Granular level spillover effect and Indian dominant firms: What does the data reveal?

Piyali Dutta* Wasim Ahmad †

Abstract

Understanding the elements driving business cycle fluctuations in any economy is challenging. It remains unclear whether these fluctuations are caused by aggregate shocks impacting all sectors simultaneously or by sector-specific shocks that collectively influence the broader economy. Conventional research shows that aggregate fluctuations are caused only by macroeconomic shocks. Gabaix (2011) suggests that in advanced economies, idiosyncratic shocks to firms play a crucial role in driving aggregate economic fluctuations (granular hypothesis). Building on this approach, this paper investigates whether idiosyncratic shocks contribute to macroeconomic fluctuations in an emerging market such as India. We observe that the size distribution of Indian manufacturing firms follows a power-law distribution, which is essential for the granular hypothesis. Further, we demonstrate how productivity shocks among top firms (i.e., dominant firms) significantly predict future productivity shocks in other non-dominant firms and may collectively influence the national economy. In addition, we find that government firms play an important role in this dynamic.

JEL Classification: E32, D20, C33, G30

Keywords: Business cycles, Emerging market, Granular residual, firm level shocks

*Department of Economic Sciences, Indian Institute of Technology Kanpur, India 208016; Email: piyalid21@iitk.ac.in

†Department of Economic Sciences, Indian Institute of Technology Kanpur, India 208016; Email: wasimad@iitk.ac.in

1 Introduction

Understanding what causes business cycle fluctuations in any economy is at the core of the modern-day macroeconomics. There has been a lot of discussion going on about the granularity of an economy, i.e., whether the idiosyncratic shocks are as important as the macroeconomic shocks in order to explain the business cycle fluctuations. Traditionally, researchers have assumed that aggregate fluctuations are due to macroeconomic shocks, and idiosyncratic shocks were instead expected to wash out in the aggregate. Lucas (1977) put forward the argument favouring this *diversification hypothesis*. It relies on the law of large numbers: positive shocks at some firms are offset by negative shocks at others. Thus, such microeconomic shocks would only have negligible aggregate effects.

In recent years, researchers have increasingly emphasized the micro-foundations of macroeconomics, recognizing that not only aggregate shocks but also firm-level (or granular) shocks play a critical role. Gabaix (2011) is the first to develop a firm-level theory on the origins of aggregate fluctuations; the so-called *granular hypothesis*, which states that idiosyncratic firm-level shocks can explain a significant portion of fluctuations in economic aggregates. He demonstrates that in the United States, the one hundred largest non-financial firms have a substantial impact on business cycle fluctuations, contributing to nearly one-third of GDP variations. It is supported by the fact that the asymmetrical distribution of firm sizes prevents the averaging out of shocks caused by individual firms at the aggregate level. Like in the U.S., the ten largest firms represent one-fourth of the overall market capitalization, (Jannati, 2020). This suggests that the firm size distribution in the U.S. is fat-tailed and follows a power law distribution (Axtell, 2001).

After Gabaix (2011), a multitude of studies have identified several macroeconomic aggregates with granular fluctuations, such as exports (del Rosal, 2013), investments (Grullon et al., 2013), credit network (Altinoglu, 2021), and many more

(Friberg and Sanctuary, 2016; Ebeke and Eklou, 2017). Di Giovanni and Levchenko (2012) demonstrates that trade openness can enhance granular performance, thereby increasing aggregate volatility. Similarly, though from a different perspective, Acemoglu et al. (2012) argue that idiosyncratic microeconomic shocks and inter-sectoral input-output linkages can contribute to aggregate fluctuations. These firm linkages enable microeconomic shocks to propagate across firms, creating positive comovements in the economy (Jannati, 2020). Research shows that the indirect effects of these shocks are approximately three times more impactful than the direct effects (Di Giovanni et al., 2014). Granular behavior has also been identified at the sectoral level, including within the banking sector (Blank et al., 2009). However, Stella (2015) shows that after properly controlling for aggregate shocks, idiosyncratic shocks have little role in explaining U.S. business cycle fluctuations. Therefore, it needs further investigation to understand which of these aforementioned hypotheses works for a country i.e., *diversification hypothesis* or *granular hypothesis*.

Taken as a whole, existing research indicates that idiosyncratic shocks play a crucial role in explaining business cycle fluctuations. Moreover, these microeconomic shocks generate aggregate fluctuations through the following two mechanisms (direct effect). First, where idiosyncratic shocks to large firms directly contribute to aggregate fluctuations. Second, aggregate fluctuations can arise from idiosyncratic shocks due to input–output linkages across the economy (indirect effect). In advanced economies, firm-specific idiosyncratic shocks are a major driver of aggregate fluctuations (Gabaix, 2011; Di Giovanni and Levchenko, 2012; Acemoglu et al., 2012). While similar effects are observed in emerging markets, their contribution to macroeconomic fluctuations is comparatively smaller than what has been reported in earlier studies focused on developed economies (Grigoli et al., 2023). This study also draws on data from an emerging market, specifically India, with a focus on the indirect effects. There is limited data available on firm-level employment, which limits our ability to directly assess the impact of idiosyncratic shocks on macroe-

conomic volatility. This study addresses this gap by examining the indirect effects of firm-level idiosyncratic shocks in the economy. Understanding this relationship provides a fresh insight into how firm-level dynamics shape aggregate fluctuations in an emerging market context.

Therefore, this study addresses two key research questions: first, whether the distribution of firm's size follows a *power law distribution* which is essential for *granular hypothesis* (Gabaix, 2016), and second, following Jannati (2020), we try to see if there is evidence for the indirect effect of idiosyncratic shocks present in the economy by capturing the impact of dominant firm's shocks on non-dominant firm's shocks. Jannati (2020) identifies the top 100 firms based on their sales as the dominant firms, while identifies non-dominant firms as those headquartered in the same state as the dominant firms. Due to a lack of data availability, we identify the dominant firms as the top one-third of firms and the bottom one-third of firms as the non-dominant firms based on their sales each year.

For firm-specific productivity shocks, we use two measures. First one is given by Gabaix (2011), where we measure the shock as the de-meaned labour productivity and second one is given by Jannati (2020), which measures a firm's productivity shocks as the residual of the firm's productivity growth, over and above the firm's exposure to the average productivity growth of the economy and the average productivity growth of its state. We find that the distribution of firm size in India exhibits a power-law distribution featuring a heavy right-skewed tail. It offers a complementary view for the granular hypothesis, highlighting the importance of the top firms. Moreover, the study reveals that the dominant firm's shocks have a significant impact on the non-dominant firms, which indicates that there is evidence of propagation of shocks from top firms to bottom. Furthermore, after excluding the non-financial services from the data set, the results remain robust. However, when we exclude government firms from the data set the results are coming out to

be negligible.

The rest of the paper is organized as follows. Section 2 describes the data and methodology used and the distributions. Section 3 describes the dynamics of Indian firms. Section 4 reports the baseline results as well as some additional results from the sub-sample analysis. Section 5 concludes.

