Granular Origin of Comovement in Fluctuations*

Anindya S. Chakrabarti† Shekhar Tomar‡

November 16, 2021

Abstract

We study the origin of comovement in economic fluctuations across regions in India using a unique administrative dataset. Regional outputs comove significantly, which can be traced to a small number of large plants located in different regions. The top 10 plants from each region explain almost two-thirds of the total variation in comovement. We test three channels for comovement, viz. regional trade, granular plants, or granular multi-plant firms. Inter-region sales of multi-plant firms subsume the first two channels. The explanatory power of idiosyncratic shocks to these multi-plant firms is equivalent to that of aggregate shocks in explaining comovement.

Keywords: Granularity, comovement, multi-plant firms.

JEL codes: E32, E23, O47.

*We thank Thomas Chaney, Sampreet Singh Goraya, Basile Grassi, Jean Imbs, Krishnamurthy Subramanian and Fernando Vega-Redondo for helpful comments and discussions. We thank the conference and seminar participants at the Bocconi University, PSE, and the NASMES (2021) for their feedback and comments. Jalshayin Bhachech provided excellent research support. This research was partially supported by R&P institute grant from IIM Ahmedabad. All remaining errors are ours.

†Economics Area and Misra Centre for Financial Markets and Economy, Indian Institute of Management Ahmedabad, Gujarat 380015, INDIA. Email: anindyac@iima.ac.in

‡(Corresponding author) Economics and Public Policy, Indian School of Business, Hyderabad, Telangana 500032, INDIA. Email: shekhar_tomar@isb.edu

1 Introduction

What explains GDP comovement across regions? At least three major channels have been proposed in the literature, ranging from common aggregate shocks (Imbs (2004)) to trade linkages at the aggregate country (Kose and Yi (2006)) or firm level (Di Giovanni et al. (2018)) to the presence of multi-region firms (Kleinert et al. (2015)). However, exactly how individual plants or firms contribute to comovement is not well understood. This paper shows that a substantial part of the comovement can be traced to idiosyncratic shocks to a few large firms with plants in multiple regions. We label it as the granular origin of comovement.[1]

To set up the key point, we present a striking result here. The top ten plants across regions in our dataset explain nearly two-thirds of the total variation in the regional GDP comovement. Interestingly, a significant fraction of them belong to multi-plant firms present in multiple regions. In the paper, we quantify the contribution of these firms to overall comovement. The previous work, such as Di Giovanni et al. (2018), Kleinert et al. (2015), has studied the role of multi-national firms in GDP comovement across countries. However, the evidence in the prior work is based on firm-level data from a single country and matches individual firms' dynamics with the trading partner countries' GDP. Therefore, it does not capture the intensive margin effect of the multi-region firm channel. We ameliorate this problem by utilizing information on plant-level data from regions across India. This allows us to pin down the role of big plants and firms on both sides of the trading partners.

We use India, a large economy with 35 regions (Figure 1), as a laboratory for our analyses. We utilize two novel administrative datasets documenting the production side of the economy at the plant×region×month level and the trade side of the economy at the region×region×month level. Both datasets come from the Goods and Services Tax Network

[1]Our usage of the term *granularity* stems from Gabaix (2011), who emphasizes the disproportionate role of large firms in explaining aggregate fluctuations. Over the last decade, a large literature has emerged that emphasizes the role of heterogeneity in size and influence of firms and sectors in the production networks, in driving aggregate fluctuations (see Gabaix (2011), Acemoglu et al. (2012) and Baqaee and Farhi (2019)).

(GSTN), which records the sales of Indian plants from all regions. The trade dataset provides region-wise aggregate inter-region (across) and intra-region (within) sales in India. These data are available at a monthly frequency for two years (April 2018–March 2020) and give us the complete trade network across the regions. The plant-level dataset provides monthly sales of the top 1,000 plants from each region over the same period.[2] We observe intra-region and inter-region sales of each plant and the parent firm if multiple plants belong to the same firm.

We first establish the role of large plants in explaining regional comovement. As a measure of aggregate comovement, we use the principal component analysis to extract the first factor from the covariance matrix of regional GDP. The decomposition of the corresponding covariance matrix of regional GDP into principal components shows that the first component explains more than half of the variation, suggesting high comovement across regions in India. As its granular counterpart, we construct regional disaggregated GDP time series arising out of top k plants, based on sales, and call it granular GDP. We then extract factors summarizing the comovement of the regional granular GDP. We find that even for a small $k = 10$ (top ten plants), the first factor of granular GDP explains almost two-thirds of the variation in regional economic fluctuations. These top k plants' explanatory power is much higher than their percentage contribution to regional GDP. For instance, the top ten plants contribute, on average, 18 percent to the GDP, but their first factor explains around 64 percent of the comovement. As we increase k, the explanatory power increases monotonically with the top $k = 1,000$ plants, explaining almost all variation in comovement as these plants start to cover a significant proportion of GDP in each region. After establishing that the regional comovement can be traced back to correlations in the total sales of the top plants from each region, we unpack the underlying mechanism.

We test three main channels. As in the international business-cycle literature, aggregate

[2]Both datasets are based on the physical movement of goods collected through Electronic Way (E-Way) Bills, a regulatory requirement for shipment of goods. Therefore, we mainly capture sales in the manufacturing sector. We provide more details in Section 2.

inter-region trade is our first candidate for explaining comovement. Kose and Yi (2006) and Baxter and Kouparitsas (2005) document that an increase in international trade can lead to an increase in comovement across countries. In our case, aggregate inter-region trade can increase comovement across regions within the country. The second candidate is granularity in plant-level sales, i.e., a few big plants can explain this comovement. Here, comovement can occur through two different channels. The inter-region sales by granular plants can subsume the aggregate trade channel, as in Di Giovanni et al. (2018), and explain comovement. Simultaneously, comovement can arise from common aggregate shocks affecting all plants in all regions or as a spillover from inter-region sales of plants. The third candidate is multi-plant firms with plants located in different regions. We denote it as firm-level granularity as around 5 percent of the plants belong to these multi-plant firms, but their contribution to aggregate sales is disproportionately high, around 40 percent. This is analogous to the multi-national firm channel for international comovement described in Kleinert et al. (2015), and Di Giovanni et al. (2018).

We first test the importance of aggregate trade relative to plant-level granularity in driving regional comovement. We regress regional GDP correlation on aggregate inter-region trade and granular GDP correlations constructed from the top k plants. We include region fixed-effects to control for unobserved regional characteristics and account for aggregate India-level shocks contributing to GDP correlation, which addresses the critique in Imbs (2004). We find that regional trade drives correlations in regional GDP, as long as the granular GDP is constructed from intra-region plant sales. Once we include the granular GDP correlation based on inter-region plant sales in our specification, the aggregate trade effect becomes weaker and disappears for $k > 200$. Consistent with the literature on international comovement (Di Giovanni et al. (2018)), the impact of aggregate trade on regional comovement is subsumed by inter-region sales of large plants. However, Di Giovanni et al. (2018) use unilateral firm-level data from France and correlate it with the aggregate GDP of the trading partners. In our case, we observe plant sales from each region and can conclude that the inter-region sales of the largest plants from each region are sufficient to explain regional comovement.

Next, we detail the contribution of firm-level granularity, i.e., multi-plant firms with plants in more than one region, or multi-plant firms (henceforth), to comovement. Since we see plant-level sales for each plant of these firms, we can capture the strength of their connection in each region, unlike Kleinert et al. (2015) and Di Giovanni et al. (2018), who only observe the presence/absence of multi-national linkages. We find that the correlation of these multi-plant firms' inter-region sales significantly explains the aggregate comovement. Additionally, the effect of aggregate regional trade is completely subsumed under inter-region sales of multi-plant firms.

The above results suggest the role of both plant-level granularity (inter-region sales by large plants) and firm-level granularity (inter-region sales by large multi-plant firms) in driving comovement. To estimate their relative importance, we conduct a horse race between the two. Since the test for plant-level granularity includes all plants, both plant-level inter-region sales and multi-plant firm linkages can impact comovement. To separate these two channels, we construct a sample of firms that have one plant each (or single-plant firms). If plant-level granularity matters, then the granular GDP from single-plant firms should continue to explain comovement. However, the single-plant firms fail to significantly impact comovement once we include the correlation constructed from multi-plant firms' sales in our specification. This test leaves firm-level granularity, through multi-plant firms, as the main channel in explaining comovement.

It leads us to quantify the relative contribution of aggregate vs. idiosyncratic shocks to multi-plant firms in driving comovement. We construct a measure of idiosyncratic shocks to multi-plant firms by filtering out regional fluctuations common across all firms. This exercise gives us a series of residuals, that captures idiosyncratic shock component of multi-plant firms, for each region. We find that the correlation between these residuals continues to explain regional comovement. A one-standard-deviation increase in the correlation of residuals, leads to an increase of 25 percent in the average regional correlation. In comparison, a one-standard-deviation increase in the correlation of regional aggregate shocks leads to a 27 percent increase in the average regional correlation. These results suggest that both channels, idiosyncratic shocks to multi-plant firms and aggregate shocks, play an equivalent role in explaining aggregate comovement.[3] This significant explanatory power indicates the *granular* origin of regional comovement through multi-plant firms.

Our paper is related to the *trade-comovement puzzle* (Kose and Yi (2006)) and regional business cycle comovement (Owyang et al. (2005)). The *trade-comovement puzzle* is based on the seminal work by Frankel and Rose (1998), who show that countries that trade more experience a higher degree of synchronization in their business cycles. Multiple studies confirmed the observation. Notably, Baxter and Kouparitsas (2005) perform a cross-country study with over 100 countries and consider mechanisms including bilateral trade, total trade, industrial structure, export-import portfolios, factor endowments, and gravity variables. The mechanism of such synchronization is not very evident in these studies. In particular, Imbs (2004) proposes that common shocks across countries might drive higher trade as well as synchronized business cycles and lead to observations such as in Frankel and Rose (1998).

More recently, Di Giovanni et al. (2018) use microdata on French firms to show how firm-level trade and multi-national firms contribute to GDP comovement; however, the former channel is more dominant. Using similar data, Kleinert et al. (2015)) show that the presence of multi-national affiliates within a French region increases its GDP's comovement with the origin country. Since, we account for the intensive margin effect for multi-plant firms (equivalent to multi-national firms in the international context) in both regions, inter-region sales of these firms turn out to be highly significant in explaining comovement. The absence of intensive margin and bilateral firm-level data in the prior work has possibly led to underestimating the role of multi-national firms in comovement.

More broadly, our work connects to the literature that focuses on understanding the

[3] Di Giovanni et al. (2018) documents that one-third of the observed correlation across France and its trading partner countries can be attributed to multinational firms.

within-country regional comovement (e.g., Owyang et al. (2005), Hamilton and Owyang (2012), Beraja et al. (2019)) and sectoral business cycles (e.g., Foerster et al. (2011)). To the best of our knowledge, our paper is the first to quantify the role of multi-plant firms in explaining comovement across regions within a country. It complements the research that studies how within-firm linkages lead to spillover of shocks from one region to another (Giroud and Mueller (2019)). While we do not observe vertical linkages across plants of multi-plant firms, Garg et al. (2021) show a high prevalence of within-firm sourcing in India, suggesting some role for vertical linkages behind our results. Finally, since country-level fluctuations arise from the aggregation of regional fluctuations, a higher degree of comovement in the latter would increase the former. Given the role played by multi-plant firms in causing regional comovement, they can be an important contributor to generating country-level fluctuations as well.

