Export-Induced Spatial Divergence

Jonas Casper* Lei Li† Jinfeng Luo‡

First draft: March 2023
This version: October 2023

Abstract

How does export liberalization affect firm location choice and the spatial concentration of economic activity? We address these questions using the geo-coordinates of Chinese manufacturing firms and find that export widens inter-city and intra-city spatial disparities by reinforcing initially large industry centers. We first show that there has been an increased spatial concentration across cities in response to improved foreign market access. Only industry city pairs that were large initially increase their employment density following trade liberalization. Second, there has also been an increased spatial concentration within cities. For a given industry, districts closer to city centers are getting denser, mainly driven by the extensive margin. Third, the above effects are not exclusive to industries directly exposed to export shocks but also spill over positively to upstream and downstream industries and negatively to industries competing for the same workers locally.[0]

Keywords: firm location, localization, spatial concentration, regional inequality, export, comparative advantage

JEL Codes: F6, F14, R12

*Department of Economics, LMU Munich, Ludwigstraße 28, 80539, Munich, Germany. E-mail: jonas.casper@econ.lmu.de.

†Department of Economics, University of Mannheim, L7, 3-5, 68161, Mannheim, Germany. E-mail: lei.li@uni-mannheim.de.

‡Deparment of Economics, Lingnan University, 8 Castle Peak Road, Tuen Mun, Hong Kong. Email: jinfengluo@ln.edu.hk

[0]We are very grateful to Jan Bakker, Loren Brandt, David Dorn, Carsten Eckel, Harald Fadinger, Jingting Fan, Lisandra Flach, Inga Heiland, Ines Helm, Martina Magli, Isabela Manelici, Andreas Moxnes, Ralph Ossa, Claudia Steinwender, Karen Helene Ulltveit-Moe, Nico Voigtländer, Horng Chern Wong, Yuan Zi, and seminar and workshop participants for their valuable comments and suggestions. Support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged.

1 Introduction

The increasingly uneven distribution of production and trade across geographic space is one of the most remarkable features of the economy (Marshall 1890; Duranton and Overman 2005; Combes et al. 2012; Redding 2020). Given the importance of spatial concentration, there has been a large literature exploring the underlying determinants. However, there is little evidence on the impact of international trade on spatial concentration with a few exceptions (Fajgelbaum and Redding 2022; Bakker et al. 2021).

In this paper, we explore the role of international trade in shaping the spatial distribution of manufacturing firms. Two plausible mechanisms can explain the phenomenon - market access and regional comparative advantage. The former suggests that reducing trade costs leads to increased population in regions with greater access to foreign markets, regardless of location fundamentals (Redding and Sturm 2008; Redding and Rossi-Hansberg 2017; Fajgelbaum and Redding 2022). However, location fundamentals such as productivity do matter in shaping regional comparative advantages. The latter mechanism implies that when two regions have the same access to foreign markets, the one with a comparative advantage in the liberalizing industry will benefit more. Trade liberalization reinforces within-country regional comparative advantages, inducing spatial concentration.

Despite the large literature on the determinants of spatial concentration, a complete answer has so far proven elusive. This is because it has been exceedingly difficult to measure spatial concentration. As firm-level geo-location information is hard to obtain, past works have mostly relied on inter-regional (or sector-by-region) variations in employment or output.[1] The regional aggregated data has several limitations. First, little is known about the intra- city spatial concentration due to the lack of firm-level geo-coordinates. Second, when researchers measure inter-city spatial concentration, city-level aggregated data creates spatial discontinuity regarding the scale and boundary of administrative units. If one industry cluster happens to locate along the borders of two administrative units, aggregation by administrative units splits the industry cluster. Lastly, the same inability to directly observe the changes of firm locations over time makes it extremely difficult to understand the determinants of firm location dynamics and thus limits previous research from

[1] A few exception includes Alfaro and M. Chen (2014) and Alfaro, M. Chen, and Fadinger (2019) using the firm-level geocode data in 2005 and Duranton and Overman (2005).

painting a complete picture of the role of international trade.

