Domestic vs Foreign Superstars: Granular Comparative Advantage, Pro-competitive Effects, and Productivity Spillovers

Thomas Rowley*

This version: August 7, 2023†

Abstract

China's immense market potential, export-led development, and policy encouragement of foreign direct investment (FDI) have coincided with the presence of superstar foreign-funded firms. This paper shows how these very large, granular operations of foreign multinationals in China are systematically different from their domestically-funded Chinese counterparts over the period 1998-2007. These differences carry over to heterogeneous associations with industry exports (granular comparative advantage) and to the degree of competitive pressure these two groups of firms place on non-granular firms. More industry granularity, or concentration among the top-few firms, sees less industry exports, the opposite of what is found for the case of France by Gaubert and Itskhoki (2021). However, using an IV strategy, the effect of granularity owing to FDI is positive, and the behavior of granularity owing to domestically-funded firms converges towards that of FDI over time. An investigation into firm-level markups then shows that granularity originating from FDI induces a greater degree of pro-competitive effects on non-granular firms. Furthermore, by utilizing official documents on FDI encouragement to build an industry-varying shock, I detect positive horizontal spillover effects on domestic firms' TFP and labor productivity coming from the granularity component of FDI, a finding that stands against the existing, non-granular specific FDI literature for China. These results underscore the paramount importance of the dichotomy of firm ownership for the granularity literature in its application to developing countries.

Keywords: Foreign Direct Investment (FDI), China, Granular Firms, Granular Comparative Advantage, Markups, Firm Ownership, Horizontal Spillovers.

JEL Codes: F14, F21, F23, F63

---

*PhD candidate at ESSEC Business School, France. Email: thomas.rowley@essec.edu

†Work-in-progress. Questions or comments are much appreciated. Email for latest version.

1 Introduction

The impact of superstar firms on the export patterns of emerging economies has garnered significant attention[1]. Notably, China's export-led growth and the comparable significance of foreign and domestic firms in terms of gross export volumes during the early 2000s offer an exceptional context for investigating superstar firms, where the dichotomy of firm ownership holds paramount importance. By utilizing Chinese firm-level panel data spanning the period 1998 to 2007, this paper segregates the largest local operations of foreign multinationals from the most prominent domestic firms to assess their heterogeneous impact on industry-level comparative advantage in this emerging economy. The analysis then shifts focus to the implications of their presence for the Chinese economy, particularly examining how the pro-competitive effects from superstar FDI differs from that of domestic superstars, as well as investigating the unique horizontal spillovers resulting from this form of FDI.

China's FDI inflows began to pick up significantly in 1992 when the country accelerated the opening up of its economy. Elevated FDI inflows as a percentage of GDP were maintained throughout the following two decades, even as the size of its economy grew rapidly. By 2001 FDI accounted for over 50% of Chinese exports and by 2007 China's trade balance had skyrocketed to over 300 billion USD. This was accompanied by high levels of gross fixed capital formation that aided in converting FDI into further growth. The story of China's FDI is of vital importance, especially for developing countries or regions positioned today similarly to China was then and seeking to replicate the success of China's industrial policy during these instrumental years of its development. In this context, the role of superstar foreign firms as well as superstar domestic firms has yet to be explored.

The Chinese Industrial Enterprise Survey data provides information on the sources of firms' capital funding that allows identification of what I refer to in this paper as Foreign-Funded-Firms (FFFs), or firms that are funded by at least 50% FDI. FFFs are shown to not only be important in the economy as a whole, but also at the industry-level, accounting for over 40% of industry exports on average across 420 industry codes. Looking closer reveals that the group of FFF superstars (here defined as the top-3 FFFs by exports) contribute more to this figure than all other FFFs combined, and nearly as much as the similarly defined set of superstar Domestically-Funded-Firms (DFFs). In the context of China, the first focus of this paper is to explore how these superstar firms impact industry export dynamics. By examining this aspect, I aim to contribute to our understanding of the dynamics and implications of these firms' presence in the Chinese market.

In their recent work, Gaubert and Itskhoki (2021) distinguish between a "fundamental com-

[1] Freund and Pierola (2015) highlighted that across 32 developing countries between 2006 and 2008, the leading firm accounted for an average of nearly 15% of total (non-oil) exports

parative advantage," which applies to all firms within an industry, and a "granular comparative advantage," (GCA) which arises from unique market forces or know-how possessed by industry leaders. The latter is attributed to the heavy-tailed distribution of firm sizes and the limited number of firms in each industry, which allows outlier firms to significantly impact industry aggregates. An industry's granularity is proxied for by the combined market share of the top-few firms per industry, which is typically many times larger than that of average firms, and varies widely across industries. As cross-sectional variation in industry exports maps one-to-one with variation in comparative advantage, significance on the granularity proxy when explaining exports is interpreted as being associated with GCA, with non-GCA interpretations being ruled out in robustness checks. Using French data, Gaubert and Itskhoki (2021) find that industry granularity, without mention of FDI, is positively associated with industry export competitiveness. However, the current paper's first finding is that this relationship is not true for a major emerging economy during the same period, and that it is, in fact, negative. Thus, the paper suggests caution in assuming that the empirical literature following the work of Gabaix (2011) on how granular firms[2] impact aggregate outcomes applies equally to countries in earlier stages of development.

In economies at earlier stages of development, there are not only fundamental comparative advantages at the industry level, but also at least two fundamentally different types of firms. In this paper, it is argued that granularity should be separated into foreign and domestic channels. This approach is justified by systematic differences in the power-law exponents estimated for the two groups' size and productivity distributions. The observed differences cannot be explained solely by additional barriers of entry faced by foreign firms as the distribution estimates have a lower bound cut-off on the right tail, whereby only the relatively larger firms from both groups are compared. Both of these groups of firms in China are described by granularity, and combining the top-3 exporters that are FFFs and the top-3 exporters that are DFFs captures over 50% of industry exports on average in each year of the sample. These two groups of granular firms ultimately lead to heterogeneous and practically opposite associations with GCA.

By partitioning granularity into two categories, I can also highlight the type of foreign direct investment (FDI) that is most represented in the granularity measure for FFFs. I demonstrate that large foreign multinationals established in China for market access purposes, whereby they are granular in terms of domestic market share, have no bearing on GCA. However, if employment is used to define granularity, yielding a more unbiased set of multinationals that includes those with efficiency-seeking objectives, a positive influence on GCA emerges in China, and this group of firms offers a similarly strong source of GCA to that observed by Gaubert and Itskhoki (2021). Furthermore, the detailed panel data allows for an instrumental variable analysis that enhances the

[2]The term granular here refers to the existence of superstar firms and the degree to which they dominate the industries in which they are part of.

credibility of the causal inference in this regard, while a variety of robustness checks placed in the appendix further address other concerns, both with respect to causality and also the application to China. In all cases, the coefficient on granularity from DFFs remains negative. Nevertheless, over the 10-year period covered by the panel data, there is evidence of convergence in the behavior of Chinese GCA to that of a developed economy scenario, where the adverse relationship between GCA from DFF granularity and industry export variation diminishes year after year.

A primary concern of this analysis is that Chinese industries which are more comparatively advantaged attract more FFF granularity, and industries which are more comparatively disadvantaged attract more DFF granularity. Besides the mentioned comprehensive robustness checks, heterogeneity results also point away from this being the full explanation. For example, FFF granularity is more strongly associated with GCA in industries that are in the middle of the technology spectrum, dominated by state-owned firms, less penetrated by other foreign firms, have lower aggregate TFP, and a higher ratio of young firms. This suggests that FFF granularity generates positive GCA in industries that are not too mature, yet are well-positioned for growth, rather than those industries which were already well-established. DFF granularity is also more negatively associated with GCA in industries with a higher presence of FDI or of younger firms, whereas if it were only that comparatively disadvantaged industries attracted DFF granularity, these stronger results would not be expected here. Regardless of the origin of DFF granularity, it seems clear that its presence holds back industry exports during the earlier years of China's development.

At the level of the granular firms themselves, both groups generate relatively more GCA when they have higher added value relative to real assets, lower labor intensity measured in wages to real assets, or pay higher average wages per worker. Though even when accounting for granular DFFs that may have all of these firm-level attributes, their net effect remains negative.

To further explore the contribution of granular firms to industry competitiveness, I delve into their impact on the non-granular firms to investigate the impact of these superstar firms on the domestic market, and by doing so contribute to two very important streams of the literature in international trade; one being the central issue of pro-competitive effects, and other being the literature on FDI spillovers.

If granularity impacts comparative advantage beyond the granular firms' own contribution to industry exports, then impacts on firm-level competitiveness or productivity spillovers would be likely channels. This notion is supported by Edmond et al. (2015), who establish that international trade reduces markup distortions when it exposes firms to effective competition, depending on the contrast between domestic and foreign producers.

The current paper finds that granular firms can also be a pro-competitive force, but with heterogeneity depending on whether or not the granularity owes to FDI. In showing as much, I model firm-level markups as a function of both firm and industry-level variables and include terms to capture the presence and nature of granular firms as well as their interactions with the non-granular firm. Only FFF granularity induces pro-competitive effects on both other large firms and also small firms.

Finally I am able to establish casual inference on horizontal spillovers from FFFs granularity on TFP and labor productivity by using the industry varying change in official guidelines on FDI encouragement upon China's WTO accession. These guidelines are detailed in the Catalog of Encouraged Industries for FDI for the years 1997 and 2002. It is the change in many of these industries that allows for a dif-in-dif IV type exercise. This industry varying shock significantly explains only FFF granularity as defined by domestic market sales, not by labor or even exports, and by no definition of granularity for DFFs. This makes sense as this catalog update was mainly concerned with access to China's domestic market by foreign firms. I further show that it was primarily the granular FFFs who were able to take advantage of these new guidelines on encouragement of FDI, and FDI penetration increased mostly along intensive margin that owes to these granular firms. I then go on to use the shock as an IV and the positive spillovers on the set of all DFFs, while if instrumenting FDI penetration instead the effect is negative. FFF granularity can thus be interpreted as a new source of horizontal spillover effects, which also assists the story of FFF granularity being a source of GCA.

These findings have significant implications for trade policy. For developing countries seeking to boost their comparative advantage in certain industries, the presence of large efficiency-seeking multinationals may be beneficial. However, encouraging market-seeking firms to become excessively dominant could be detrimental. Heterogeneity in the characteristics of superstar firms and the industries in which they operate provides a checklist for assessing the appropriateness of particular firms. This research aims to ensure that developments of theory and empirical evidence on GCA are inclusive of a more diverse set of countries.

The following section gives a brief review of the recent literature on granular firms as well as literature relating FDI to comparative advantage and export competitiveness. The current paper combines these two branches of literature in the context of China. Section 3 presents the data, discusses the nature of granularity in China, and gives descriptive statistics on the two groups of granular firms. The methodology for detecting GCA with FDI is presented in section 4 while results on GCA with FDI, relevant robustness checks and an investigation into how granular firms heterogeneously affect exports are presented in section 5. Section 6 then explores how both groups of granular firms influence non-granular firm markups, followed causal inference on superstar FDI spillovers in section 7. Section 8 concludes.

2 Related Literature

The first strand of literature that this paper relates to is that on large firms. Since Bernard et al. (1995) there have been numerous literatures on the unique characteristics of large exporting firms. Along this line of literature, Eaton et al. (2004) find that sensitivity of trade to trade costs is mostly along the extensive margin. This may imply that large exporting firms are relatively more persist and thus bring stability to export competitiveness of the industries in which they operate. Further empirical evidence presented by Freund and Pierola (2015) shows the single largest firm within each of 32 developing countries as accounting for 15 percent of total exports on average, with the top 1 percent of firms taking up 53 percent of exports. Naturally it follows that a country is also exposed to additional business cycle volatility in the presence of large firms and multinationals in general. Cravino and Levchenko (2016) determine that 10 percent of aggregate productivity shocks owe to the presence of foreign affiliates transmitting shocks from abroad. Bernard et al. (2018) find that firms that are more productive take increasingly larger market shares, with costs differences magnified by productivity differences via extensive and intensive margins of trade and FDI. Taken together, these papers suggest that superstar firms have an out-sized impact on industry export competitiveness and that this relationship may crucially depend on the mix of foreign versus domestic superstar firms. Testing this hypothesis is the first point of analysis of this paper.

In looking closely at the influence individual firms have on aggregate economy or industry characteristics, Gabaix (2011) establishes the “Granular” hypothesis; that heavy-tailed firm size distributions go against the diversification argument, and therefore shocks to such large firms do not average out in the aggregate as they would for smaller, more numerous firms, but instead propagate throughout the economy. Specifically, he shows firm-level shocks from the 100 largest US firms explain one third of US variation in output growth. In a similarly motivated paper, Acemoglu et al. (2012) establish a channel whereby idiosyncratic shocks may cascade across an economy due to networks of intersectoral input–output linkages. These network effects explain how some well-known industries consist of countless suppliers or other smaller firms that are dependent on the business and are practically at the mercy of a few oversized firms with complex supply chains.

Carvalho and Gabaix (2013) puts granularity to the test to show how the U.S. great moderation (80s-90s) owes to diversification following a decrease in granular volatility (heavy manufacturing). Grassi (2018) details how the structural importance of a firm, measured by the elasticity of aggregate output to the particular firm's productivity, depends on the interaction between the competition intensity of the firm's industry, the input-output network, and the firm size. The channel through which firm-level shocks propagate through the economy is labor augmenting productivity shock's impact on oligopolistic markups. As mentioned in the introduction, in Gaubert and Itskhoki (2021) the largest firms impact their industry's comparative advantage through their granularity, which can be separated from fundamental comparative advantage common to all firms. They use French firm-level data and show that about 20 percent of the variation in realized export intensity across sectors comes specifically from the added grandness of the country's largest firms. They also show that a 10 percentage point greater sales share concentrated in the top-3 firms (as opposed to being spread out among other firms) is associated with a 9% (log points) increase in the aggregate level of industry exports. That is to say that superstar firms in France impact the country's set of comparative advantages in a positive manner according to the industry in which the superstar firm is located. Their methodology for detecting the presence of a granular comparative advantage (GCA) is detailed section 4 as the starting point of the current paper's empirical analysis. The point of departure is immediate, as initial results show the exact opposite relationship exists when applied to the developing country case of China, and that distinguishing between foreign and domestic superstars is vital for drawing insight.

The second strand of literature motivating the current paper is that on FDI and its role in international trade. Almfraji and Almsafir (2014) gather that most literature relating FDI to economic growth indicate that FDI, especially in the manufacturing sector, aids growth and development. It is then a logical extension to look for a positive source of GCA in the form of FDI if firms in general are indicating a negative GCA on average, as is described below.

FDI was nearly nonexistent in China prior to the 1978 opening-up initiated by Deng Xiaoping. Throughout the 1980s and 1990s FDI became not only allowed, but foreign firms were often given preferential treatment on multiple fronts, largely through the implementation of special economic zones (SEZs) scattered about the country's coastal areas. Restrictions remained, however, in many industries, such as requiring joint-ventures, technology sharing or some degree of local-content requirement. China's bid to enter the WTO brought a further opening of the economy throughout the 1990s in anticipation of joining in 2001. By 1997 average tariffs had been cut and many additional industries opened, while the effective tariff rate, or the percentage of imports actually collected, was only 3 percent (Tseng and Zebregs, 2002).

Young (1995) shows that the high growth rates experienced by East Asian countries owes to capital accumulation, not rapid technological progress and Whalley and Xin (2006) show that 90 percent of recent growth in local Chinese firms owes to capital accumulation. While domestic resource mobilization is the most important source of capital formation for economic growth Nunnenkamp (2004), FDI both directly adds to this and has potential to catalyze domestic capital formation (Amighini et al. (2017)) and alleviate domestic credit constraints (Poncet et al. (2010)). Taken in combination with the literature on granularity, there is no reason not to expect that superstar foreign funded firms operating in China contribute to aggregate outcomes beyond that of FDI more generally, and in a manner different than that of superstar domestically funded firms.

In addition to the effect of superstar firms on industry exports, the current paper delves into the their impact on the non-granular firms of the economy. Edmond et al. (2015) discusses the significance of the type of foreign firms engaging in international trade, highlighting how, depending on how foreign and domestic firms compare, pro-competitive effects can lead to lower overall markups when a country opens to trade. Section 6 of the current paper extends this perspective to examine how non-granular firms in China respond to granularity in terms of their competitive behavior.

Iacovone et al. (2015) and Bloom et al. (2013) give case study evidence of particular superstar multinationals entering a developing country and producing positive spillovers on domestic suppliers. Amiti et al. (2023) looks directly at vertical spillovers from multinationals and then superstar firms of any ownership, where they use Belgian firm-to-firm transaction data to show that beginning to supply a multinational or beginning to supply a superstar firm, regardless or not of whether it is a multinational, produce similar positive spillovers on the domestic supplier's Total Factor Productivity (TFP). However, their event study approach caries a concern that in most results there is a significant increase in TFP 2 years before a new relationship with either a multinational or superstar firm is formed, suggesting that TFP growth, either pre-existing or soon to come, may be what enables firms to secure larger contracts, rather than the contracts causing the TFP increase. To analyze horizontal spillovers from superstar FDI in 7 of the current paper I instead rely on industry level shocks for identification of the treatment group and thus avoid concerns of signalling by individual firms. Additionally, while there exists much conclusive literature on vertical spillovers on suppliers or buyers of foreign multinationals, the literature on horizontal spillovers is inconclusive, giving additional importance to analyzing the particular channel of horizontal superstar FDI spillovers on local firms in the same industry.

Several studies, such as Du et al. (2012), Lin et al. (2009), and Lu et al. (2017), have examined spillover effects of foreign direct investment (FDI) using the same dataset as the present paper, though none address the presence of superstar firms. These authors generally find evidence of positive spillovers only in forward or backward effects on suppliers or buyers, while horizontal spillovers on firms within the same industry are typically found to be negative. Lu et al. (2017) uses the same industry level shock as the current paper uses to identify the treatment group for analyzing spillovers, but arrives at a more narrowly defined list of treated industries whereas the current paper seemingly maintains a more strict criteria for an industry being in the control group, and thus the results of Lu et al. (2017) cannot be reproduced as their data has not been made available. Further comparison of their study to the present paper is left to section 7.

3 Data and Context

3.1 Description of the panel data

The Chinese Industrial Enterprise Survey (CIES) conducted by the National Bureau of Statistics of China (NBS) is a large firm-level dataset with consistent data on firm funding and various firm-reported figures for the ten years spanning 1998 through 2007. It is meant to include all firms with total sales above five million RMB as well as all state-owned firms for these years. The survey continues past 2007 but no longer includes all firms and some key variables. Variables reported throughout these years include total firm sales, exports, input costs, total employment, wage bill, current and fixed assets, paid-in capital, and many others. Brandt et al. (2014) provide a good description of the dataset. Largely following their suggestions and using additional text analysis, I match firms that may have had a change in ID, eliminate firms and industries that are not actually in manufacturing or are in highly restricted industries such as tobacco or weapons, and create some additional variables such as real physical capital stocks. The result is 2,035,995 observations across the 10 years, consisting of 533,677 firms in an unbalanced panel sorting into 420 Chinese industry codes that sometimes vary within a firm from one year to the next.