The rest of the paper is arranged as follows. We provide a description of the data and GDP construction in section 2, followed by the framework to measure regional comovement and contribution of plant-level granularity in section 3. Section 4 provides the evidence for granular comovement. Section 5 describes the empirical methodology to conduct the horse race between various determinants of regional comovement and main results. Section 6 summarizes the paper and concludes.

2 Data

We utilize two unique administrative datasets over the time period of April 2018–March 2020, constructed from the Electronic-way (E-way) Bills information collected by the GSTN.[4] The first one contains aggregate regional trade flow information and the second one contains plant-level sales information. Since April 2018, under the Goods and Services Tax regime, plants are legally required to generate an E-way bill before transporting goods above INR

[4]The GSTN started collecting the data from April 2018. We do not use the data after March 2020 as economic activities were severely disrupted in India due to the COVID-19 pandemic.

50,000 (around USD 700). This allows the GSTN to collect information on the sales of goods in real time. Given that an E-way Bill is required only for physical movement of goods, our datasets exclude all sales of services. Our regional- and plant-level datasets are based on the aggregation of E-way Bills, as provided by GSTN.

Region-level data: This dataset provides sales information aggregated at the region-level across 35 states and union territories (region refers to both state and union territories hereafter; see figure 1).[5] It provides total sales from one region to the other (inter-region) as well as within-region (intra-region) sales. For each month, this dataset consists of a 35 × 35 matrix of region-to-region sales.[6]

Plant-level data: This dataset provides monthly sales information for the top 1,000 plants for each region. The sales consist of both intermediate goods and final goods, with the former contributing almost two-thirds of the total sales.[7] Each plant has a unique identifier at the region level and can be tracked over time, as long as it falls within the top 1,000 plants for that month in a given region. We also have a breakdown on how much each plant sells within its own region as well as outside it. This dataset has two features. First, for each region, the top 1,000 plants' sales data are reported while in some cases there are more than 1,000 plants. Also, the set of the top 1,000 plants can change every month. However, none of our analyses is affected as large plants consistently feature in the set of top 1000 plants. Second, we do not observe the destination of the plants' sales. For each plant, we know intra-region and total inter-region sales within India. We do not know to which region they are selling

[5]Some of the Indian states are comparable to countries in terms of total production, e.g. Maharashtra, Tamil Nadu and Uttar Pradesh have state GDP equivalent of the GDP of Singapore, Finland and New Zealand respectively. Union Territories (UTs) are administrative divisions in India. Unlike the states, the UTs are directly governed by the central government of India. For our purpose, they are equivalent to states as far as geographical boundaries of production are concerned. There are 28 states and 7 UTs in India as of August 2020. We exclude Lakshadweep from our analysis due to its negligible size.

[6]This large number of regions allow us to exploit the regional heterogeneity as opposed to countries with very limited regional dispersion. For instance, Basile et al. (2014) study the effects of firm heterogeneity on regional business cycles in the Italian data using only northern and southern geographical split.

[7]The ratio is quite similar in the context of global trade. In 2019, intermediate products worth of USD 8.3 trillion were sold as opposed to consumer products worth USD 4.8 trillion (UNCTAD, 2020).

the products, as out-of-region sales data is aggregated together into inter-region sales.

To summarize, there are two differences between the aggregate region and plant level data. First, the region-level dataset provides region-to-region flow, as opposed to the plant-level dataset, which provides only the origin but not the destination region. Second, the aggregate regional trade dataset is exhaustive in nature, as it gives the total domestic trade flow and productions as opposed to the sales of the top 1,000 plants from each region.

2.1 Ranking Plants at the Region Level

For the purpose of our analyses, we classify plants in two ways:

1. **Non-ranked plants:** As described above, we observe the sales of top 1,000 plants for each region×month pair. We arrange the plants in decreasing order of sales for each pair. We call them "non-ranked" plants since the identities of the plants belonging to the monthly set of top 1,000 plants vary over the 24-month period in each region. Thus this dataset picks up both the intensive (same plants are tracked if they belong to the top 1,000 plants over months) and extensive margin (there is exit and entry of plants from the top 1,000 plants).

2. **Ranked plants:** In this case, we fix the plants' ranks based on their total sales over the two-year period. If a plant does not belong to a given region×month's list of top 1,000 plants, we impute zero sales for it. This dataset specifically picks up the intensive margin of sales by construction.

We use the ranked plants for our main analyses and report the results with non-ranked plants for robustness.

2.2 Aggregate Sales as a Surrogate Measure of Regional GDP

As described above, our datasets are based on plant-level sales and do not provide information on the value added. Therefore, we do not have a measure of regional GDP in a strict national accounting sense. Instead, we construct a surrogate measure of GDP by adding the plant-level sales. It would clearly lead to double-counting as there would be pairs of plants that would supply inputs to others and, hence, total sales would be effectively a scaled-up version of the actual GDP. However, this construction imposes minimal restrictions on our analyses. First, in our case the nominal level of the GDP is not important. Our main hypothesis is about comovement of the regional GDP. Thus, even if the scaling factors are different across regions, it would not impact the measure of comovement. This follows from the fact that the correlation between variables x and y is the same as the correlation between variables $a.x$ and $b.y$, where a and b are constants. Second, sales is a consistent variable across both regional and plant level datasets at monthly frequency. Therefore, the two datasets can be compared without making any adjustments. Finally, there is a very strong correlation between the surrogate and official GDP (2018-19) for the cross-section of regions in India (Figure 2).

Now we describe the construction of aggregate and granular regional GDP. The surrogate measure of aggregate GDP series comes from the regional trade flow dataset by summing over the intra- and inter-region sales at the region$\times$month level. This gives us a 24 month GDP series for each of the 35 regions. The granular GDP series is based on the sales of top-k plants. For the k-th level granular GDP series, we sum over the sales of top k plants at the region-time level, where $k \in \{1, 2, 5, 10, 20, 50, 100, 200, 500, 1,000\}$. Clearly, as k approaches 1,000, we get closer to the aggregate GDP series created from the regional trade flow data.

In the main paper, we report results based on the above constructed GDP series. Additionally, we provide results with HP-filtered GDP series in the Appendix as robustness.[8] Now, we describe coverage and distributional properties of plant-level data.

[8] In our baseline specification, we do not use growth rates, as our data covers 24 months, implying that using year-on-year growth rate would leave us with only 12 data points for each region. This would make inference difficult. However, we have reported the results with year-on-year growth rates as well as with HP-filtered data to rule out any concerns that might arise due to seasonality. In all cases, the results are robust.

2.3 Granularity: Plant Size Distribution

We calculate the average fraction of sales in the total GDP that can be captured by the top k plants. We take an average over region×month, given by:

$$\text{Average Share}_k^s = \frac{1}{T} \times \sum_{m=1}^{T} \left(\frac{\sum_{r=1}^{N} \sum_{p=1}^{k} SalesGranular_{p,r,m}^s}{\sum_{r=1}^{N} SalesAgrregate_{r,m}} \right) \tag{1}$$

where p denotes the identity of the plant (total number of plants considered is $k \leq 10^3$), r denotes region (the total number of regions $N = 35$), m denotes the month (the total number of months $T = 24$), and $s \in \{\text{intra-region, inter-region, all sales}\}$ is the sales type. The variable in the numerator $SalesGranular_{p,r,m}^s$ denotes sales s of the p-th largest plant in the r-th region in the m-th month as obtained from the plant-level data. The variable in the denominator $SalesAggregate_{r,m}$ denotes the total regional sales in the r-th region in the m-th month (obtained from the regional data).

We report the average shares in Table 1. In the non-ranked case, we see that the top $k = 1$ plant across all regions explain 7 percent of inter-region sales (column (1)), but only 1 percent of intra-region sales (column (2)). For All Sales, sum of inter- and intra-region sales, the average share goes up to 9 percent (column (3)). If we increase k, i.e., the number of plants, then the average share increases. Finally, when we consider the largest sample $k = 1{,}000$, then the average share in the inter-region sales goes up to 41 percent and in intra-region sales the share goes up to 25 percent. For All Sales, this number is 75 percent. In the last three columns (4)–(6), we present the same statistics for the rank ordered plants. Since in this case the plant identities are fixed, the average shares would be lower than the non-ranked case. However, the shares are similar, suggesting that larger plants contribute disproportionately more to the aggregate region-level sales and do not suffer frequent entry/exit from our sample.[9]

Multi-plant Firms: Here we provide summary statistics of multi-plant firms which we

[9] The plant size distribution in our data follows Zipf's law (Axtell (2001), Gabaix (2009), Di Giovanni et al. (2011)) both at the national and regional level.

later show to cause regional comovement. Table 2 describes the distribution of number of plants per firm and their average size (in INR 1 Crore $\sim$ 133,548 USD). We find that 94 percent of the plants belong to single-plant firms in our sample and have the lowest average plant-size. In terms of contribution to total sales, multi-plant firms account for more than 40 percent although they are only 6 percent in count.

3 Framework: Regional Comovement

In this section, we first present a theoretical framework to show how plant-level sales are connected to regional comovement, followed by a non-parametric framework to quantify the degree of comovement. We use the latter to measure comovement based on aggregate and granular GDP as well as describe the connection between the two.

3.1 Correlation Decomposition: Top k Plants

Let there be two regions i and j with n_i and n_j plants respectively. The aggregate regional sales correlation, r_{ij}, is given by:

$$r_{ij} = Corr\left(\sum_{p_i=1}^{n_i} s_{i,p_i}, \sum_{p_j=1}^{n_j} s_{j,p_j}\right) \tag{2}$$

where $Corr$ is the correlation function and s_{i,p_i} is the sales of plant p_i in region i and similarly s_{j,p_j} is the sales of plant p_j in region j. Each term in the above equation also has a time subscript which we have suppressed to simplify notation. To derive the contribution of the top k plants to regional correlation, we divide the set of plants from each region into two groups, those that belong to the top k plants by sales and others that do not. We denote $p_i \in k$, if p_i belongs to the top k plants by sales in region i, else $p_i \in -k$, where k is a positive integer. We can now rewrite the above equation as:

$$r_{ij} = r\left(\sum_{p_i \in k} s_{i,p_i} + \sum_{p_i \in -k} s_{i,p_i}, \sum_{p_j \in k} s_{j,p_j} + \sum_{p_j \in -k} s_{j,p_j}\right)$$

which further gives:

$$r_{ij} = \frac{1}{\sigma_i \sigma_j}\left[\underbrace{Cov\left(\sum_{p_i \in k} s_{i,p_i}, \sum_{p_j \in k} s_{j,p_j}\right)}_{Top\ k\ Plants} + Cov\left(\sum_{p_i \in k} s_{i,p_i}, \sum_{p_j \in -k} s_{j,p_j}\right)\right.$$
$$\left. + Cov\left(\sum_{p_i \in -k} s_{i,p_i}, \sum_{p_j \in k} s_{j,p_j}\right) + Cov\left(\sum_{p_i \in -k} s_{i,p_i}, \sum_{p_j \in -k} s_{j,p_j}\right)\right]. \quad (3)$$

The first term on the right-hand side (RHS) corresponds to the contribution of top k plants from each region to the aggregate correlation between the regions i and j. This decomposition therefore captures how a subset of plants from each region can contribute to aggregate regional correlation. For our analysis, the first term on the RHS of Equation 3 describes the contribution of top k plants from each region to r_{ij}. Similae to total sales, we can generate the list of top k plants based on sales type, intra- and inter-region, and find their contribution to r_{ij}. Before unpacking the contribution of plants in overall comovement, we first describe how we measure average comovement across regions in our data.