We explore the role that regional comparative advantage (proxied by initial industry-city cluster size) plays by using novel Chinese firm-level data with geo-location information. The novel firm-level data on geo-coordinates allows us to explore the dynamic impact of trade liberalization on spatial concentration. We exploit differential reductions in export uncertainty for Chinese industries when China joined the WTO in 2001 in a dif-in-dif type setting. Our results show that increasing access to foreign markets increases both inter-regional and intra-regional spatial concentration. Inter-regional spatial concentration increases because only clusters that were already large initially expand, while intra-regional concentration increases because this expansion is strongest around the cluster center. These patterns hold even if we hold the traditional market access channel fixed by controlling for distance to the coast and changes in internal trade costs. Hence, trade liberalization reinforces regional comparative advantages and in doing so increases the spatial concentration of economic activity by reinforcing clusters that were large initially.

Specifically, we ask three questions. First, what's the impact of export liberalization on inter-city spatial concentration? Second, what's the impact of export liberalization on intra-city spatial concentration? Third, how do industry-specific export shocks spill over to other industries? To measure export shocks, we utilize a large reduction in trade uncertainty for Chinese firms after the US granted China permanent normal trade relations (NTR) in accordance with China's accession to the WTO in 2001 (Pierce and Schott 2016; Handley and Limão 2017).

We start the analysis by studying the direct effect of export liberalization in shaping the spatial clustering of firms across cities. We provide evidence on the existence of a Darwin effect where advantage begets further advantage. We find that large industry-city pairs are getting larger in response to increased access to foreign markets, while smaller and medium size clusters do not get larger. This reinforcement of initially large agglomeration clusters leads to a sizeable increase in the spatial concentration of economic activity. Contrary to theoretical predictions, these reinforcing agglomeration effects are primarily driven by entry rather than an expansion of the original firms in dense areas. While in principle incumbents also benefit from liberalization, in initially dense areas these benefits are outweighed by increased competition from new entrants.

Furthermore, we show that the Darwin effect not only holds across cities but holds across districts within cities. Districts close to the cluster center expand more as a consequence of trade liberalization, while employment density increases attenuate as one moves away from the center. This implies that within-city economic development has become more unequal over time. Moreover, the positive impact of increasing market size on agglomeration are sizeable but attenuates with distance. For an average cluster in the fourth quartile of the initial size distribution facing one standard deviation higher reduction in trade uncertainty, employment density increases by 21.2% 10 to 20km from the cluster center. We find positive effects on density up to 100km away from the initial cluster center, but not in its immediate vicinity (0 to 10km).

The direct effect does not, however, correspond to the full impact of trade liberalization on agglomeration, which also encompasses several indirect channels. We study the spillover effects and find that these direct density expansions spill over locally to related industries. The spatial concentration of industries upstream and downstream of the liberalizing industry also expand, but only if the initial cluster of the liberalizing industry was already large. On the other hand, industries that use a similar labour pool as the liberalizing industry contract, indicating local competition for labor and imperfect labor mobility.

Our paper relates to the large literature on the driving forces of agglomeration, such as input-output linkage, labor pool, technology diffusion (Ellison, Glaeser, and Kerr 2010; Combes et al. 2012; Alfaro and M. X. Chen 2018; Gaubert 2018), and trade cost. In addition, Alfaro and M. Chen (2014) and Alfaro, M. Chen, and Fadinger (2019) study the spillover effect of multinational firms on the agglomeration of domestic firms, and Lu et al. (2019) examine the role of special economic zone, where both of them use detailed geo-location information. Regarding the role of international trade, Curuk and Vannoorenberghe (2017) and Helm (2020) use region-industry level data to study how the export growth in one industry affects the employment growth in other related industries. They show that trade's spillover effect mainly works through the channels of the labor pool and vertical linkages. Bakker et al. (2021) shows that export-specific agglomeration spillovers arise due to the self-selection of productive firms into exporting and economically dense areas, which subsequently lead to larger welfare gains in dense areas following trade liberalization.

