mentioned before, that was not suitable for the current analyses.

Other avenues include studying the impact of various policy reforms, such as the dereservation of industries and labor regulations, on our variables of interest, that is, productivity and producer-specific demand. Earlier studies by Martin, Nataraj, and Harrison (2017) and Besley and Burgess (2004) respectively have examined their effect on surface-level variables such as establishment size and output so such an analysis would add to that body of work. One could evaluate the causal impact of this idiosyncratic component on firm growth by identifying an exogenous shock that changes the demand that these establishments face, such as infrastructural developments. For example, there have been initiatives to increase road and internet accessibility in India over the years and data on this is available as well. A starting point would be to use the framework provided by Foster, Haltiwanger, and Syverson (2008) for decomposing aggregate productivity growth into various components such as within-firm growth, across-firm reallocation of market share and entry-exit effects. This could be undertaken for the Indian data in order to uncover the contribution of demand reallocation in the growth of its manufacturing sector. A low contribution would add more weight to the findings of this paper. I hope to take these ideas up in future research.

Appendix A Data

The data obtained from ASI was in completely raw form (text files delimited by spaces) but came along with supporting documents such as the description of the various files and the data structure. There were separate folders for each year and each of these contained ten text files for different blocks of data (A-J). Each of these blocks corresponds to a different set of information, for example, Block A for each year contains information on factory identification while Block J included data on output manufactured by the factory. Using the data structure, I developed a loop for each block of data that ran for all years, converting the text file corresponding to the block into a STATA dataset with all required variables and saving it in a separate folder. Then all these datasets were appended to create a single dataset for each block which was cleaned, that is, keeping only the variables relevant for our analysis, and saved. Finally, these blocks were merged to create the main data set to be used for analysis.

It is important to note that the ASI contains detailed information on all establishments. Except the first two blocks which correspond to factory identification, all other blocks include several rows of data for each firm every year. For example, Block C relates to Fixed Assets and contains information on the gross value and depreciation of various types of assets like land, building, machinery, etc., block D collects data on opening and closing values of different Working Capital items like raw materials, fuels, cash in hand, etc., and Block E relates to Employment Costs with data on wages and days worked by category of staff. In such cases, I kept the total values for each variable which gave unique plant-level observations for each year.

For classifying plants into industries, I use the 5-digit codes based on the National Industrial Classification (NIC) developed by the Central Statistical Organization (CSO) of India along the lines of the United Nations International Standard Industrial Classification (ISIC). These classifications go through revisions, with the latest one coming in the form of NIC-2008 (previously NIC-2004 and NIC-1998). For cleaning these codes, I took the help of the replication package provided by Bau and Matray (2023), which also provides concordance tables, in developing a unique 4-digit NIC-98 classification. The choice of 1998 as the base year is convenient as the revisions generally result in further detailing of existing codes and so, it is easier to aggregate them back. Further, the deflators that we use also come from the same replication package and are at the 3-digit NIC-1987, along with a concordance between NIC-1998 and NIC-1987.

Cleaning the product codes was even more complicated as they also needed to be harmonized but concordance tables for them were not readily available. Very few researchers have used the ASI for product-level analysis (see, e.g., Barua (2020) and Martin, Nataraj, and Harrison (2017)) as the Prowess database is the preferred choice for that (however, Prowess focuses on large firms and not openly available) so, replication packages for ASI product-level data are not available like those for plant-level data. With regards to classifying products, the ASI relies on the National Product Classification for Manufacturing Sector (NPCMS) from 2011 onward and the Annual Survey of Industries Commodity Classification (ASICC) before that. The former was developed by the CSO as a 7-digit classification made up of a 5-digit Central Product Classification (CPC) code, which is the reference classification for all product classifications within the international system of economic classifications put in place by the United Nations, plus a 2-digit Indian requirement. The ASICC is a 5-digit classification that was in use prior to 2011 and has two main versions: ASICC-1998 and ASICC-2008-09. I aggregate all the product codes up to NPCMS in two steps. First, I map all the ASICC-1998 product codes into their corresponding ASICC-2009 counterpart to identify the products under a unique ASICC version. In the second stage, I use the concordance from ASICC-2009 to NPCMS-2011 published by CSO to develop a unique 5-digit NPCMS-11 classification.