3.2 Measuring Regional Comovement

To measure the average degree of comovement, we employ the principle component analysis (PCA). We use the surrogate GDP matrix of size $T \times N$ where T is the number of months and N is the number of regions. From this matrix, we create a correlation matrix of size $N \times N$ and conduct an eigendecomposition to generate N eigenvalues and their corresponding eigenvectors. Without loss of generalization, we sort the eigenvalues in decreasing order in terms of absolute values and label them λ_1, λ_2, ..., λ_N.

The explanatory power of P principal components (PC) is given by:

$$f_P = \frac{\sum_{j=1}^{P} \lambda_j}{\sum_{i=1}^{N} \lambda_i} \tag{4}$$

where f_P is the share explained by the first P components. For correlated regional GDP, the top principal components would explain a sufficiently large degree of the total variation. Alternatively, one can also calculate the average pairwise correlation across regions as a measure of comovement and we report it later in our analyses.

We define granular comovement as the comovement in granular GDP constructed from the sales of top k plants. For a given k, we get a matrix of the size $T \times N$. From this matrix, we create a correlation matrix to generate N number of eigenvalues and the corresponding eigenvectors. Once again, the explanatory power of P principal components is given by:

$$f_{Pk} = \frac{\sum_{j=1}^{P} \lambda_{jk}}{\sum_{i=1}^{N} \lambda_{ik}} \tag{5}$$

where the subscript k denotes the number of top plants accounted in construction of granular GDP.

3.3 Explaining Comovement through a Factor Model

We explore the relationship between the aggregate and granular GDP comovement using a regression model. As described above, we construct the GDP series arising out of top k plants for trade flow (Inter), domestic sales (Intra), and the sum of them (All). The constructed datasets, both aggregate and granular, have a $T \times N$ dimension, where $T = 24$ and $N = 35$. Given that $T < N$, we cannot use a standard factor model to extract the factors, as the covariance matrix does not have a full rank. Instead, we apply the NIPALS (Nonlinear

Iterative Partial Least Squares)[10] algorithm, which works even with $T < N$.

As an alternative to NIPALS, we can also run the analyses with simple unweighted average of the sales time series across regions. The idea is that instead of the weight chosen by the NIPALS algorithm we can use equal weights for each time series of sales, and see if the results hold. If they do not, then the results would be crucially dependent on the weights implied by the NIPALS algorithm. All the results that we report later hold for the case with equal weights and therefore our results are not dependent on our choice of NIPALS algorithm.

Let us denote the resulting dominant factor for the aggregate regional GDP by z^1_{aggr} which captures the common component across regional GDP. The superscript 1 in z^1_{aggr} denotes the utilization of only one dominant factor. We focus on the first dominant factor in our main results as it explains more than 40 percent of variation for each GDP series as shown in the Appendix Figure A.1. As a robustness check, we also report the case with two factors. Similarly, we construct $z^1_{granular,k}$, dominant factor in granular GDP, for each k. Once again, we have suppressed the time subscript for both z^1_{aggr} and $z^1_{granular,k}$ for simplifying our notation. We run the following regression model to measure the role of $z^1_{granular,k}$ in explaining z^1_{aggr}:

$$z^1_{aggr} = \phi + \rho z^1_{granular,k} + \xi_k. \tag{6}$$

The variable $z^1_{granular,k}$ is the the dominant factor calculated using k plants and for different types of sales in Inter, Intra, and All. We use the R^2 of this regression to quantify the degree of explanatory power of the first dominant factor created out of granular GDP, in explaining the first dominant factor created out of aggregate GDP. Intuitively it answers the question

[10]The NIPALS algorithm does not compute the covariance matrix for construction of the factors. Instead, it proceeds via an iterative approximation to the factors and their loadings via the power iteration method. For the present purpose, we have utilized the *chemometric* package in R programming (available here: https://cran.r-project.org/web/packages/chemometrics/; accessed on September, 2020) environment for calculating the factors. The results are identical in Matlab implementation of NIPALS. A textbook treatment of the NIPALS algortihm can be found in Vermuza and Filzomoser (2016).

of how much of the comovement in the regional economic fluctuations can be attributed to the comovement in the top k plants' sales.

4 Granular Buildup of Comovement

In this section, we provide the quantitative measure of contribution of plant-level granularity to regional comovement. We begin by providing the estimates of the latter.

4.1 High Comovement across Regions

To measure the comovement, we use Equation 4 which is based on the PCA. The results are reported in Table 3, using (i) $P = 1$ and (ii) $P = 2$ for robustness, which explains two-thirds of the total variation (Figure A.1). We also present the results for each of the sales component, across region (Inter), domestic (Intra), and their sum (All) sales. We find that the first PC alone explains more than 50 percent of the fluctuations in region-level sales for all three types of sales. In case of inter-region sales (column (1)), the first PC explains 53 percent, while it jumps to 74 percent in the case of intra-region sales. These results suggest that intra-region sales across regions display a higher degree of comovement relative to inter-region sales. If we use two PCs (last row), the explanatory power in Intra, Inter, and All sales jump to 66 percent, 90 percent, and 71 percent, respectively.

These results suggest a high degree of regional comovement and it exists at all levels: within regions (Intra), across regions (Inter), and the aggregate (All).[11] Since Table 3 uses raw sales, it might raise a concern that seasonality in sales can contribute to high degree of comovement. However, the results do not change even with HP-filtered data. These results

[11]The official regional GDP information is available at an yearly frequency as opposed to the baseline dataset. As a check for consistency, we collect the official yearly data available from 2011-2020 across regions within India. Ignoring missing data for some regions, we obtained complete data on 32 regions. Applying PCA on the yearly data, we see that the first principal component explains 69 percent of the total variation. With the baseline monthly data, the first principal component explains around 53 percent of the total variation. Therefore, our dataset is representative of the comovement seen in the regional GDP, and in fact, our estimates provide a lower bound for the level of comovement.

are reported in Appendix Table A.1, and we consistently find that the first (or first two) PC is able to explain more than 50 percent of the variation in the regional GDP. We next quantify the level of comovement in granular GDP.

4.2 High Comovement in Granular GDP

In Table 4, we present the share of region-wise granular GDP (constructed from top k plants) explained by the first principle component f_{1k}, as described in Equation 5. We report results for Inter, Intra, and All sales by non-ranked (columns (1)–(3)) and ranked plants (columns (4)–(6)). In general, the first principle component has a sizeable explanatory power for all k-s. Even for the top 1 plant, i.e., for $k = 1$, we see that the first eigenmode explains 29 percent of the aggregate comovement in terms of inter-region sales (column (1)). For intra-region sales (column (2)), the corresponding value is 55 percent, while for all sales this number is 28 percent (column (3)). The explanatory power goes up as we increase k. If we take all 1,000 plants, then the explanatory power goes beyond 50 percent with only one eigenmode. These findings indicate very strong comovement in granular GDP. The results remain unchanged when we use rank ordered of plants, as shown in columns (4)–(6).

To benchmark the degree of comovement quantitatively, we compare the above results with the contribution of plants to regional GDP, provided in Table 1. If we focus on column (1), the top 1 plant on average contributes 7 percent to inter-region sales. However, a single PC based on $k = 1$ is able to explain 29 percent of the regional comovement. In the case of intra-region sales, the contribution of the top 1 plant to sales is only 1 percent, but 55 percent of their comovement is explained by a single PC. The results are similar for all sales in column (3). With an increase in k, both contribution of k plants to regional sales (Table 1) and the comovement across them rise (Table 4). At $k = 20$, the top 20 plants cover 28 percent of sales, but the first PC is able to explain 46 percent of the movement in granular GDP (column (3)) whereas for $k = 1{,}000$, the first PC explains 54 percent (sales share is 59 percent).

These results suggest a granular buildup of comovement. For instance, the degree of comovement for $k = 20$ is much larger than the contribution of these plants to the regional GDP. With an increase in k, both the numbers rise but the comovement plateaus around $k = 20$. Each subsequent set of plants contributes more to the sales (Table 1) but does not lead to an equivalent increase in comovement. Therefore, we can surmise that comovement in sales across the largest plants across regions is driving the aggregate regional comovement. We test this hypothesis formally in the next part.

4.3 Granular Buildup of Regional Comovement

The first key result of our paper is presented in Figure 3, where we show how the common component arising from the top k plants across regions, $z^1_{granular,k}$, is able to explain a significant fraction of common component in the aggregate regional GDP, $z^1_{granular,k}$. The x-axis gives the number of plants k, while the y-axis gives the R^2 based on regression Equation 6 for the corresponding k. For $k = 1$, the R^2 is more than 30 percent, indicating that the factor constructed from the total sales of the top 1 plant across regions explain more than 30 percent variation in the first factor constructed from aggregate regional GDP.[12] The explanatory power, R^2, rises fast as we increase the number of plants k. The R^2 reaches 55 percent with only the top 5 plants. In contrast, these top 5 plants on average contribute 14 percent to the total sales. Therefore, the top plants' contribution to comovement is much larger in proportion to their size. At $k = 50$, the R^2 goes to 80 percent and gradually approaches 100 percent as we include the top 1,000 plants.

To further establish the granular buildup of comovement, we show that the largest plants are fundamentally different in capturing comovement than a randomly selected set of plants from each region. We compare the top $k = 10$ plants against sets of 10 plants randomly chosen from those with a rank $k > 100$.[13] We perform a similar analysis as in Figure 3,

[12]These estimates can also be used to explain the overall variation in the regional GDP series. From Table 3 we know that the first PC of regional GDP explains 53 percent variation in aggregate GDP. Therefore, the first PC based on $k = 1$ explains 17 percent (33×0.53) of the total variation in those series.

[13]The rank $k > 100$ is chosen to exclude the largest plants from each region. For simulation, we draw the

i.e., calculate granular GDP from the set of randomly chosen 10 plants and construct their granular comovement by extracting the first factor, $z^1_{granular,k}$. We then test their explanatory power in explaining aggregate comovement by calculating R^2 from equation 6. We simulate 1,000 draws and calculate the corresponding R^2 values. The density plot for R^2 based on these simulations is reported in Figure 4. On average, the set of randomly chosen 10 plants captures the comovement much less relative to the top $k = 10$ plants. The mean R^2 for simulated cases is only 22 percent, while it is 64 percent for the top $k = 10$ plants. In fact, the R^2 from simulated draws are larger than the R^2 from the top $k = 10$ plants only in less than 7.5 percent of the cases.

Robustness: We also report R^2 for each sales component separately in Appendix Table A.2. The dependent variable z^1_{aggr} in this table is still based on aggregate region-level sales, while the independent variables are based on various aggregation (Intra, Inter, and All sales) of plant-level sales. We continue to find similar results. We get similar results when we use the first two factors constructed from granular GDP as explanatory variables for aggregate comovement. The R^2 jumps to more than 80 percent with just $k = 1$ in the all sales case (Appendix Table A.3). Similarly, the results are robust to using HP-filtered series instead of the raw sales. The corresponding results are provided in Table A.4 and they only get stronger with HP-filtered data. Finally, we also check if the above results are robust to using equal weighted case, instead of NIPALS algorithm, to construct the dominant factor. All results continue to hold (results available on request).

Overall, these results demonstrate granular origin of comovement. Next, we explore the channels that explain this phenomenon.

sample of 10 plants from the remaining plants with equal probability.

5 What Matters for Comovement?

In this section, we conduct a horse-race to find the dominant channel behind regional comovement. It might arise due to aggregate inter-region trade, correlated granular plants, or correlated granular firms (having plants in multiple regions).