We contribute to the literature on international trade and urban economics by exploring both the inter-city and intra-city effects. No consensus has been reached regarding the impact of trade on the inter-regional agglomeration and there is little evidence on the impact of trade on intra-regional agglomeration due to data limitation. Using firm-level geo-location data, we find that export widens inter-city and intra-city spatial disparities in economic activity. There has been an increased spatial concentration across cities in response to improved foreign market accesses. For a given industry, cities that were initially populous are getting more populous over time. Regarding the within-city effect, districts that are closer to industry-specific city centers are getting more populous, which implies export also plays a reinforcing role for initially populous districts.

Second, this paper relates to the literature on agglomeration and agglomeration formation. We show that trade liberalization reinforces within industry agglomeration and co-agglomeration between industries. We contribute to the literature on the measurement of agglomeration (Ellison and Glaeser 1997; Duranton and Overman 2005) by measuring agglomeration at a fine spatial scale. The geo-coded firm level data allows us to capture one key features of agglomeration, namely employment density. By using a concentric ring approach, we show that the effect of trade liberalization on agglomeration attenuates with distance. The geo-coded firm-level data further enables us to capture the intensive and extensive margin of agglomeration. We show that the positive effect of trade liberalization on agglomeration is mainly driven by new entrants. We then show that the impact of trade on agglomeration attenuates with firms' distances to cluster centers.

1.1 Literature

This paper contributes to two strands of literature and highlights their connection. The first strand is the literature on agglomeration, which is concerned with why firms agglomerate and whether agglomeration spillovers exist. The second strand is the recent literature on spatial dimensions along which trade shocks disseminate.

Agglomeration

Marshall (1890) already noticed the severe geographic concentration of economic activity and attributed it to three main reasons. The first are spillovers due to input-output linkages. Proximity to suppliers and customers reduces transaction and transportation costs. The second reason is knowledge spillovers between firms in the same industry, while the third reason is the pooling of specialized labor. More recently, Duranton and Puga (2004) proposed that there are also sharing effects, which relate to the pooling of risk, the use of local public goods, and gains due to a greater input variety.

Krugman (1991) picked up on this notion but heavily emphasized input-output linkages. In his model, increasing returns to scale and transport cost play a key role. The key argument is that firms benefit from locating close to suppliers and customers, as transport is costly. This in turn attracts labor, further increasing local product demand and thus economic concentration (Krugman 1991). The model maintains identical firms, which Baldwin and Okubo (2005) extend to a model with heterogeneous firms. The key result from Baldwin and Okubo (2005) is that the most productive firms self-select into clusters as lower productivity firms cannot survive in areas with high wages, high land rents, and heavy competition.

The self-selection mechanism highlighted by Baldwin and Okubo (2005) and also Nocke (2006), is potentially at odds with the existence of agglomeration spillovers. If economically dense areas form due to path dependency, first order geography, or local amenities, then the productivity advantage of dense areas might solely be driven by self selection, calling into question whether agglomeration spillovers exist at all.

Recent evidence, however, suggests that both forces play a role. Firms sort into economically dense areas, and in turn, agglomeration spillovers make them more productive. Greenstone et al. (2010) use a dataset of US counties that won or barely lost a large plant opening as a quasi-natural experiment. They find that large plant openings increased incumbents' total factor productivity (TFP) by 12% over a six year period. These gains were larger for incumbents that were economically close to the new plant, especially in terms of technology usage and labor pooling (Greenstone et al. 2010). In fact it seems like all of Marshall's three reasons for agglomeration drive the co-agglomeration of industries, as Ellison, Glaeser, and Kerr (2010) find. The prime reason for co-agglomeration between industries is input-output linkages, while externalities from knowledge spillovers and labor pooling play a lesser role (Ellison, Glaeser, and Kerr 2010). Dense areas, economically as well as in terms of population, are usually cities. Combes et al. (2012) specifically try to test whether the productivity advantage of large cities can be explained entirely by self-selection. In their data on French establishments, they conclude that it cannot. Also using French data, Gaubert (2018) tries to quantify to what extent self-selection or agglomeration spillovers drive the productivity advantage of large cities. She finds that about half of the productivity advantage is due to self-selection, while the other half stems from agglomeration spillovers.