Appendix B Estimation

B.1 Output Elasticities

Table 21: Summary Statistics for Output Elasticities (plant level)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Labor Share	0.113	0.070	0.018	0.034	0.073	0.104	0.135	0.184	0.368
Materials Share	0.825	0.079	0.588	0.742	0.787	0.830	0.872	0.930	0.965
Capital Share	0.062	0.024	0.017	0.035	0.046	0.060	0.075	0.091	0.139

Notes: This table displays summary statistics for the plant-level output elasticities with respect to each input estimated using the cost share approach ($N = 560845$).

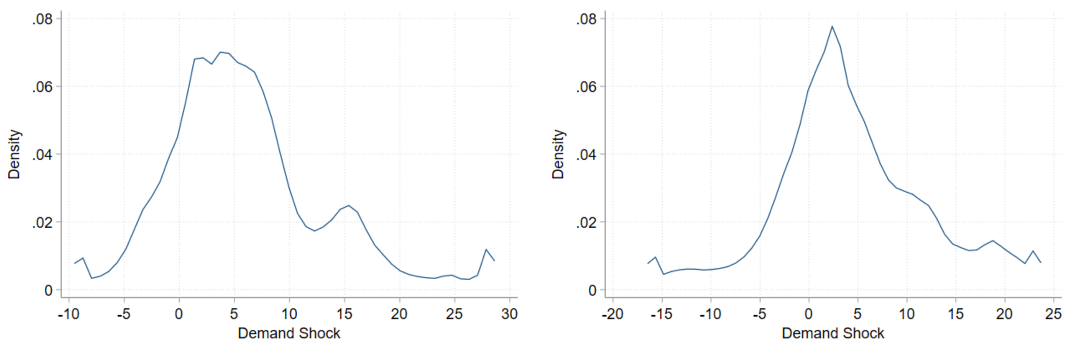
Table 22: Summary Statistics for Output Elasticities (product level)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Labor Share	0.103	0.061	0.016	0.030	0.073	0.092	0.127	0.166	0.366
Materials Share	0.837	0.073	0.589	0.754	0.802	0.833	0.874	0.949	0.968
Capital Share	0.060	0.025	0.016	0.019	0.045	0.059	0.075	0.092	0.141

Notes: This table displays summary statistics for the product-level output elasticities with respect to each input estimated using the cost share approach ($N = 973984$).

B.2 Demand Shocks

Figure 3: Distribution of Demand shocks



(a) Plant-level Demand Shock

(b) Product-level Demand Shock

Notes: These figures plot the distribution of the idiosyncratic demand measure estimated from the plant- and product-level samples. The variable is winsorized at the 1st and the 99th percentile.

Table 23: Summary Statistics for Demand shock (plant level)

	Obs.	S.D.	P1	P10	P25	P50	P75	P90	P99
Demand shock	542656	7.092	-9.145	-1.828	1.377	5.003	9.230	15.825	28.246

Notes: This table displays summary statistics for the plant-level idiosyncratic demand measurs. The variable is winsorized at the 1st and the 99th percentile.

Table 24: Summary Statistics for Demand shock (product level)

	Obs.	S.D.	P1	P10	P25	P50	P75	P90	P99
Demand shock	926222	7.800	-16.115	-4.266	-0.232	3.331	8.700	15.114	23.308

Notes: This table displays summary statistics for the product-level idiosyncratic demand measure (N = 926222). The variable is winsorized at the 1st and the 99th percentile.