2 Data and Methods

This study uses the CMIE Prowess database for the firm-level annual data between 2006 and 2020. Next, we begin by explaining the sampling framework and estimation strategies, followed by an overview of the firm specific productivity measures, outlining how they are constructed and applied in the study. Finally, we illustrate the main variables and presents a summary statistic for the controls.

2.1 Sampling Framework and Estimation Strategies

It takes several steps to structure the final sample for this study. To begin with, we have excluded some specific firms like financial services, irrigation and diversified. The reason behind this exclusion is that sales from these firms do not accord with the underlying economic significance of the measure used in the study. Further, we refine our dataset by excluding those firms which enters or exits and got merged or acquired during the considered time period. Then, we exclude all the firms which has missing or negative values for the sales variable. It is worth noting that, given the two central research questions of this study, we are using two distinct samples accordingly.

First, to test for the *power-law distribution* we have a total of 18,890 firms and 11,029 firms for 2006 and 2020, respectively. The distinction between the number of firms taken each year is merely due to a lack of data availability. According to Gabaix (2016), the *granular hypothesis* holds if the firm size distribution is suffi-

ciently skewed towards the right. In such cases, a *power law distribution* of firm size renders the law of large numbers inapplicable which is the basis of the *diversification hypothesis*. This indicates that shocks at the firm level can have the potential to make substantial aggregate effects on the economy. To analyse the firm size distribution, the next step involves the estimation of the power-law parameter. To estimate the power law exponent ζ, we regress $log(rank - 1/2)$ on $log(sales)$ where, $\zeta \approx 1$ suppose to hold. We substantiate the results by running the regression on the years 2010 and 2015 as well, which can be seen in Table A.1 and Table A.2.

Second, to test the indirect effect, our sample contains data on 524 non-financial firms during the period 2006-2020. Due to limited data on the number of employees in the Prowess database, we must drop a sizeable number of observations. Only the non-financial firms with sales and employee data available for the current and previous years are included, which led to a 15-year data restriction. We sorted the firms based on their sales each year and considered the top one-third of the total firms in our sample as *dominant firms* and the bottom one-third as *non-dominant firms*. So, we have taken the top 175 firms as the *dominant firms* and the bottom 175 firms as the *non-dominant firms*. We did this for each year and then pooled each year's data together to make a pooled dataset. We call it a pooled data because even though the number of cross-sections is the same each year, the firms are different because of the sorting. This pooled panel setting allows us to correctly study the effects of a *non-dominant firms* on the *non-dominant firms*.

Moreover, to address the concern of a within-unit error correlation that may arise due to construction of the panel in this manner, we use the Fama and MacBeth (1973) two-step regression method and adjust the standard errors using the Newey and West (1987) method with a 6-year lag (Jannati, 2020). We also substantiate the robustness of the results by using different lag values to adjust the standard errors (Table A.4 and Table A.5). Furthermore, we conducted a sub-sample analysis

by focusing on core manufacturing sectors (excluding non-financial services) and considering only private firms (excluding government firms) as well.

2.2 Productivity Measures

In the study the productivity of a firm is taken as the ratio of its sales per employee. Using the measure suggested by Gabaix (2011) we define productivity growth as follows:

$$ProductivityGrowth_{j,t} = \log\left(\frac{Sales_{j,t}}{Employees_{j,t}}\right) - \log\left(\frac{Sales_{j,t-1}}{Employees_{j,t-1}}\right) \quad (1)$$

Then the firm specific productivity shocks can be written as the demeaned productivity growth:

$$ProductivityShocks_{j,t} = ProductivityGrowth_{j,t} - \frac{1}{N}\sum_{j=1}^{N} ProductivityGrowth_{j,t} \quad (2)$$

Additionally, to robust the analysis we use a second measure proposed by Jannati (2020) using the residual $\epsilon_{j,s,t}$ of the following pooled-panel regression:

$$\begin{aligned} Productivity\ Growth_{j,s,t} =& \alpha_j + \beta_j^s ProductivityGrowth_{s,t} + \beta_j ProductivityGrowth_t \\ &+ \epsilon_{j,s,t} \end{aligned} \quad (3)$$

In the above equation (3), $Productivity\ Growth_{j,s,t}$ is the productivity growth of firm j, headquartered in state s at time t. $Productivity\ Growth_{s,t}$ indicates the average productivity growth of firms headquartered in state s at time t and $Productivity\ Growth_t$ specifies the average productivity growth of all firms in the sample at time t. The estimated residual i.e., $\epsilon_{j,s,t}$ depicts firm-level idiosyncratic productivity shocks over and above the combined geography- and economy-wide shocks. We call the productivity shocks from equation (2) and (3), *Measure 1* and *Measure 2*, respectively.

Figure 1 represents a snapshot of both the productivity measure viz, *Measure 1* and *Measure 2* during the period 2006-2020. Here, we have used our full sample data to calculate the firm-level productivity shocks and then compute the annual average to construct the graph. We can see that both the measures have followed a similar pattern throughout the time period. It shows a little deviation from the zero in 2007-2008 likely due to the global financial crisis at that time. Thereafter, it remains more or less stable until 2014. However, the period between 2014 and 2020 demonstrates significant volatility in the graph. This could be influenced by several key domestic and global events like, demonetization in 2016, new indirect tax implementation in 2017 and global trade uncertainties in 2018-2019.

2.3 Variable description

To control the profitability and growth opportunities of a firm, we use the following variables: *Size, Leverage, Loss, Market-to-Book Ratio, Cash Flows*, and *Dividend Yield* (Addoum and Murfin, 2017). We also control for the lagged productivity shocks of the non-dominant firm to ensure that the shock propagation is beyond the effects of the firm's own shocks. *Size* is the three-year average of the total income and total assets of the firm. *Total assets* are the sum of all current and non-current

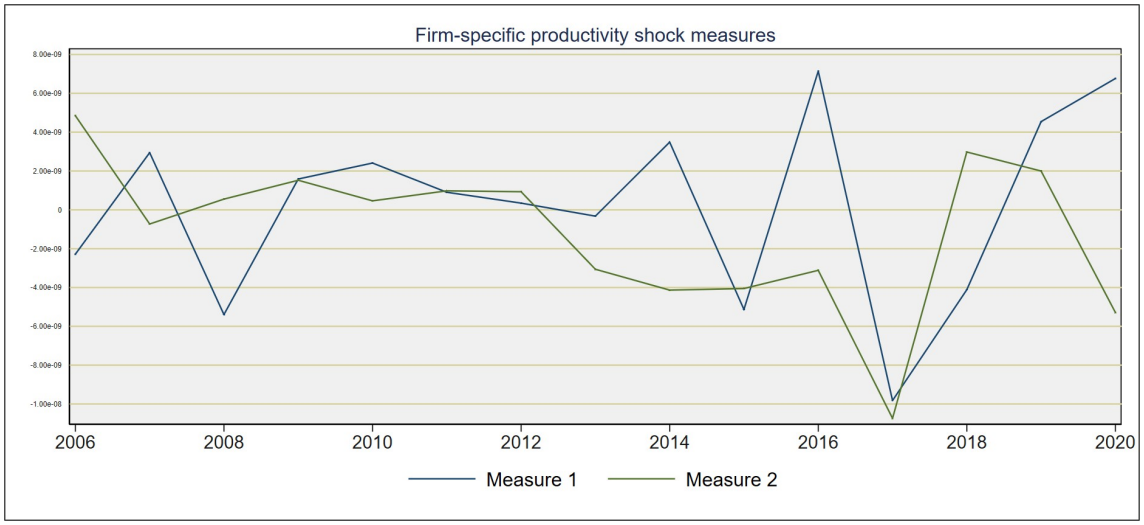


Figure 1: Firm-specific shock measures Snapshot of average productivity shock measures used in the study from 2006 to 2020 annually. The blue line corresponds to *Measure 1*, and the green line corresponds to *Measure 2*.