5.1 Empirical Strategy

Our main dependent variable is the correlation coefficient r_{ij} between aggregate GDP of regions i and j. We estimate its dependence on aggregate trade and plant-level granularity through the following regression specification:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_2 r^k_{ij,s} + \epsilon_{ij} \tag{7}$$

where $trade_{ij}$ is the log of the average trade between the regions i and j over 24 months. If trade increases regional comovement, then the coefficient β_1 would be positive and significant. The second-term $r^k_{ij,s}$ captures the correlation between granular GDP constructed from k plants from regions i and j and sales type $s \in \{$intra-region, inter-region, all sales$\}$. We use decomposition of sales type s to gauge the relative importance of intra- vs. inter-region sales in explaining aggregate regional comovement. If the correlation across plants drives aggregate regional comovement, then β_2 will be positive and significant. The summary statistics for these variables are presented in Table 5. We estimate the above equation using OLS where each observation is constructed at the region-pair level. We report robust standard errors for each regression.

Similarly, we test how much of the comovement is driven by multi-plant firms with plants located in more than one region. As mentioned earlier, 5 percent of all plants belong to multi-plant firms (Table 2). We update the above equation to test this hypothesis:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_3 r^{mf}_{ij,s} + \epsilon_{ij} \tag{8}$$

where the term $r^{mf}_{ij,s}$ corresponds to the correlation in the total sales of plants that belong to multi-plant firms, mf. If the multi-region multi-plant firms matter for comovement, then β_3 will be positive and significant. The construction of $r^{mf}_{ij,s}$ as a measure of affiliates of a multi-plant firm is different relative to what Di Giovanni et al. (2018) and Kleinert et al. (2015) use for multi-national affiliates. While the former uses a binary indicator for multi-national affiliates, the latter uses the share of employment contributed by the multi-national firms in a region. In contrast, we can observe firm identities in both regions i and j, as opposed to observing them for plants in one region. Since we observe sales of each of the plants of these multi-plant firms, we can construct a normalized measure of the correlation of total sales contributed by these firms across regions. Once again, we can disaggregate these sales into $s \in \{\text{intra-region, inter-region, all sales}\}$ and construct three different measures of $r^{mf}_{ij,s}$.

We estimate the above equations and describe our results in the next subsections.

5.2 Comovement: Trade vs. Granular Plants

We estimate Equation 7 for different s and report the results in Table 6. We first test the relative importance of aggregate trade vs. sales of granular plants in explaining comovement. Each column corresponds to the granular plants, k used for constructing the granular correlation $r^k_{ij,s}$. In Panel (a), we study the role of intra-region sales. We find that both trade and intra-region sales are significant in all columns. The coefficient on trade in each column is around 0.023 and significant, i.e., more trade across regions is associated with a higher degree of comovement. A one-standard-deviation change in trade leads to an increase in correlation by 0.7 (or 25 percent over the average). The coefficient on $r^k_{ij,s}$ is also positive and significant for all k. It increases sharply as we go from $k = 1$ to $k = 10$ and then remains around 0.3. Therefore, having more correlated intra-region sales of granular plants is associated with higher aggregate comovement across regions. For $k = 10$, a one-standard-deviation change in plant-level correlation is associated with an almost one-third increase over the mean region-level correlation.

Next, we report results based on inter-region sales in Panel (b). In this case, the results are in sharp contrast to those found with intra-region sales. For small values of k, the coefficients on trade and $r^k_{ij,s}$ are both positive and significant. However, as we include more plants in calculating granular correlation, $r^k_{ij,s}$, the coefficient on trade starts to decline and becomes insignificant as k approaches 200. At the same time, the coefficient on $r^k_{ij,s}$ becomes larger with increase in k. Since $r^k_{ij,s}$ is based on the inter-region sales of granular plants, these results show that the trade channel is active through these large plants. As one includes more plants, i.e., k approaches 1000, the entire effect of aggregate regional trade is subsumed within the term driven by granular plants.

These results become clearer when juxtaposed against those obtained using intra-region sales. For intra-region sales, the coefficient on trade continues to remain positive and significant for all k. Since the intra-region sales do not subsume the across-region aggregate trade, the coefficients on both remain significant, unlike in the case of inter-region sales. These results are in line with Di Giovanni et al. (2018), who suggest that trade by big firms leads to higher comovement across countries. In our case, inter-region trade by big plants drives comovement through trade.

Next, we include correlation of all sales, sum of inter- and intra-region sales, in our regression, and the results are reported in Panel (c). Here, the results are similar to inter-region sales. For low k, both trade and $r^k_{ij,s}$ have positive and significant coefficients. However, the coefficient on trade becomes insignificant as soon as $k = 20$. To summarize, the comovement in sales of large plants show a high degree of correlation with the aggregate regional sales comovement.

5.3 Comovement: Trade vs. Granular Firms

In this part, we test the role of granular firms, i.e., firms with plants in at least two regions. We estimate the equation 8 and report the results in Table 7.

We first look at column (1) with results based on intra-region sales of multi-plant firms.

r_{ij}^{mf} is based on the correlation between the intra-region sales of these firms. In this case, the coefficients on trade and intra-region sales of big firms are positive and significant. Therefore, both trade and within-region sales of these firms are associated with higher comovement. In column (2), we introduce the correlation of multi-plant firms based on their inter-region sales. The coefficient on inter-region sales of multi-plant firms remains positive and significant, while the coefficient on trade becomes insignificant. Similar to previous results on plants, the importance of regional trade diminishes as soon as we include inter-region sales of multi-plant firms.

The inter-region sales in case of multi-plant firms also include within-firm sales and point to supply-chain connection between their plants. We demonstrate by negation that the inter-region sales of multi-plant firms are crucial in capturing regional comovement. We look at inter-region sales of single-plant firms, i.e., those with plants in only one region.[14] We denote these firms by $-mf$ and their correlation by $r_{ij,s}^{-mf}$. These single-plant firms, by definition, exclude any supply-chain connections as they are not part of multi-plant firms. Therefore, if the within-firm supply chain channel of multi-plant firms is not crucial, then the regression with trade and inter-region sales of single-plant firms should once again lead to trade channel getting subsumed under single-plant firms.

We re-estimate our regressions with $r_{ij,s}^{-mf}$ as an independent variable and report the results in Table 7 (columns (3) and (4)). We find that the coefficients on both trade and $r_{ij,s}^{-mf}$ are positive and significant in all cases. Therefore, the granularity of plants explains aggregate comovement but so does trade. However, the regression to test the above-mentioned hypothesis corresponds to the regression with inter-region sales of single-plant firms (column (4)). For these firms, aggregate regional trade continues to remain significant. In contrast, the coefficient on trade in column (2) is insignificant. A possible explanation is that the supply chain linkages of multi-plant firms completely subsume the aggregate trade effect. Together, these results suggest that the inter-region sales of multi-plant firms is a crucial

[14] While some firms can have more than one plant in one region, the fraction of such firms is very minimal. Hence, we call the complementary set of mf firms as single-plant firms.

factor behind the regional comovement.

While we do not have exact information on within-firm supply chains, Garg et al. (2021) show that within-firm supply chain linkages account for a large share of the total sales of multi-plant firms in Karnataka, one of the 35 regions in India. In fact, they show that around 36.8 percent of multi-plant firms source all their inputs from within the firm. Therefore, a significant fraction of the multi-plant firms' effect in our case would possibly be on account of within-firm supply chain linkages. However, such linkages among plants of multi-plant firms could be a feature of Indian economy. Atalay et al. (2014) shows that such within-firm linkages across plants are smaller in the case of the US.

5.4 Comovement: Trade vs. Granular Plants vs. Granular Firms

Lastly, we jointly estimate the impact of aggregate trade, sales of multi-plant firms and single-plant firms on comovement, i.e., we include $trade_{ij}$, r_{ij}^{mf}, and r_{ij}^{-mf} as explanatory variables in the equation 8. These results are shown in columns (5) and (6) of Table 7.

In column (5), we evaluate the correlation of intra-region sales as the dependent variable. As expected, the coefficient on trade is significant. The coefficient on the correlation of sales of multi-plant firms, r_{ij}^{mf}, is also positive and significant. Since these are multi-plant firms, it is possible that their within-region sales are also affected due to interconnection with other plants of the same firm. Giroud and Mueller (2019) document how within-firm linkages can lead to spillover of local economy shocks even for firms dealing in non-tradable sectors. Since our firms are mostly in tradable sectors, this channel is bound to be stronger in our case.

We report the results with inter-region sales in column (6). Here, only the correlation of sales of multi-plant firms remains positive and significant. Once again, the coefficient on trade is insignificant, as is the coefficient on inter-region sales of single-plant firms. Therefore, the supply-chain linkage between multi-plant firms, or firm-level granularity, is the primary channel driving regional comovement. These results also highlight how, in the absence of data on inter-region sales of multi-plant firms, aggregate trade between regions can indirectly explain regional comovement. At the same time, our results also show how information on the few largest multi-plant firms is sufficient to capture regional comovement, justifying the choice of denoting multi-plant firms as granular firms.

5.5 Robustness

Incorporating Region Fixed-effects: Our main results using Equations 7 and 8 do not control for unobserved regional characteristics that might influence regional comovement. For instance, aggregate shocks at all-India level can impact all regions and lead to high co-movement. Similarly, other non-time-varying regional factors can lead to high comovement. Therefore, we re-estimate the above equations to control for such biases at the region level by including region fixed-effects.[15] We cannot include region-pair fixed-effects in our regressions as the correlation, r_{ij}, is available at the cross-section level. The first set of results with this modification are reported in Tables A.5. Once we include region fixed-effects, the coefficient on $trade_{ij}$ becomes insignificant for all cases. It is not surprising since average trade between regions is relatively constant over our sample period and its effect is completely captured through region fixed-effects. However, the coefficient on the correlation across the top plants continues to remain positive and significant. It is similar to our main results from the previous subsections. For the horse race among aggregate trade, granular plants, and granular firms, the results are reported in Table A.6. Once again, the results are similar to those reported in the previous sub-section. The coefficient on the correlation in the sales of multi-plant firms remains positive and significant in all cases. Most importantly, the coefficient on aggregate trade as well as on correlation of single-plant firms is insignificant in columns (5) and (6). These results are stronger than those without region fixed-effects and continue to support that multi-plant firms are the main factor explaining correlation in regional comovement.

[15] This specification allows us to remove the effect of aggregate shocks in the final regression. An alternate approach, as in Di Giovanni et al. (2018) and Gabaix (2011), filters out aggregate shocks at the firm level and then performs aggregation.

De-trended Series: It is possible that our results are influenced by our choice of constructing the main dependent and independent variables. We therefore check if our results are robust to using HP-filtered data and year-on-year growth. The results with HP-filtered data are reported in Appendix Tables A.7 and A.8. The results with growth rates are reported in Appendix Tables A.9 and A.10. Our main results go through in each case.

5.6 Comovement and Idiosyncratic Shocks to Granular Firms

The above results establish that the granular multi-plant multi-region firms strongly influence regional comovement. However, this phenomenon can be driven by idiosyncratic shocks to these firms, and/or by aggregate shocks impacting all such firms in all regions. Given the granularity of multi-plant firms, it is likely that their idiosyncratic shocks would not fully cancel out upon aggregation (Gabaix (2011)).