Appendix C Summary Statistics by Category of Plants

Table 25: Product-level Summary Statistics (Category 1 plants)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	-0.011	0.411	-1.756	-0.375	-0.138	0.017	0.182	0.386	0.947
Physical Productivity	0.001	2.082	-5.983	-2.371	-0.898	-0.015	0.879	2.462	5.804
Price	-0.020	2.034	-5.764	-2.430	-0.841	0.010	0.826	2.252	5.927
Demand shock	0.296	2.061	-5.063	-2.181	-0.906	0.258	1.560	2.864	5.187

Notes: This table displays summary statistics for product-level TFP, price and demand measures in single-product plants (Category 1, N=349326 observations). The variables are in logs, winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

Table 26: Product-level Summary Statistics (Category 2 plants)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	0.007	0.361	-1.394	-0.340	-0.125	0.021	0.175	0.368	0.944
Physical Productivity	0.023	2.132	-5.889	-2.424	-0.896	-0.033	0.890	2.653	5.884
Price	-0.010	2.097	-5.816	-2.590	-0.826	0.048	0.848	2.385	5.877
Demand shock	-0.073	2.375	-6.253	-3.131	-1.514	0.035	1.498	2.791	5.096

Notes: This table displays summary statistics for product-level TFP, price and demand measures in plants producing 2-3 products only (Category 2, N=369514). The variables are in logs, winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

Table 27: Product-level Summary Statistics (Category 3 plants)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	0.003	0.362	-1.370	-0.350	-0.133	0.015	0.173	0.377	0.929
Physical Productivity	-0.030	2.359	-6.326	-2.964	-1.066	-0.049	1.050	3.006	6.031
Price	0.036	2.324	-5.933	-2.953	-1.018	0.048	1.024	2.933	6.281
Demand shock	-0.302	2.616	-7.000	-3.703	-1.933	-0.132	1.426	2.839	5.499

Notes: This table displays summary statistics for product-level TFP, price and demand measures in plants producing 4-10 products only (Category 3, N=250080). The variables are in logs, winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

Table 28: Product-level Summary Statistics (Category 4 plants)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	0.181	0.394	-0.485	-0.231	-0.079	0.074	0.371	0.908	1.126
Physical Productivity	-0.269	2.317	-6.293	-3.110	-1.592	-0.233	0.916	2.596	5.494
Price	0.481	2.302	-5.378	-2.370	-0.674	0.403	1.858	3.280	6.416
Demand shock	-0.288	2.834	-6.859	-3.997	-2.150	-0.237	1.630	3.169	6.797

Notes: This table displays summary statistics for product-level TFP, price and demand measures in plants producing over 10 products only (Category 4, N=3689). The variables are in logs, winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

Appendix D Summary Statistics by Age Group

D.1 Plant-level Data

Table 29: Young plants (< 5 years)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	-0.016	0.414	-1.818	-0.358	-0.120	0.015	0.159	0.353	0.985
Physical Productivity	-0.047	2.502	-6.328	-3.435	-1.288	-0.005	1.364	3.062	5.956
Price	0.031	2.464	-5.925	-3.042	-1.357	0.007	1.241	3.361	6.225
Demand shock	-0.023	2.037	-5.243	-2.515	-1.244	0.000	1.255	2.423	4.757

Notes: This table displays summary statistics for plant-level TFP, price and demand measures in young plants (less than 5 years old, N = 117526). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude industry-year fixed effects.

Table 30: Medium-aged plants (5-20 years old)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	0.005	0.381	-1.700	-0.287	-0.094	0.029	0.160	0.338	0.949
Physical Productivity	-0.034	2.515	-6.352	-3.462	-1.302	0.036	1.416	3.064	5.972
Price	0.039	2.480	-5.919	-3.013	-1.371	-0.013	1.273	3.409	6.272
Demand shock	0.147	2.133	-5.231	-2.514	-1.169	0.167	1.515	2.735	5.087

Notes: This table displays summary statistics for plant-level TFP, price and demand measures in medium-aged plants (5-20 years old, N = 213467). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude industry-year fixed effects.

Table 31: Old plants (> 20 years)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	0.015	0.395	-1.722	-0.298	-0.090	0.038	0.180	0.368	0.983
Physical Productivity	-0.005	2.584	-6.309	-3.531	-1.448	0.050	1.615	3.189	6.019
Price	0.021	2.543	-5.940	-3.126	-1.557	-0.008	1.417	3.492	6.200
Demand shock	0.159	2.337	-5.501	-2.740	-1.392	0.167	1.712	3.055	5.518

Notes: This table displays summary statistics for plant-level TFP, price and demand measures in old plants (over 20 years old, N = 136204). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude industry-year fixed effects.