assets held by the firm. *Leverage* is the Debt-to-Equity ratio. *Loss* is a dummy variable that takes the value of 1 if *PBDITA* is negative, and 0 otherwise. Here, *PBDITA* refers to profit before depreciation, interest, tax, and amortisation. *Market-to-Book Ratio* is the Price to book value as per the Bombay Stock Exchange (BSE). *Cash Flows* are the cash flows from operating activities, divided by total assets, and *Dividend Yield* is dividends, divided by shareholder's equities. Table A.3 consists of the detailed description of the variables used in the study.

Subsequently, Table 1 reports the descriptive statistics of the main variables for the dominant and non-dominant firms obtained after the sorting process mentioned in section 2.1. We can see the non-dominant firms have a higher level of volatility in productivity shocks (57% vs. 33% for dominant firms). This pattern holds true for firm's cash flows and dividend yield measures. Also, sales and number of employees of the dominant firms are well above those of non-dominant firms. On average, dominant firms have sales of INR 153.77 billion (USD 1.80 billion)[1], but this number is INR 1.03 billion (USD 0.01 billion) for non-dominant firms.

Table 1: Summary statistics of dominant and non-dominant firms This table represents the summary statistics of the main variables used in the analysis. Panels A and B show the summary statistics for the dominant and non-dominant firms, respectively. The specific definition of each variable is reported in Table A.3. The sample period is from 2006 to 2020.

	Panel A: Dominant Firms			Panel B: Non-Dominant Firms		
Main variables	Mean	Std.	Obs.	Mean	Std.	Obs.
Productivity shocks (Measure 1)	0.0132	0.3381	2625	-0.0162	0.5707	2625
Productivity shocks (Measure 2)	0.0123	0.3362	2625	-0.0159	0.5436	2625
Sales (Rs. Million)	153771.9	460459.1	2625	1026.39	850.03	2625
Employees	14184.51	29979.07	2625	447.99	857.19	2625
Size	10.8814	1.2949	2625	6.7897	1.1973	2625
Cash flows	0.9292	0.1455	2625	0.0517	0.2014	2625
Loss	0.0228	0.1494	2625	0.1253	0.3311	2625
Leverage	1.0608	28.4581	2584	2.8984	20.2446	2507
Dividend yield	0.0746	0.2584	2476	0.0236	0.3969	1904
Market to book value	3.7784	7.32	2312	5.8189	39.1563	2325

[1]USD 1 = INR 83 (approx), conversion based on exchange rate as of May 17th, 2025. This exchange rate is consistently used throughout the paper.

Additionally, to better understand the distribution of firms, we present the summary statistics utilising the full sample data on sales and employees, in Table 2. As demonstrated, the average sales (Rs. 53,957.66 million ≈ USD 650.09 million) are approximately 9 times greater than the median sales (Rs. 5725.3 million ≈ USD 68.97 million), indicating the existence of large firms. Similarly, the average size is 5813.82 employees, which is much higher than the median size, 1090 employees. Both variables exhibit right-skewed distributions, meaning a few firms have extremely high values. The sample of firms with comprehensive employee data is relatively small, but is clearly dominated by bigger firms.

Table 2: Summary statistics of the full sample This table represents the summary statistics of the variables sales and employees using the full sample data i.e., 524 firms from 2006-2020. Here, the unit of sales is in Rs. millions and pctl refers to percentiles.

	All Firms					
Main variables	Obs.	Mean	25th pctl	Median	75th pctl	95th pctl
Sales	7860	53957.66	1507.15	5725.3	20411.55	169115
Employees	7860	5813.82	369	1090	3499	24316

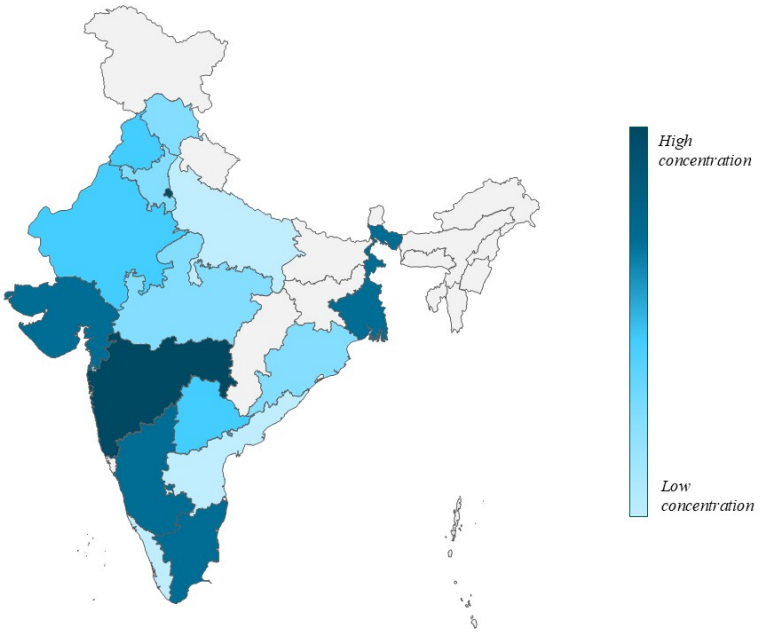


Figure 2: Geographic Distribution of Dominant Firms This figure shows the geographic distribution of dominant firms across Indian states in 2020. Dominant firms are the top firms defined based on firm's sales. The darker blue shades imply the most concentrated region, and the concentration weakens as the shade becomes lighter.

Table 3: Top 10 Dominant Firms This table represents the distribution of the top 10 dominant firms in India after sorting the firms based on sales in a five-year interval from 2006 to 2020. See Table A.6 for the complete names of the companies.