To quantify the role of idiosyncratic shocks to multi-plant firms in driving comovement, we first construct a measure of shocks to these firms at the regional level sans any aggregate component. We regress granular GDP of multi-plant firms on granular GDP of single-plant firms for each region. From each regression, we extract residuals giving us a total of 35 series. By construction, each residual series filters out any common shocks affecting single- and multi-plant firms alike, as well as any spillovers across them at the regional level. In addition, such filtering also accounts for India-level aggregate shocks. As a result, the residuals would be free of demand shocks at the regional and all-India level. Finally, we calculate the correlation between residual series of region i and j that gives us the measure of correlation in multi-plant firm sales arising purely from their idiosyncratic shocks. Its impact on aggregate regional comovement is captured through the following regression:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_4 r_{ij}^{resid} + \epsilon_{ij}. \tag{9}$$

The right hand side variables include aggregate inter-region trade and pairwise correlations across the residuals, r_{ij}^{resid} with β_4 being our main coefficient of interest. We estimate this model and report the results in Table 8. Column (1) shows that the coefficient on the correlation of residuals is positive and significant. It shows that the correlation in the idiosyncratic component of the multi-plant firms' sales across regions increases regional co-movement. These results are robust to including region fixed-effects (Column (2)). In terms of magnitude, a one-standard-deviation increase in the correlation of the residuals leads to around 27 percent increase in regional GDP correlation over the mean value (based on β_4 value in column (2)).[16]

Next, we quantify the relative importance of regional aggregate shocks vis-á-vis shocks to the granular firms in driving regional comovement by including r_{ij}^{-mf} in the regression model:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_4 r_{ij}^{resid} + \beta_5 r_{ij}^{-mf} + \epsilon_{ij}. \tag{10}$$

The coefficient β_4 on r_{ij}^{resid} captures the effect of the shocks to granular firms on the regional comovement, whereas the coefficient β_5 on r_{ij}^{-mf} captures the effect of the aggregate shocks on regional comovement. The term r_{ij}^{-mf} denotes the correlation between aggregate sales of single-plant firms across regions i and j. It captures the correlation on account of aggregate shocks (both India-level and regional) as well as idiosyncratic shocks to single plant firms. For ease of exposition, we denote the effect through the term r_{ij}^{-mf} as aggregate shocks. Column (3) in Table 8 shows the estimates of the model with $trade_{ij}$ and r_{ij}^{-mf} as explanatory variables. We find that the coefficient on r_{ij}^{-mf} is positive and significant, showing that correlation in sales of single-plant firms (or aggregate shocks) explain regional comovement.

Column (4) provides the estimates for the full model presented in Equation 10. We find that both β_4 and β_5 are positive and significant, showing that both idiosyncratic shocks to multi-plant firms and aggregate shocks matter for comovement. The inclusion of region fixed-effects in the model (Column (5)) gives similar results. In terms of magnitude, a one-

[16]The effect is in the order of $(0.20 \times 0.38)/0.28 \approx 27$ percent, where 0.28 is the mean of the regional correlations.

standard-deviation increase in the correlation of the residuals, leads to 25 percent increase in regional GDP correlation over the mean value (based on β_4 value in column (5)). Similarly, a one-standard-deviation increase in the correlation of the regional aggregate shocks leads to 27 percent increase in regional GDP correlation over the mean value (based on β_5 value in column (5)).[17] These results show that the role of idiosyncratic shocks to multi-plant firms is equivalent to that of aggregate shocks in explaining regional comovement.

Robustness: In the above results, the construction of residuals does not directly control for the role of India-level aggregate shocks, except what is implicitly captured through the granular GDP of single-plant firms. For robustness, we construct alternate residual series by filtering out the India-level aggregate shock component directly in the residual construction stage. We do this by including the first dominant factor from NIPALS algorithm (as in Section 3) or aggregate India Sales of all plants as an additional variable in the regression. Using these alternate residual series, we then re-estimate the most saturated specifications of Equation 10 and report the results in Table A.11. All results go through and idiosyncratic shocks to multi-plant firms continue to play a significant role in explaining comovement.

6 Conclusion

In this paper, we explore the origin of the regional comovement. We exploit a unique administrative dataset that provides plant-level sales with region- and time-level variations in India. We find that regional economic fluctuations exhibit high levels of comovement. We quantify the effects of three sources of comovement, viz. aggregate trade, plant-level granularity, and firm-level (multi-region, multi-plant) granularity. While all three effects are present with different intensities, granular firms emerge as the dominant channel for comovement. In fact, idiosyncratic shocks to these firms not only matter for explaining regional comovement, they have explanatory power similar to the aggregate shocks. This

[17] For β_4, the effect is in the order of $(0.19 \times 0.38)/0.28 \approx 25\%$ and for β_5, the effect is in the order of $(0.28 \times 0.27)/0.28 \approx 27\%$, where 0.28 is the mean of the regional correlations.

points toward the *granular* origin of comovement, similar to the idea proposed by Gabaix (2011) on the role of large firms to explain aggregate fluctuations.

This paper extends the literature on granular nature of economic fluctuations by identifying the contribution coming from intensive margin of multi-region, multi-plant firms. Lack of bilateral data on such firms has led to underestimating their role in explaining comovement. Future work may further explore the role of vertical, financial, or managerial linkages across plants belonging to such firms in driving comovement. Beyond the scope of this paper, our proposed framework utilizing bilateral firm-level data, can also be used to quantify the role of multinational firms in driving international comovement.

References

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

Atalay, E., Hortaçsu, A., and Syverson, C. (2014). Vertical integration and input flows. *American Economic Review*, 104(4):1120–1148.

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

Baqaee, D. R. and Farhi, E. (2019). The macroeconomic impact of microeconomic shocks: beyond hulten's theorem. *Econometrica*, 87(4):1155–1203.

Basile, R., de Nardis, S., and Pappalardo, C. (2014). Firm heterogeneity and regional business cycles differentials. *Journal of Economic Geography*, 14(6):1087–1115.

Baxter, M. and Kouparitsas, M. A. (2005). Determinants of business cycle comovement: a robust analysis. *Journal of Monetary Economics*, 52(1):113–157.

Beraja, M., Hurst, E., and Ospina, J. (2019). The aggregate implications of regional business cycles. *Econometrica*, 87(6):1789–1833.

Di Giovanni, J., Levchenko, A. A., and Mejean, I. (2018). The micro origins of international business-cycle comovement. *American Economic Review*, 108(1):82–108.

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.

Foerster, A. T., Sarte, P.-D. G., and Watson, M. W. (2011). Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy*, 119(1):1–38.

Frankel, J. A. and Rose, A. K. (1998). The endogeneity of the optimum currency area criteria. *The Economic Journal*, 108(449):1009–1025.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294.

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

Garg, S., Tan, B., and Ghosh, P. (2021). Within firm supply chains: Evidence from india. *Working paper*.

Giroud, X. and Mueller, H. M. (2019). Firms' internal networks and local economic shocks. *American Economic Review*, 109(10):3617–49.

Hamilton, J. D. and Owyang, M. T. (2012). The propagation of regional recessions. *Review of Economics and Statistics*, 94(4):935–947.

Imbs, J. (2004). Trade, finance, specialization, and synchronization. *Review of Economics and Statistics*, 86(3):723–734.

Kleinert, J., Martin, J., and Toubal, F. (2015). The few leading the many: Foreign affiliates and business cycle comovement. *American Economic Journal: Macroeconomics*, 7(4):134–59.

Kose, M. A. and Yi, K.-M. (2006). Can the standard international business cycle model explain the relation between trade and comovement? *Journal of International Economics*, 68(2):267–295.

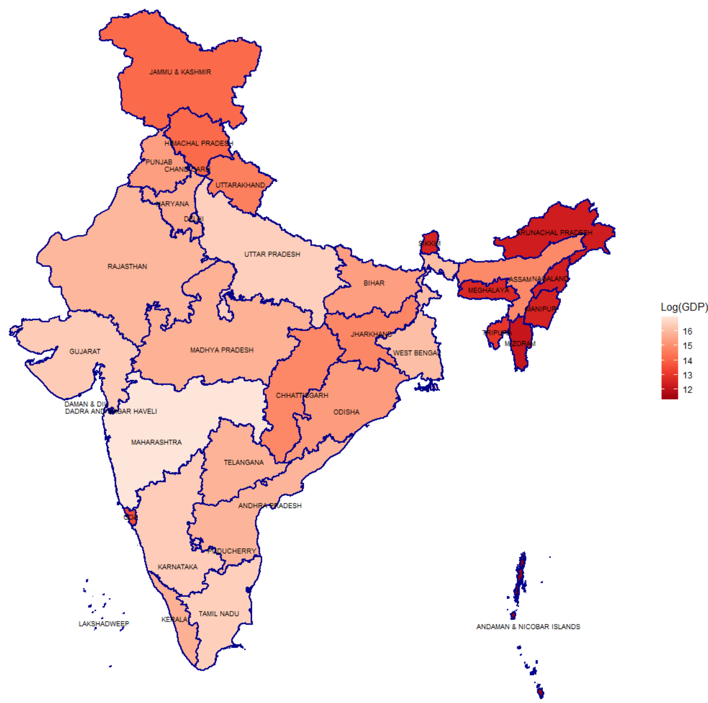
Owyang, M. T., Piger, J., and Wall, H. J. (2005). Business cycle phases in us states. *Review of Economics and Statistics*, 87(4):604–616.

UNCTAD (2020). Key statistics and trends in international trade 2019. *United Nations Conference on Trade and Development*.

Vermuza, K. and Filzomoser, P. (2016). *Introduction to multivariate statistical analysis in chemometrics*. CRC Press.

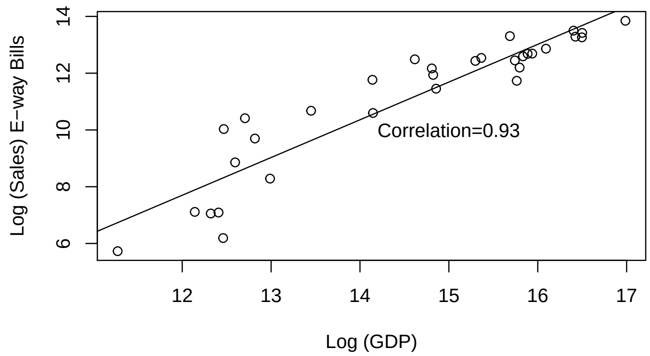
Figures and Tables

Figure 1: GDP of Regions in India (2018-19)



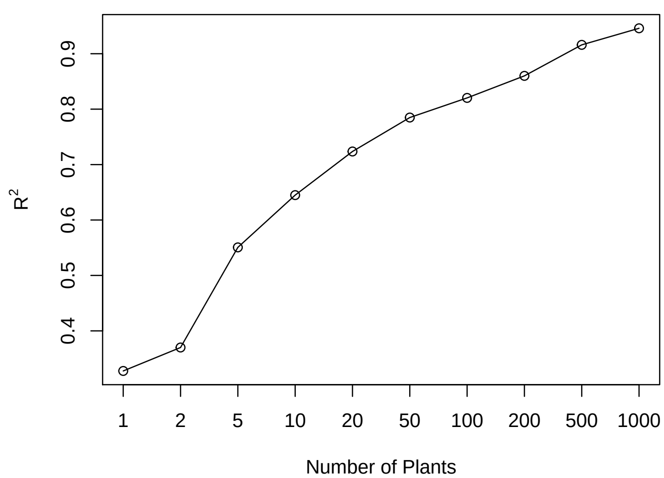
Notes: The above map gives the GDP (in INR millions) of the states and union territories of India (described as regions in the text) for the financial year April 2018-March 2019.