Table 32: Exiting plants

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	-0.012	0.386	-1.703	-0.317	-0.106	0.017	0.142	0.319	0.935
Physical Productivity	0.067	2.456	-6.134	-3.252	-1.124	0.076	1.483	3.058	5.998
Price	-0.079	2.417	-5.951	-3.039	-1.449	-0.071	1.077	3.171	6.063
Demand shock	-0.757	1.907	-5.858	-3.056	-1.866	-0.691	0.399	1.474	3.790

Notes: This table displays summary statistics for plant-level TFP, price and demand measures in exiting plants (N = 73258). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude industry-year fixed effects.

D.2 Product-level Data

Table 33: Young plants (< 5 years)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	-0.019	0.404	-1.660	-0.404	-0.155	0.005	0.171	0.381	0.978
Physical Productivity	0.021	2.161	-6.069	-2.434	-0.919	-0.030	0.936	2.667	5.916
Price	-0.040	2.120	-5.884	-2.643	-0.903	0.022	0.836	2.349	6.052
Demand shock	0.040	2.235	-6.028	-2.784	-1.216	0.135	1.463	2.720	4.975

Notes: This table displays summary statistics for product-level TF, price and demand measures in young plants (less than 5 years old, N = 179914). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

Table 34: Medium-aged plants (5-20 years)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	0.003	0.367	-1.479	-0.337	-0.126	0.018	0.174	0.368	0.935
Physical Productivity	0.015	2.170	-6.034	-2.516	-0.921	-0.018	0.931	2.683	5.914
Price	-0.010	2.134	-5.824	-2.640	-0.885	0.028	0.875	2.460	6.036
Demand shock	0.132	2.327	-6.047	-2.827	-1.227	0.226	1.645	2.927	5.258

Notes: This table displays summary statistics for product-level TFP, price and demand measures in medium-aged plants (5-20 years old, N = 368024). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

Table 35: Old plants (> 20 years)

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	0.018	0.372	-1.474	-0.331	-0.118	0.031	0.195	0.397	0.926
Physical Productivity	-0.057	2.252	-6.179	-2.791	-1.062	-0.055	0.932	2.747	5.886
Price	0.075	2.212	-5.814	-2.674	-0.868	0.067	1.030	2.742	6.153
Demand shock	0.061	2.488	-6.352	-3.137	-1.438	0.152	1.690	3.091	5.541

Notes: This table displays summary statistics for product-level TFP, price and demand measures in old plants (over 20 years old, N = 270619). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

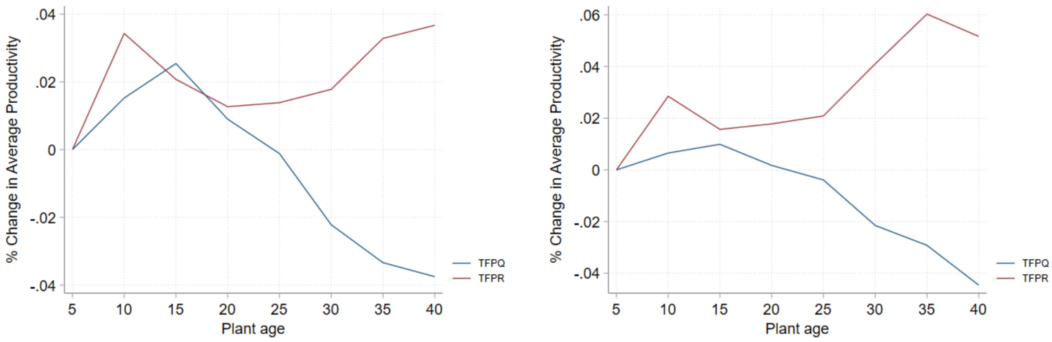
Table 36: Exiting plants

	Mean	S.D.	P1	P10	P25	P50	P75	P90	P99
Revenue Productivity	-0.018	0.376	-1.532	-0.369	-0.141	0.008	0.152	0.346	0.919
Physical Productivity	0.066	2.064	-5.803	-2.228	-0.807	-0.005	0.924	2.548	5.862
Price	-0.084	2.024	-5.828	-2.515	-0.887	0.000	0.745	2.147	5.729
Demand shock	-0.650	2.141	-6.401	-3.388	-1.880	-0.500	0.683	1.863	4.148