2006	2010	2015	2020
Indian Oil	Indian Oil	Indian Oil	Indian Oil
Reliance Indus.	Reliance Indus.	Reliance Indus.	Reliance Indus.
Bharat Petro.	Bharat Petro.	Bharat Petro.	Bharat Petro.
Hindustan Petro.	Hindustan Petro.	Hindustan Petro.	Hindustan Petro.
ONGC	FCI	FCI	FCI
FCI	ONGC	Nayara Energy	Nayara Energy
SAIL	NTPC	ONGC	NTPC
NTPC	SAIL	NTPC	ONGC
Chennai Petroleum	Nayara Energy	Larsen and Toubro	Larsen and Toubro
Tata Motors	Tata Motors	GAIL	Infosys

3 Dynamics of Indian Firms

This section presents various figures to better understand the dynamics of Indian firms. India's leading non-financial firms play a major role in contributing to GDP across sectors like manufacturing, telecommunications, IT, consumer goods, and energy. These sectors significantly contribute to GDP growth, employment, export revenues, and technological advancement. Leading industries within the non-financial sector are driven by some of India's largest corporations, such as the Tata Group, Reliance Industries, Hindustan Unilever, and Infosys, each playing a unique role in their respective fields. We have done some distributional analysis of our sample. Figure 2 shows the geographical distribution of the dominant firms in India in 2020. Most of the dominant firms are located in NCT of Delhi and Maharashtra. The darker blue shades imply the most concentrated region, and the concentration weakens as the shade becomes lighter. Therefore, NCT of Delhi and Maharashtra are the most concentrated ones, and then there are Gujarat, Karnataka, Tamil Nadu, and West Bengal. In the 3rd phase, there are Rajasthan, Punjab, Telangana, and Odisha. The least concentrated regions are Andhra Pradesh, Kerala, and Uttar Pradesh.

Table 3 and Table 4 represent the distribution of Top 10 and Bottom 10 Firms based on their sales respectively from 2006 to 2020 in five-year intervals. Indian Oil,

Table 4: Bottom 10 Non Dominant Firms This table represents the distribution of the bottom 10 non-dominant firms in India after sorting the firms based on sales in a five-year interval from 2006 to 2020. See Table A.6 for the complete names of the companies.

2006	2010	2015	2020
Unjha Formulations	Kernex Microsys.	Gujarat Hotels	CSGS Ltd.
Antarctica Ltd.	Antarctica Ltd.	Geecee Ventures	Jayshree Chem.
National Textile	Gujarat Hotels	Antarctica Ltd.	Jaysynth Org.
HMTC Watches	LALW Ltd.	Facor Alloys	Accel
Gujarat Hotels	Saven Technologies	SAE Ltd.	Antarctica Ltd.
Wep Solutions	CSGS Ltd.	Jaysynth Org.	Bihar Sponge Iron
CSGS Ltd.	SAE Ltd.	HMTC Watches	SAE Ltd.
Vintron Info.	Jaysynth Org.	Accel	PCM Inds. Ltd.
SAE Ltd.	HMTC Watches	Chrome Silicon	Pankaj Polymers
Starlite Comp.	Wep Solutions	IEL Ltd.	Sri Vajra Granites

Reliance Industries, Bharat Petroleum, and Hindustan Petroleum are at the top of the list and remain at the top over time. Indian Oil Corporation (IOC), Bharat Petroleum Corporation Limited (BPCL), and Hindustan Petroleum Corporation Limited (HPCL) are the state-owned oil companies dominate India's refining and fuel distribution. Together, they contribute around 3-4% of GDP by providing crucial fuels for transportation, power generation, and industrial use. Then, there are other firms, like FCI, SAIL, ONGC, NTPC, Tata Motors, etc., have also made it to the top ten dominant firms list. However, the names and positions of the firms in the bottomm 10 non-dominant firms changes over the period.

Furthermore, we have sorted the firms based on sales, productivity and number of employees over five-year intervals and presented the top 10 and Bottom 10 distribution of firms in Figure 3 and Figure 4 respectively. It illustrates the allocations across three dimensions viz., Private vs. Public, Non-oil vs. Oil, and Services vs. Manufacturing (Mfg.). Looking at the first section, where firms have been sorted based on sales, in both the table, we can see that most of the top firms are state-owned and oil companies and most of the bottom firms are private and non-oil and this distribution remains the same over the period.

Now looking at the second section of Figure 3 where we have sorted the firms based on number of employees, we can observe that state-owned and oil companies

have a lower number of employees than the private and non-oil companies, and this number decreases over time. Figure 3 also shows that most of the firms are manufacturing firms in the top 10 companies, while the number of services firms remains low across all three categories. Again, when we have sorted the firms based on productivity, we observe that the number of non-oil companies is greater than the oil companies in the top 10 places, i.e., non-oil companies are more productive. Figure 4 exhibits that all the bottom 10 firms are private and non-oil companies across all years when sorting is done based on sales and number of employees. The distribution between manufacturing and services in terms of sales and number of employees shows slight fluctuations over the year. Productivity scores are relatively evenly distributed between public & private firms and show consistent values in Oil vs non-oil categories throughout time. Also, in the bottom ten firms, the number of manufacturing firms is more significant than the services, which signifies that manufacturing firms are less productive than the services.

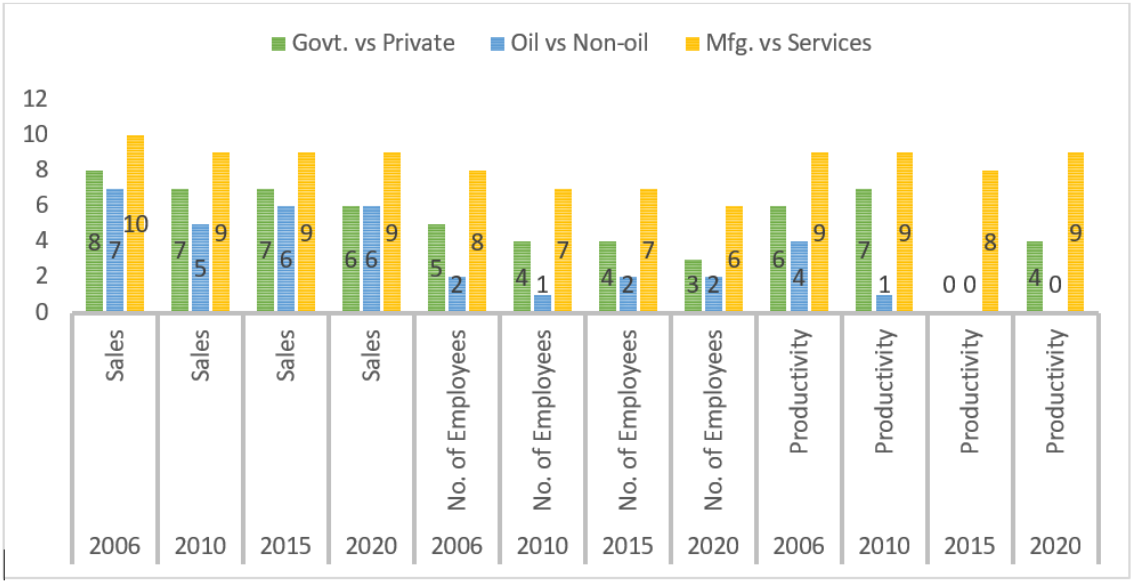


Figure 3: Distribution of Top 10 Dominant Firms This table uses a stacked bar chart to represent the distribution of the top 10 firms in India when sorting is done based on sales, No. of employees and productivity in a five-year interval from 2006 to 2020. The numbers inside the green, sky, and yellow bars indicate the count of government, oil, and manufacturing firms, respectively, in the top 10 list.