There are 6 variables describing different origins of paid-in capital, each with a corresponding running total that is used to construct ownership percentages. These 6 variables are paid-in capital from each of Hong Kong/Macao/Taiwan (henceforth HKMT), foreign (all non-HKMT rest of world funding), private, corporate, state, and local collective. I first combine private and corporate into a single private category, as well as state and local collective into a single state category. A firm can reasonably have paid-in capital under each category if it is a joint-venture among many participants. I combine the categories of HKMT and foreign, which then contains all non-mainland China funding sources, and label a firm as an FFF if these two categories make up at least 50 percent of total paid-in capital. All other firms are therefore DFFs, meaning that more than 50% of their ownership is traced to mainland China. A DFF can further be called private if the combined categories of HKMT, foreign, and private make up 50 percent or more of total paid-in capital, given that the firm was not already classified as an FFF. Any remaining DFFs with reported capital funding are therefore dominated by some form of state capital. While I primarily maintain only two groups throughout the analysis, robustness checks are done for separating FFFs to HKMT FFFs and non-HKMT FFFs as well as for separating private DFFs from state-owned DFFs.

Further details on how funding sources are identified and why the 50% financing threshold is maintained for defining FFFs are left for appendix section A1. Table A1 in the appendix shows the percentage of FFFs and DFFs among all firms and that by 2007 FFFs accounted for 27.4% of the annual aggregate total sales and 61.4% of total exports. The next section discusses how the largest firms are remarkably important for these aggregate figures.

3.2 Granularity in China

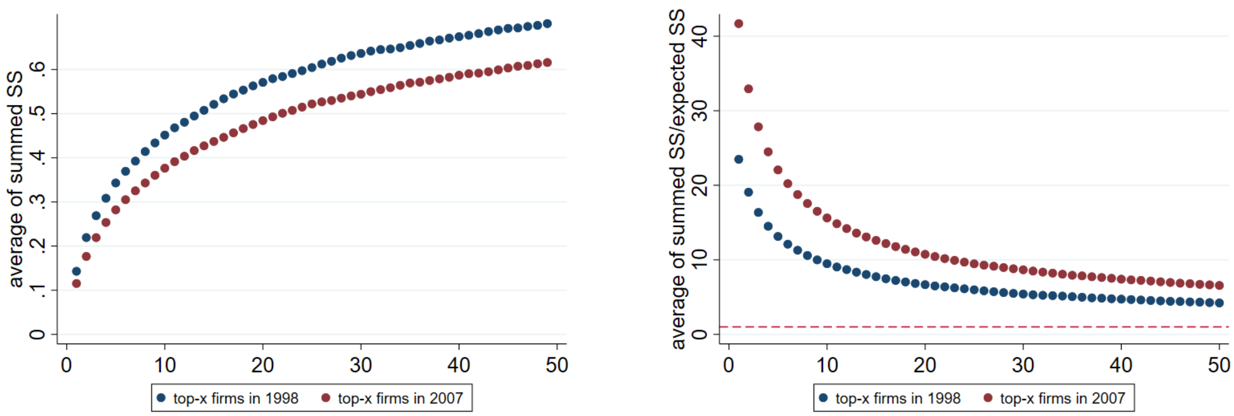
Chinese industries are indeed very granular. Across the 420 industries, the average total sales share of of the largest firm per year is roughly between 15% and 11%. For the subset of export shares this figure ranges from about 30% to 22%, as shown in the summary statistics on top-firms presented in table A2 of the appendix. If all firms took an equal share of sales in 1998, then they would expect a total sales share of only 0.3% across industries on average, decreasing to about 0.13% by 2007 given the growth in total number of firms. There is not an equally large decrease in the average share going to the top firm, implying an increase in granularity as the Chinese economy grew and new firms entered the market. Moreover, the single largest firm is not the only possible source of granularity in a typical industry, as there may be a number of super large firms that together take outlier shares of the market.

Figure 1 displays as much, where all possible groupings of top-x firms are laid out up to the grouping of the top-50 firms. For the single largest firm per industry, or group sizes of 1, the average sales share is about 14.3% across industries in the year 1998, as listed in table A2 and plotted in figure 1a. The second largest firm on average takes about 7.7% of total sales, making a group total of 22%. The group of top-3 firms then averages at 27%, still a substantial increase. Soon this marginal increase from adding the next largest firm becomes less substantial, and the plot in figure 1a tappers off as less and less granular firms (firms that are much more average in size) are included.

While the average shares are lower per respective group of granular firms in 2007, there are more firms in a typical industry, and normalizing the average shares for this in figure 1b shows that granularity is in fact becoming more pronounced in relative terms over the time period. In 2007, the single largest firm per industry was on average 41.7 times larger than what would be expected if all firms were of equal size. The cotton textile industry had the largest firm relative to its industry, at 827 times the average of 10,308 firms in 2007, and its top-3 firms had a combined size 350 times the average per 3 firms. The auto parts and accessories industry had the second most disproportionately large firm at 435 times the average of 7,579 firms in 2007, though its top-3 firms had a combined relative size 189 times the average per 3 firms, only the 4th highest of industries for that grouping. It should also be noted that this is among firms with greater than 5 million RMB of sales, meaning the smallest firms are cut out and thus these relative size figures are lower-bound estimates.

This paper settles on the threshold of top-3 firms for each group of FFFs and DFFs to build proxies for an industry's degree of granularity according to whichever metric is under consideration.

Figure 1: Grouping granular firms



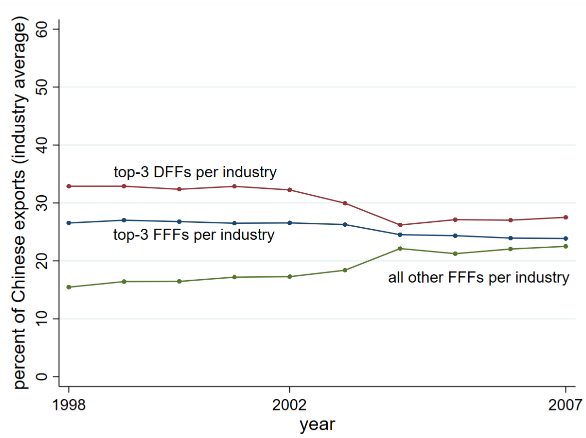
(a) Mean SS of top X firms

(b) Number of times sector mean

Note: Each X-axis denotes the group of top-x firms, per industry where there are at least as many firms. Left shows averaged summed sales shares (SS) across all sectors: $\sum_{z=1}^{Z}\sum_{i=1}^{x} SS_{z,i,t}/Z$. Right normalizes this by dividing it by x/I. It is the number of times greater a group's share is compared to the hypothetical homogeneous share, giving a lower bound of 1.

Figure 2 shows that on average across the 420 industries, the top-3 FFFs by exports contribute more than all other FFFs combined, and are almost at the same level as the top-3 DFFs, indicating the importance of superstar FFFs across a broad set of industries. Table A2 of the appendix shows that together the top-3 FFFs and top-3 DFFs account for over 50% of industry exports on average. These figures may suggest that China's FDI success story may more specifically be a success story of superstar FDI.

Figure 2: Accounting for China's industry-level exports



Note: The graph shows average combined industry export share of the top-3 DFFs, top-3 FFFs, and all other FFFs not in the top-3 per industry.

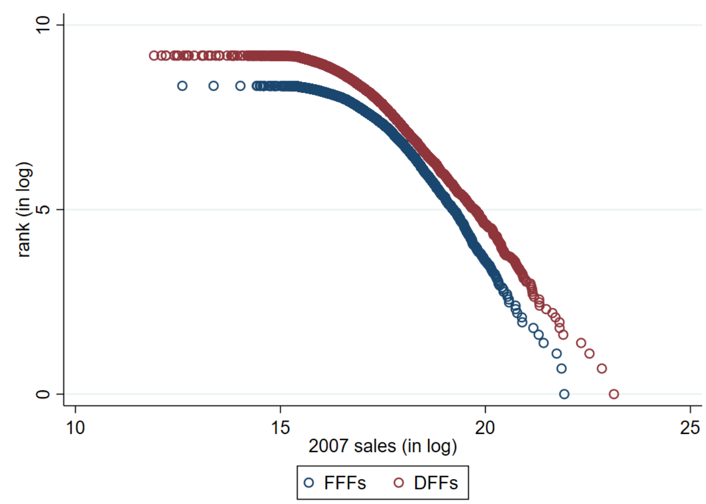
3.3 Heterogeneity in firm size distributions

To justify the separation of granularity into the two groups of DFFs and FFFs so to facilitate the identification of a granularity mechanism behind industry export performance, as well as the estimation of the impact of granularity on non-granular firms in later analysis, this section presents evidence of their systematically differing firm size distributions across industries. Section 3.4 then introduces the precise definitions of granularity used for regression analysis and discusses further the differing characteristics of the two types of superstar firms.

There is a strong power-law distribution of firm size for most industries. Assuming the power law probability distribution $p(x) = \Pr(X = x) = Cx^{\ \alpha}$ from lower bound $x_{min} > 0$ and normalization constant $C = (\alpha - 1)x_{min}^{\alpha\ 1}$, α is estimated separately in the sub-samples of DFFs vs FFFs. A lower α implies stronger potential of granularity via more inequality, in the sense that small differences in firms gives large differences in size, exports, sales, etc. In the cumulative distribution function (cCDF), a higher power law exponent is reflected by a sharper drop in the right end, where as lower exponents stretch out further in their right ends.

A pattern emerges of statistically different alpha estimates for DFFs and FFFs, with alpha estimates for DFFs systematically lower than that of FFFs, meaning that there is more inequality among DFFs than FFFs in a typical industry and greater dispersion at the fat-tail end. Details on how the estimates compare across industries and at different levels of aggregation are discussed in section A3 of the appendix.

Figure 3: Example of firm sales rank to sales cCDF: Textile and apparel industry



Note: shows the cumulative distribution function (cCDF) of sales, equivalent to the descending rank (log value) of firms by sales plotted against the firms' sales (log value).

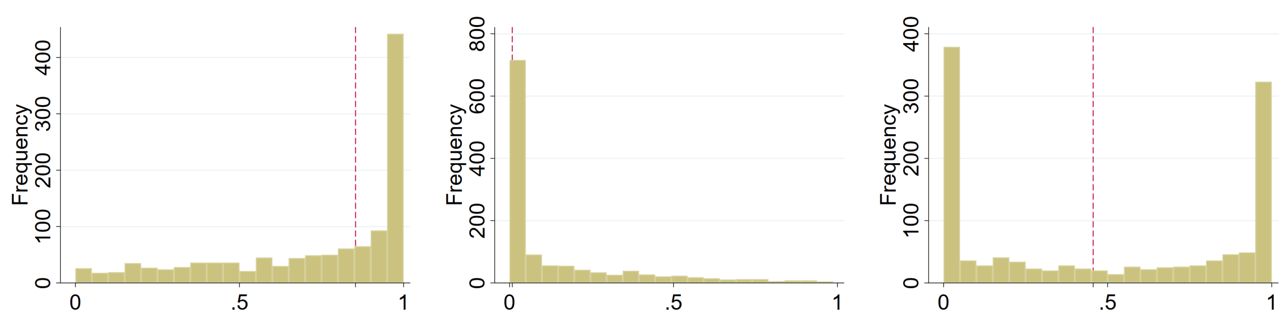
As an example, figure 3 shows the cCDF of total sales, equivalent to the descending rank (logged) of firms by sales plotted against firms' log of sales, for textile and apparel manufacturing, a typical four-digit industry. The near-straight line following some minimum level of sales on the log-log scale indicates a strong power-law distribution in firm size. The estimated power-law exponents are 2.31 and 2.51 for DFFs and FFFs, respectively. This industry is also typical in that there are roughly 3 firms at the far right tail for both FFFs and DFFs that are visually spaced apart from the next few largest firms on the log scale in their respective groups, indicating their granularity or distinct large size.

3.4 The granularity proxies

The panel data provides at least 5 logical metrics for sorting firms according to size: by total sales, exports, domestic market sales, total employment, and total wage bill. Section A2 of the appendix highlights the changing dominance of FFFs and composition with respect to HKMT or non-HKMT ownership, the declining presence of state-owned firms, and the increasing presence of private domestic firms in the top-6 positions across these different metrics.

For detecting granular comparative advantage (GCA), the methodology for which is presented in section 4 below, the preferred measure of granularity is in terms of domestic market sales, as it avoids systematic correlation with exports. The issue with applying this approach to a developing country is that there are many large exporters that do not sell to the domestic market, but instead use the country only as an export base to other markets. At the same time, China's large market is able to sustain manufacturing firms that are very large despite zero exports. Thus, many granular FFFs either only export or only focus on local market. Figure 4 shows histograms for the export intensity of all top-3 FFFs across industries as defined by exports (left), domestic market sales (middle), and by employment (right). There is a mass of top exporters with an export intensity of 1, many of which are indeed very large firms. Conversely, there is a mass of top domestic market sellers with an export intensity of 0. Top employers represent a more equal mix.

Figure 4: Export intensities of top-3 FFFs in 2007



(a) FFFs granular in terms of exports

(b) FFFs granular in terms of domestic market sales

(c) FFFs granular in terms of employment

Note: The red dashed line indicates the median among the top firms.

Thus, defining granularity by domestic sales creates a horizontal FDI (domestic market access) bias. Defining granularity by employment is more neutral and captures large exporters, large domestic market sellers, as well as more firms in the middle of the spectrum. I therefore rely on two primary proxies for granularity; both that by domestic market sales and that by employment.

A comparison between the two groups of granular DFFs and FFFs as defined by employment is shown in table 1, where firm characteristics are regressed on dummies for top-3 FFFs and top-3 DFFs, while including industry and year fixed effects. For more insight, I also include dummies for large, but not "granular" FFFs and DFFs, defined as those firms with more than 100 employees but not in their respective top-3 spot. All other firms are excluded. The granular DFF dummy is set as the reference group, meaning that the estimated coefficients are relative to granular DFFs. Firm age and a state-owned firm dummy are also included, but coefficients generally see only small changes in the expected directions if these two control variables are left out.

When comparing a typical granular FFF to a typical DFF, the granular FFF has a higher export intensity by 0.26 on the range 0 to 1, lower markups by 0.017 log points, a higher real capital-labor ratio by 0.297 log points, a higher labor productivity by 0.184 log points, and a higher average wage rate by 0.268 log points. Not controlling for state-owned firms leads to a doubling of the labor productivity coefficient on the dummy for granular FFFs, but all other coefficients see only minor adjustments (not shown). Compared to non-granular DFFs, granular DFFs have a higher real capital-labor ratio by 0.289 log points, higher labor productivity by 0.126 log points, and higher average wage rate by 0.119 log points. Additionally, both granular groups and non-granular but large FFFs employed slightly more college graduates compared to large but not granular DFFs. Notably, only granular FFFs have an average labor productivity significantly higher than granular DFFs, setting them apart from non-granular FFFs or FDI in general.

As mentioned, the following sections primarily rely on defining granularity by domestic market sales or by employment. Let $\widetilde{s}_{z,i,t}$ denote some share of industry z activity concentrated in firm i in year t. These shares are summed over the top-3 firms, either without respect to funding or separately for each group of FFFs and DFFs, depending on the application. The corresponding measure for granularity used in the following analysis is then:

$$\sum_{i=1}^{3} \widetilde{s}_{z,i,t} = \sum_{i=1}^{3} \textit{domestic sales share}_{z,i,t} \quad \textit{or} \quad \sum_{i=1}^{3} \textit{employment share}_{z,i,t} \tag{1}$$

Deriving the measures of granularity as pertains to the methodology of detecting GCA with FDI is discussed in more detail in section 4 below. I also use a definition by wage bill share for additional robustness checks. The heterogeneous characteristics of granular firms captured by employment shares versus wage bill shares is discussed in section A4 of the appendix, where I argue that defining granularity by employment shares is most representative of vertical (efficiency

Table 1: Characteristics of granular and other large firms (defining $\widetilde{s}$ by employment)

	(1) export intensity	(2) ln markups	(3) ln KL ratio	(4) ln labor productivity	(5) ln mean wage rate	(6) ratio grads to staff (2004)
top-3 FFFs	0.260***	-0.017***	0.297***	0.184***	0.268***	0.004
other large FFFs	0.219***	-0.072***	0.200***	0.021	0.176***	-0.005
other large DFFs	-0.057***	-0.086***	-0.289***	-0.126***	-0.119***	-0.020***
State-owned dummy	-0.033***	-0.016***	-0.003	-0.266***	-0.109***	0.004***
firm age	0.000	-0.000	0.008***	-0.011***	0.002***	-0.000***
Constant	0.239***	0.355***	3.678***	3.907***	2.472***	0.060***
Industry & Year FE	Y	Y	Y	Y	Y	Y
Observations	1058852	1058852	1055558	1058852	119773	86794

Granular DFF is the reference dummy. KL ratio is real capital / workers, labor productivity is real value added / workers. Not controlling for state-owned firms leads to a doubling of the labor productivity coefficient on the dummy for granular FFFs, but all other coefficients see only minor adjustments (not shown).

seeking) granular FDI, or least biased towards horizontal granular FDI.

Note that in the regression analysis of the following sections, granularity is defined at the industry level for each group of DFFs and FFFs separately. However, the calculated shares for the top-3 FFFs or top-3 DFFs are out of the industry total, with all firms included in the total without respect to funding type. For example, if the top-3 FFFs together account for 10% of industry employment, and the top-3 DFFs together account for 15%, then the remaining non-granular firms account for 75% of industry employment. The granularity variables (by employment) in this case would be equal to 0.10 for FFF granularity and 0.15 for DFF granularity.

4 GCA with FDI: Methodology

This section motivates the baseline empirical specification for analyzing industry-level granular comparative advantage from the two groups of DFFs and FFFs, results for which are presented in section 5. Methods utilized for detecting pro-competitive effects of granular firms in China onto non-granular firms and for detecting a new source of horizontal spillover effects from FDI on firm-level TFP and labor productivity are presented in sections 6 and 7, respectively.

Gaubert and Itskhoki (2021) develop a Ricardian comparative advantage model building on Dornbusch et al. (1977) with Melitz (2003) firm heterogeneity but with finite firms as in Eaton et al. (2012). Their model combines Fundamental CA across sectors (things common to all firms i.e. availability of specific human capital, infrastructure, and technology) with Granular CA of individual firms (idiosyncratic contribution of a firm, i.e. specific know-how).