The interaction between agglomeration and productivity also has implications for labor markets.

Agglomeration spillovers can potentially magnify local economic shocks. For example, Gathmann et al. (2020) find considerable local spillovers to mass lay-offs that amplify regional employment losses. Similarly, Helm (2020) shows that trade shocks propagate locally to other industries that are connected through labour pooling.

We add to the literature by providing evidence in line with agglomeration spillovers magnifying local economic shocks that result in further spatial concentration of the industry directly affected by the shock and industries upstream and downstream in the supply chain. Additionally, we show that these agglomeration effects attenuate over distances up to 100km. Furthermore, we decompose these concentration effects to show that they are primarily driven by new entrants rather than incumbents.

Lastly, we add to the literature on the measurement of agglomeration. The lack of geo-coded firm level data induces researchers to use more aggregate levels such as city level or industry-city level data (Ellison and Glaeser 1997; Helm 2020; Bakker et al. 2021; Bakker 2021). This can lead to measurement error as firms that are virtually next to one another might still be on opposing sides of administrative regional borders. This problem was first addressed by Duranton and Overman (2005), who use geo-coded data on UK firms to investigate the extent to which firms in different industries tend to agglomerate. Their measure, which is based on the density of firms over distance captures what we refer to as "closeness", i.e., the extent to which firms cluster close to one another. However, their measure tries to capture how localized different industries are compared to one another and cannot speak to the strength of agglomeration of any given cluster. To expand on this we propose a measurement of agglomeration that varies by industry-location pairs and captures two dimensions of agglomeration: "closeness" - to what extent do firms cluster together; and "thickness" - the employment density of a location. Our measure captures the strength of the two dimensions over space around the cluster, and can thus show how agglomeration effects attenuate over distance.

Spatial dimensions of trade shocks

The literature on trade has only recently begun to disentangle the spatial dimensions by which trade shocks ripple through the economy. Most of this literature has focused on the effects of Chinese import competition on local labor markets, which drastically increased with China's accession to the WTO in 2001. Autor et al. (2013) explore the relationship between the US manufacturing employment decline and a local labor market's exposure to Chinese imports. They find sizeable effects and attribute a quarter of the manufacturing employment decline in the US between 1990 and 2007 to increased competition with Chinese imports. A range of studies has since confirmed similar effects for other countries. For example, Dauth et al. (2014) find similar effects on German manufacturing employment in exposed local labor markets, though they note that the fall of the Iron Curtain and trade integration with Eastern Europe played a larger role. Balsvik et al. (2015) and Donoso et al. (2015) also find significant manufacturing employment declines related to Chinese import competition in Norway and Spain, respectively.

Acemoglu et al. (2016) subsequently follow up on Autor et al. (2013) and include indirect competition through input-output linkages on a national level, and find that it accounts for about half of the total employment losses. More recently, Bloom et al. (2019) highlight neglected firm dynamics behind the impact of Chinese imports on US local labor markets. Specifically, they find that Chinese import competition generated no net job losses, but instead firms switched out of manufacturing into services. This reallocation happened along a spatial dimension, as employment shifted from manufacturing in the US heartland towards services at the coasts (Bloom et al. 2019). This paper contributes to this literature by showing that trade shocks not only have varying regional implications because of pre-existing industry compositions, but also that they increase the spatial concentration of economic activity within sectors due to pre-existing agglomeration externalities.