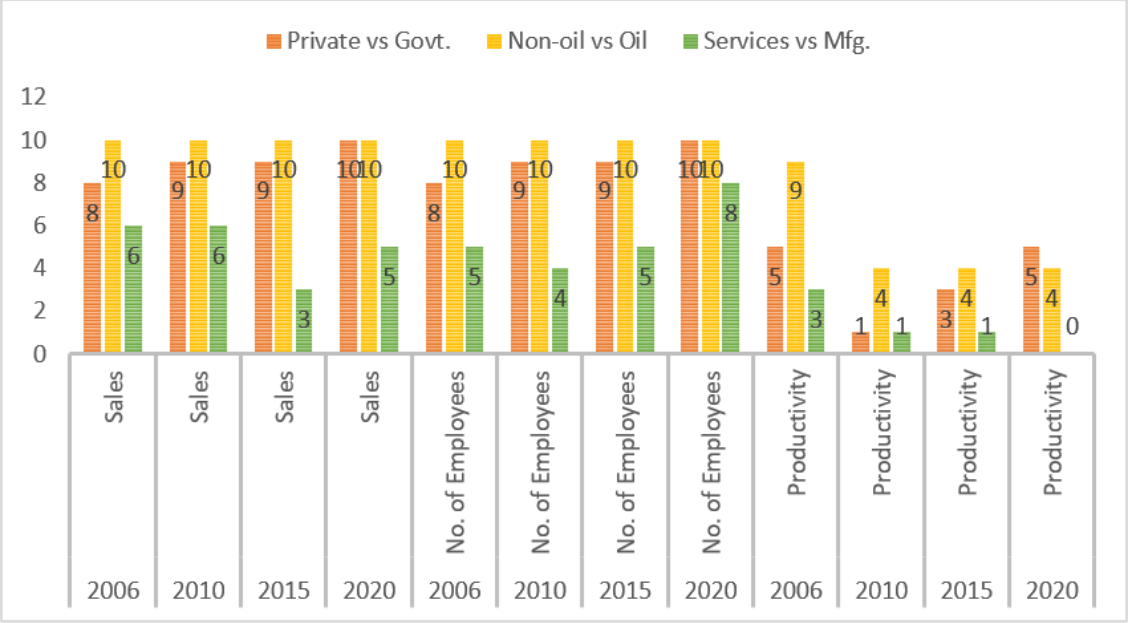


Figure 4: Distribution of Bottom 10 Non-Dominant Firms This table uses a stacked bar chart to represent the distribution of the bottom 10 firms in India when sorting is done based on sales, No. of employees and productivity in a five-year interval from 2006 to 2020. The numbers inside the orange, yellow, and green bars indicate the count of private, non-oil, and services firms, respectively, in the bottom 10 list.

4 Results and findings

This section exhibits the key empirical findings of the paper. First, we show sample firms follow a power law distribution, providing initial evidence supporting the granular hypothesis. Next, we present the baseline regression results demonstrating the indirect effect i.e., shocks to the dominant firms do affect the non-dominant firm's shocks in India. In the end, we present the regression results from the sub-sample analysis as the robustness checks.

4.1 Testing for power law

Following Gabaix (2016), a power-law distribution can be represented as the following:

$$P(Size > x) = k/x^{\zeta} \tag{4}$$

Here, x is inversely proportional to x^{ζ}, where ζ is the powecr-law parameter

Table 5: Rank vs sales regression 2006 This table presents the results from regressing log-rank on log sales for the year 2006. Here, (1) uses the full sample data, then (2), (3), (4), and (5) retain firms with revenues exceeding INR 100, 500, 1000 and 1500 million (USD $\approx 1, 6, 12$ and 18 million), respectively. Standard errors are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	log rank				
	(1)	(2)	(3)	(4)	(5)
log sales	-0.3101***	-0.6953***	-0.8570***	-0.9207***	-0.9537***
	(0.0018)	(0.0018	(0.0017)	(0.0017)	(0.0017)
Constant	9.7818***	12.3616***	13.6647***	14.2256***	14.5306***
	(0.0101)	(0.0125	(0.0131)	(0.0145)	(0.0174)
Observation	11029	6357	3488	2331	1720
Adjusted R^2	0.718	0.956	0.986	0.991	0.992
Prob>F	0.000	0.000	0.000	0.000	0.000

and k is a constant. ζ having a value between (and including) 0 and 2 indicates a fat-tailed distribution whilst, a value of 1 corresponds to Zipf's law. Following Gabaix and Ibragimov (2011), we regress the shape parameter as follows:

$$log(rank - 1/2) = logk + \beta log(sales) \tag{5}$$

In the given equation (5), *rank* is estimated based on the firm's sales and β is the primary coefficient which is expected to be negative. Moreover, ζ is the absolute value of the estimated coefficient β indicating the probability of finding large firms is inversely proportional to the magnitude of ζ. Given that the power law distribution fails to adequately characterize the distribution of small firms, this paper follows the methodology of Gabaix (2009) and Di Giovanni et al. (2011). Consequently, it performs the regression analysis after exclusion of small firms that fall below predefined thresholds of annual sales.

Table 5 and Table 6 present the OLS estimates of the above equation utilizing sales and sales ranking in 2006 and 2020, respectively. Columns (1) – (5) present the regression results using a distinct sample that employs a unique threshold for eliminating small firms. Column (1) uses the full sample, followed by columns (2) – (5), which possess firms with sales exceeding INR $100, 500, 1000$, and 1500 million (USD $\approx 1, 6, 12$ and 18 million), respectively. Across the columns, the magnitude of

the coefficient ranges from 0.31 to 0.95 in Table 5, likewise, 0.31 to 0.91 in Table 6, respectively. This indicates a power-law distribution in Indian firms. Moreover, moving from left to right, the magnitude of the coefficient is increasing, suggesting the power-law parameter approaching Zipf's law parameter ($\zeta = 1$). In cases where the sample is limited to larger corporations, the power-law tends to approach Zipf's law (Gabaix, 1999). Furthermore, the coefficient of determination (R^2) in the regression analysis exhibits a significant magnitude, particularly in (2) – (5), which suggests a highly accurate representation of the power-law distribution in characterizing the upper end of firm size.

Table 6: Rank vs sales regression 2020 This table presents the results from regressing log rank on log sales for the year 2020. Here, (1) uses the full sample data, then (2), (3), (4), and (5) retain firms with revenues exceeding INR 100, 500, 1000 and 1500 million (USD $\approx 1, 6, 12$ and 18 million), respectively. Standard errors are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	log rank				
	(1)	(2)	(3)	(4)	(5)
log sales	-0.3178***	-0.6058***	-0.7709***	-0.8551***	-0.9098***
	(0.0013)	(0.0015	(0.0016)	(0.0017)	(0.0017)
Constant	10.2060***	12.7078***	14.1068***	14.8703***	15.3855***
	(0.0085)	(0.0112	(0.0133)	(0.0144)	(0.0145)
Observation	18890	12581	8147	6090	4953
Adjusted R^2	0.734	0.923	0.963	0.976	0.983
Prob>F	0.000	0.000	0.000	0.000	0.000

Figure 5 and Figure 6 provides straightforward evidence of the suitability of the Pareto distribution for characterizing large firms in India. It presents a scatter plot depicting the observed log(rank-1/2) v/s log(sales) data for the year 2006 and 2020. Additionally, it also illustrates the signified relationship estimated in Table 5 and Table 6. Although the *Pareto distribution* may not accurately model firms with logarithmic sales values below eight, it is a suitable fit for firms with revenues exceeding INR 1,500 million (USD 18 million).