For industry exports, X_z, and domestic industry expenditure, Y_z, which is the sum of domestic sales and imports, define $\Lambda_z = X_z/Y_z$, the export intensity of industry z that varies one-to-one with CA. This industry-level measure can be rewritten as the sum of the products of firm domestic market shares, $s_{zi} = d_{zi}/Y_z$, where d_{zi} is the sales of firm i to the local or domestic market in which it produces goods, and export intensities as defined by the firm's exports from that market relative to what is sold locally, $\lambda_{zi} = x_{zi}/d_{zi}$:

$$\Lambda_z = \sum_{i=1}^{N_z} s_{zi}\lambda_{zi} = \mathbf{s_z}^{\prime}\boldsymbol{\lambda}_\mathbf{z} \tag{2}$$

where N_z is the finite number of firms in industry z. This summation makes clear the exposure of an industry's export intensity to granular elements at the firm level, whereby a disproportionately large firm may have a disproportionately large impact on industry level Λ_z as s_{zi} becomes larger, and without a continuum of firms to guarantee it averages out in the aggregate. If such a large firm has the same export intensity as all other firms in its industry, then its existence does not directly alter Λ_z. However, if the large firm's export intensity is different from a typical firm in its industry, as is often found in empirical observations, then industry level Λ_z in the presence of that granular firm is different from Λ_z if that firm were to disappear.

There is also the possibility that the presence of a large, granular firm boosts or diminishes the export performance of related firms beyond what would be expected if the granular firm were replaced by an equivalent number of non-granular firms. This could owe to crowding out smaller firms from the domestic market, forcing them to seek opportunities in international markets, or it could owe to spillover effects either affecting productivity or supply chains. These additional reasons are especially relevant to the case of a developing country.

A stochastic data generating process for each industry, $(\lambda, \mathbf{s}) \sim F_z(\cdot)$, gives realizations of observed firm-level market shares and export intensities. The population mean of Λ_z has an expected value defined by the distribution of $F_z(\cdot)$, which therefore describes industry characteristics, or benefits of comparative advantage, accessible to all firms in the industry z. Deviations from this expected value owe to the granular components of industry export intensity provided by outlier firms that do no average out in the aggregate. This decomposition is written as:

$$\Lambda_z = \Phi_z + \Gamma_z \quad \text{where} \quad \Phi_z = \mathbb{E}_z\{\Lambda_z\} = \int s^{\prime}\lambda \; dF_z(\lambda, s). \tag{3}$$

Φ_z is the expected value of industry export intensity, and $\Gamma_z = \Lambda_z - \Phi_z$ is the granular component, also called the granular residual. Φ_z and Γ_z are orthogonal in the cross section of sectors z, thus variance decomposition holds:

$$var\ (\Lambda_z) = var\ (\Phi_z) + var\ (\Gamma_z). \tag{4}$$

That is, the variance of industry-level export intensity, which varies one-to-one with comparative advantage, splits between variance in the expected value of export intensity and variance in the granular residual. Φ_z is therefore representative of a fundamental comparative advantage (FCA), common to all firms, and Γ_z represents granular comparative advantage (GCA).

In the literature on firm productivity, draws are typically from a single distribution for some level of disaggregation, as is the case in Gaubert and Itskhoki (2021). The current paper argues that, especially in the context of developing economies, there are at least two fundamentally different categories of firms within each industry. Specifically, there is the group of local domestically-funded firms and the group of foreign-funded firms that result from FDI. Foreign multinationals enter a country to take advantage of local comparative advantages (vertical FDI), or to capitalize on a growing domestic market (horizontal FDI). Upon entry, they often already have huge supply and demand networks, and years of expertise. Their utilization of technology may be vastly different from local firms, as is evident in the literature on FDI spillovers. Assuming multinationals that enter the host country are comprised of similar fundamentals and grouped together, variance decomposition becomes:

$$var\ (\Lambda_z)\ = var\ (\Phi_z^{DFF}) + var\ (\Phi_z^{FFF}) +\ var\ (\Gamma_z^{DFF}) + var\ (\Gamma_z^{FFF}), \tag{5}$$

with $\Lambda_z = \sum_{i=1}^{N_z^{DFF}} s_{zi}\lambda_{zi} + \sum_{i=N_z^{DFF}+1}^{N_z^{DFF}+N_z^{FFF}} s_{zi}\lambda_{zi} = s_z^{0DFF}\lambda_z^{DFF} + s_z^{0FFF}\lambda_z^{FFF}$. The pairs $s_z^{0DFF}\lambda_z^{DFF}$ and $s_z^{0FFF}\lambda_z^{FFF}$ are realizations of two separate stochastic DGPs: $(s_z^{0DFF}, \lambda_z^{DFF}) \sim F_z^{DFF}(\cdot)$ and $(s_z^{0FFF}, \lambda_z^{FFF}) \sim F_z^{FFF}(\cdot)$, possibly with different population means. Λ_z then writes as the sum of integrals for both groups of firms following their respective distributions as $\Lambda_z = \Phi_z^{DFF} + \Phi_z^{FFF} + \Gamma_z^{DFF} + \Gamma_z^{FFF}$, which yields the variance decomposition above. Even if $\Phi_z^{DFF} = \Phi_z^{FFF}$ for the case of China, there is strong evidence presented throughout this paper that Γ_z^{DFF} and Γ_z^{FFF} are orthogonal.

The existence of granular comparative advantage in China would see the degree of industry granularity correlated with aggregate industry export intensity. This can be detected by regressing industry exports on granularity proxies; namely the concentration ratio describing how dominant the top-few DFFs and FFFs are per industry. The domestic market share s_{zi} defined above requires imports data, which is not available at the level of disaggregation desired for this paper, as is also the case in Gaubert and Itskhoki (2021). Thus $\widetilde{s}_{z,i,t}$ denotes an alternative share of industry activity or size, which is additionally given the superscript FFF or DFF to describe FFFs or DFFs, respectively. As mentioned in section 3.4, I define $\widetilde{s}_{z,i,t}$ in different ways that are relevant to building granularity proxies for the application to China.

First, I follow Gaubert and Itskhoki (2021) by computing $\widetilde{s}_{z,i,t}$ from domestic market sales. This gives $\widetilde{s}_{z,i,t} = domestic\ sales\ share = d_{z,i,t}/\sum_{i=1}^{N_z} d_{z,i,t}$, where $d_{z,i,t}$ is a firm's domestic market sales and is equal to their total sales originating from their mainland-China production minus what is exported out of main-land China, and N_z is the total number of firms in industry z. This definition of $\widetilde{s}_{z,i,t}$ is not mechanically correlated with exports and thus is the preferred share to use in building granularity proxies. However, there are many large, export-intensive firms in China do not sell much to the domestic market, and will therefore be underrepresented by this measure. More specifically, if thinking of FDI as being either horizontal or vertical in nature, ranking firms by domestic sales leads to a proxy of granularity more relevant to FFFs established in the emerging economy for horizontal-type, market access reasons, as was discussed in section 3.4. Therefore, I also define $\widetilde{s}_{z,i,t}$ as a firm's share of industry employment (or number of workers), whereby $\widetilde{s}_{z,i,t} = \textit{employment share} = e_{z,i,t} / \sum_{i=1}^{N_z} e_{z,i,t}$ and $e_{z,i,t}$ is a firms' total number of workers. Industry granularity is then defined as in equation 1 of section 3.4. A firm's share of the industry wage bill is also used in robustness checks, whereby $\widetilde{s}_{z,i,t} = \textit{wage bill share} = wb_{z,i,t} / \sum_{i=1}^{N_z} wb_{z,i,t}$ and $wb_{z,i,t}$ is total the wage bill paid out by the firm, with a contrast of this definition and that by employment given in section A4 of the appendix. I maintain the notation of $\widetilde{s}_{z,i,t}$ for the three definitions for simplicity and to make clear that only one is used at a time with the idea to compare result when using the different definitions. To summarize, $\widetilde{s}_{z,i,t} = \textit{domestic sales share}$ or *wage bill share* or *employment share*, as specified per regression.

The granularity proxies are then defined to be the concentration ratio among the top-3 DFFs, $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$ and the top-3 FFFs, $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$. The maximum for both of these concentration ratios added together is 1. Then the baseline regression for detecting GCA in China is:

$$\log X_{z,t} = \alpha + \beta^{DFF} \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF} + \beta^{FFF} \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \log D_{z,t} + \delta_z + \delta_t + \varepsilon_{z,t}, \tag{6}$$

where β^{FFF} and β^{DFF} are the estimates of interest that relate granularity owing to FFFs and DFFs, respectively, to industry exports, and $D_{z,t}$ is a control for the industry size. $D_{z,t}$ is defined according to whichever definition of $\widetilde{s}$ is used, so either the aggregate domestic sales among firms in mainland China, the aggregate wage bill, or the aggregate number of workers. As $\widetilde{s}$ is a share, the granularity proxies are not necessarily correlated with industry exports unless through a granular residual. In other words, if industry exports are not correlated with the size of the largest firms relative to other firms or relative to the industry norm in a manner that does not extend beyond the direct effect on $D_{z,t}$, then the coefficient on these shares would be zero.

Each concentration ratio in equation 6 is over the top-3 firms per industry-funding group, thus yielding 6 firms per industry for the case of two groups, which seems ideal for capturing any obvious presence of granularity from these two groups of firms for the case of China, as discussed in sections 3.2 and 3.3. There is a trade-off in that industries are also very granular and thus the larger ones may have many more firms that can still be considered granular compared to a larger mass of non-granular firms, but empirical results, all of which include industry and year fixed effects, remain similar if reducing or slightly enlarging the number of top firms. The granular terms are left as shares given that taking logs does not improve the normality of residuals in the base regressions, and would go against the point of granularity that there are expected outliers resulting from the fat tail Pareto distributions.

As the Chinese economy was in earlier stages of development during the years 1998-2007, the limitations of running regressions on equation 6 quickly become apparent in the the comparatively less diversified economy has some industry-years with zero exports. Thus taking logs for these industry exports in the dependent variable eliminates the observation, leading to sample selection bias. There are 47 such industry-year observations out of 4,200, though every industry still has at least one year with positive exports. As these zeros are the worst performing industries in terms of export competitiveness, the bias can be significant if excluding them. The non-linear Poisson pseudo-maximum likelihood estimator is appropriate for handling zeros or near zero values, for which the regression equation is:

$$X_{z,t} = \alpha + \beta^{DFF} \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF} + \beta^{FFF} \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \log D_{z,t} + \delta_z + \delta_t + \varepsilon_{z,t}, \tag{7}$$

for the conditional mean of the exponential form $E(X|regressors) = exp(regressors^{\prime}\beta)$, which is true as long as industry exports are non-negative. The first order conditions on the corresponding log-likelihood function are of the form $\sum_{i=1}^{N} X_{z,t} - exp(regressors^{\prime}_{z,t}\beta) = 0$, where $\hat{\beta}$ is found numerically as there is no explicit solution. The partial elasticity for the FFF granularity term is then

$$\frac{\partial \log E(X_{z,t}|regressors_{z,t})}{\partial \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}} = \beta^{FFF} \tag{8}$$

and similar for the partial elasticity of the DFF granularity term.

Indeed the Poisson estimator has been shown to be more appropriate for analyzing exports under the presence of heteroskedasticity, regardless whether or not zero values are present, since Santos Silva and Tenreyro (2006). As it is a pseudo-maximum likelihood estimator, the data is not required to be count data or follow a Poisson distribution and in fact remains consistent for any distribution on the data. Additionally, the Poisson estimator in particular is consistent in the presence of two-way fixed effects, whereas, for example, the tobit estimator is not due to the incidental parameters problem. The Poison estimator is also scale invariant.

The three forms of $\widetilde{s}_{z,i,t}$ yield three distinct empirical specifications. Comparisons between these empirical specifications allows looking at the association of GCA with a set of relatively market seeking foreign firms versus a less biased set including more efficiency seeking firms. To address endogeneity between the granularity proxies and industry exports, a Poisson 2sls exercise

Table 2: Granular CA via domestic sales share

	1998-2007 full FEs			first-diff 2-dig FE			Poisson full FEs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}$	-1.952***			-1.088***			-0.814***		
	(0.73)			(0.41)			(0.24)		
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		-1.370**			-0.710*			-0.809***	
		(0.65)			(0.37)			(0.23)	
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			0.358			0.124			-0.471
			(0.60)			(0.43)			(0.34)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			-2.070***			-1.219**			-1.194***
			(0.76)			(0.49)			(0.29)
$D_{z,t}$	1.023***	1.034***	0.964***				0.396***	0.384***	0.359***
	(0.20)	(0.21)	(0.20)				(0.07)	(0.07)	(0.07)
Industry & Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-sqr	0.445	0.439	0.450	0.012	0.009	0.016			
Observations	4200	4200	4200	3780	3780	3780	4200	4200	4200

Note: DV is industry exports. In all columns $\widetilde{s}$ is defined by domestic sales and $D_{z,t}$ is industry domestic sales. Columns (4) - (6) takes first differences and includes year and two-digit industry fixed effects. Columns (7) - (9) repeat (1) - (3) (all with year and 4-digit industry fixed effects) using a Poisson regression, where the DV is not logged so that zero-export industries are included without needing to add 1 unit before taking logs as is done in columns (1) - (6). There are generally more DFFs among an industry's top 3 firms than FFFs. Thus, the first row with granular firms by any funding for columns (1), (4) and (7) is biased towards DFFs. Columns (2), (5), and (8) put together the industry shares of the top 3 DFFs and top 3 FFFs, so to be directly comparable to columns (3), (6), and (9), where these two groups are separated. Coefficients are similar if specifying granularity as the top 2 or 4 firms' industry shares in each group. Robust errors; * p<0.10, ** p<0.05, *** p<0.01.

using capital investments as an IV as well as an out-of-sample prediction exercise are presented after the baseline results and discussion of other robustness checks specific to the Chinese data in section 5.

5 GCA with FDI: Results

5.1 Base results

Table 2 shows results from estimating equation 6 using domestic sales share to identify and sum together the top firms in each industry. As argued in section 3.4 and discussed further in section A4 of the appendix, defining granularity by domestic sales shares biases the analysis towards a representation of horizontal FDI for the case of China, but at the same time carries the least empirical concerns as domestic sales shares are not systematically correlated with exports. Furthermore, the variables for DFF domestic sales granularity and FFF domestic sales granularity have a near zero correlation of -0.05.

$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}$ in columns (1), (4), and (7) is for the top-3 firms by domestic sales regardless of funding type. Columns (1) and (4) are directly comparable to columns (4) and (5) in table 1 of Gaubert and Itskhoki (2021). Whereas they found a positive association for the application to France, the current paper finds evidence of strong negative granular residual in describing variation in exports among Chinese industries. That is, larger industry concentration among top-firms in China correlates with lower total exports in their respective industries, thereby providing evidence that granularity hinders comparative advantage in China, and GCA in developing economies may indeed have the opposite impact on over all comparative advantage compared with developed economies.

Some industry-years have zero exports in China over the panel of analysis, and thus exports were incremented by one unit to obtain the results shown in columns (1) through (6) of table 2. Columns (7) - (9) show the equivalent Poisson specification to that of columns (1) - (3), which accommodates for zero-export industry-years without needing to increment the data. Including these zero export observations is important as they may be zero or break away from zero for both fundamental and granular reasons. The concern that such industries may be unusually constrained or too small is addressed by the fixed effects and the industry size parameter $D_{z,t}$. Poisson regression is therefore the specification of choice for this dataset whenever industry exports are the dependent variable, as was discussed in section 4. For the Possion regression in column (7) the interpretation is that an additional 10% of total domestic market sales (from operations located within mainland China) concentrated among the top-3 firms (regardless of funding-type) associates with a 7.82% (exp(-0.814x0.10)-1) decrease in industry exports.

Columns (2), (5), and (8) see the variable of interest created by adding the domestic sales shares of the top-3 FFFs and top-3 DFFs together, or $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$. This is done simply to show what happens when the two groups of aggregated top-firms are then split into two different sources of granularity, $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$ and $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$, in columns (3), (6), and (9). Looking specifically at the Poisson specifications; from the variable of interest made from only 3 firms per industry-year in column (7) to expanding to 6 firms in column (8), there almost no difference in the estimated coefficients. Now when splitting the 6 firms into two groups in column (9) it is clear that the bulk of negative association of the granularity proxy with exports is coming from the largest DFFs, with the coefficient on the granularity proxy for FFFs being indistinguishable from zero.

Columns (4) - (6) capture a more dynamic story regressing first-differences of all variables and implement 2-digit industry fixed effects. The dynamic specification strongly maintains that as the degree of granularity in an industry increases there is an associated decrease in that industry's export stance, still the opposite of findings presented in Gaubert and Itskhoki (2021) for France. Given the only difference between columns (1) and (7) are that (1) is a log-linear regression while

Table 3: Granular CA via industry wage and worker shares

	Defining $\widetilde{s}$ by wage bill			Defining $\widetilde{s}$ by employment		
	(1)	(2)	(3)	(4)	(5)	(6)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}$	-0.697***			0.120		
	(0.24)			(0.29)		
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		-0.692***			0.272	
		(0.23)			(0.27)	
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			-0.033			1.495***
			(0.35)			(0.40)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			-1.350***			-0.807**
			(0.26)			(0.39)
$D_{z,t}$	0.860***	0.838***	0.804***	0.793***	0.795***	0.728***
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Industry & Year FE	Y	Y	Y	Y	Y	Y
pseudo-R-sqr	0.980	0.980	0.981	0.979	0.979	0.980
Obs	4200	4200	4200	4200	4200	4200

Note: All are Poisson regressions, comparable to columns (7) - (9) of previous table. $\widetilde{s}$ is wage bill shares for columns (1) - (3) but changed to employment shares for columns (4) - (6). Both resulting proxies capture vertical FDI more than domestic market share, with worker share more so than wage bill share. $D_{z,t}$ is industry total wage bill for columns (1) – (3) and total number of workers for columns (4) – (6). Robust errors; * p<0.10, ** p<0.05, *** p<0.01.

(7) is a Poisson regression, the largely different coefficient estimates of -1.952 and -0.814 indicate that the log-linear model may suffer from heterogeneity-induced bias as discussed in Santos Silva and Tenreyro (2006). Perhaps more importantly, if excluding the 47 zero export industry-year observations the coefficient of interest in column (7) hardly changes, as expected given that Poisson regressions have been shown to be robust to zeros in the literature, while the coefficient when re-estimating column (1) adjusts drastically to -0.590 with a standard error of 0.343. In other words, the Poisson estimator is stable and the log-linear OLS model is not.

While table 2 shows no association of granularity from FFFs with exports, table 3 offers a different perspective by using the two alternative methods of sorting firms to identify granularity in an industry, which as mentioned, allow better representation of vertical FDI.

Firms are ranked by wage bill shares for columns (1) through (3) and ranked by employment shares for columns (4) through (6), where these three columns for each group repeat the Poisson regressions of columns (7) through (9) in table 2. The coefficient on granularity from a single source remains negative when creating the proxy from wage bill shares, but zeros out when using employment shares. Similarly, the coefficient on granularity from FFFs remains indistinguishable from zero when using wage bill shares but turns strongly positive at 1.495 when using employment shares. Thus, an additional 10% of total employment concentrated among top-3 FFFs associates with a 16.1% ([exp(1.495x0.10)-1]) increase in industry exports. The coefficient on granularity from DFFs, however, remains negative in either specification, though less negative in the latter. This demonstrates a positive trend in detected GCA as the set of granular firms transitions to firms that are more efficiency-seeking or relatively oriented towards vertical FDI, as argued in sections A2 and A4 where sorting by employment shares is shown to be particularly representative of efficiency-type motives. The IV strategy discussed later and shown in table 7 strengthens these results.