Notes: This table displays summary statistics for product-level TFP, price and demand measures in exiting plants (N = 118445). The variables are in logs (except Demand shock), winsorized at the 1st and the 99th percentile and exclude product-year fixed effects.

Appendix E Continuous Data

Figure 4: Productivity over Age (Continuous Data)

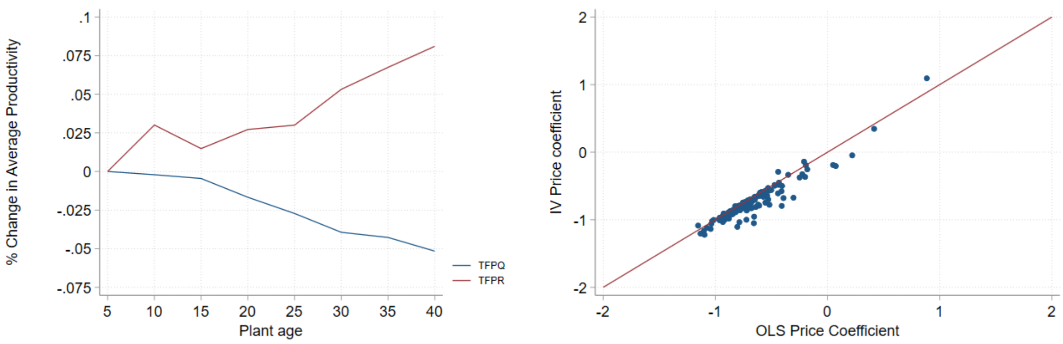


(a) Plant-level Productivity (b) Product-level Productivity

Notes: These figures plot the percentage change in the average productivity (relative to the youngest age group) for each age group in plant- and product-level samples using only continuous data. The red line corresponds to TFPR while the blue line represents TFPQ. Both variables are in logs and winsorized at the 1st and the 99th percentile.

Appendix F Alternate Estimation Strategy

Figure 5: Alternate Estimation of Productivity and Demand



(a) Productivity over Age (b) Demand Estimates

Notes: These figures plot the percentage change in the average productivity (relative to the youngest age group) over the life cycle and the estimated price elasticities using product-level data. Both TFP variables are in logs and winsorized at the 1st and the 99th percentile. In (a), the red line corresponds to TFPR while the blue line represents TFPQ. In (b), the 45-degree line is shown in red for comparison.

References

Ackerberg, D., K. Caves, and G. Frazer (2015). “Identification Properties of Recent Production Function Estimators”. In: *Econometrica* 83 (6), pp. 2411–2451.

Aghion, P. et al. (2008). “The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India”. In: *American Economic Review* 98 (4), pp. 1397–1412.

Alfaro, L. and Anusha C. (2014). “Deregulation, Misallocation, and Size: Evidence from India”. In: *Journal of Law Economics* 57 (4), pp. 897–936.

Aw, B.Y., X. Chen, and M. Roberts (2001). “Firm-level evidence on productivity differentials and turnover in Taiwanese manufacturing”. In: *Journal of Development Economics* 66 (1), pp. 51–86.

Baily, M.N., C. Hulten, and D. Campbell (1992). “Productivity Dynamics in Manufacturing Plants”. In: *Brookings Papers on Economic Activity* 23 (Microeconomics), pp. 187–267.

Bartelsman, E.J. and M. Doms (2000). “Understanding Productivity: Lessons from Longitudinal Microdata”. In: *Journal of Economic Literature* 38, pp. 569–594.