The empirical data presented here provides indications that the firm size distribution within the Indian context unveils significant asymmetry. It precisely characterises a power law distribution featuring a heavy right-skewed tail, with a cor-

respondingly low value of the scaling exponent ζ. The aforementioned outcome serves as the foundation of this research, as it presents the possibility of identifying the specific sources of aggregate fluctuations within the Indian economy. In the subsequent section, we establish the presence of indirect effect by showing that the shocks to dominant firms in India has a significant impact on the non-dominant firm's productivity shocks.

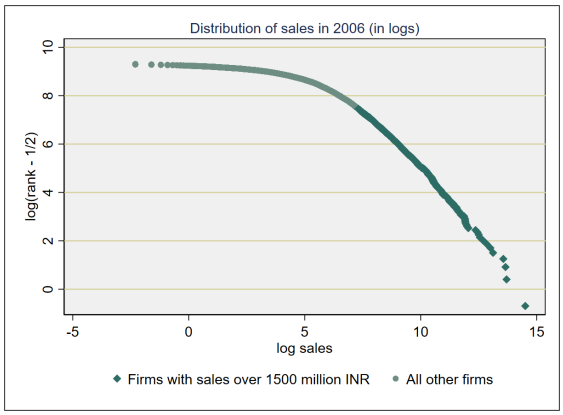


Figure 5: Distribution of sales in 2006 (in logs)

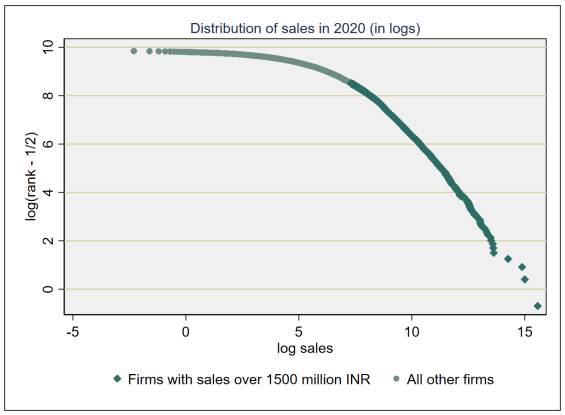


Figure 6: Distribution of sales in 2020 (in logs)

4.2 Baseline results

The baseline model of the study is given by the following equation where we have performed the predictive Fama and MacBeth (1973) regression. Here β_1, the coefficient of interest, shows the predictive power of a dominant firm's shocks on future shocks to a non-dominant firm. The Fama-MacBeth method is chosen as it runs a cross-sectional regression for each year, providing an average estimate suited to our data. The $X_{j,t}$ include the following control variables, Lagged Productivity shock, Size, Cash Flows, Leverage, Dividend Yield, Market to Book value and Loss. The Fama and MacBeth (1973) regression is given as:

$$Non-Dominant\ firm's\ shock_{j,t+1} = \alpha_j + \beta_1 Dominant\ firm's\ shock_{j,t} + \beta_2 X_{j,t} + \epsilon_{j,t+1} \tag{6}$$

where $X_{j,t}$ are the control variables.

Now, we have presented the regression results in Table 7 where the first column represents the results using *Measure 1*, i.e., the Gabaix method and the second

column represents the results using the *Measure 2*. From both the measures, we found similar kinds of results. We have found that shocks to a dominant firm predict productivity shocks to the local non-dominant firm. However, the coefficient is small, and the significance level is 10 %. We have taken all the factors that can affect the productivity shock of the non-dominant firms as control variables. It indicates a negative and statistically significant coefficient for the lagged value of Non-Dominant firm's productivity shocks. The reason could be additional profits of a firm may attract competitors. When entry and exit are sufficiently free, this attraction leads to a decline in the profit margin of the existing firm (Jannati, 2020). *Size* is negatively related to productivity shock, as the more significant the firm's size, the greater the shock will have a more negligible effect on it. Again, *Cash flows* is negatively related to the productivity shock, i.e., if cash flow is larger in the firm at t, then the firm will experience a smaller productivity shock in the future. Now, *Leverage* is positively related to the productivity shock as we have taken debt to equity as leverage if the ratio increases, there is a high chance that the firm can experience larger shocks in the future. The rest of the controls i.e., *Dividend yield*, *Market to book value* and *Loss* can also be explained in the equivalent manner.

Table 7: Fama-Macbeth regression of full sample This table represents the regression results of equation 6 using full sample data. Column (1) and (2) represents the results using *Measure 1* and *Measure 2* explained in the section 2.2 respectively. The sample period is from 2005 to 2020. Standard errors that are adjusted using the Newey and West (1987) method are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	Non-Dominant Firm's Productivity Shocks (t+1)	
	(1)	(2)
Dominant firm's productivity shocks(t)	0.0691**	0.0612*
	(0.031)	(0.0289)
Non-Dominant firm's productivity shocks(t)	-0.1395***	-0.1317***
	(0.0306)	(0.0306)
Size (t)	-0.0728***	-0.0651***
	(0.019)	(0.0182)
Cash flows (t)	-0.5590*	-0.5713*
	(0.3171)	(0.3095)
Leverage (t)	0.0064	0.0063**
	(0.0038)	(0.0025)
Dividend yield (t)	0.0691	0.1031
	(0.1393)	(0.1344)
Market to book value (t)	-0.0087***	-0.0076**
	(0.0032)	(0.0027)
Loss (t)	-0.0611	-0.0358
	(0.0950)	(0.0871)
Constant	0.5331***	0.4794***
	(0.1184)	(0.1111)
Observation	1639	1639
Average R^2	0.180	0.166
Prob>F	0.000	0.001

We have performed the regression process on different sub-samples to make the results more robust. Table 8 have shown the results after we exclude all the non-financial services firms like, business services and consultancy, shipping and transport, education, tourism, hotel etc. from our dataset. We have then repeated the sorting process for the reduced dataset and run the same regression again on the new dataset. The result from the analysis also indicates that our main explanatory variable i.e., Dominant firm's productivity shocks does have a significant effect on the Non-Dominant firm's productivity shocks. Results from the Table 8 suggests that even when we are looking at the core manufacturing firms in India, we find evidence for the propagation of productivity shocks across firms. All other variable

Table 8: Fama-Macbeth regression after excluding services This table represents the regression results of equation 6 using data after excluding non-financial services. Column (1) and (2) represents the results using *Measure 1* and *Measure 2* explained in the section 2.2 respectively. The sample period is from 2005 to 2020. Standard errors that are adjusted using the Newey and West (1987) method are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	Non-Dominant Firm's Productivity Shocks (t+1)	
	(1)	(2)
Dominant firm's productivity shocks(t)	0.1258*	0.1396**
	(0.0635)	(0.0609)
Non-Dominant firm's productivity shocks(t)	-0.2106***	-0.1761***
	(0.024)	(0.0184)
Size (t)	-0.0604***	-0.0475***
	(0.0154)	(0.0151)
Cash flows (t)	-0.2747	-0.3105
	(0.2289)	(0.2407)
Leverage (t)	0.0076*	0.0076*
	(0.0044)	(0.0036)
Dividend yield (t)	-0.0648	0.0089
	(0.0538)	(0.0689)
Market to book value (t)	-0.0105*	-0.0105*
	(0.0061)	(0.0058)
Loss (t)	0.0080	0.0254
	(0.1775)	(0.1492)
Constant	0.4234***	0.3331***
	(0.0960)	(0.0867)
Observation	1326	1326
Average R^2	0.226	0.209
Prob>F	0.000	0.000

has the similar effects as we have discussed for the Table 7. Average r-square is around 23 and 21 per cent for both the models, respectively.