These results indicate large firms locating in mainland China for vertical FDI reasons can offer a positive granular residual, thereby providing a boost to comparative advantage. Taken all together, the variations of regressions for equation 6 suggest that letting in the largest firms with market access intentions does not aid the global competitiveness of domestic industries, while letting in the largest efficiency seeking firms is the only source of positive GCA.

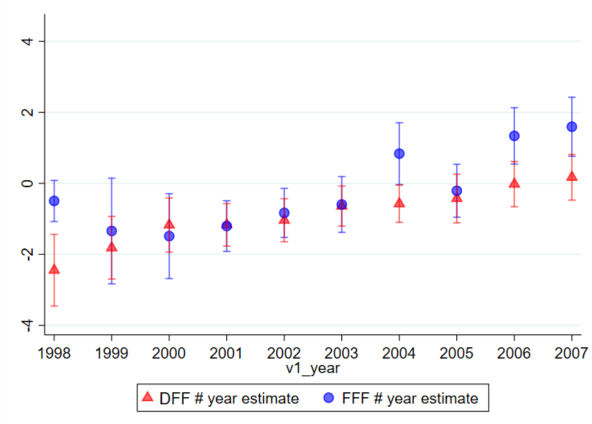
Granularity shows signs of changing its relationship with comparative advantage over the years of the panel in China. Indeed, when modifying the regression in column (6) of table 3 by introducing the interaction of year dummies with the granularity terms while retaining both year and industry fixed effects, a notable convergence becomes evident. Figure 5b illustrates this result, where by 2007 the estimated coefficients on industry DFF and FFF granularity (by employment) are indistinguishable. Meanwhile, Figure 5a shows that while both groups see a negative association with exports of their granularity when proxying by domestic sales in the initial years of the survey data, there seems to be an upward trend. In particular, this relatively domestic market-oriented proxy for granularity sees significantly positive coefficients by the end of the sample for FFFs, with coefficients becoming very similar to when proxying by employment shares.

5.2 Ruling out bias from state-owned firms

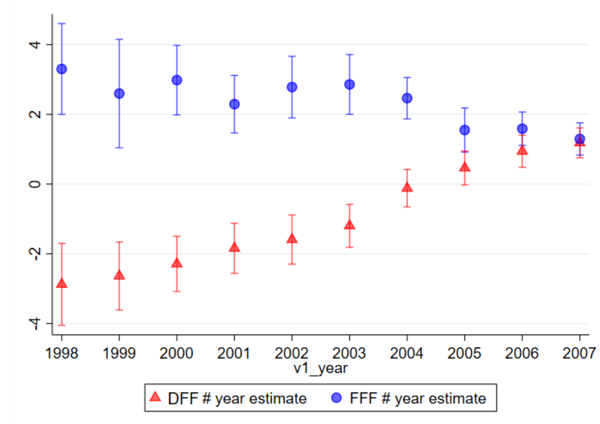
Granularity owing to state-owned firms could arguably go in both directions. It could help build positive granular residual in comparative advantage due to the fact that state-owned firms often prevail in industries where large private firms may face fundamental comparative disadvantage, especially in this era of privatization of more competitive industries. However, state-owned firms may potentially constrain industries if they absorb available funding or underbid others, for example. It is therefore a useful robustness check to adjust the analysis by excluding state-owned firms from the ranking of top DFFs.

Table 4 presents the same Poisson regressions as tables 2 and 3 for the two sources of granularity, and also adds to equation 6 the term $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{SDFF}$ to control for granularity from state-owned

Figure 5: Convergence of coefficients on granularity proxies



(a) defining $\widetilde{s}$ by domestic sales



(b) defining $\widetilde{s}$ by employment

Note: Plots coefficients of a single Poisson regression equivalent to column (9) in table 2 except for adding year interactions on each of the two granularity proxies. Industry and year fixed effects included. Robust errors with solid fill indicates significance past the 90 percent level.

Chinese-funded firms, where $\widetilde{s}^{SDFF}_{z,i,t}$ is a given state-owned firms' share and $n = 1$, 2, and 3 are the three largest state-owned firms by the specified method of sorting and calculating industry shares. The term $\widetilde{s}^{DFF}_{z,i,t}$ is therefore now calculated excluding any state-owned firms, and thus represents private domestically funded firms in China. The coefficients on the proxies for granularity from FFFs and the altered DFFs remain very similar to before, indicating state-owned firms did not bias results, and indeed granularity from them performs rather similarly to that of private DFFs.

5.3 Predictive power out of sample

Reverse causality is a potential source of endogeneity for granularity explaining exports. Export growth at the industry level could fuel consolidation by the fact that opening an industry to trade brings import competition and more economies of scale to those firms that begin to export. Though while openness is certainly a prerequisite of foreign firms entering a country and scaling production, it is unclear whether the granular firms are primarily responsible for generating exports or if their operations are heavily influenced by the presence or absence of exports. In contrast, granular DFFs may have experienced import competition, resulting in reallocation to more productive firms.

Thus, if there is reverse causality from industry exports onto either granularity from FFFs or DFFs, it should be that variation in granularity is explained by past exports or some combination of past exports and industry characteristics beyond dynamics from exports' mean reversion. This is tested in table 5, where in column (1) I first regress 10-year change in worker granularity from DFFs on the 1998 levels of log of industry exports, the size control of log industry-workers, the two

Table 4: Separating state-owned firms from granular proxies

	$\widetilde{s}$ by domestic sales		$\widetilde{s}$ by wage bill		$\widetilde{s}$ by employment	
	(1)	(2)	(3)	(4)	(5)	(6)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$	-0.109	-0.601*	0.204	-0.110	1.642***	1.400***
	(0.41)	(0.33)	(0.36)	(0.35)	(0.40)	(0.40)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$, private	-0.598**	-1.168***	-0.811***	-1.452***	-0.476**	-0.843***
	(0.25)	(0.25)	(0.18)	(0.24)	(0.24)	(0.29)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$, state		-1.396***		-1.320***		-1.024**
		(0.40)		(0.31)		(0.44)
$D_{z,t}$	0.358***	0.358***	0.788***	0.786***	0.729***	0.716***
	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)
Industry & Year FE	Y	Y	Y	Y	Y	Y
pseudo-R-sqr	0.970	0.971	0.980	0.981	0.980	0.980
Obs	4200	4200	4200	4200	4200	4200

Note: All are Poisson regressions. State-owned firms are excluded from computation of the granularity proxies and given their own granularity proxy in (1), (3), and (5), while a control is added for total industry share owing to state-owned firms in (2), (4), and (6). The first two columns use domestic market shares to proxy for granularity, the next two use wage bill shares, and the last two use employment shares. Industry and year fixed effects included. Robust errors; * p<0.10, ** p<0.05, *** p<0.01

granularity variables, a full set of 2-digit industry fixed effects. 10-year change in granularity from DFFs is explained only by its partial mean reversion. Of note is that only granularity from DFFs is predictive of granularity from DFFs, with not even the size of the industry by workers having a significant coefficient within industries, an indication that granularity is not simply a feature of industry size or lack thereof. Column (2) repeats for granular FFFs. There is still no predictive power from exports onto the 10-year change in granularity. However, granularity from FFFs is pulled down by industry size in terms of workers, meaning any bias from a correlated size variable is likely negative rather than positive, giving more validity to its positive coefficient in explaining industry exports. Adding industry export intensity and its interaction with industry exports in column (3) still does not help in predicting 10-year change in granularity from FFFs. Finally, even when adding the 10-year change in exports and its interaction with 1998 exports in column (4), with these two explanatory variables being the only contemporaneous ones, all measures of exports remain insignificant in explaining variation in granularity from FFFs.

Conversely, column (5) shows granularity from FFFs is predictive of positive export growth over the 10-year period, even after controlling for mean reversion and the positive effect from industry worker size. This is an indication that any feedback from exports to granularity is controlled for when conducting a within-industry analysis with year fixed effects, as exports would be significant in columns (1) through (4) if not including 2-digit, though still not nearly as significant

Table 5: Predictive power of key variables

Dependent Variable:	$gDFFs_{z,07}-gDFFs_{z,98}$	$gFFFs_{z,07}-gFFFs_{z,98}$			$X_{z,07}-X_{z,98}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$X_{z,98}$	-0.005	0.004	-0.001	0.002	-0.501***	-0.489***	-0.568***
	(0.003)	(0.002)	(0.003)	(0.004)	(0.043)	(0.059)	(0.060)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,98}^{FFF}$	0.031	-0.741***	-0.755***	-0.768***	1.644*	1.530*	6.207***
	(0.073)	(0.046)	(0.047)	(0.048)	(0.909)	(0.898)	(2.022)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,98}^{DFF}$	-0.455***	0.044*	0.045*	0.060**	0.341	0.647	1.128**
	(0.042)	(0.026)	(0.027)	(0.027)	(0.532)	(0.523)	(0.523)
$log\ workers_{z,98}$	0.008	-0.017***	-0.012**	-0.013***	0.311***	0.174*	0.310***
	(0.006)	(0.004)	(0.005)	(0.005)	(0.078)	(0.095)	(0.097)
$export\ intensity_{z,98}$			0.141	0.168		-12.574***	-10.813***
			(0.153)	(0.157)		(2.958)	(2.944)
$\times X_{z,98}$			-0.006	-0.008		0.841***	0.733***
			(0.010)	(0.010)		(0.187)	(0.187)
$X_{z,07}-X_{z,98}$				0.004			
				(0.008)			
$\times X_{z,98}$				0.000			
				(0.001)			
$FDI\ cap\ share_{z,98}$							1.971***
							(0.477)
$\times \sum_{i=1}^{3} \widetilde{s}_{z,i,98}^{FFF}$							-11.200***
							(3.394)
2-digit industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-sqr	0.453	0.481	0.487	0.509	0.431	0.462	0.490
Observations	413	413	413	412	412	412	412
dfres	380	380	378	375	379	377	375

Note: All include 2-digit fixed effects. Granularity proxy from $\widetilde{s}$ defined by employment only. Regresses 1998 values on 10-year change for indicated DV. Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

as granularity from FFFs would be in explaining change in exports (not shown). Controlling for industry export intensity and its interaction with log industry exports does not take away from the explanatory power of granularity in workers from FFFs, with it remaining significant at the 10% confidence level. Further, adding the share of aggregate industry invested capital that is from FDI and its interaction with granularity from FFFs brings much significance to the main variable of interest, and it is revealed that granularity when not accompanied by existing high levels of FDI penetration in particular is predictive of positive export growth at the industry level. Note that if standardizing the beta coefficients here, granularity from FFFs and percent of industry invested capital from FDI both have equally large coefficients of around 0.18, while their interaction would be -0.12, meaning the net effect of granularity from FFFs is still very positive on 10-year change in industry exports.

5.4 Instrumental variable approach

Granularity in financial capital is correlated with granularity in domestic sales shares, wage bill, or employment. In order to address endogeneity with a 2SLS approach, I argue that injecting financial capital into a firm causes growth in scalables, such as wage bill or employment levels, and only through these scalables does financial capital translate to variation in exports. This logic is more direct for wage bill and employment than for domestic sales, but results when instrumenting for domestic sales and wage bill are very similar, and the fact the coefficient on DFF granularity remains negative is convincing.

The first stage thus regresses each granularity proxy on the available exogenous variables according to $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$ *and* $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} = \sum_{i=1}^{3} \widetilde{k}_{z,i,t}^{DFF} + \sum_{i=1}^{3} \widetilde{k}_{z,i,t}^{FFF} + \log D_{z,t}$, where $\widetilde{k}_{z,i,t} = \frac{financial\ capital_{z,i,t}}{financial\ capital_{z,t}}$, the share of industry financial capital belonging to firm i, whether it is a DFF or FFF. The set of firms used to build each granularity proxy are the same set used to build the corresponding instrument.

2SLS with log industry exports will exclude 47 zero-export observations, and likely bias estimates. Yet a non-Poisson 2SLS regression allows easy computation of test statistics, and thus these results are presented first in table 6, which is then followed by table 7 that shows results of a Poisson 2SLS regression using the technique suggested in Lin and Wooldridge (2017). The Poisson regression includes the residuals of each first stage as controls in the second stage. Shown here are the results for granularity by domestic sales shares and by employment shares are displayed here, with wage bill shares shown in table A10 of the appendix.

Columns (1) and (4) of each table maintain that overall granularity negatively impacts industry exports, with only the Poisson 2sls column (4) when defining granularity by employment shares being not significant. Only when splitting firms into the two groups of DFFs and FFFs in column (6) is there a positive and significant coefficient on granularity from FFFs. Column (7) removes state-owned firms before building the granularity proxy for DFFs to further show that the result of superstar DFFs negatively impacting GCA is not driven by state-owned firms, but the fundamental differences between superstar DFFs and superstar FFFs.

FFFs are the only source of positive association of granularity with comparative advantage in China. These results are also robust to including real fixed assets as a firm-level control (not shown), something else financial capital might influence exports through. They are also robust to including the percentage of firms that are FFFs in an industry (not shown), meaning the granularity terms are not simply proxying for FDI other than in the sense of granular FDI. Furthermore, adding industry total paid-in capital as a control does not affect results, and thus only the relevant industry size control, D_z, t, is included.

Table 6: 2sls using invested capital shares as IVs

	Defining $\widetilde{s}$ by domestic sales			Defining $\widetilde{s}$ by employment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}$	-1.430***			-1.173***			
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		-1.812***			-1.774***		
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			-0.317			1.421**	2.210***
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			-1.567***			-1.836***	
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$; private only							-0.647**
$D_{z,t}$	0.688***	0.679***	0.657***	0.903***	0.844***	0.884***	0.958***
LM-stat	836	598	406	879	675	595	688
CD-F-stat	938	626	201	1000	723	311	370
Industry & Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	4153	4153	4153	4153	4153	4153	4153

Note: Capital shares serve as the instruments for the granularity variables. Adding industry total paid-in capital as a control does not change signs or significance levels. Underidentification test rejection of the null with a p-value < 0.001 all indicates that the instruments are relevant, and the model is identified. Cragg-Donald Wald F stat of above 100 imply instruments are not weak. As regression is not Poisson here, log exports excludes zero export industry-years. Robust errors; * p<0.10, ** p<0.05, *** p<0.01

Table 7: Poisson 2sls using invested capital shares as IVs

	Defining $\widetilde{s}$ as dom sales share			Defining $\widetilde{s}$ as employment share			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}$	-3.236**			-0.778			
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		-4.874**			-1.337		
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			-2.810			3.053*	3.002*
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			-3.921**			-1.537	
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$; private only							-1.531**
$D_{z,t}$	0.387***	0.393***	0.348**	0.753***	0.703***	0.698***	0.710***
s1 v	2.623*			1.007			
s1 v		4.344*			1.732		
FFFs s1 v			2.571			-1.668	-1.459
DFFs s1 v			3.050*			0.874	1.423*
Industry & Year FE	Y	Y	Y	Y	Y	Y	Y
Observations	4200	4200	4200	4200	4200	4200	4200

Note: Poisson IV version of table 7 with fixed effects procedure from Lin and Wooldridge (2017); estimates the reduced form (first stage) for endogenous variables and includes the residuals as controls (s1 v) in a Poisson regression. Robust errors; * p<0.10, ** p<0.05, *** p<0.01

5.5 Heterogeneity in GCA

Industry estimates describing GCA are heterogeneous in a number of ways that can be captured with the Chinese firm-level database. The following discusses three groups of interaction variables.

First are industry-level technology interactions. I match the Chinese industry codes to ISIC revision 3 codes which have OECD classifications of R&D intensities, where industries are divided into the four groups of high, mid-high, mid-low, and low. Here I combine mid-high and mid-low to make results easier to interpret, and also because the cutoff of classifications is less pronounced in the middle.

Table 8 shows results for each of the three granularity proxies, with the combined mid-tech classification used as the reference group for granularity form each FFFs and DFFs. There are 236 industries out of the 420 that are classified as mid-tech, 131 that are low-tech, and 53 that are high-tech. For the reference group of mid-tech, the coefficients on the granularity proxies move from indistinguishable from zero when using domestic sales shares to a significant 0.761 when using wage bill shares, and finally jumps to a significant 2.532 using employment shares. The relationship is the opposite for DFFs, though with less pronounced jumps.

FFFs see weaker association of exports with the granularity proxies in high-tech industries, which becomes significantly weaker when proxying by wage bill shares. Thus, granularity in high-tech industries does not add to GCA, and may indeed work against it. Results are more mixed for low-tech industries, with a weak significantly positive coefficient when proxying by domestic market shares, but strongly negative when proxying by employment shares. Overall, however, it seems not much, if any, GCA comes from FFFs in low-tech industries. DFFs are again practically the opposite, with associations of granularity with exports relatively positive at the ends of the technology spectrum – and perhaps net positive for low-tech. Possible explanations for the heterogeneity in tech-levels may include varying economies of scale, absorptive capacity, barriers of entry or exporting, or crowding out of smaller firms by larger firms.

Proxying granularity by employment shares generally gives the most positive association of granularity with exports. Table 9 shows results from interacting the granularity proxy by employment shares with various industry level controls. Column (1) reproduces column (6) of table 3, with subsequent columns then interacting each of the two groups' granularity proxy with a specific industry level measure. In brief, the presence of superstar FFFs associate more strongly with higher industry export competitiveness in in industries that have a relatively higher portion of young firms and even state-owned firms, while more weakly in industries that are described by a higher portion of FFFs of any size and in industries that already have a high level of TFP. Granular DFFs then carry the opposite associations with industry exports in every case. The below describes these interactions as well as few others as shown in table 3.

First is an interaction with the ratio of firms per industry that are younger than 5 years old. This variable is meant to proxy for industry dynamics, with a higher ratio implying the industry has relatively more entrants and thus more open to competition. There is a positive and significant association of this ratio with exports on average, with an estimated coefficient of 1.399. When

Table 8: Interacting with industry tech dummies

	$\widetilde{s}$ by dom sales (1)	$\widetilde{s}$ by wage bill (2)	$\widetilde{s}$ by employment (3)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$	-0.245	0.761***	2.532***
x high-tech	-0.628	-1.137**	-1.044
x low-tech	0.864*	-0.167	-1.725***
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$	-1.157***	-1.381***	-1.765***
x high-tech	-1.001	-0.355	1.635**
x low-tech	1.829***	0.956*	1.994***
$D_{z,t}$	0.356***	0.810**	0.760***
Industry & Year FE	Y	Y	Y
R-sqr	0.972	0.981	0.981
Obs	4200	4200	4200

Note: The granularity proxies are interacted with a low tech and high-tech dummy. Of the 420 industries, 131 are classified at low-tech and 53 are classified as high-tech, leaving 236 as mid-tech. Column (1) uses wage bill shares while column (2) uses employment shares. DFFs may consist of either private or state-owned firms as in earlier tables. All models implement Poisson regression. Industry and year fixed effects included. Robust errors; * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

interacting with granularity from FFFs the association becomes more positive, implying that industries with large granular FFFs in combination with relatively more young firms see stronger GCA.