Figure 2: Comparison: Official vs. Surrogate GDP



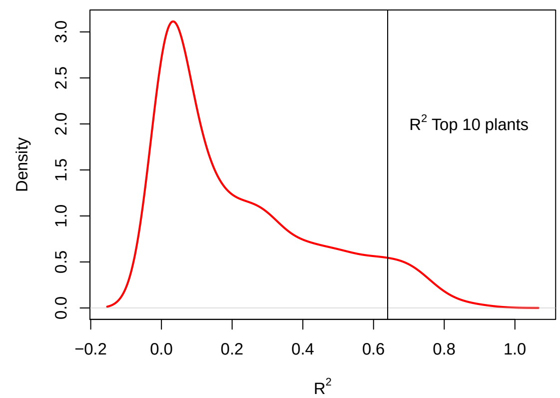
Notes: The figure plots the log values of surrogate GDP (inter- plus intra-region sales) constructed from the E-Way Bills data for the financial year April 2018–March 2019 against the official regional GDP. Each dot on the plot corresponds to a region. A very high value of the correlation coefficient indicates that the surrogate GDP is quantitatively a good estimate of the official GDP. Data Source for Official Regional GDP: Central Statistics Office, India.

Figure 3: Granular Buildup: Explanatory Power of All Sales of Top k Plants



Notes: The figure gives the R^2 of the regression Equation 6 for ranked top k plants based on all sales, i.e., Inter- plus Intra-region sales (last column in Table A.2). It gives the explanatory power of the first factor constructed from top k plants across regions over the first factor based on aggregate region-level sales. The first factor of top 10 plants across all regions explain about two-thirds of the total variation of the first factor of regional GDP.

Figure 4: Density Plot: Explanatory Power of All Sales of 10 Randomly Chosen Plants



Notes: The figure gives the density plot of R^2 based on $1,000$ simulations. It is calculated for the regression Equation 6 where the independent variable, $z^1_{granular,k}$ is constructed for a set of 10 randomly chosen plants from the population of plants with rank $k >$ 100. It gives the explanatory power of the first factor constructed from the total sales of 10 plants across regions over the first factor based on aggregate region-level sales. The vertical line corresponds to the R^2 value in the case of granular GDP constructed from the top $k = 10$ plants by sales.

Table 1: Average Shares of Top k Plants as Fractions of Total Sales

	Non-Ranked			Ranked		
	Inter (1)	Intra (2)	All (3)	Inter (4)	Intra (5)	All (6)
$k = 1$	0.07	0.01	0.09	0.07	0.01	0.08
$k = 2$	0.09	0.02	0.12	0.09	0.01	0.10
$k = 5$	0.12	0.03	0.18	0.12	0.03	0.14
$k = 10$	0.15	0.04	0.23	0.15	0.04	0.18
$k = 20$	0.18	0.06	0.28	0.18	0.06	0.22
$k = 50$	0.23	0.09	0.38	0.23	0.09	0.29
$k = 100$	0.27	0.12	0.46	0.27	0.11	0.35
$k = 200$	0.32	0.16	0.55	0.31	0.15	0.42
$k = 500$	0.37	0.21	0.67	0.36	0.19	0.52
$k = 1{,}000$	0.41	0.25	0.75	0.39	0.23	0.59

Notes: The table gives the average share of top k plants over the period April 2018 to March 2020 in terms of inter-regional sales (Inter), intra-regional sales (Intra), and the sum of them (All) based on Equation 1. The plant identities are not fixed in the non-ranked cases presented in columns (1)–(3), and they are fixed in columns (4)–(6) based on total sales for given plants over 24 months. The top 1,000 plants in each region explain more than 40% of total trade flow and 25% of domestic sales and, when combined together, explain 75% of the total sales (last row).

Table 2: Number of Plants belonging to the Firms

Number of plants per firm	Frequency	Percentage	Average size (sd)
1	111,103	94.06	23.28 (142.33)
2	4,339	3.67	68.56 (366.18)
3	1,037	0.9	109.83 (364.89)
4	451	0.4	146.03 (456.93)
5	245	0.2	268.93 (1165.34)
6	183	0.1	132.79 (190.70)
7	126	0.1	162.09 (347.71)
>7	633	0.5	-

Notes: Distribution of firms with a given number of plants (range: 1-7 and beyond; max. 51). Most of the plants ($\sim$ 94%) correspond to single-plant firms. The last column give the average sizes for each cohort of multi-plant firms in Crores (INR 1 Crore $\sim$ 133,548.00 USD). The multi-region multi-plant firms, i.e. firms with plants in at least two regions, account for 5 percent of the plants.

Table 3: Explanatory Power of Common Components in Regional Fluctuations (Share)

	Inter	Intra	All
	(1)	(2)	(3)
One Principal Component	0.53	0.74	0.53
Two Principal Components	0.66	0.90	0.71

Notes: This table gives the average share of comovement of the regional sales that can be explained by the common components. The exercise is performed for inter-regional sales (column (1)), intra-regional sales (column (2)), and all sales (column (3)). The share of common component is calculated by using Equation 4 and is based on eigen-decomposition of regional sales. The above table reports the results based on raw sales.

Table 4: Comovement in Granular Components of Regional Fluctuations

	Non-Ranked			Ranked		
	Inter (1)	Intra (2)	All (3)	Inter (4)	Intra (5)	All (6)
$k = 1$	0.29	0.55	0.28	0.28	0.60	0.29
$k = 2$	0.34	0.57	0.32	0.34	0.65	0.33
$k = 5$	0.40	0.61	0.37	0.38	0.69	0.38
$k = 10$	0.43	0.63	0.42	0.42	0.67	0.40
$k = 20$	0.46	0.66	0.46	0.46	0.69	0.46
$k = 50$	0.48	0.68	0.49	0.49	0.71	0.49
$k = 100$	0.50	0.69	0.50	0.50	0.70	0.50
$k = 200$	0.51	0.71	0.52	0.52	0.72	0.51
$k = 500$	0.53	0.72	0.53	0.54	0.73	0.53
$k = 1{,}000$	0.53	0.74	0.54	0.54	0.74	0.53

Notes: This table gives the average share of comovement in the granular regional sales (constructed from the sales of top k firms) that can be explained by the common component. The exercise is performed for Inter-regional, Intra-regional and All Sales, for top k plants in each row. The share of common component is calculated by using Equation 5 with $P = 1$ and is based on eigen-decomposition of granular region-level sales. The above table reports the results based on raw sales.

Table 5: Summary of Main Regression Variables

Variable	Mean (1)	Standard Deviation (2)
r_{ij}	0.285	0.361
$log(trade_{ij})$	3.467	3.095
r_{ij}^{1}	0.139	0.300
r_{ij}^{2}	0.173	0.324
r_{ij}^{5}	0.256	0.324
r_{ij}^{10}	0.282	0.327
r_{ij}^{20}	0.336	0.331
r_{ij}^{50}	0.369	0.328
r_{ij}^{100}	0.384	0.335
r_{ij}^{200}	0.393	0.338
r_{ij}^{500}	0.408	0.342
r_{ij}^{1000}	0.414	0.341
r_{ij}^{mf}	0.410	0.343
r_{ij}^{-mf}	0.444	0.297

Notes: The above table gives the mean and standard deviation of main regression variables. The r_{ij} captures the correlation between regional GDP, while r_{ij}^{k} is the correlation between granular GDP of k plants. r_{ij}^{mf} denotes correlations between sales of multi-plant firms, r_{ij}^{-mf} refer to correlations between sales of single-plant firms.

Table 6: Regional Comovement: Trade vs. Granular Plants

	Dependent Variable: Regional GDP correlation (r_{ij})							
# of Plants	k=1	k=5	k=10	k=50	k=100	k=200	k=500	k=1000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel (a): With Intra-Region Plant Sales				
$trade_{ij}$	0.030***	0.025***	0.023***	0.023***	0.023***	0.023***	0.023***	0.022***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$r^k_{ij,s}$	0.172***	0.280***	0.318***	0.289***	0.275***	0.288***	0.298***	0.306***
	(0.039)	(0.038)	(0.037)	(0.035)	(0.035)	(0.036)	(0.036)	(0.036)
Observations	595	595	595	595	595	595	595	595
R^2	0.016	0.015	0.069	0.045	0.048	0.060	0.076	0.079
				Panel (b): With Inter-Region Plant Sales				
$trade_{ij}$	0.026***	0.024***	0.022***	0.017***	0.013***	0.010**	0.007	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$r^k_{ij,s}$	0.145***	0.157***	0.176***	0.247***	0.292***	0.319***	0.342***	0.360***
	(0.052)	(0.049)	(0.046)	(0.044)	(0.044)	(0.044)	(0.043)	(0.044)
Observations	595	595	595	595	595	595	595	595
R^2	0.040	0.046	0.053	0.139	0.157	0.161	0.168	0.169
				Panel (c): With All Plant Sales				
$trade_{ij}$	0.024***	0.018***	0.012***	−0.002	−0.003	−0.005	−0.007*	−0.006*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$r^k_{ij,s}$	0.224***	0.357***	0.448***	0.621***	0.614***	0.619***	0.627***	0.634***
	(0.047)	(0.046)	(0.043)	(0.041)	(0.039)	(0.039)	(0.037)	(0.037)
Observations	595	595	595	595	595	595	595	595
R^2	0.039	0.121	0.154	0.290	0.296	0.295	0.311	0.314

Notes: The table is based on the following regressions:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_2 r^k_{ij,s} + \epsilon_{ij}$$

where r_{ij} captures the correlation between regions i and j, $trade_{ij}$ is the regional trade, while $r^k_{ij,s}$ variable is based on sales type s at the plant-level for the top k plants. Panel (a)-(c) gives results based on intra-region, inter-region, and all sales respectively. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Regional Comovement: Trade vs. Granular Plants vs. Granular Firms

	Dependent Variable: Regional GDP correlation (r_{ij})					
	(1)	(2)	(3)	(4)	(5)	(6)
$s =$	Intra	Inter	Intra	Inter	Intra	Inter
$trade_{ij}$	0.019***	0.007	0.023***	0.021***	0.018***	0.005
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Multi-plant Firms $r_{ij,s}^{mf}$	0.314***	0.296***			0.292***	0.271***
	(0.040)	(0.045)			(0.044)	(0.048)
Single-plant Firms $r_{ij,s}^{-mf}$			0.253***	0.163***	0.065	0.084
			(0.052)	(0.050)	(0.054)	(0.052)
Observations	595	595	595	595	595	595
R^2	0.14	0.11	0.092	0.076	0.155	0.117

Notes: The table is based on the following regressions:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_3 r_{ij,s}^{mf} + \beta_4 r_{ij,s}^{-mf} + \epsilon_{ij}$$

where r_{ij} captures the correlation between regions i and j and $trade_{ij}$ is the log regional trade. $r_{ij,s}^{mf}$ corresponds to granular plant-level correlation corresponding to multi-plant firms that belong to at least two regions and sales type s, while $r_{ij,s}^{-mf}$ is the correlation for the remaining set of plants. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Shocks to Granular Firms and Regional Comovement

	Dependent variable: Regional GDP correlation (r_{ij})				
	(1)	(2)	(3)	(4)	(5)
$trade_{ij}$	0.005	−0.005	0.014***	−0.004	−0.011
	(0.005)	(0.010)	(0.005)	(0.006)	(0.010)
r_{ij}^{resid}	0.317***	0.205***		0.285***	0.190***
	(0.047)	(0.045)		(0.046)	(0.045)
r_{ij}^{-mf}			0.308***	0.247***	0.281***
			(0.053)	(0.051)	(0.076)
Observations	595	595	595	595	595
R^2	0.129	0.501	0.099	0.153	0.513
Region FE		Y			Y

Notes: All models above use variation of the following regression:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_4 r_{ij}^{resid} + \beta_5 r_{ij}^{-mf} + \epsilon_{ij}$$

where r_{ij} corresponds to regional correlations, $trade_{ij}$ corresponds to log regional trade, r_{ij}^{resid} is the regional correlation between residuals of multi-region multi-plant firms' sales over and above single-plant firms' sales and r_{ij}^{-mf} is the regional correlation for the total regional sales of single-plants firms. Columns (3) and (5) estimates the same models as in columns (2) and (4) respectively, with region fixed–effects for both i and j. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix

A Additional Tables

Table A.1: Explanatory Power of Common Components in Regional Business Cycles (HP-filtered)

	Inter (1)	Intra (2)	All (3)
One Principal Component	0.56	0.78	0.61
Two Principal Components	0.69	0.90	0.72

Notes: This table gives the average share of business-cycle movement at the region-level that can be explained by the common component. The exercise is performed for inter-regional sales (column (1)), intra-regional sales (column (2)) and all sales (column (3)). The share of common component is calculated by using Equation 4 and is based on eigen-decomposition of region-level sales. The above table reports the results based on HP-filtered sales.