As we have already discussed in section 3 that government firms play an important role in India. Most of the firms in the top 10 list of Indian firms are government when sorting is done based on sales. Therefore, it is crucial to investigate whether we find a similar results after excluding them from the dataset. Table 9 represented the results where we exclude the government firms from the dataset. We again repeat the entire process with this reduced dataset. Here, we could not find significant results that shows that there might be propagation of productivity shocks

Table 9: Fama-Macbeth regression after excluding government firms This table represents the regression results of equation 6 using data after excluding government firms. Column (1) and (2) represents the results using *Measure 1* and *Measure 2* explained in the section 2.2 respectively. The sample period is from 2005 to 2020. Standard errors that are adjusted using the Newey and West (1987) method are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	Non-Dominant Firm's Productivity Shocks (t+1)	
	(1)	(2)
Dominant firm's productivity shocks(t)	-0.0008 (0.0216)	0.0662 (0.0439)
Non-Dominant firm's productivity shocks(t)	-0.1706*** (0.0334)	-0.1587*** (0.0324)
Size (t)	-0.0684** (0.0258)	-0.0590** (0.0245)
Cash flows (t)	-0.5915* (0.3244)	-0.6038* (0.3096)
Leverage (t)	0.0105*** (0.0024)	0.0092*** (0.0021)
Dividend yield (t)	0.0257 (0.1359)	0.0773 (0.1349)
Market to book value (t)	-0.0082** (0.0030)	-0.0066** (0.0024)
Loss (t)	-0.0361 (0.1134)	-0.0049 (0.1018)
Constant	0.4847*** (0.1589)	0.4214*** (0.1503)
Observation	1415	1415
Average R^2	0.208	0.193
Prob>F	0.000	0.000

across firms. The average r-square is around 20 per cent for both the models and the F-statistics suggests that the model is overall significant. This suggests another evidence for the importance of government firms in India.

5 Conclusion

Literature in modern-day macroeconomics has increased its concentration on the role of large firms. Gabaix (2011) demonstrates that if the distribution of firm size follows a power law with an exponent close to negative one—which is the case in the data—the economy is granular. It implies idiosyncratic shocks to the largest

firms can lead to aggregate fluctuations. This paper has examined whether there is evidence that supports the granular hypothesis in an emerging economy like, India. As we have discussed earlier that we cannot directly test for the granular hypothesis due to data limitation for number of employees. Therefore, we answer the main two research question of the study as follows.

First, we test for power-law distribution across Indian firms which gives us a foundation for the granular hypothesis. Following Gabaix (2016) we find that the distribution of firm size in India exhibits power-law distribution featuring a heavy right-skewed tail. It offers a complementary view for the granular hypothesis, with highlighting the importance of the top firms in India. Second, following Jannati (2020) we try to examine the indirect effect by capturing the impact of dominant firm's shocks to non-dominant firm's shocks. The study reveals that the dominant firm's shocks have a small but significant impact on the non-dominant firms in India. It indicates that there is evidence of propagation of shocks from top firms to bottom. Moreover, after excluding the non-financial services from the data set the results remain robust. However, when we exclude government firms from the data set the results are coming out to be negligible.

Overall, we find that the Indian manufacturing sector shows the evidence for the granular hypothesis and that government firms play a significant role in it.

Acknowledgment

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, one of the authors, Piyali Dutta, is a recipient of a Ph.D. fellowship from the Indian Institute of Technology Kanpur.

Conflict of interest: The authors declare that there are no conflicts of interest related to the publication of this paper.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Addoum, J. M. and Murfin, J. R. (2017). Hidden in plain sight: Equity price discovery with informed private debt. *Working Paper*.

Altinoglu, L. (2021). The origins of aggregate fluctuations in a credit network economy. *Journal of Monetary Economics*, 117:316–334.

Axtell, R. L. (2001). Zipf distribution of us firm sizes. *science*, 293(5536):1818–1820.

Blank, S., Buch, C. M., and Neugebauer, K. (2009). Shocks at large banks and banking sector distress: The banking granular residual. *Journal of Financial Stability*, 5(4):353–373.

del Rosal, I. (2013). The granular hypothesis in eu country exports. *Economics Letters*, 120(3):433–436.

Di Giovanni, J. and Levchenko, A. A. (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy*, 120(6):1083–1132.

Di Giovanni, J., Levchenko, A. A., and Mejean, I. (2014). Firms, destinations, and aggregate fluctuations. *Econometrica*, 82(4):1303–1340.

Di Giovanni, J., Levchenko, A. A., and Ranciere, R. (2011). Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics*, 85(1):42–52.

Ebeke, M. C. H. and Eklou, K. M. (2017). *The granular origins of macroeconomic fluctuations in Europe*. International Monetary Fund.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Friberg, R. and Sanctuary, M. (2016). The contribution of firm-level shocks to aggregate fluctuations: The case of sweden. *Economics Letters*, 147:8–11.

Gabaix, X. (1999). Zipf's law and the growth of cities. *American Economic Review*, 89(2):129–132.

Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, 1(1):255–294.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772.

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1):185–206.

Gabaix, X. and Ibragimov, R. (2011). Rank- 1/2: a simple way to improve the ols estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

Grigoli, F., Luttini, E., and Sandri, D. (2023). Idiosyncratic shocks and aggregate fluctuations in an emerging market. *Journal of Development Economics*, 160:102949.

Grullon, G., Hund, J., and Weston, J. (2013). *A granular analysis of corporate investment*. SSRN.

Jannati, S. (2020). Geographic spillover of dominant firms' shocks. *Journal of Banking & Finance*, 118:105844.

Lucas, R. E. (1977). Understanding business cycles. *Carnegie-Rochester Conference Series on Public Policy*, 5:7–29.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Stella, A. (2015). Firm dynamics and the origins of aggregate fluctuations. *Journal of Economic Dynamics and Control*, 55:71–88.