A possible explanation for this is that startups are more quick to respond to granular FFFs, or that they exist due to a boost in demand for intermediates, close substitutes, or complimentary products resulting from the presence of granular FFFs, and subsequently contribute more to exports. Granularity from DFFs, however, is more mixed in its association with exports, with the coefficient on the corresponding interaction term being negative and significant, although the main effect is now positive at 1.296. Thus, granularity from DFFs aids exports only in stagnant industries with relatively less young firms. It could be that in such industries economies of scale play a bigger factor.

Interacting with a variable for the ratio of firms that state-owned, with may indicate a level of industry restrictiveness as well as comparative disadvantage, results in a similar pattern of coefficient estimates other than the negative association on this ratio on average, at a significant -1.061. The relatively strong coefficient on the interaction term with granularity from FFFs indicates the value of allowing large FFFs to exist in industries relatively more dominated by state-owned firms if wishing to boost exports. If assuming industries dominated by state-owned firms are comparatively disadvantaged then GCA via the external channel of FFFs may be easily magnified.

Table 9: Interacting with various industry level controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$	1.495***	1.284**	0.730*	4.189***	1.156***	6.205***	0.708	-7.302***	2.470**
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$	-0.807**	1.296**	0.974***	-1.694***	-0.832**	-4.704***	-2.706	5.366***	3.644***
$D_{z,t}$ (staff)	0.728***	0.765***	1.015***	0.788***	0.734***	0.681***	0.193***	0.559***	0.975***
ind. ratio ¡ 5 yrs		1.399***						0.514**	0.681**
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$		2.892**						3.444**	1.938
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		-5.675***						-7.527***	-6.051***
ind. ratio state			-1.061***					-0.807***	-0.912***
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			4.088***					1.293	1.045
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			-8.015***					-7.145***	-7.701***
ind. ratio FFFs				2.763***				1.514***	2.035***
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$				-7.857***				-6.613***	-5.307***
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$				4.104***				-2.877**	-2.306**
ind. dom. growth					-0.047			-0.073**	
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$					0.009			-0.037	
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$					0.135*			0.273***	
ind. TFP						0.417***		0.350***	
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$						-0.822***		-0.782***	
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$						0.526***		0.082	
ind. real FA (log)							0.660***	0.342***	
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$							0.078	0.886***	
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$							0.134	-0.132	
Industry & Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-sqr	0.980	0.982	0.984	0.982	0.982	0.983	0.983	0.989	0.986
Obs	4200	4200	4200	4200	3780	4168	4200	3751	4200

Note: Granularity proxy from $\widetilde{s}$ defined by employment only. All models implement Poisson regression. Robust errors; * p<0.10, ** p<0.05, *** p<0.01

Column (4) takes a somewhat opposite approach with the ratio of firms that are FFFs. A higher ratio would imply more openness to foreign entrants, and this main effect carries a positive association with industry exports. Industries with relatively more FFFs in combination with higher levels of granularity from FFFs, however, see less exports, with this interaction term having a coefficient of -7.857. This implies that granularity from FFFs boost exports only in industries with a low presence of FFFs in general, and once there are many FFFs granularity begins to drag on industry exports.

Domestic sales growth does not associate strongly with the level of exports, on average, and only the interaction term with granularity from DFFs is slightly significant, with a positive coefficient of 0.135. This coefficient does become stronger when including all terms simultaneously in column (8). Thus, industries with granularity from DFFs that are also growing in their domestic sales see stronger exports, somewhat reducing the overall negative associate of granularity from DFFs with exports.

Total factor productivity (TFP), estimated following Wooldridge (2009) separately per industry and aggregated to the industry level following Olley and Pakes (1996), carries a positive association with exports on average which becomes negative when interacting with granularity from FFFs. This implies that GCA from FFFs, which is the only channel consistently giving any positive association of granularity with exports, is magnified in industries with relatively lower TFP. These industries would arguably have more to gain from the presence of a large FFF, as their low TFP makes them less competitive in international export markets, on average.

Finally, real fixed assets, aggregated to the industry level, carry positive association with exports for obvious reasons of correlation with size. More interesting is the robustness of most of the estimates when including all these terms in column (8), where the interaction of real fixed assets with granularity from FFFs also becomes significantly positive. Industries with more real fixed assets likely have more absorptive capacity and benefit more from the presence of large FFFs in terms of higher exports.

At the level of the group of firms making up the granularity proxies, column (1) of table 10 shows that higher granularity from DFFs associates relatively more positive with exports when those granular firms have higher ratios of intermediate inputs (deflated) to total real fixed assets, with this not making a difference for granular FFFs. The case is similar for interacting granularity with the ratio real value added to real fixed assets, and also average worker real wage. Interacting with the ratio of wages to real fixed assets, however, sees much more negative association of granularity with exports for FFFs, implying that FFFs paying disproportionately more wages relative to their amount of fixed assets they have invested in may hinder industry exports. This may be due to those firms siphoning workers from other firms without translating that into more investment.

Column (6) includes these firm-level controls simultaneously, with coefficients on the interaction terms with granularity from FFFs and DFFs becoming more similar, though the main effects moving even further apart. This shows that certain firm-level characteristics adjust GCA similarly regardless of whether it comes from FFFs or DFFs, but the fact is these two groups are made up of very different kinds of firms and hence the large difference in the main effects. Column (7) then adds the industry level controls from column (9) of table 9. Only the inputs to assets ratio ends up switching significance among granular FFFs and DFFs compared to when the terms are regressed separately.

Table 10: Interacting with various firm level controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$D_{z,t}$ (staff)	0.731***	0.739***	0.853***	0.750***	0.753***	0.815***	1.002***
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$	1.714***	1.657***	1.738***	2.479***	1.260***	0.126	0.095
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$	-1.326***	-1.514***	-4.907***	-0.872***	-2.217***	-3.785***	-2.487***
inputs/assets x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$	-0.013					0.124***	0.163***
inputs/assets x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$	0.155***					0.133***	0.118***
VA/assets x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$		0.059*				0.124***	0.067*
VA/assets x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		0.483***				0.090	0.047
labor productivity x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			0.232**			0.371***	0.438***
labor productivity x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			1.065***			0.564***	0.335***
wages/assets x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$				-1.014***		-2.178***	-1.525***
wages/assets x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$				0.480		-1.498***	-1.329***
wages/workers x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$					0.002	0.014***	0.015***
wages/workers x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$					0.039***	0.029***	0.018***
ind. ratio firms age ¡ 5 yrs							-0.034***
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$							0.062
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$							0.135**
ind. ratio firms state-owned							-0.807***
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$							1.607
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$							-5.803***
ind. ratio FFFs							2.058***
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$							-7.021***
x $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$							-0.091
Industry & Year FE	Y	Y	Y	Y	Y	Y	Y
R-sqr	0.980	0.981	0.981	0.981	0.982	0.983	0.987
Observations	4061	4061	3985	4060	4060	3984	3984

Note: DV is industry exports. Granularity proxy with $\widetilde{s}$ defined by employment only. The terms being interacted with each of the granularity proxies are computed from the corresponding group of top-3 FFFs or DFFs separately. For example, the inputs/assets ratio interacted with the proxy for granularity from FFFs sums inputs from these 3 FFFs only and divides by the sum of assets from the 3 FFFs only. All models implement Poisson regression. Robust errors; * p<0.10, ** p<0.05, *** p<0.01

In summary, large, granular FFFs associate more with GCA when simultaneously high in their ratio of value added to real fixed assets or have a higher real fixed assets to wage bill ratio. Large DFFs also associate more with GCA when simultaneously high in their ratio of value added to real fixed assets and also pay a higher average worker wage. These two groups of granularity see a convergence of their estimated coefficients on the firm-level interaction terms when all are included together along side industry level controls, though their total effects remain different.

The fact that granular FFFs actually export less intensively on average compared to non-granular FFFs, with respective average export intensities of 0.40 and 0.49, and that coefficients remain robust to including the ratio of FFFs making up an industry, hints that the granular FFFs affect aggregate outcomes not just by adding to figures in their own right, but more likely through these complex interactions and spillover effects on other non-granular firms. Similarly, the fact that granular DFFs export more on average compared with non-granular DFFs, at respective mean export intensities of 0.18 and 0.11 across the panel, yet typically associate negatively with industry exports, shows that this type of granularity also affects aggregate outcomes via its impact on the non-granular firms.

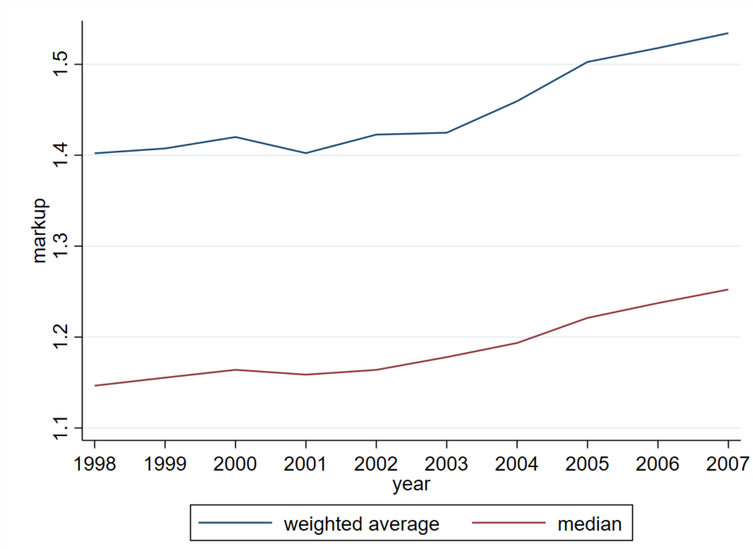
To investigate more whether the granular firms affect outcomes of non-granular firms, the following section presents results on associating granularity with firm-level estimated markups.

6 Granularity and Pro-competitive effects

Atkeson and Burstein (2008) give a oligopolistic competition model where markups are increasing a firms market share. Edmond et al. (2015) extend this model and show that international trade reduces markups when there is both existing misallocation in the sense of markup dispersion and trade puts domestic producers under more effective competitive pressure. In the following I give an empirical specification for testing whether granularity from the two groups of either DFFs or FFFs operating within mainland China associates with more or less markups at the firm level. While foreign entrants would logically reduce markup dispersion by adding to competition, the impact of existing out-sized firms is less obvious. As a firm becomes larger relative to others, it takes their market share, and while this may result in higher markups for the large firm, it should also result in lower markups by all other firms if considering the linear relationship between market share and markups. However, if a larger firm displaces smaller firms entirely or causes inequality in the distribution of productivity then this effect may be reversed. In particular if the presence of Granularity is only from consolidating industries whereby other firms are also becoming larger, then there may be an association of industry granularity with markups of other firms.

I estimate firm-level markups, $mu_{z,i,t}$, following De Loecker and Warzynski (2012) with Ackerberg et al. (2015) value-added translog production function elasticities. Control variables are taken from the literature include both firm-level controls as well as industry level controls, all varying with time. Perhaps the most important control is one for market share. Oligopoly models of competition suggest markups increase in market share. For export-intensive Chinese manufacturing it is sensible to proxy for this with $SS_{z,i,t} = ts_{z,i,t}/\sum_{i=1}^{N_z} ts_{z,i,t}$, the firms' total sales shares, where $ts_{z,i,t}$ comprises of both domestic sales and exports coming from the firm's operations inside China.

Figure 6: Markups



Note: Markups estimated as in De Loecker and Warzynski (2012) with Ackerberg et al. (2015) translog production function elasticities.

Price of labor, $w_{z,i,t}$, is defined as total real wages over the number of workers (employment). A higher price of labor cuts negatively into markups as long as labor productivity is controlled for. At the same time, industry-level price of labor, $w_{z,t}$, or the logged ratio of total wages to total number of workers, should create barriers to entry, indicating tight labor supply for that sector, and thus raise markups. Labor productivity, $lp_{z,i,t}$, is defined as the logged ratio of real value added to the number of workers. Firms charge relatively higher markups when they are more productive as a result of lower unit costs. More imports implies more competition all the way through to domestic market sales, thereby lowering markups of firms based in the country. Import penetration, $IP_{z,t}$, is calculated at the more aggregate ISIC revision 3 level in order to match with international datasets, with 111 industries rather than the 420 CIC codes used for other industry-level variables. Higher export intensity at the firm level, $EI_{z,i,t}$, exposes the firm to more international competition causing markups to be lower. A higher export intensity at the industry level, $EI_{z,t}$, implies that firms are more competitive in international markets, and thus may have higher markups. Capital intensity, $ci_{z,i,t}$ is defined as the logged ratio of real fixed assets to the number of workers, and is expected to be positive with markups due to subsequently greater costs of capital accumulation. it is defined by taking real fixed assets divided by the number of workers. Capital intensity at the industry level, $ci_{z,t}$, is expect to associate positively with markups due to barriers of entry resulting from industry-level costs of capital. Positive variation in TFP, $tfp_{z,i,t}$, should associate with higher markups, as a firm is able to produce more for less than its competitors. Conversely, positive variation in industry level TFP, $tfp_{z,t}$, means the industry becomes more competitive on average, which should lower markups. Finally, a larger number of firms means more competition, and thus lower markups. Therefore the number of firms is included in log form, $obs_{z,t}$.

The variables of interests are the proxies for granularity from each the DFF and FFF groups. Here I use employment shares to build the granularity proxies, and maintain the notation $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$ and $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$, for each group respectively. The association of granularity with markups of any given firm within an industry will depend on how the granularity proxy interacts with key descriptors of the nature of competition facing the firms. Thus each of the industry level granularity proxies are interacted with the individual firm sales share, $SS_{z,i,t}$. A low $SS_{z,i,t}$ but high $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$ or $\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$ should see firm i as having lower markups in the sense that the larger granular firms are taking sales share from the smaller firm i.

Lower markups of course may attract more market share, which is a channel of reverse causality here for the term $SS_{z,i,t}$. Including firm-level fixed effects at least partially addresses this. Differentiated pricing or growth strategies, such as when a firm systematically undercuts its otherwise equal competitors in order to gain market share, will therefore be controlled for. Additionally, all variables are lagged one year in order to minimize endogeneity and reverse causality. The model estimated for all firms i not ranked in the top-3 of either DFFs or FFFs is then:

$$
\begin{aligned}
mu_{z,i,t+1} = \alpha + SS_{z,i,t} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF} + SS_{z,i,t} * \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + SS_{z,i,t} * \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF} \\
+ lp_{z,i,t} + w_{z,i,t} + w_{z,t} + ci_{z,i,t} + ci_{z,t} + obs_{z,t} + EI_{z,i,t} + EI_{z,t} + IP_{z,t} + \delta_i + \delta_t + \varepsilon_{z,i,t}.
\end{aligned} \tag{9}
$$

and includes with both year and firm fixed effects. All firms included in calculation of the granularity proxies are excluded. Firms-year observations that are outliers in either the computed firm-level markup or estimated firm-level TFP are also omitted, where outliers are identified under the strict criteria of being both more than three standard deviations from the mean as well as in the top or bottom 1 percent of observations in that industry-year.

Results are presented in table 11. Column (1) regresses firm-level markups on their the definitions of granularity by employment shares, with the two groups giving opposite signs. An additional 10 percent of total industry number of workers belonging to the top-3 FFFs associates with a 0.6 percent decrease in markups of the non-granular firm, while the same increase in workers belonging to the top-3 DFFs associates with a 0.33 percent increase in markups. This is robust to controlling for the non-granular firm's sales share and the interaction of the firm-level sales shares the the two definitions of industry granularity. The interaction terms in column (3), however, are both negative, implying that there are strong pro-competitive effects from granular firms onto non-granular firms in their industry if those non-granular firms are also large (having a higher sales share).

Table 11: Granularity and Pro-competitive effects

	(1)	(2)	(3)	(4)
$SS_{z,i,t-1}$; sales share		0.631***	1.577***	1.172***
		(0.06)	(0.11)	(0.11)
$\sum_{i=1}^{3} \widetilde{s}_{z,t-1}^{FFF}$	-0.060***		-0.072***	-0.056**
	(0.02)		(0.02)	(0.02)
$\sum_{i=1}^{3} \widetilde{s}_{z,t-1}^{DFF}$	0.033***		0.026***	-0.006
	(0.01)		(0.01)	(0.01)
$SS_{z,i,t-1} \times \sum_{i=1}^{3} \widetilde{s}_{z,t-1}^{FFF}$			-1.496***	-1.114***
			(0.35)	(0.33)
$SS_{z,i,t-1} \times \sum_{i=1}^{3} \widetilde{s}_{z,t-1}^{DFF}$			-2.085***	-1.467***
			(0.20)	(0.17)
$FDI_{z,t-1}$; firms that are FFFs				-0.005***
				(0.00)
$lp_{z,i,t-1}$; labor productivity, log				0.011***
				(0.00)
$w_{z,i,t-1}$; price of labor, log				0.002***
				(0.00)
$w_{z,t-1}$; ind price of labor, log				0.005
				(0.00)
$ci_{z,i,t-1}$; capital intensity, log				-0.013***
				(0.00)
$ci_{z,t-1}$; ind capital intensity, log				0.000***
				(0.00)
$obs_{z,t-1}$; number of firms, log				0.000
				(0.00)
$EI_{z,i,t-1}$; export intensity				0.006***
				(0.00)
$EI_{z,t-1}$; ind export intensity				-0.039***
				(0.01)
$IP_{z,t-1}$; import penetration				0.004
				(0.01)
constant	0.263***	0.262***	0.261***	0.257***
	(0.00)	(0.00)	(0.00)	(0.02)
Firm & Year FEs	Y	Y	Y	Y
Adjusted R_sq	0.861	0.861	0.861	0.862
Observations	1220553	1220553	1220553	1218377

Note: Dependant variable is log of estimated firm markup. Granularity proxy with $\widetilde{s}$ defined by employment only. All explanatory variables are lagged one year. Lowercase denotes log values. Column (4) implements the full markups regression of equation 9. Summary statistics in table A12 of the appendix. Industry×Year clustered errors; * p<0.10, ** p<0.05, *** p<0.01.

Column (4) introduces all of the firm and industry-level controls of equation 9. The estimated partial effect of FFFs' granularity on a firms markup is $\partial E(mu_{z,i,t})/\partial worker\ share^{FFF}_{z,t\ 1} = -0.056 - 1.114 * SS_{z,i,t\ 1}$. The estimated partial effect of granularity from the group of DFFs on a firms markup is $\partial E(mu_{z,i,t})/\partial worker\ share^{DFF}_{z,t\ 1} = 0 - 1.467 * SS_{z,i,t\ 1}$. Thus, there is a pro-competitive effect in the sense that granularity form FFFs reduces markups of non-granular firms of any size, while the effect from granular DFFs only kicks in for larger non-granular firms described by higher $SS_{z,i,t\ 1}$.

It should be noted, however, that sales shares are quite low for a typical, non-granular firm, and thus granularity form FFFs is more economically significant. Among the firms included in the regressions for table 11 the mean sales share is 0.0038 across all industries and years with a standard deviation of 0.004. The mean sales share reaches a high of 0.136 at the industry-year level, however.