Table A.2: Explanatory Power of Top k Plants in Granular Comovement

	Non-Ranked			Ranked		
	Inter (1)	Intra (2)	All (3)	Inter (4)	Intra (5)	All (6)
$k = 1$	0.21	0.71	0.53	0.24	0.73	0.33
$k = 2$	0.21	0.67	0.57	0.25	0.81	0.37
$k = 5$	0.27	0.68	0.72	0.35	0.82	0.55
$k = 10$	0.31	0.72	0.79	0.37	0.80	0.64
$k = 20$	0.37	0.75	0.85	0.43	0.80	0.72
$k = 50$	0.43	0.78	0.90	0.46	0.80	0.78
$k = 100$	0.48	0.79	0.93	0.52	0.80	0.82
$k = 200$	0.53	0.81	0.96	0.56	0.82	0.86
$k = 500$	0.58	0.82	0.98	0.59	0.83	0.92
$k = 1{,}000$	0.60	0.82	0.99	0.62	0.83	0.95

Notes: The table gives the R^2 of the regression Equation 6 for top k plants based on different sales values (Inter, Intra, and All). It gives the explanatory power of the first factor constructed from a given sales type of top k plants across regions over the first factor based on aggregate region-level sales.

Table A.3: Explanatory Power of Top k Plants in Granular Comovement (Two Factors)

	Non-Ranked			Ranked		
	Inter	Intra	All	Inter	Intra	All
	(1)	(2)	(3)	(4)	(5)	(6)
$k = 1$	0.33	0.71	0.84	0.45	0.81	0.86
$k = 2$	0.27	0.72	0.88	0.54	0.81	0.87
$k = 5$	0.43	0.72	0.91	0.65	0.82	0.88
$k = 10$	0.51	0.75	0.94	0.6	0.81	0.94
$k = 20$	0.58	0.76	0.96	0.67	0.81	0.95
$k = 50$	0.6	0.78	0.97	0.69	0.8	0.97
$k = 100$	0.61	0.79	0.98	0.66	0.8	0.97
$k = 200$	0.64	0.81	0.98	0.65	0.82	0.98
$k = 500$	0.67	0.83	0.99	0.69	0.83	0.98
$k = 1000$	0.67	0.84	0.99	0.7	0.84	0.99

Notes: The table gives the R^2 of the regression similar to Equation 6 for top k plants based on different sales values (Inter, Intra and All). It gives the explanatory power of the first two factors constructed from given sales type of top k plants across regions over the first factor based on aggregate region-level sales.

$$z^1_{aggr} = \phi + \rho_1 z^1_{granular,k} + \rho_2 z^2_{granular,k} + \xi_k$$

where the right hand side variable $z^1_{granular,k}$ and $z^2_{granular,k}$ are the first two dominant factors calculated using sales of top k plants.

Table A.4: Explanatory Power of Top k Plants in Granular Comovement (HP filtered)

	Non-Ranked			Ranked		
	Inter (1)	Intra (2)	All (3)	Inter (4)	Intra (5)	All (6)
$k = 1$	0.54	0.70	0.87	0.54	0.71	0.85
$k = 2$	0.58	0.71	0.88	0.68	0.73	0.88
$k = 5$	0.66	0.72	0.90	0.75	0.74	0.93
$k = 10$	0.73	0.73	0.91	0.77	0.74	0.93
$k = 20$	0.74	0.73	0.92	0.80	0.73	0.94
$k = 50$	0.74	0.73	0.94	0.78	0.73	0.95
$k = 100$	0.73	0.74	0.95	0.78	0.73	0.96
$k = 200$	0.73	0.74	0.96	0.76	0.74	0.96
$k = 500$	0.72	0.75	0.97	0.75	0.75	0.97
$k = 1000$	0.72	0.75	0.98	0.75	0.75	0.98

Notes: The table gives the R^2 of the regression Equation 6 for top k plants based on HP-filtered sales values, Inter, Intra and All. It gives the explanatory power of the first factor constructed from given sales type of top k plants across regions over the first factor based on aggregate region-level sales.

Table A.5: Regional Comovement: Trade vs. Granular Plants (Region Fixed-effects)

	Dependent Variable: Regional GDP correlation (r_{ij})							
# of Plants	k=1	k=5	k=10	k=50	k=100	k=200	k=500	k=1000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel (a): With Intra-Region Plant Sales							
$trade_{ij}$	−0.001	−0.003	−0.002	−0.007	−0.008	−0.008	−0.009	−0.009
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$r^k_{ij,s}$	0.124***	0.246***	0.245***	0.250***	0.246***	0.264***	0.277***	0.283***
	(0.042)	(0.049)	(0.047)	(0.046)	(0.044)	(0.047)	(0.048)	(0.050)
Observations	595	595	595	595	595	595	595	595
R^2	0.493	0.508	0.510	0.513	0.513	0.514	0.515	0.515
Region FE	Y	Y	Y	Y	Y	Y	Y	Y
	Panel (b): With Inter-Region Plant Sales							
$trade_{ij}$	−0.001	−0.001	−0.001	−0.002	−0.003	−0.004	−0.005	−0.005
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$r^k_{ij,s}$	0.078	0.046	0.092*	0.140***	0.183***	0.209***	0.230***	0.247***
	(0.052)	(0.054)	(0.051)	(0.053)	(0.055)	(0.057)	(0.058)	(0.059)
Observations	595	595	595	595	595	595	595	595
R^2	0.487	0.485	0.488	0.492	0.495	0.498	0.500	0.501
Region FE	Y	Y	Y	Y	Y	Y	Y	Y
	Panel (c): With All Plant Sales							
$trade_{ij}$	−0.001	−0.002	−0.005	−0.011	−0.014*	−0.017**	−0.020**	−0.020**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$r^k_{ij,s}$	0.090*	0.140***	0.247***	0.459***	0.477***	0.498***	0.515***	0.531***
	(0.047)	(0.050)	(0.048)	(0.056)	(0.055)	(0.054)	(0.054)	(0.055)
Observations	595	595	595	595	595	595	595	595
R^2	0.488	0.492	0.509	0.543	0.549	0.556	0.560	0.563
Region FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table is based on the following regressions:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_2 r^k_{ij,s} + \delta_i + \delta_j + \epsilon_{ij}$$

where r_{ij} captures the correlation between regions i and j, $trade_{ij}$ is the log regional trade, while $r^k_{ij,s}$ variable is based on sales type s at the plant-level for the top k plants. Panel (a)-(c) gives results based on intra-region, inter-region, and all sales respectively. δ_i and δ_j correspond to the region-fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: Regional Comovement: Trade vs. Granular Plants vs. Granular Firms (Region Fixed-effects)

	Dependent Variable: Regional GDP correlation (r_{ij})					
	(1)	(2)	(3)	(4)	(5)	(6)
$s =$	Intra	Inter	Intra	Inter	Intra	Inter
$trade_{ij}$	-0.005	-0.006	-0.006	-0.002	-0.007	-0.007
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Multi-plant Firms: $r_{ij,s}^{mf}$	0.238***	0.195***			0.209***	0.186***
	(0.047)	(0.054)			(0.051)	(0.054)
Single-plant Firms: $r_{ij,s}^{-mf}$			0.249***	0.119*	0.123	0.096
			(0.075)	(0.064)	(0.080)	(0.064)
Observations	595	595	595	595	595	595
R^2	0.509	0.497	0.495	0.488	0.511	0.499
Region FE	Y	Y	Y	Y	Y	Y

Notes: The table is based on the following regressions:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_3 r_{ij,s}^{mf} + \beta_4 r_{ij,s}^{-mf} + \delta_i + \delta_j + \epsilon_{ij}$$

where r_{ij} captures the correlation between regions i and j and $trade_{ij}$ is the log regional trade. $r_{ij,s}^{mf}$ corresponds to granular plant-level correlation corresponding to multi-plant firms and sales type s, while $r_{ij,s}^{-mf}$ is the correlation for the remaining set of plants. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Regional Comovement: Trade vs. Granular Plants (HP-filtered Sales)

	Dependent Variable: Regional GDP correlation (r_{ij})							
# of Plants	k=1	k=5	k=10	k=50	k=100	k=200	k=500	k=1000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (a): With Intra-Region Plant Sales								
$trade_{ij}$	0.024***	0.025***	0.028***	0.025***	0.025***	0.025***	0.025***	0.024***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$r^k_{ij,s}$	0.117***	0.134***	0.186***	0.240***	0.255***	0.312***	0.342***	0.352***
	(0.040)	(0.040)	(0.042)	(0.042)	(0.042)	(0.040)	(0.040)	(0.040)
Observations	595	595	595	595	595	595	595	595
R^2	0.016	0.015	0.069	0.045	0.048	0.060	0.076	0.079
Panel (b): With Inter-Region Plant Sales								
$trade_{ij}$	0.021***	0.018***	0.015***	0.005	0.002	−0.001	−0.005	−0.007
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$r^k_{ij,s}$	0.240***	0.246***	0.288***	0.469***	0.500***	0.511***	0.521***	0.526***
	(0.059)	(0.058)	(0.056)	(0.052)	(0.049)	(0.047)	(0.045)	(0.045)
Observations	595	595	595	595	595	595	595	595
R^2	0.040	0.046	0.053	0.139	0.157	0.161	0.168	0.169
Panel (c): With All Plant Sales								
$trade_{ij}$	0.021***	0.014***	0.011**	−0.002	−0.002	−0.004	−0.007	−0.008*
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$r^k_{ij,s}$	0.266***	0.440***	0.532***	0.678***	0.665***	0.667***	0.686***	0.692***
	(0.055)	(0.050)	(0.047)	(0.037)	(0.035)	(0.036)	(0.035)	(0.035)
Observations	595	595	595	595	595	595	595	595
R^2	0.039	0.121	0.154	0.290	0.296	0.295	0.311	0.314

Notes: The table uses HP-filtered data on sales and is based on the following regressions:

$$r_{ij}^{HP} = \beta_0 + \beta_1 trade_{ij} + \beta_2 {r^k_{ij,s}}^{HP} + \epsilon_{ij}$$