Appendix

Table A.1: This table presents the results from regressing log rank on log sales for the year 2010. Here, (1) uses the full sample data, then (2), (3), (4), and (5) retain firms with revenues exceeding INR 100, 500, 1000 and 1500 million (USD $\approx 1, 6, 12$ and 18 million), respectively. Standard errors are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	log rank				
	(1)	(2)	(3)	(4)	(5)
log sales	-0.3007***	-0.6501***	-0.8073***	-0.8763***	-0.9212***
	(0.0016)	(0.0016	(0.0016)	(0.0018)	(0.0019)
Constant	10.0848***	12.5334***	13.8308***	14.4488***	14.8682***
	(0.0097)	(0.0118	(0.0128)	(0.0153)	(0.0169)
Observation	13748	8858	5431	3761	2913
Adjusted R^2	0.74	0.943	0.977	0.983	0.986
Prob>F	0.000	0.000	0.000	0.000	0.000

Table A.2: This table presents the results from regressing log rank on log sales for the year 2015. Here, (1) uses the full sample data, then (2), (3), (4), and (5) retain firms with revenues exceeding INR 100, 500, 1000 and 1500 million (USD $\approx 1, 6, 12$ and 18 million), respectively. Standard errors are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	log rank				
	(1)	(2)	(3)	(4)	(5)
log sales	-0.3310***	-0.6534***	-0.8042***	-0.8625***	-0.8965***
	(0.0015)	(0.0014	(0.0013)	(0.0014)	(0.0015)
Constant	10.4757***	12.7333***	13.9760***	14.4982***	14.8177***
	(0.0089)	(0.0099	(0.0102)	(0.0116)	(0.0136)
Observation	16819	10852	6455	4515	3438
Adjusted R^2	0.736	0.950	0.983	0.988	0.989
Prob>F	0.000	0.000	0.000	0.000	0.000

Table A.3: This table represents the definition of the main variables used in the empirical analysis. The main data source is CMIE Prowess database.

Variable Name	Definition
Sales	Income generated by companies from clearly identifiable sale of goods and from non-financial services
No of employees	The number of people employed by the firm
Total Assets	Sum of all current and non-current assets held by the firm
Cash Flows	Cash flows from operating activities, divided by total assets
Dividend Yield	Dividend, divided by shareholder's equities
Leverage	Debt to Equity ratio
Market to Book value	Price to book value as per the BSE
PBDITA	PBDITA is profits before depreciation, interest, tax and amortisation
Size	Three-year average of the total income and total assets of the firm
Loss	Set to 1 if PBDITA is negative, 0 otherwise

Table A.4: This table represents the regression results of equation 6 using full sample data. Column (1) and (2) represents the results using *Measure 1* and *Measure 2* explained in the section 2.2 respectively. The sample period is from 2005 to 2020. Standard errors that are adjusted using the Newey and West (1987) method with 3 year lag are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	Non-Dominant Firm's Productivity Shocks (t+1)	
	(1)	(2)
Dominant firm's productivity shocks(t)	0.0691**	0.0612**
	(0.031)	(0.0272)
Non-Dominant firm's productivity shocks(t)	-0.1395***	-0.1317***
	(0.0317)	(0.0348)
Size (t)	-0.0728***	-0.0651**
	(0.0295)	(0.0287)
Cash flows (t)	-0.5590*	-0.5712*
	(0.3167)	(0.3035)
Leverage (t)	0.0064	0.0063*
	(0.0051)	(0.0039)
Dividend yield (t)	0.0691	0.1030
	(0.1758)	(0.1618)
Market to book value (t)	-0.0087***	-0.0076**
	(0.0033)	(0.0029)
Loss (t)	-0.0611	-0.0358
	(0.1308)	(0.1206)
Constant	0.5331***	0.4794**
	(0.1920)	(0.1850)
Observation	1639	1639
Average R^2	0.180	0.166
Prob>F	0.004	0.006

Table A.5: This table represents the regression results of equation 6 using full sample data. Column (1) and (2) represents the results using *Measure 1* and *Measure 2* explained in the section 2.2 respectively. The sample period is from 2005 to 2020. Standard errors that are adjusted using the Newey and West (1987) method with 1 year lag are in parentheses. $***p < 0.01; **p < 0.05; *p < 0.1$.

	Non-Dominant Firm's Productivity Shocks (t+1)	
	(1)	(2)
Dominant firm's productivity shocks(t)	0.0691**	0.0612**
	(0.030)	(0.0267)
Non-Dominant firm's productivity shocks(t)	-0.1395***	-0.1317***
	(0.0342)	(0.0374)
Size (t)	-0.0728***	-0.0651**
	(0.0304)	(0.0293)
Cash flows (t)	-0.5590*	-0.5712*
	(0.2751)	(0.2639)
Leverage (t)	0.0064	0.0063
	(0.0059)	(0.0052)
Dividend yield (t)	0.0691	0.1030
	(0.1824)	(0.1647)
Market to book value (t)	-0.0087**	-0.0076*
	(0.0046)	(0.0042)
Loss (t)	-0.0611	-0.0358
	(0.1565)	(0.1395)
Constant	0.5331***	0.4794**
	(0.2066)	(0.1987)
Observation	1639	1639
Average R^2	0.180	0.166
Prob>F	0.007	0.012

Table A.6: This table represents the complete names of the top 10 and bottom 10 firms used in the analysis. The main data source is the CMIE Prowess database.

Companies	**Complete names of the companies**
Indian Oil	Indian Oil Corpn. Ltd.
Reliance Indus.	Reliance Industries Ltd.
Bharat Petro.	Bharat Petroleum Corpn. Ltd.
Hindustan Petro.	Hindustan Petroleum Corpn. Ltd.
ONGC	Oil & Natural Gas Corpn. Ltd.
FCI	Food Corpn. Of India
SAIL	Steel Authority Of India Ltd.
NTPC	National Thermal Powe Corpn. Ltd.
Chennai Petroleum	Chennai Petroleum Corpn. Ltd.
Tata Motors	Tata Motors Ltd.
Nayara Energy	Nayara Energy Ltd.
Larsen & Toubro	Larsen & Toubro Ltd.
GAIL	GAIL (India) Ltd.
Infosys	Infosys Ltd.
Unjha Formulations	Unjha Formulations Ltd.
Antarctica Ltd.	Antarctica Ltd.
National Textile	National Textile Corpn. Ltd.
HMTC Watches	H M T Chinar Watches Ltd.
Gujarat Hotels	Gujarat Hotels Ltd.
Wep Solutions	Wep Solutions Ltd.
CSGS Ltd.	Capricorn Systems Global Solutions Ltd.
Vintron Info.	Vintron Informatics Ltd.
SAE Ltd.	South Asian Enterprises Ltd.
Starlite Comp.	Starlite Components Ltd.
Kernex Microsys.	Kernex Microsystems (India) Ltd.
LALW Ltd.	Lakshmi Automatic Loom Works Ltd.
Saven Technologies	Saven Technologies Ltd.
Jaysynth Org.	Jaysynth Orgochem Ltd.
Geecee Ventures	Geecee Ventures Ltd.
Facor Alloys	Facor Alloys Ltd.
Accel Ltd.	Accel Ltd.
Chrome Silicon	Chrome Silicon Ltd.
Jayshree Chem.	Jayshree Chemicals Ltd.
Bihar Sponge Iron	Bihar Sponge Iron Ltd.
PCM Indus. Ltd.	Panyam Cements & Mineral Inds. Ltd.
Pankaj Polymers	Pankaj Polymers Ltd.
Sri Vajra Granites	Sri Vajra Granites Ltd.