7 Granular FDI and Horizontal Spillovers

Spillover effects from FDI on local firms is a popular topic, but one difficult to make progress on due to FDI fundamentally being an endogenous decision, whereby firms may enter a developing country and its particular industries after observing or anticipating productivity growth among local firms. There is a source of exogenous variation in industry level granularity when China implemented reforms on the guidelines for FDI encouragement in late 2002, upon its accession to the WTO. This shock can be traced industry-by-industry when comparing the official 1997 Catalog of FDI Encouragement document to the 2002 updated version. These two documents are lists of industries or specific areas of manufacturing divided into the three categories of "FDI Encouraged," "FDI Restricted," and "FDI Prohibited." A fourth category, "Permitted", therefore consists of all manufacturing items not listed in the catalog, where FDI is not given special treatment as in the "FDI Encouraged" category but not restricted either. Investment in items belonging to the encouraged category, for example, may be entitled to an "Import and Export Goods Tax Exemption Certificate" after submitting a request to a local Foreign Economic and Trade Commission.

There are 386 lines in the 1997 catalog and 449 lines in 2002 catalog, both including category headers, and with a majority, but not all, of the items relating to some manufacturing activity. The mentioned activities are sometimes specific, encompassing what matches to single 8-digit CIC codes, and sometime more broad, covering multiple 4-digit CIC codes. I proceed by manually mapping each year's catalog to Chinese Industry Codes by identifying which 8-digit product descriptions match the descriptions given in the catalogs. While this strategy is also used in Lu et al. (2017) when investigating general FDI spillovers, they do not make publicly available their constructed dataset and so I redo the matching for the current paper.

The goal is to then identify which of the 420 4-digit CICs have a change to becoming more encouraging of FDI from 2002 compared to before 2002. To this aim, if at least one 8-digit CIC within a 4-digit industry experiences a move towards more encouragement of FDI while no 8-digit codes in that 4-digit industry experiences a lessening of encouragement, then that industry is marked as in the treatment group, receiving a positive shock on FDI encouragement. 4-digit industries that see no 8-digit code as being affected by the updating of the catalog from the 1997 version to the 2002 version are marked as the control group. All industries that experienced a worsening of FDI encouragement or a mixture of some 8-digit codes improving while some worsening are then excluded for the following analysis.

In total, I identify 121 4-digit industries in the treatment group that experienced an improvement of FDI encouragement, and 283 industries in the control group that saw no change, meaning that out of the 420 industries in total, 16 saw either a mixed change or worsening of encouragement. 121 industries in the treatment group is slightly more than found in Lu et al. (2017). I view more industries in the treatment group as a relatively conservative approach, making differences between the treatment group and the control group more difficult to detect.

An FDI encouragement dummy variable is constructed and the treatment group is assigned a value of 1 for years 2003 and later, and a value of 0.75 for 2002, as the 2002 catalogue was implemented 3 months into the year. The variable is zero for years before 2002 in the treatment group and for all observations of the control group.

This FDI encouragement shock can then be used an IV for industry level granularity. As granularity bounded on the interval zero to one with a mass closer to zero and with few values above 0.5 for FFFs domestic market granularity, it can be approximately described by an exponential Poisson distribution, which helps in more accurately determining which variables this shock is correlated with. I run the diff-in-diff style Poisson regression on all definitions granularity and also a generic term for domestic market FDI penetration, as it may be expected that such an FDI encouragement shock affects all FDI equally, or perhaps even smaller firms more than larger ones. FDI domestic market penetration is defined as

$$\sum_{i=1}^{N_z} domestic\ sales * FDI\ share\ of\ invested\ capital_{z,i,1998} \tag{10}$$

Table 12 presents results regressing these variables on the corresponding $treatment_z \times post_t$ term, with treatment defined by the FDI encouragement dummy and post being 0 for years before 2002 and 1 for 2002 and after. Year and industry fixed effects are also included. There is a dramatic estimated increase in FFF granularity for $\widetilde{s}$ defined by domestic sales, with an estimated coefficient of 0.199 in table 12, meaning the WTO shock induced roughly 20 percent of the domestic market to go to top FFFs in impacted industries. There is no effect, however, on FFF employment granularity, or even exports granularity (if defining $\widetilde{s}$ by exports), and also no effect for any form of DFF granularity (only DFF domestic sales granularity is shown but results are similar for DFF employment and exports granularity). This seems logical in that exports of foreign firms were already more liberally encouraged both generally from the existing 1997 catalogue and within numerous special economic zones, and China's WTO accession instead largely focused on improving domestic market access to foreign firms. Regardless, these results indicate that the WTO shock would make for a week IV for any variable other than FFF domestic market granularity. Column (5) of table 12 then introduces the FDI domestic market penetration (FDI dom penetration). While its coefficient is significant at the 10% confidence level, it is half the magnitude as the granular component in column (1), and its significance as well as magnitude are lost when removing granular firms from the calculation of FDI dom penetration. This indicates that the boost to FDI caused by this shock was only among granular firms serving the domestic market. That is, only firms positioned relatively well in the domestic market were able to benefit from the marginal changes in this shock[3]. Additionally, many of these regulation changes had minimum scale requirements and were largely in marginal applications given that FDI was already quite established in China at that point.

As labor productivities and TFP of granular FFFs are the highest among all other groups and easy to reliably measure given the constraints of Chinese Industrial Enterprise survey, I proceed by regressing labor productivity and TFP of all DFFs on the granular proxy and a series of controls according to:

$$y_{z,i,t} = \alpha + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \mathbf{X}_{z,i,t} + \delta_i + \delta_t + \varepsilon_{z,i,t}, \tag{11}$$

with the first stage regression:

$$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} = \alpha + treatment_z \times post_t + \mathbf{X}_{z,i,t} + \delta_z + \delta_t + \varepsilon_{z,i,t}. \tag{12}$$

$y_{z,i,t}$ in equation 11 is firm level productivity, $\mathbf{X}_{z,i,t}$ is a series of firm-level and industry-level controls, to be discussed below, and $\widetilde{s}_{z,i,t}^{FFF}$ is FFF domestic market granularity for $\widetilde{s}$ again denoting domestic market sales shares.

Using the FDI 1997 to 2002 catalogue change as an IV requires that it impacts productivity of domestic firms only through its effect on FDI, and not through other channels, such as tariffs. In terms of types of FDI, the previous paragraph establishes that the IV is most relevant for the specific type of FFFs granular in domestic sales. As Lu et al. (2017) mention, concerns on the exclusion restriction are partially alleviated in that the FDI deregulation in 2002 came after a very lengthy and process leading up to China's accession to the WTO, described by much uncertainty

[3]The struggle of small firms to capitalize from the 2002 catalog change has been documented in case studies (http://www.cnfi.org.tw/cnfi/ssnb/155-428-20.htm)

and thus the specifics of the changes were not anticipated. Additionally, these catalogue changes are not particularly correlated with changes in tariffs, and by 2007 there was indeed little variation left in China's weighted import and export tariffs as mentioned by Brandt et al. (2017).

To further alleviate concerns coming from tariff policy, I include in $\mathbf{X}_{z,i,t}$ the 1998 level of import and export tariffs obtained from the World Integrated Trade Solution (WITS) website at the ISIC Revision-3 product code-level, for which there are 110 industry codes match to the 420 CIC codes, and interact it with year dummies to allow CIC-level variation that controls for WTO induced variation. I also tested the non-time varying tariff levels in 2001 as well as the time and industry varying tariff levels with no interaction. The results on the variables of interest are not affect regardless of the chosen strategy, but any significance on estimated coefficients for the tariff variables in explaining firm-level productivity is only found when taking the 1998 values and interacting with year dummies, with the expected estimations of tariff reductions being positively correlated with firm-level productivity. Utilizing WITS trade data, I additionally control for China's domestic market import penetration in a similar manner.

An additional concern for the exclusion restriction is China's privatization of state-owned firms at the same time of China's WTO accession which may then bias results. I therefore also include in $\mathbf{X}_{z,i,t}$ the 1998 total share of industry sales owing to state-owned firms, interacted with year dummies to allow for each industry's own time-evolution. To control for non-FDI granularity, I also include the 1998 values of DFF domestic market granularity interacted with year dummies. Additionally, 1998 FDI domestic sales sales penetration interacted with year dummies is included to control for direct effects of FDI in general.

There are then other identified determinants of FDI. Following Lu et al. (2017), I control for new product intensity, export intensity, number of firms, and average age of firms, all by interacting the 1998 industry level values with year dummies. With respect to China's special economic zones, as also noted by Lu et al. (2017), there were no changes regarding the regional aspects of FDI entry regulations, and a 1997 law forbid further location discretion. Finally, the time-varying firm-level controls of logged output, logged capital-labor ratio, a state firm dummy, and an exporter dummy are all included in $\mathbf{X}_{z,i,t}$.

Table 13 presents results using the FDI encouragement shock as an IV for domestic sales granularity from FFFs, with both first and second stage regressions for the entire set of DFFs, whether granular or not. Note that column (1) is the firm-level equivalent of column (6) in table 12, and continues to show that the FDI encouragement shock did not enable market gains for non-granular FDI. Column (2) then shows the firm-level equivalent of column (1) in table 12, and indicates that industries in the treatment group saw an additional 1% of domestic market sales concentrated in the top-3 FFFs. Taking account of all the controls, industries in the treatment group saw an increase in average concentration among top-3 domestic market FFFs from 0.113 to

Table 12: China's WTO accession effect on industry granularity and FDI

	(1) FFF dom granularity	(2) FFF emp granularity	(3) FFF exp granularity	(4) DFF dom granularity	(5) FDI dom penetration	(6) non-granular FDI dom pen.
FDI encouragement	0.199**	0.131	.003	0.030	0.087*	-0.033
	(0.083)	(0.093)	(0.083)	(0.051)	(0.050)	(0.056)
Industry & Year FEs	Y	Y	Y	Y	Y	Y
Pseudo R-sqr	0.115	0.111	0.099	0.097	0.108	0.113
Observations	4040	4040	3990	4040	4040	4020

Note: Industry-level Poisson regressions with industry and year fixed effects. The DV for columns (1) - (3) are differnt measures of FFF granularity. Column (4) repeats column (1) for DFFs, while the repetition of columns (2) and (3) for DFFs is excluded (similar insignificant result). Column (5) is the FDI weighted share of domestic sales. Column (6) repeats column (5) but excludes granular FFFs. Industry clustered errors; * p<0.10, ** p<0.05, *** p<0.01

Table 13: Granular FDI and Horizontal Spillovers

	1st stage		IV		OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	FDI dom penetration	FFF dom granularity	ln TFP	ln labor productivity	ln TFP	ln labor productivity
$treatment_z \times post_t$	-0.001	0.010***				
	(0.006)	(0.003)				
$\sum_{i=1}^{3} \widetilde{s}_{k,i,t}^{FFF}$ (FFF dom granularity)			3.570*	3.697*	-0.012	0.154***
			(2.089)	(2.065)	(0.064)	(0.040)
$\sum_{i=1}^{N_z} \widetilde{s}_{k,i,1998}^{FFF}$ * yr (FDI dom pen.)		0.000***	-0.001***	-0.000**	-0.000**	-0.000***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\sum_{i=1}^{3} \widetilde{s}_{k,i,1998}^{DFF}$ * yr	0.000	0.000***	-0.000**	-0.000*	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Kleibergen-Paap Wald rk F statistic			12.34	12.34		
Anderson-Rubin Wald test			4.95**	5.13**		
Firm & Year FEs	Y	Y	Y	Y	Y	Y
Industry controls × year dummies	Y	Y	Y	Y	Y	Y
Time-varying firm controls	Y	Y	Y	Y	Y	Y
Observations	1200640	1200640	1200640	1200640	1200640	1200640

Note: Column (2) shows the first stage estimation of equation 12 for the IV regressions of equation 11 presented in column (3) for TFP and column (4) for labor productivity. $\widetilde{s}$ is defined by domestic sales. Firm-level controls are logged output, logged capital-labor ratio, a state firm dummy, and an exporter dummy. CIC industry controls are the year dummy interactions with 1998 values of DFF domestic market granularity (shown), FDI domestic sales sales penetration (shown), share of sales by state firms, export intensity, new product output intensity, mean firm age, and number of firms. Domestic market import penetration is at the more aggregate level of 110 ISIC industries. Robust standard errors clustered at the CIC industry level shown in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.

0.123, or from of 11.3% to 12.3% of total domestic market sales belonging to all firms located in China.

Second stage results in column (3) shows that an additional 1% of domestic sales by firms in China concentrating in the top-3 FFFs per industry induces a gain in TFP for DFFs of 3.57%. Column (4) changes the dependant variable to labor productivity and gives a surprisingly similar coefficient. The IV passes relevant weak instrument testing, and gives coefficients much higher than the OLS regression shown in columns (5) and (6). While the coefficients in columns (3) and (4) may look large, the above mentioned interpretation combined with the fact that first stage results indeed indicate that exactly an additional 1% of domestic sales by firms in China became concentrated in the top-3 FFFs as a result of this regulation change leads to an end result that is quite modest, and thus it seems it is the OLS coefficients that are likely underestimated due to endogeneity.

These results are in stark contrast to the existing non-granular FDI literature on horizontal spillovers, which are typically estimated to be negative for the case of China. These results imply that granular FDI behaves differently from non-granular, or the other bulk of FDI. This may be due their better market position, more differentiated products, or general ability to bring new market activity rather than competing at lower more crowded ends of the spectrum.

8 Conclusion

This paper examined the distinct impacts of the largest foreign and domestic firms on industry-level comparative advantage in China between 1998 and 2007, investigated pro-competitive effects resulting from superstar FDI compared to domestic superstars, and presented casual evidence of horizontal spillovers from superstar FDI.

The association of granularity with industry exports in China over the years 1998 through 2007 is nearly the exact opposite of what Gaubert and Itskhoki (2021) find for France over a similar time-frame when defining granular firms by domestic sales shares. The negative estimate on the granularity proxy for these out-sized firms implies that they hinder industry exports, and therefore generally describe a negative granular comparative advantage, whereby China's net comparative advantage at the industry level becomes dampened. There is, however, huge heterogeneity in the types of firms operating inside mainland China, and once treating granularity from domestically-funded Chinese firms (DFFs) separately from mainland China operations of foreign-funded-firms' (FFFs), there emerges signs that granularity among FFFs does not associate with negative GCA, but may instead be the only source of positive GCA in China. One requirement for detecting this positive granular residual is the sorting firms according to industry employment shares rather than domestic market sales. This is important for the case of developing countries as many out-sized firms remain export intensive, rarely touching the domestic market despite their immense size, and therefore have a small domestic market share even though they may be hugely influential in the workings of the domestic economy.

These results are robust to separating out state-owned firms from DFFs, as well as Poisson 2sls regression using capital shares as instrumental variables. That is, there remains strong evidence of negative GCA coming from the largest DFFs and positive GCA coming from the largest FFFs. However, there is evidence of a transition taking place. While FFF granularity in terms of employment started out in 1998 with strong positive GCA and the corresponding group of granular DFFs strongly negative, they have been converging over the years, with the coefficients on FFFs slowly moving down and that on DFFs more rapidly moving up to both being positive and indistinguishable from each other by 2007. This attests to developing countries being dynamic and unique, taking on completely different characteristics from developed countries in their early stages, but remain on a path of convergence.

The association of granularity with exports is heterogeneous according to a number of key channels. Firstly, industry technology intensity seems to play an important role. While overall granularity from FFFs seems to be beneficial to industry export performance, granular residual from FFFs in the high-tech sector shows signs of a negative association. This may be due to the exaggerated productivity gaps in this sector when large foreign enterprises are present, resulting in lack of sufficient absorptive capacity and suppression of other firms' growth. On the other hand, the granularity from DFFs turns positive in its association with exports in low-tech sectors.

Interacting with other industry level controls shows granularity from FFFs more positively associated with exports in industries that have a higher ratio of young firms, a lower ratio of other FFFs, lower levels of aggregated TFP, and more real fixed assets. This paints a picture of granularity from FFFs pushing up industry exports when there is a healthy startup environment that is not already dominated by many FFFs, and where firms can compete within a relatively lower initial industry-productivity level. Granularity from DFFs is relatively worse for exports in industries that are growing slowly in domestic market sales or have a high ratio of young firms, state-owned firms, or FFFs. Firms in such industries need to develop in a more equitable manner than out-sized DFFs seem to allow.

Policy makers in developing countries may wish to determine what type of superstar firms they should allow to conduct FDI in their economy, or whether they should be concerned about outlier-firm growth or encourage it, if hoping to boost exports. Interacting the granularity proxies with variables describing the granular firms making up these proxies attempts to address this. Granular firms that more intensively use intermediate inputs relative to fixed assets, add more value added relative to fixed assets, or pay higher average wages, associate more with positive GCA. However, higher granularity among labor intensive firms, measured as total wages paid to fixed assets, leads to negative association with exports.

Industry granularity plays an important role in determining firm-level markups, which indicate potential welfare implications beyond variation in exports. The extent of this relationship depends on whether the source of granularity is from DFFs or FFFs and on the non-granular firm's size. Granularity may be beneficial in reducing markups in industries described by firms with higher market shares, or less firms overall. While granularity from both groups induces pro-competitive effects, granular DFFs only do so for other large but non-granular firms, while granular FFFs induce pro-competitive effects for non-granular firms of any size.

Finally, the plausibly exogenous 2002 shock to industries in China where FDI is encouraged by official guidelines allows testing new sources of horizontal spillover effects as these new guidelines directly impacted domestic market granularity owing to FDI. Results show that an additional 1% of domestic sales by firms in China concentrating in the top-3 FFFs per industry induces a gain in TFP for DFFs of 3.57% and a gain in labor productivity of 3.70%, despite the existing literature on FDI in China generally finding negative results when focusing on more general definitions of FDI penetration. This creates a story that superstar foreign firms are beneficial for China's industries in terms of both aggregate and micro-level outcomes.

References

Acemoglu, D., Carvalho, V., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5):1977–2016. Publisher: Econometric Society.

Almfraji, M. A. and Almsafir, M. K. (2014). Foreign Direct Investment and Economic Growth Literature Review from 1994 to 2012. *Procedia - Social and Behavioral Sciences*, 129:206–213.

Amighini, A., McMillan, M., and Sanfilippo, M. (2017). FDI and Capital Formation in Developing Economies: New Evidence from Industry-level Data. NBER Working Paper 23049, National Bureau of Economic Research, Inc.

Amiti, M., Duprez, C., Konings, J., and Van Reenen, J. (2023). FDI and Superstar Spillovers: Evidence from Firm-to-Firm Transactions.

Atkeson, A. and Burstein, A. (2008). Pricing-to-Market, Trade Costs, and International Relative Prices. *American Economic Review*, 98(5):1998–2031.

Bernard, A. B., Jensen, J. B., and Lawrence, R. Z. (1995). Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987. *Brookings Papers on Economic Activity. Microeconomics*, 1995:67–119. Publisher: Brookings Institution Press.

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2018). Global Firms. *Journal of Economic Literature*, 56(2):565–619. Number: 2 Publisher: American Economic Association.