Agglomeration and Trade

After establishing the importance of spatial dimensions to trade shocks, the literature has started to consider the role agglomeration economics plays in disseminating these trade shocks. Curuk and Vannoorenberghe (2017) and Helm (2020), pick up on the role agglomeration economics play in the impact of trade shocks. Helm (2020) looks at Germany's initial local industry composition and finds that positive net trade shocks to a national industry had positive spillovers to other industries in the same labor market. These spillovers were primarily transmitted through labor sharing, in line with one of Marshall's three forces (Helm 2020). Dix-Carneiro and Kovak (2017) study local labor market adjustments to trade liberalizations in Brazil and find dynamic differences between regions that increased over time. These dynamic effects provide supportive evidence on the existence of the agglomeration economies forces.

Other studies that explore the connection between trade and agglomeration take two diverging views. The first view is that agglomeration or the density of economic activity is an outcome of trade and internal geography. Agglomeration externalities do not exist and the spatial distribution of economic activity is based on transportation cost and comparative advantage. Coşar and Fajgelbaum (2016) and Fajgelbaum and Redding (2022) explore this perspective in the context of China and Argentina, respectively. Sectors with a comparative advantage benefit more from access to global markets through ports and hence distribute themselves spatially in a way that reduces internal transport costs. This leads to spatial clustering, and hence agglomeration. However, agglomeration is purely an outcome of internal trade cost and comparative advantage and does not itself contribute to the spatial effects of trade.

A different perspective to which this paper belongs emphasizes the interconnections between trade and agglomeration due to heterogeneous firms. Trade theory suggests that firms sort into exporting according to their productivity, and that subsequent trade liberalization benefits the relatively more productive and exporting firms the most (Melitz 2003). Agglomeration theory follows a similar logic. The most productive firms self-select into economically dense areas (Baldwin and Okubo 2005; Gaubert 2018) which in turn increase their productivity (Gaubert 2018). Combining trade theory and agglomeration theory implies that exporters should disproportionately cluster in economically dense areas, which Bakker et al. (2021) confirm using Chinese firm data. Furthermore, Bakker (2021) shows that export specific agglomeration spillovers arise due to the self selection of productive firms into exporting and economically dense areas. These subsequently lead to larger welfare gains in dense areas following trade liberalization.

Greenaway and Kneller (2008) already found export specific agglomeration spillovers by investigating the relationship between agglomeration and the likelihood that firms export. They obtain evidence consistent with agglomeration spillovers raising the probability of exporting in their dataset of UK manufacturing firms. Entering the export market further increases the productivity of the firms that entered (Greenaway and Kneller 2008). Similarly, Koenig (2009) and Koenig et al. (2010) use firm level data and firm export transaction data from France to show that proximity to other exporting firms raises the likelihood that a firm will also start to export. These export spillovers are industry and destination country specific, implying that firms within the same industry begin to export to the same export markets that other local firms are already serving.

One important implication of combining trade theory as in Melitz (2003) and agglomeration theory as in Gaubert (2018) is that trade liberalization should benefit the densest agglomeration clusters the most, as this is where productive exporting firms are located that can take the greatest advantage of increased access to foreign markets (Bakker et al. 2021; Bakker 2021). We directly add to this literature by empirically showing that indeed it is only the largest pre-existing clusters that expand following trade liberalization whereas smaller clusters stagnate. We further show that these expansions propagate to related industries and attenuate over distances up until 100km. Lastly, in contrast to predictions in line with Melitz (2003), we show that this is mainly driven by new entrants and not by an expansion of productive incumbents.

The rest of the paper is organized as follows. Section 2 discussed the related literature. Section 3 introduces the data, while section 4 describes the empirical approach. Section 5 presents the results and section 6 concludes.