Panel (a)-(c) gives results based on intra-region, inter-region, and all sales respectively. k gives number of top plants used in the construction of granular plant-level correlation, $r^k_{ij,s}$. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Regional Comovement: Trade vs. Granular Plants vs. Granular Firms (HP-filtered Sales)

	Dependent Variable: Regional GDP correlation (r_{ij})					
	(1)	(2)	(3)	(4)	(5)	(6)
$s =$	Intra	Inter	Intra	Inter	Intra	Inter
$trade_{ij}$	0.021***	−0.004	0.017***	0.007	0.017***	−0.009
	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Multi-plant Firms: $r_{ij,s}^{mf}$	0.299***	0.453***			0.233***	0.346***
	(0.043)	(0.058)			(0.047)	(0.066)
Single-plant Firms: $r_{ij,s}^{-mf}$			0.348***	0.364***	0.200***	0.236***
			(0.052)	(0.045)	(0.056)	(0.053)
Observations	595	595	595	595	595	595
R^2	0.108	0.115	0.087	0.099	0.121	0.136

Notes: The table uses HP-filtered sales data and is based on the following regressions:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_3 r_{ij,s}^{mf} + \beta_4 r_{ij,s}^{-mf} + \epsilon_{ij}$$

r_{ij}^{mf} corresponds to correlation between sales of multi-plant firms, while r_{ij}^{-mf} is the correlation for the remaining set of plants. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Regional Comovement: Trade vs. Granular Plants (Growth Rate)

	Dependent Variable: Regional GDP correlation (r_{ij})							
# of Plants	k=1	k=5	k=10	k=50	k=100	k=200	k=500	k=1000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel (a): With Intra-Region Plant Sales							
$trade_{ij}$	0.017**	0.017**	0.019***	0.018***	0.018***	0.018***	0.018***	0.018***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$r^k_{ij,s}$	0.026	0.098*	0.173***	0.182***	0.203***	0.263***	0.303***	0.312***
	(0.046)	(0.050)	(0.051)	(0.055)	(0.057)	(0.057)	(0.057)	(0.057)
Observations	595	595	595	595	595	595	595	595
R^2	0.01	0.016	0.027	0.026	0.029	0.039	0.046	0.047
	Panel (b): With Inter-Region Plant Sales							
$trade_{ij}$	0.009	0.009	0.008	−0.001	−0.005	−0.009	−0.013*	−0.014*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
$r^k_{ij,s}$	0.401***	0.257***	0.257***	0.469***	0.575***	0.581***	0.573***	0.562***
	(0.064)	(0.067)	(0.065)	(0.071)	(0.065)	(0.062)	(0.061)	(0.061)
Observations	595	595	595	595	595	595	595	595
R^2	0.075	0.034	0.032	0.065	0.091	0.091	0.092	0.088
	Panel (c): With All Plant Sales							
$trade_{ij}$	0.006	0.009	0.009	−0.001	−0.001	−0.003	−0.005	−0.006
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$r^k_{ij,s}$	0.432***	0.277***	0.358***	0.574***	0.594***	0.613***	0.643***	0.658***
	(0.057)	(0.052)	(0.056)	(0.053)	(0.051)	(0.052)	(0.051)	(0.052)
Observations	595	595	595	595	595	595	595	595
R^2	0.098	0.047	0.061	0.124	0.135	0.139	0.151	0.154

Notes: The table uses year-on-year growth rate of sales to calculate correlations (over 12 observations each) and is based on the following regressions:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_2 r^k_{ij,s} + \epsilon_{ij}.$$

Panel (a)-(c) gives results based on intra-region, inter-region, and all sales respectively. k gives number of top plants used in the construction of granular plant-level correlation, $r^k_{ij,s}$. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.10: Regional Comovement: Trade vs. Granular Plants vs. Granular Firms (Growth Rate)

	Dependent Variable: Regional GDP correlation (r_{ij})					
	(1)	(2)	(3)	(4)	(5)	(6)
$s =$	Intra	Inter	Intra	Inter	Intra	Inter
$trade_{ij}$	0.016**	−0.023**	0.011	0.004	0.012*	−0.025**
	(0.007)	(0.010)	(0.007)	(0.008)	(0.007)	(0.010)
Multi-plant Firms: $r_{ij,s}^{mf}$	0.204***	0.562***			0.134**	0.519***
	(0.058)	(0.076)			(0.062)	(0.081)
Single-plant Firms: $r_{ij,s}^{-mf}$			0.311***	0.262***	0.211***	0.099
			(0.076)	(0.069)	(0.077)	(0.074)
Observations	595	595	595	595	595	595
R^2	0.028	0.073	0.029	0.029	0.035	0.076

Notes: The table uses year-on-year growth rate of sales to calculate correlations (over 12 observations each) and is based on the following regressions:

$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_3 r_{ij,s}^{mf} + \beta_4 r_{ij,s}^{-mf} + \epsilon_{ij}.$$

r_{ij}^{mf} corresponds to correlation between sales of multi-plant firms, while r_{ij}^{-mf} is the correlation for the remaining set of plants. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.11: Robustness: Shocks to Granular Firms and Regional Comovement

	Dependent variable: Regional GDP correlation (r_{ij})			
	(1)	(2)	(3)	(4)
$trade_{ij}$	-0.001	-0.013	-0.004	-0.011
	(0.005)	(0.010)	(0.006)	(0.010)
r_{ij}^{resid}	0.282***	0.207***	0.288***	0.194***
	(0.044)	(0.040)	(0.046)	(0.044)
r_{ij}^{-mf}	0.238***	0.271***	0.252***	0.284***
	(0.051)	(0.076)	(0.051)	(0.076)
Observations	595	595	595	595
R^2	0.157	0.520	0.154	0.514
Region FE		Y		Y

Notes: All models above use variation of the following regression:

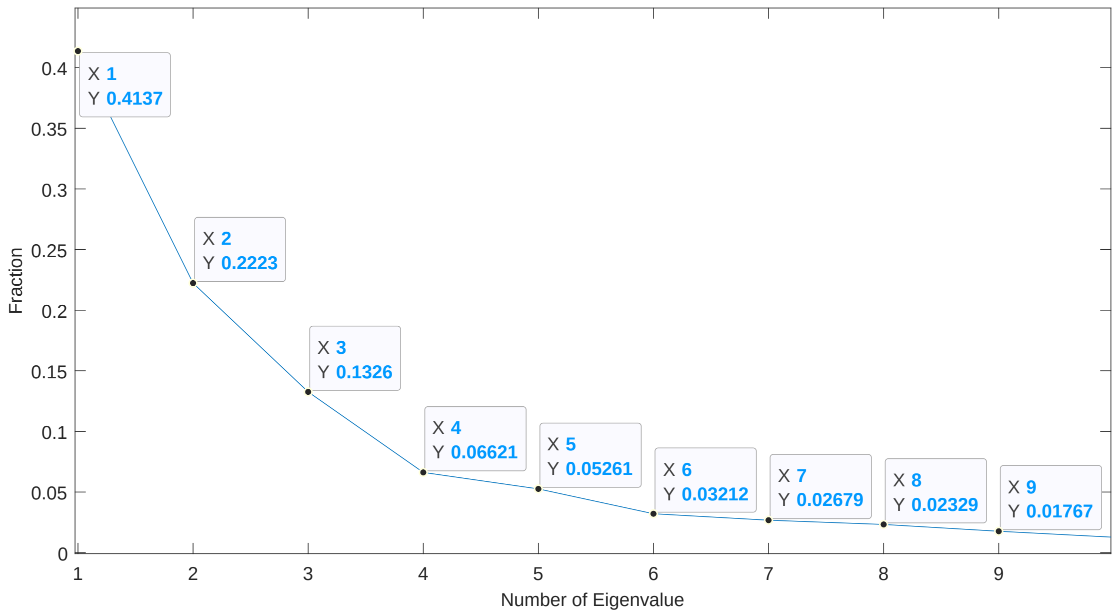
$$r_{ij} = \beta_0 + \beta_1 trade_{ij} + \beta_4 r_{ij}^{resid} + \beta_5 r_{ij}^{-mf} + \epsilon_{ij}$$

where r_{ij} corresponds to regional correlations, $trade_{ij}$ corresponds to log regional trade, and r_{ij}^{-mf} is the regional correlation for the total regional sales of single-plants firms. r_{ij}^{resid} is the regional correlation between residuals of multi-region multi-plant firms' sales over and above single-plant firms' sales. In columns (1) and (2) the residual used to construct r_{ij}^{resid} also filters out aggregate shocks by including the first factor as extracted from the NIPALS algorithm in the residual construction stage. Similarly, in columns (3) and (4), the residual is constructed by filtering out aggregate All India Sales of all plants. Columns (2) and (4) estimates the same models as in columns (1) and (3) respectively, with region fixed–effects for both regions i and j. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

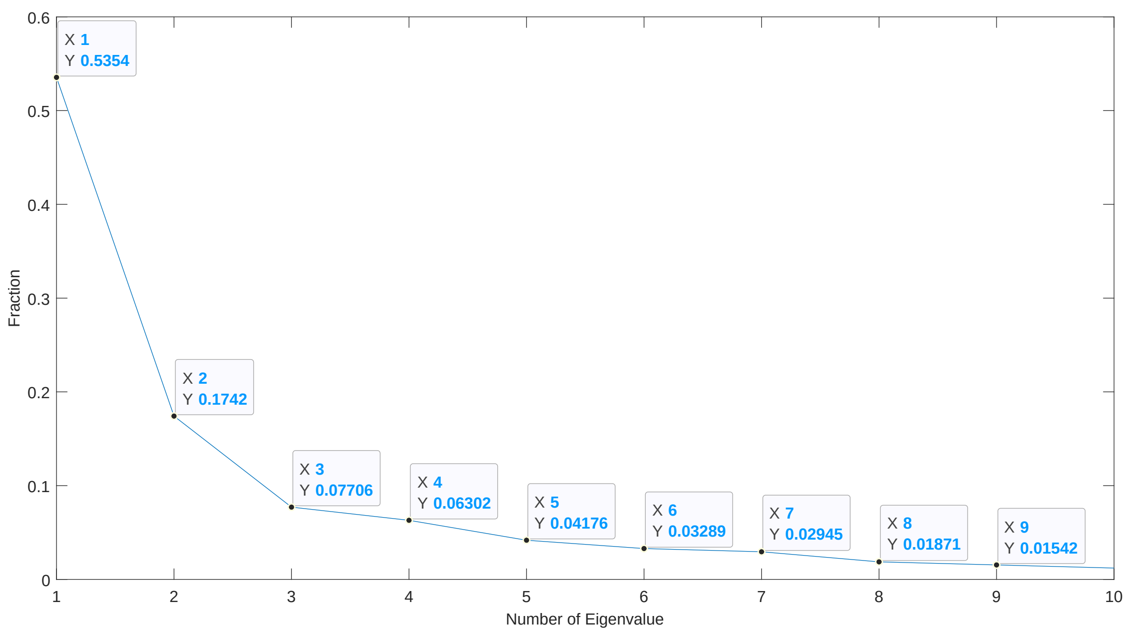
B Additional Figures

Figure A.1: Explanatory Power of Eigenvalues

(a) Scree Plot for Regional GDP



(b) Scree Plot for Granular GDP for $k = 1,000$ Plants



Notes: The figure reports explanatory power of the top 10 eigenvalues in explaining the variation of sales constructed from region-level sales. Panel (a) corresponds to the aggregate region-level sales, while Panel (b) is based on granular GDP constructed from the sales of top $k = 1,000$ plants from each region.