Bloom, N., Van Reenen, J., and Melvin, S. (2013). Gokaldas Exports (A): The Challenge of Change. Technical Report No SM213A, Stanford.

Brandt, L., Van Biesebroeck, J., Wang, L., and Zhang, Y. (2017). WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, 107(9):2784–2820.

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2014). Challenges of working with the Chinese NBS firm-level data. *China Economic Review*, 30(C):339–352. Publisher: Elsevier.

Carvalho, V. and Gabaix, X. (2013). The Great Diversification and Its Undoing. *American Economic Review*, 103(5):1697–1727.

Cravino, J. and Levchenko, A. A. (2016). Multinational Firms and International Business Cycle Transmission. Working Paper 22498, National Bureau of Economic Research. Issue: 22498 Series: Working Paper Series.

Dornbusch, R., Fischer, S., and Samuelson, P. A. (1977). Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods. *The American Economic Review*, 67(5):823–839. Publisher: American Economic Association.

Du, L., Harrison, A., and Jefferson, G. H. (2012). Testing for horizontal and vertical foreign investment spillovers in China, 1998–2007. *Journal of Asian Economics*, 23(3):234–243. Publisher: Elsevier.

Eaton, J., Kortum, S., and Kramarz, F. (2004). Dissecting Trade: Firms, Industries, and Export Destinations. *American Economic Review*, 94(2):150–154. Number: 2 Publisher: American Economic Association.

Eaton, J., Kortum, S., and Sotelo, S. (2012). International Trade: Linking Micro and Macro. NBER Working Paper 17864, National Bureau of Economic Research, Inc.

Edmond, C., Midrigan, V., and Xu, D. Y. (2015). Competition, Markups, and the Gains from International Trade. *American Economic Review*, 105(10):3183–3221.

Freund, C. and Pierola, M. D. (2015). Export Superstars. *Review of Economics & Statistics*, 97(5):1023–1032. Number: 5 Publisher: MIT Press.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3):733–772.

Number: 3 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA8769.

Gaubert, C. and Itskhoki, O. (2021). Granular Comparative Advantage. *Journal of Political Economy*, 129(3):871–939. Publisher: The University of Chicago Press.

Grassi, B. (2018). IO in I-O: Size, Industrial Organization, and the Input-Output NetworkMake a Firm Structurally Important. Technical Report 619, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University. Issue: 619 Publication Title: Working Papers.

Iacovone, L., Javorcik, B., Keller, W., and Tybout, J. (2015). Supplier responses to Walmart's invasion in Mexico. *Journal of International Economics*, 95(1):1–15. Publisher: Elsevier.

Lin, P., Liu, Z., and Zhang, Y. (2009). Do Chinese domestic firms benefit from FDI inflow?: Evidence of horizontal and vertical spillovers. *China Economic Review*, 20(4):677–691. Publisher: Elsevier.

Lu, Y., Tao, Z., and Zhu, L. (2017). Identifying FDI spillovers. *Journal of International Economics*, 107(C):75–90. Publisher: Elsevier.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725. Publisher: [Wiley, Econometric Society].

Nunnenkamp, P. (2004). To What Extent Can Foreign Direct Investment Help Achieve International Development Goals? *The World Economy*, 27(5):657–677. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0378-5920.2004.00620.x.

Olley, G. S. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297. Publisher: [Wiley, Econometric Society].

Poncet, S., Steingress, W., and Vandenbussche, H. (2010). Financial constraints in China: Firm-level evidence. *China Economic Review*, 21(3):411–422.

Santos Silva, J. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658. Publisher: MIT Press.

Tseng, W. and Zebregs, H. (2002). Foreign Direct Investment in China: Some Lessons for Other Countries. *IMF Policy Discussion Paper*.

Whalley, J. and Xin, X. (2006). China's FDI and Non-FDI Economies and the Sustainability of Future High Chinese Growth. Working Paper 12249, National Bureau of Economic Research. Series: Working Paper Series.

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

Young, A. (1995). The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience. *The Quarterly Journal of Economics*, 110(3):641–680. Publisher: Oxford University Press.

9 Appendix

A1 Identifying FFFs and DFFs

The distinction of foreign funded firms and private or state Chinese funded firms is made primarily from firms' reported paid-in capital (shí shōu zīběn), and secondarily through their registration-type (dēngjì zhùcè lèixíng) for a small number of observations with missing paid-in capital data. There are then a small number of firms that do not report capital funding but ownership can be identified by the registration-type, including 0.8 percent of firms that are among the top-3 employers in their industry.

A firm is officially recognized as a foreign invested enterprise in China if its non-mainland China capital investments are 25% or more than its total capital investments. The main analysis, however, leaves the definition at 50%, as this paper is concerned with comparing the fundamental differences between foreign affiliates and local Chinese firms, and FFFs with FDI amounting to 25% to 50% of firm capital may behave more like DFFs rather than FFFs. This range of FDI capital also has a higher likelihood of owing only to "round-tripping," whereby local capital resources are rerouted through Hong Kong or other markets in order to benefit from tax incentives given to foreign invested enterprises. Additionally, labeling FFFs as including those firms with FDI in the range of 25% to 50% would take away from the most productive DFFs, thus inflating the differences between FFFs and DFFs and may bias results for applications that rely on direct comparison between these two types of ownership. Regardless, the main results of this paper hold whether using the 50% definition or the 25% definition. Additionally, results hold if excluding FFFs that have more HKMT capital than non-HKMT capital. In other words, potential round-tripping through HKMT is not a concern; likely due to the fact that I focus only on granular FFFs and not the bulk of non-granular FFFs.

The first three columns of table A1 shows what percentage of all firms are FFFs versus DFFs, and how many firms fall into each of 4 categories - for the years 1998, 2007 and across the whole 10 years of the panel. FFFs made up 11.8% of firms in 1998 and 16.0% of firms in 2007. Also of note is the considerable reduction in the number of state-owned firms, which is discussed in further detail in the next subsection within the context of large granular firms. The right-4 columns of table A1 give a breakdown of total sales and exports according to funding-type for 1998 and 2007. FFFs of any origin saw a more than an 8-fold increase in total sales from a combined 1,138 billion RMB in 1998 to 9,454 billion RMB in 2007, corresponding to an increase from 20.0% to 27.4% of the annual aggregate total. Total sales here includes both production sold to the Chinese domestic market as well as exports. There is also an increase of FFFs' share of aggregate exports from 49.7% in 1998 to 61.4% in 2007.

The prevalence of FFFs in terms of employment and total wage bill also grew significantly over the period, employing 10.9% of workers in 1998 at 15.3% of the wage bill and 26.8% of workers in 2007 at 31.4% of the wage bill. Notably, by 2007 FFFs did not produce a disproportionate amount given their labor input relative to DFFs. Thus, the importance of FDI to emerging economies, in the sense of making up significant portions of aggregate values, is well demonstrated in China over the panel of consideration in this paper.

Table A1: Break-down of panel data according to dominate firm funding type

	# of firms in survey data			total sales, b. RMB		exports, b. RMB	
	1998	2007	Total	1998	2007	1998	2007
% FFFs	11.8%	16.0%	15.0%	20.0%	27.4%	49.7%	61.4%
Foreign FFFs	8,038	25,707	147,553	684	6,601	284	3,124
HKMT FFFs	9,348	24,036	157,647	454	2,853	236	1,344
% DFFs	88.2%	84.0%	85.0%	80.0%	72.6%	50.3%	38.6%
Private DFFs	40,751	241,678	1,234,208	1,475	20,163	201	2,332
State DFFs	89,433	19,988	493,489	3,083	4,893	326	482
Total	147,570	311,409	2,032,897	5,696	34,511	1,046	7,282

Note: Percentages compare all FFFs to all DFFs, which are each made up of two subgroups as shown. The first 3 columns count the number of firms, while the last 4 aggregate their total sales and exports (a subset of total sales) and are in current billions of RMB. Firms kept in dataset are in one of 420 industries included in the analysis, and for which dominant origin of firm funding (ownership) can be deduced either capital funding or registration.

A2 Changing composition and relative dominance of granular firms

Table A2 below shows how ranking firms separately per funding type identifies each groups' different patterns of granularity. In each year, the top DFF takes both a larger industry sales share and a larger export share on average across industries compared to the averages of the largest FFFs. This still holds when looking at groups of the top-3 firms per industry. The difference is less pronounced, however, in terms of export shares, and adding these two groups together accounts for over 50% of industry exports on average for each year in the sample.

Table A3 shows how the top-6 firms from each of the 420 industries are distributed across the different funding-types and for 5 different methods of ranking firms. 32.1% of top-6 firms were FFFs in 2007 when ranking by total sales. When ranking by firm exports, 50.2% of top-6 firms were FFFs by 2007. However, in terms of sales to the domestic market, a much lower 24.3% of firms were FFFs.

For comparison, table A5 shows number of industries where FFFs take the number 1 spot by either total sales, exports, or sales to the domestic market. In 1998 112 out of the 420 were FFFs, increasing to 125 by 2007. A FFF was the top exporter in 190 industries in 1998, and 205 in 2007. FFFs are not only dominate at the industry level, but also take 299 places in the top-500 exporters in 1998, increasing to 336 places by 2007.

The dominance of FFFs as the largest firms grew within industries, although mostly through the foreign category, as the number of top-6 firms that are HKMT remained mostly unchanged from 1998 to 2007, both in terms of total production and exports. The number of top-6 firms that are private DFFs also grew in terms of all ranking methods, more than doubling their presence. What gave was the number of state-owned firms in a top-6 position, which from 1998 to 2007 declined from 1,226 spots to 319 spots in terms of total sales, or just 12.7%. Table A5 shows a similar decline in the number of industries where state-owned firms remained the largest. This is largely due to their privatization and also partly responsible for the increase in private DFF

Table A2: top-firms summary statistics - industry averages

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
# of firms	352	346	349	369	394	429	614	592	661	742
of which are FFFs	41	43	46	51	57	65	99	98	108	118
top firms SS	.143	.144	.147	.144	.14	.132	.117	.12	.118	.115
	(.131)	(.138)	(.140)	(.136)	(.124)	(.126)	(.108)	(.113)	(.116)	(.114)
top firms ES	.299	.295	.29	.285	.281	.262	.224	.235	.232	.232
	(.253)	(.236)	(.226)	(.226)	(.225)	(.218)	(.193)	(.188)	(.198)	(.201)
top FFFs SS	.064	.06	.064	.062	.064	.064	.061	.061	.057	.058
	(.089)	(.079)	(.078)	(.079)	(.079)	(.086)	(.074)	(.074)	(.072)	(.077)
top FFFs ES	.183	.18	.177	.166	.171	.164	.148	.147	.144	.139
	(.212)	(.189)	(.179)	(.171)	(.178)	(.172)	(.161)	(.152)	(.149)	(.149)
top DFFs SS	.122	.127	.128	.124	.118	.11	.096	.099	.098	.093
	(.121)	(.133)	(.135)	(.131)	(.117)	(.113)	(.101)	(.105)	(.108)	(.104)
top DFFs ES	.205	.208	.205	.205	.196	.183	.15	.161	.161	.166
	(.224)	(.221)	(.216)	(.219)	(.21)	(.199)	(.165)	(.172)	(.183)	(.187)
$\sum_{i=1}^{3} SS_{z,i,t}^{FFF}$	.11	.107	.114	.113	.114	.113	.11	.11	.105	.105
	(.12)	(.111)	(.114)	(.116)	(.116)	(.116)	(.11)	(.11)	(.109)	(.112)
$\sum_{i=1}^{3} ES_{z,i,t}^{FFF}$	.282	.288	.287	.275	.276	.269	.249	.248	.247	.245
	(.242)	(.229)	(.224)	(.223)	(.221)	(.216)	(.205)	(.203)	(.198)	(.203)
$\sum_{i=1}^{3} SS_{z,i,t}^{DFF}$	.231	.235	.236	.232	.223	.21	.184	.187	.184	.174
	(.185)	(.188)	(.192)	(.19)	(.177)	(.169)	(.15)	(.15)	(.152)	(.148)
$\sum_{i=1}^{3} ES_{z,i,t}^{DFF}$	.339	.34	.334	.337	.326	.305	.263	.273	.271	.277
	(.278)	(.279)	(.273)	(.275)	(.271)	(.251)	(.231)	(.237)	(.235)	(.241)

Note: SS denotes total sales shares, where total sales is the sum of exports and domestic sales. ES denotes export shares. Figures show industry-level data averaged across all 420 industries. FFFs include those of both HKMT or foreign funding. DFFs include private as well as state-owned firms. The summation $\sum_{i=1}^{3}$ takes the top-3 firms according to SS or ES, as specified, and within each funding group.

firms, though those also grew in their own right. Table A4 shows how firms that made it into a top-6 position at least once over the 10 years changed their primary funding type. There were a total of 8,303 firms that made it to a top-6 position at least once in an industry. 2,160 of these firms changed to a new dominant ownership (funding) group either before, while, or after being in a top-6 position, with respect to the four funding-types. Large state-owed firms saw the most movement, with 1,118 being reclassified, 1,030 of which were to domestic private funding.

Table A3 shows 35.7% of top-6 firms in terms of ranking by wages paid out were FFFs in 2007, and 29.2% of top-6 firms were FFFs when ranking by number of employees; respectively 2.6% more and 2.9% less than when by total sales. Ranking by domestic sales on the other hand sees FFFs taking only 24.3% of the top-6 spots in 2007, or 7.8% less than by total sales.

Calculation of domestic sales shares cannot include imports at the 420-industry level. To match with imports data these must be aggregated up to 111 ISIC industries. Combined domestic sales of DFFs and FFFs make up between 77.8% to 84.1% of the entire domestic market in each year of the panel. The average 111-digit industry across all years sees the combined domestic sales of DFFs and FFFs taking 82.75% of the domestic market that includes imports.

There is a much different mix of firms in the top-6 spots when ranking by exports than when

Table A3: Break-down of top 6 firms by proxy shares

top 6 firm by:	total sales		exports		dom sales		wages		workers	
	1998	2007	1998	2007	1998	2007	1998	2007	1998	2007
% FFFs	25.5%	32.1%	41.8%	50.2%	18.1%	24.3%	20.9%	35.7%	13.3%	29.2%
Foreign FFFs	377	546	562	822	276	423	310	528	162	402
HKMT FFFs	262	263	418	418	178	188	214	371	171	335
% DFFs	74.5%	67.9%	58.2%	49.8%	81.9%	75.7%	79.1%	64.3%	86.7%	70.8%
Private DFFs	643	1,392	516	1,045	700	1,564	588	1,229	545	1,369
State DFFs	1,226	319	847	183	1,352	344	1,396	392	1,628	414
Total	2,508	2,520	2,343	2,468	2,506	2,519	2,508	2,520	2,506	2,520

Note: Funding types of the top-6 firms in each of the 420 industries. Percentage by main FFF or DFF group, and counts by further break-down into subgroups. In the main regressions, the top-3 firms from a funding group are used to build the granularity proxies, and may not match the overall top-6 firms described in this table.

ranking by one of the three granularity proxies of domestic sales, wages, or number of employees, with ranking by wages coming closest to that of by exports. Also, ranking by wage or employee shares means nearly every firm is ranked and every industry has at least 6 firms by 2007, to give a total of 2,520 top-6 firms across the 420 industries, while one industry is still a firm short in 2007 if looking only at domestic sales.

A3 Heterogeneity in firm size distributions

The Chinese data with 420 CIC codes can be aggregated up to 111 four-digit ISIC Revision 3 codes so to compare with international descriptions of industries. Aggregating the 111 matched four-digit ISIC industries further up to 22 two-digit industries allows for a clearer visualization of firm size distributions across the whole economy. Figure 7 shows estimated power-law exponents at this level of aggregation for DFFs and FFFs separately, in order to gain perspective on whether these two groups of firms systematically follow different parameters in their size distributions. There are 18 2-digit industries for which the power-law distribution describing the right tail for sales passes the goodness of fit test for both DFFs and FFFs. When comparing the estimated power law exponent for these 18 industries, FFFs have a higher alpha in 15, while DFFs have a higher alpha in the remaining 3. The pattern of lower alpha for DFFs may be an indicator that there is more inequality among DFFs than FFFs in a typical industry. These power-law estimates are only taken from a minimum threshold value of sales that is selected from a combination of criteria and tests. As the numerous smaller firms that fall below this minimum value are ignored in estimation, the differences in any sample selection bias between DFFs and FFFs according to how smaller firms enter the survey data is likely not a factor. This minimum threshold value also implies that the seemingly greater inequality among DFFs is actually inequality among the relatively large firms, rather than all the firms in general.

The size distributions for DFFs and FFFs are generally different even at the disaggregated four-

Table A4: Number of top firms that change dominant ownership group

change from:	to new group by last year observed					
	Foreign	HKMT	Private	State	any group	totals
Foreign FFFs	-	178	98	17	293	1,414
HKMT FFFs	176	-	82	16	274	961
Private DFFs	105	79	-	291	475	3,527
State DFFs	56	32	1,030	-	1,118	2,401
any group	337	289	1,210	324	2,160	
totals	1,458	976	4,259	1,602		8,303

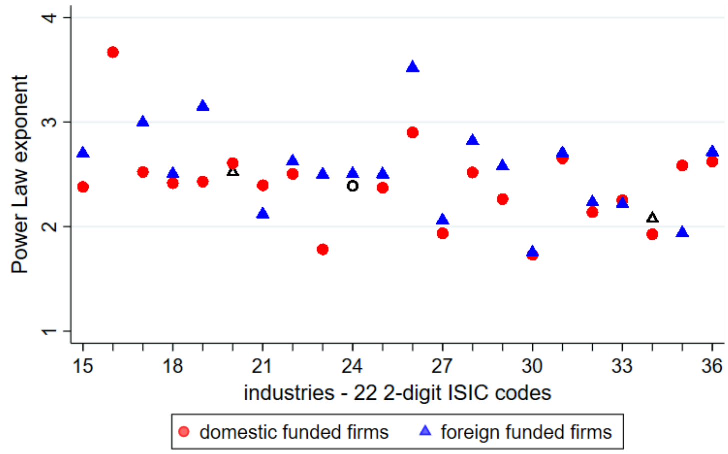
Note: There are a total of 8,303 firms that were a top-6 firm in terms of total sales for at least one year, i.e, only counting a firm once for the 10 years. of these 8,303 firms, 2,160 see a change in primary funding type. 1,030 initially state-owned firms became private DFFs, either in a year before, while, or after being a top-6 firm.