2 Data and Variable Construction

2.1 Data

The primary data source is the Annual Survey of Industrial Firms (ASIF) from the Chinese National Bureau of Statistics (NBS) covering the years 1998-2007. The survey covers “all state-owned and above-scale private-owned industrial enterprises”, where “above-scale” refers to firms with annual sales greater than or equal to 5 million RMB (about $ 770,000 at exchange rates in 2021). The data descriptions are shown in Appendix Table A1. While the unit of observation is a firm, subsidiaries of larger firms independently enter the sample as long as they are defined as a separate legal unit (Brandt, Van Biesebroeck, and Y. Zhang 2014). Additionally, in some years, the surveyed firms were asked how many of their establishments engage in industrial activity. In 1998, 88.9% of firms reported only a single production plant, a share that increased to 96.6% of firms in 2007 (Brandt, Van Biesebroeck, and Y. Zhang 2014). This suggests that ASIF comes close to an establishment-level survey. Using data from the economic census of 2004, Brandt, Van Biesebroeck, and Y. Zhang (2012) show that ASIF captures most of the industrial activity in China. Though it only accounts for 20% of all industrial firms in 2004, these employ 71.2% of the industrial workforce, create 90.7% of output and generate 97.5% of exports.

Though a unique firm identifier allows us to match firms over time, some firms' identifiers change due to mergers, acquisitions, restructuring, or because they are simply assigned a new number. To ensure that we can adequately track firms over time, we employ Brandt, Van Biesebroeck, and Y. Zhang (2012)'s matching approach. Their approach is based on five steps, where each step only uses entries not matched by the previous step. In the first step, we match firms by their unique identifier, then by the firm's name, and then by a combination of the legal representative's name, prefecture code, and industry code. In this step, we depart from Brandt, Van Biesebroeck, and Y. Zhang (2012), who use only the representative's name and the firm's prefecture code, as we found this to produce inconsistent matches frequently. Lastly, we match by combining the firm's phone number and prefecture code, and then by combining the founding year, 6-digit region code, industry code, address and primary product name.

We drop firms with missing values in employment, exports and wage bill as well as firms that report negative exports or negative gross output. Following Brandt, Van Biesebroeck, Wang, et al. (2017) we also drop firms with fewer than 8 employees as these fall under a different legal regime. Do avoid re-exporting firms we further remove firms that report and export output ratio greater than 2. The Chinese Industrial Classification (CIC) changed between 2002 and 2003, to consistently track industries we employ the crosswalk by Brandt, Van Biesebroeck, and Y. Zhang (2012). As part of this change some industries were reclassified as services, hence we also remove these from our sample[2]. We further drop other industrial non-manufacturing industries (e.g. mining or oil refining). Lastly, all firms where longitude or latitude could not be inferred from the address data are also removed. This leaves us with 142,393 firms in 1998 and 310,064 firms in 2007.

The geo-location data is obtained based on the firm address. The ASIF data provides detailed annual information on firm name and firm address, which can be used to link to locations defined by longitude and latitude coordinates through Baidu Map API v3.0, an online map and navigation system widely used in China. The Baidu Map API is based on the GCJ02 coding system which is later transfered to the WGS84 coding system.[3] With the geocodes, we are able to trace the dynamics of the geographic distribution of Chinese manufacturing firms from 1998 through 2013.

[2]These are CIC codes 1711, 1712, 1713, 1714, 2220, 3648, 3783, 4183, and 4280.

[3]There is one caveat of the geocode database. Because the retrieve was done in 2021 and was based on the most recent Baidu Map, a small number of firms that changed address may not be found. Based on firm names, the retrieved coordinates are based on firms' recent addresses. The share of firms with missing coordinate information is only xxx, which is negligible.

Data on NRT gaps is taken from Pierce and Schott (2016) and matched to the Chinese Industry Classification (CIC) based on the product-industry concordance provided by Brandt, Van Biesebroeck, and Y. Zhang (2012) and various versions of HS 6-digit product-level concordance provided by the World Integrated Trade Solution (WITS).