Barua, Shubhasish (2020). “Effects of Chinese Import Competition on Plant Productivity and Product Scope: Evidence from India”. In: *Social Science Research Network*.

Bau, N. and A. Matray (2023). “Misallocation and Capital Market Integration: Evidence from India”. In: *Econometrica* 91 (1), pp. 67–106.

Berry, Steven T. and Philip A. Haile (2021). “Foundations of Demand Estimation”. In: *National Bureau of Economic Research: http://www.nber.org/papers/w29305* 29305.

Besley, T. and R. Burgess (2004). “Can Labor Regulation Hinder Economic Performance? Evidence from India”. In: *The Quarterly Journal of Economics* 119 (1), pp. 91–134.

Boehm, J., S. Dhingra, and J. Morrow (2022). “The Comparative Advantage of Firms”. In: *Journal of Political Economy* 130 (12).

Boehm, J. and E. Oberfield (2020). “Misallocation in the Market for Inputs: Enforcement and the Organization of Production”. In: *The Quarterly Journal of Economics* 135 (4), pp. 2007–2058.

De Loecker, J., J. Eeckhout, and G. Unger (2020). “The Rise of Market Power and the Macroeconomic Implications”. In: *The Quarterly Journal of Economics* 135 (2), pp. 561–644.

De Loecker, J., P.K. Goldberg, et al. (2016). “Prices, Markups, and Trade Reform”. In: *Econometrica* 84 (2), pp. 445–510.

Foster, L., J. Haltiwanger, and C.J. Krizan (2001). “Aggregate Productivity Growth: Lessons from Microeconomic Evidence”. In: *NBER Chapters, in: New Developments in Productivity Analysis*, pp. 303–372.

— (2006). “Aggregate Productivity Growth: Lessons from Microeconomic Evidence”. In: *The Review of Economics and Statistics* 88 (4), pp. 748–758.

Foster, L., J. Haltiwanger, and C Syverson (2008). “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” In: *American Economic Review* 98 (1), pp. 394–425.

Goldberg, P.K. et al. (2010a). “Imported Intermediate Inputs and Domestic Product Growth: Evidence from India”. In: *The Quarterly Journal of Economics* 125 (4), pp. 1727–1767.

— (2010b). “Multi-product Firms and Product Turnover in the Developing World: Evidence from India”. In: *The Review of Economics and Statistics* 92 (4), pp. 1042–1049.

Hsieh, C. and P. Klenow (2009). “Misallocation and Manufacturing TFP in China and India”. In: *The Quarterly Journal of Economics* 124 (4), pp. 1403–1448.

— (2014). “The Life Cycle of Plants in India and Mexico”. In: *The Quarterly Journal of Economics* 129 (3), pp. 1035–1085.

Levinsohn, J. and A. Petrin (2003). “Estimating Production Functions Using Inputs to Control for Unobservables”. In: *The Review of Economic Studies* 70 (2), pp. 317–341.

Martin, L.A., S. Nataraj, and A.E. Harrison (2017). “In with the Big, Out with the Small: Removing Small-Scale Reservations in India”. In: *American Economic Review* 107 (2), pp. 354–386.

Muralidharan, K., P. Niehaus, and S. Sukhtankar (2017). "General equilibrium effects of (improving) public employment programs: experimental evidence from India". In: *National Bureau of Economic Research, Inc.* NBER Working Papers (23838).

Olley, S. and A. Pakes (1996). "The dynamics of productivity in the telecommunications equipment industry". In: *Econometrica* 64 (6), pp. 1263–1298.

Santangelo, G. (2019). "Firms and Farms: The Local Effects of Farm Income on Firms' Demand". In: *Cambridge Working Papers in Economics 1924, Faculty of Economics, University of Cambridge*.

Shamdasani, Y. (2021). "Rural road infrastructure agricultural production: Evidence from India". In: *Journal of Development Economics* 152 (C), p. 102686.

Wooldridge, Jeffrey M. (2009). "On estimating firm-level production functions using proxy variables to control for unobservables". In: *Economics Letters* 104 (3), pp. 112–114.