Table A5: Descriptive statistics of industry leaders

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
FFF is top sales	112	112	125	131	124	126	136	129	126	125
avg # of firms if FFF is top sales	385	403	377	377	392	363	526	445	525	644
portion FFFs	.15	.162	.169	.188	.187	.208	.216	.222	.24	.214
FFFs in top-500 sales	114	127	130	138	142	169	177	160	157	154
FFFs in top-500 sales, no state	278	255	253	230	229	243	241	211	201	198
avg # of exporters	83	82	88	96	107	120	182	177	186	187
portion FFFs	.323	.339	.342	.356	.35	.361	.389	.368	.375	.406
FFF is top exporter	190	195	201	195	213	202	221	210	213	205
avg # of firms if FFF is top exports	393	376	402	373	443	446	632	594	685	746
portion FFFs	.149	.167	.162	.164	.172	.169	.193	.197	.2	.203
FFFs in top-500 exports	299	309	316	309	321	325	352	355	341	336
FFFs in top-500 exports no state	397	386	379	370	368	370	391	380	369	368
FFF is top domestic sales	72	78	89	82	87	86	87	91	84	98
state firm is top sales	195	171	144	132	130	106	87	71	72	74
all state firms' sales share	.514	.452	.385	.319	.275	.224	.179	.135	.119	.101
state firm is top exports	133	109	98	89	79	65	47	45	44	46
all state firms' exports share	.359	.315	.27	.217	.189	.155	.114	.097	.089	.08

Note: Note: Other than top-500 figures, which are across all firms, figures show counts per industry or industry-level data averaged across either all 420 industries or the indicated subset. FFFs include those of both HKMT or foreign funding.

Figure 7: Power-law exponent estimates by firm funding type



Note: Goodness-of-fit p-value < 0.10 indicated by solid fill. Firms are aggregated up to 22 2-digit ISIC Revision 3 codes and split among DFFs and FFFs, resulting in 44 groups. Each group then has their power law exponent estimated separately. Note that DFFs generally result in lower power law estimates, implying more inequality in sales shares within industries compared to FFFs.

digit level. For each industry, the two categories of firms can be compared using the Kolmogorov-Smirnov test, where the null hypothesis is that the sales of DFFs and FFFs are drawn from the same distribution and with the same parameter values on this distribution. In order to guard against scale differences owing to superficial sampling reasons, the two sub datasets are first normalized at the aggregate level by taking $(sales - mean)/sd$ for FFFs and DFFs separately. This would dampen the ability to differentiate the distributions but provides additional robustness when stating that the distributions are different. Of the 420 industries in the year 2007, 268 have at least 30 FFFs and 30 DFFs. Among these 268 industries, 257 give Kolmogorov-Smirnov test p-values of less than 0.05, thereby rejecting the null that sales of DFFs and FFFs come from the same distribution. The null hypothesis is rejected for all industries with more than 63 FFFs. Where there is a small sample size of between 3 and 29 firms for one of the two groups, 62 additional industries exhibit significantly different distributions among DFFs and FFFs. All industries with more than 30 DFFs and 30 FFFs exporting have significantly different distributions of exports.

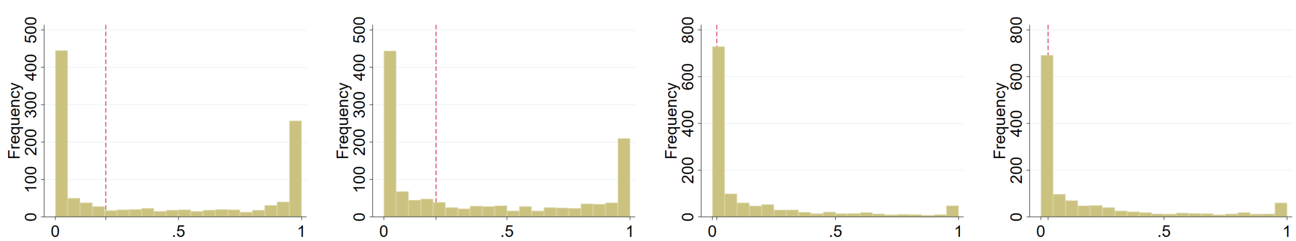
Finally, there are huge differences in export intensities of granular DFFs versus granular FFFs. This implies different exposures to trade and generally different business models that may logically translate into distinct channels of GCA, or lack thereof. Figure 8 displays histograms of all top-3 firms per group of DFFs and FFFs separately for the years 1998 and 2007. Notice the mass of export oriented firms shrinking slightly among top-3 FFFs, while it grows slightly among top-3 DFFs.

Table A6: Mean ranks of firms captured by definitions

	Top-3 dom sales		Top-3 wages		Top-3 workers	
	FFFs	DFFs	FFFs	DFFs	FFFs	DFFs
total sales rank	3.7	2.5	5.2	11.9	7.4	16.9
	(4.1)	(1.8)	(12.9)	(66.5)	(19.7)	(65.1)
exports rank	16.5	15.1	7.4	11.9	7.4	12.9
	(64.3)	(44.6)	(18.9)	(28.6)	(19.6)	(29.6)
dom sales rank			9.3	18.5	11.4	22.7
			(30.1)	(91.8)	(34.0)	(84.2)
Observations	11607	12586	11723	12573	11707	12528

Note: Mean coefficients; sd in parentheses. Groups of top-3 firms by wage bill shares and top-3 firms by employment shares have the same average exports rank of 7.4, but wage bill shares performs slightly better overall, lower average total sales ranks and lower average domestic sales rank.

Figure 8: export intensities of top-3 FFFs and DFFs by total sales



(a) top-3 FFFs in 1998 (b) top-3 FFFs in 2007 (c) top-3 DFFs in 1998 (d) top-3 DFFs in 2007

Note: Histograms show the heterogeneity in export intensities of granular firms. The red dashed line here indicates the median export intensity of all top-3 firms across industries separately for FFFs and DFFs and for years 1998 and 2007, according to the labels. Note the median export intensity of these groups of granular firms is near zero for DFFs (graph (c) and (d)), while there are masses of export-oriented granular firms pulling this median much further out for FFFs.

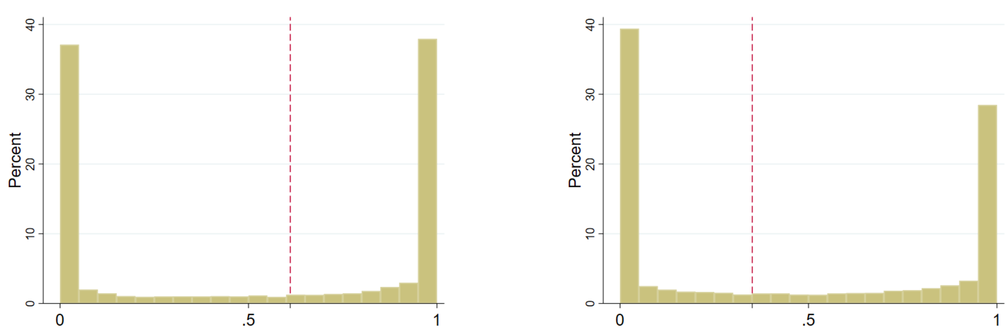
A4 Mapping employee and wage bill shares to horizontal and vertical FDI

Table A6 shows the mean total sales and export ranks of all firms identified by the three different granularity proxies. When grouping the largest 3 firms by domestic sales volume for each industry, separately for each category of FFFs and DFFs, the mean rank of total sales volume is 3.8 for FFFs, and 2.5 for DFFs, but a mean exporter rank of only 16.4 and 15.4, respectively, and thus, there is a bias towards granularity from domestic market-oriented firms. Sorting firms by wages or employment gives groups of top-3 FFFs that have an average export rank of 7.4, and each is also lower for DFFs than when sorting by domestic sales. Top-3 firms by employment, however, performs the least well in capturing firms ranked high in domestic and total sales, indicating it is the most export-biased method of grouping granular firms.

Export intensity has a low correlation with wage bill and employment shares of 0.0206 and 0.0636, respectively, among FFFs across the entire panel. In levels, wage bill, employment, log wage bill, and log employment correlate slightly more with export intensity at 0.0274, 0.1101, 0.134 and 0.257, respectively. In the following, I interpret export intensity as a spectrum mapping horizontally motivated FFFs on the left, towards an export intensity of 0, and vertically motivated FFFs on the right, towards an export intensity of 1. Figure 9 shows a histogram of FFFs' export intensities for the years 1998 and 2007. Notice the median export intensity, defined by exports over total sales, decreases from 0.61 in 1998 to 0.35 in 2007, indicating a shift of the typical FFF to focusing on the domestic market rather than on exports. With respect to the correlations, more employment should indicate a more rightward position on the spectrum, or a relatively more vertically motivated FFF, while this argument in terms of wages is much weaker.

To discuss the difference in what type of firms the similar granularity proxies of wage bill shares (WS) and employment shares (LS) tend to represent, table A7 associates export intensity with them and their log levels as controls, includes industry and year fixed effects, and clusters errors at the industry level. All models here are GLM with logit transformation, as export intensity lies between 0 and 1, with large concentrations around 0 and 1. Column (1) of table A7 includes only FFFs, giving 303,174 observations across the panel, and shows a strong negative coefficient on WS accompanied with a strong positive coefficient on LS. Thus, for a given LS, a higher industry WS is associated with firms more on the left of this horizontal-vertical spectrum. This pattern holds when including all DFFs and singling out FFFs with a dummy interaction term, denoted δ^{FFF}, in column (2). It similarly holds up with replacing shares with levels in column (3), and even when including shares and levels together as in columns (4) and (5).

Figure 9: export intensities of FFFs



(a) Export intensities in 1998 (b) Export intensities in 2007

Note: Red dashed line indicates median export intensity across all FFFs. The median was 0.61 in 1998 and decreased to 0.35 by 2007, indicating a greater proportion of FFFs' Chinese operations focusing relatively more on the domestic market rather than on exports.

Table A8 further shows firm-level regressions, still looking at all firms in the dataset, where the dependent variable is firm-level log exports or firm-level log domestic sales according to the label in the first row. The coefficient on wage bill share is twice as high in magnitude than that on total employment share in explaining exports of exporting DFFs in (1), while when interacting with δ^{FFF}, the FFF dummy, there is exactly the opposite relationship. The similarity of magnitudes on wage bill share for DFFs and FFFs but the much higher magnitude on employment share for

FFFs in explaining exports may point to FFFs' greater reliance on China's comparative advantage in labor intensive sectors for exports. The fact that the coefficient on the FFFs employment share interaction is so comparatively high attests to employment shares being a relatively vertical FDI biased proxy for sorting firms. An alternative interpretation is that foreign firms extract more productivity per additional worker, unconditional on their wage; but while this may explain why the coefficient is high for FFFs relative to DFFs, it does not explain why the coefficient is so much higher when explaining exports of FFFs compared to domestic sales of FFFs shown in column (2).

Table A7: Associating wages and worker shares with export intensity

	export int (1)	export int (2)	export int (3)	export int (4)	export int (5)
wages share	-13.364***	5.435***		-9.481***	-6.504***
x δ^{FFF}		-18.384***			-8.403***
worker share	48.926***	3.872***		9.684***	-3.416***
x δ^{FFF}		44.436***			23.202***
log wages			-0.092***	-0.061**	0.353***
x δ^{FFF}					-0.268***
log workers			0.442***	0.418***	0.078***
x δ^{FFF}					0.176***
δ^{FFF}		1.686***			2.737***
Industry & Year FE	Y	Y	Y	Y	Y
Observations	303174	1999201	302505	302505	1991085
deviance	263406	1005295	256080	255918	969208
deviance_p	18887608	30047303	225392	226655	1379778
chi2	69104	458223	73269	72841	466400

Note: Dependent variable is export intensity at the level of firm-industry-year. Exports are assumed not related to explanatory variables other than through their interaction with total sales (i.e. if multiplying both sides by the export intensity's denominator). This is plausible as the variation in exports is a subset of the variation in sales. If including log exports as an explanatory variable the estimated coefficient signs and significance levels do not change (not shown). All models are GLM with logit transformation. Industry-clustered robust errors; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

While the non-interacted-DFFs' coefficients maintain similar relative magnitudes in column (2) of table A8 as those in column (1), the FFFs' interaction terms see their relationship reversed compared to column (1), which are now more in line with the pattern for DFFs. This switch for the FFFs with the now much higher total magnitude on wage bill shares implies the relatively strong association of wages with horizontal FDI, which is manifested through more domestic sales. Including domestic sales shares along side wage bill shares and employment shares when explaining firm exports in column (3) hardly affects their coefficients compared with column (1), and domestic sales shares has a relatively low magnitude. In fact, if excluding DFFs from the regression (not shown), the coefficient is not statistically different from zero. Columns (5) and (6) each contain 3 regressions for each of the granularity shares separately, with the patterns just discussed still prevalent.

Table A8: Associating the granularity definitions with exports and domestic sales

	exports	dom sales	exports	exports	dom sales
	(1)	(2)	(3)	(4)	(5)
wage share	14.057***	21.126***	12.681***	19.594***	32.911***
x δ^{FFF}	-0.319	9.034**	0.768	13.472***	2.175
worker share	7.748***	15.652***	7.205***	22.354***	36.395***
x δ^{FFF}	26.876***	-4.744	27.538***	27.564***	11.773***
domestic sales share			2.447*	15.466***	
x δ^{FFF}			-1.885	1.731	
δ^{FFF}	0.576***	0.126***	0.576***		
R-sqr	0.169	0.186	0.169		
Obs	549458	1843091	549458		

Note: Dependent variable alternates from log exports to log domestic sales, both at the level of firm-industry-year. Columns (4) and (5) each include 3 separate regressions for each of the granularity proxies and their interaction with δ^{FFF} (δ^{FFF} also included separately in each but not shown). Industry and year fixed effects included. Industry-clustered robust errors; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Thus, I argue that ranking FFFs by domestics sales shares aligns almost entirely with ranking firms by horizontal FDI motives, employment shares best captures a ranking of vertical FDI motives, while wage bill shares, although likely more biased towards vertical FDI, is somewhere between the two. This is likely the reason wages in table A3 capture more foreign and HKMT FFFs in the top 6 industry rankings than do the relatively polarized rankings by domestic sales and employment.

Table A9: Extension of table A7: associating firm level wages and workers with export intensity

	exp int	exp int	exp int	exp int	exp int	export int	exp int	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
wages share	8.544***				7.713***			
x δ^{FFF}					2.872***			
worker share		32.272***				9.200***		
x δ^{FFF}						22.830***		
log wages			0.247***				0.387***	
x δ^{FFF}							-0.122***	
log workers				0.359***				0.380***
x δ^{FFF}								-0.048***
δ^{FFF}					1.725***	1.686***	2.510***	1.893***
Obs	303174	303174	302700	302942	1999201	1999201	1991968	1997952
deviance	265001	263692	260069	256654	1007033	1005695	972794	978072
deviance_p	226131	1396983	225658	225812	1327423	6007398	1353413	1357232
chi2	67527	69146	70996	73240	458702	458598	468017	466281

Note: See table A7. Robust errors; *** $p < 0.001$.

A5 Extended results

Table A10: 2sls using invested capital shares as IVs (Defining $\widetilde{s}$ as wage bill share)

	OLS IV			Poisson IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}$	-0.666**			-2.363*		
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		-1.741***			-1.617	
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			0.576			0.778
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			-1.853***			-2.181**
$D_{z,t}$	0.985***	0.940***	0.959***	0.836***	0.814***	0.779***
s1 v				1.818		
s1 v					0.993	
FFFs s1 v						-0.911
DFFs s1 v						0.994
Industry & Year FE	Y	Y	Y	Y	Y	Y
LM-stat	830	599	729			
CD-F-stat	930	628	396			
Observations	4153	4153	4153	4200	4200	4200

Note: Repeats the 2SLS exercise of table 6 in columns (1) - (3) and table 7 in columns (4) - (6) defining $\widetilde{s}$ by wage bill for the granularity proxy, where now wage bill granularity is instrumented for with the corresponding capital shares. Robust errors; * p<0.10, ** p<0.05, *** p<0.01

Table A11: Excluding the granular firms from industry exports

	Defining $\widetilde{s}$ by domestic sales			Defining $\widetilde{s}$ by employment		
	(1)	(2)	(3)	(4)	(5)	(6)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}$	-0.952***			-0.028		
	(0.241)			(0.289)		
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF} + \sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$		-0.953***			0.098	
		(0.229)			(0.270)	
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{FFF}$			-0.650*			1.352***
			(0.344)			(0.389)
$\sum_{i=1}^{3} \widetilde{s}_{z,i,t}^{DFF}$			-1.297***			-1.048***
			(0.297)			(0.390)
$D_{z,t}$	0.391***	0.378***	0.355***	0.817***	0.815***	0.746***
	(0.068)	(0.070)	(0.071)	(0.040)	(0.041)	(0.036)
Industry & Year FE	Y	Y	Y	Y	Y	Y
pseudo-R-sqr	0.968	0.968	0.968	0.978	0.978	0.980
Observations	4180	4180	4180	4190	4190	4190

Note: Repeats the Poisson regressions of columns (7) - (9) of table 2 and columns (4) - (6) of table 3, excluding the exports of granular firms from industry aggregates. Robust errors; * p<0.10, ** p<0.05, *** p<0.01.

Table A12: Markups regression summary stats

	mean	sd	min	max
$mu_{z,i,t}$; firm markup	.1653586	.2102257	-7.165493	4.521645
$SS_{z,i,t}$; firm sales share	.001493	.0037833	0	.1712571
$SS^2_{z,i,t}$	.0000165	.0001594	0	.029329
SS^{FFF}_{zt}; top-3 FFFs S	.0619915	.0641722	0	.876778
SS^{DFF}_{zt}; top-3 DFFs S	.1066401	.0881049	.0080249	.9919031
$SS^{FFF}_{zt} \times SS_{z,i,t}$	.0001357	.0006022	0	.058714
$SS^{DFF}_{zt} \times SS_{z,i,t}$	.0002509	.0011369	0	.1247247
mu^{FFF}_{zt}; top-3 FFFs markup avg	.2212758	.1297107	-1.511568	3.870852
mu^{DFF}_{zt}; top-3 DFFs markup avg	.2107103	.1005985	-.5063486	1.219947
$mu^{FFF}_{zt} \times SS^{FFF}_{zt}$	.0137781	.0196513	-.0855598	1.032074
$mu^{DFF}_{zt} \times SS^{DFF}_{zt}$	.0223764	.0241158	-.4259783	.4809249
$tfp_{z,i,t}$; TFP	4.847174	1.213008	-2.12472	10.6708
tfp_{zt}; ind TFP	5.92447	.8816606	.3063346	10.23052
$lp_{z,i,t}$; labor productivity	3.852993	1.161568	-5.925182	11.87841
$w_{z,i,t}$; price of labor	2.506345	.7647095	-7.81843	10.2643
w_{zt}; ind price of labor	2.710029	.4517251	1.244393	4.810623
$ci_{z,i,t}$; capital intensity	3.643026	1.248059	-7.046785	11.92603
ci_{zt}; ind capital intensity	93.63037	79.48284	9.431174	973.9621
obs_{zt}; number of firms	6.877505	1.121089	1.386294	9.540435
$EI_{z,i,t}$; export intensity	.173762	.3452372	0	1
EI_{zt}; ind export intensity	.2256916	.2036394	0	.9714133
IP_{zt}; import penetration	.5808644	.3376136	.0000563	.9998213
Observations	1824552			

Note: Summary statistics for observations included in regressions of table 11 over all years of the panel data, 1998 through 2007. Firm level TFP and markup outliers are excluded if both more than three standard deviations from the mean as well as in the top or bottom 1 percent of observations in that industry-year. Lowercase variables are in log. See text for variable descriptions.