In addition to the aforementioned datasets, we use various statistical yearbooks and collect information on road intensity, high-speed railways, and special economic zones to measure city characteristics.

2.2 Variable Construction

Spatial concentration

A key challenge in establishing a link between trade and agglomeration is measuring spatial concentration. As summarized in Duranton and Overman (2005), there are some criteria to evaluate the measurement of spatial concentration. The first one is that the measure should be comparable across spatial units. The existing literature mainly makes aggregation based on administrative units. However, the geographic coverage and population of each administrative region differ from each other, as they are defined based on administrative needs rather than economic comparability. These make the existing aggregations mix different spatial scales and make the size indicators, such as regional total employment and output, less comparable. For instance, the most populous city in China was Shanghai with a population of 23 million, more than 60 times as large as the least populous city (Shizihe), according to the 2010 population census. The second one is that the measure should be spatially continuous. For example, if a cluster happens to be located along the borders of two regions, measurements based on administrative boundaries, such as regional total employment or output, would split a cluster into two pieces. Another limitation of aggregation by administrative boundary is that it is hard to compare results across different scales. Because in many countries the number of administrative divisions is limited to two or four layers.

Using geo-coded firm data, we measure spatial concentration as the employment density in a certain radius around the industry-prefecture cluster center. This measurement is comparable across regions and industries. It also works in continuous space rather than being constrained by administrative units, which allows us to compare the results with different radii.

Using the ASIF data (1998-2007), we are able to capture the evolution of spatial concentration

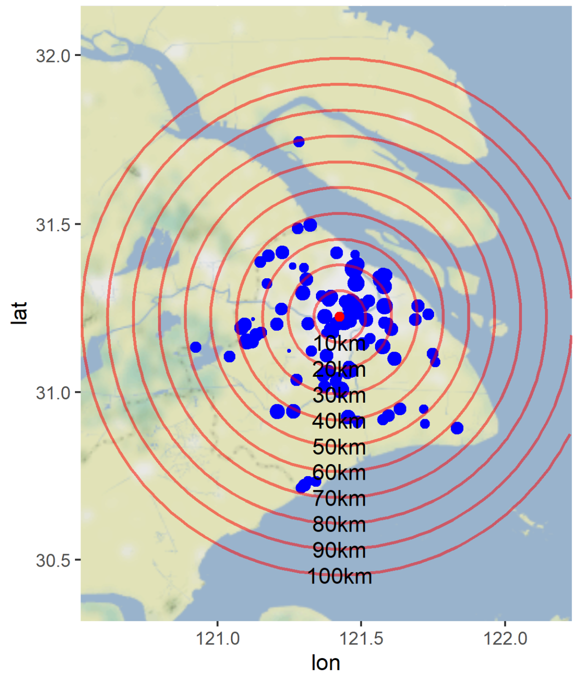


Figure 1: Concentric Rings for "Basic & Raw Chemical" industry in and around Shanghai in 1998

between regions while avoiding mismeasurement due to arbitrary administrative boundaries.

At the same time, we wish to measure concentration within clusters. Hence, we employ a concentric ring approach, where we measure employment density in 10km rings around a cluster center.

To define the cluster center, we take an employment-weighted average longitude and latitude of every firm in a given industry and prefecture in 1998. For example, figure 1 shows this approach for the "Basic & Raw Chemical" industry in Shanghai in 1998. Blue dots represent individual firms, whereas the red dot denotes the cluster center. Firm dots are scaled by their respective employment size.

Specifically, the measure for spatial concentration is defined as the employment density for each industry-prefecture-ring combination:

$$concentration^t_{jpr} = \frac{L^t_{jpr}}{Area_{pr}} = \frac{\sum_f L^t_{jprf}}{Area_{pr}} \tag{1}$$

Where j denotes a 3-digit industry, p denotes a prefecture, r is a ring, and f denotes individual firms. Years are indexed by t. The area is measured in $10,000m^2$.

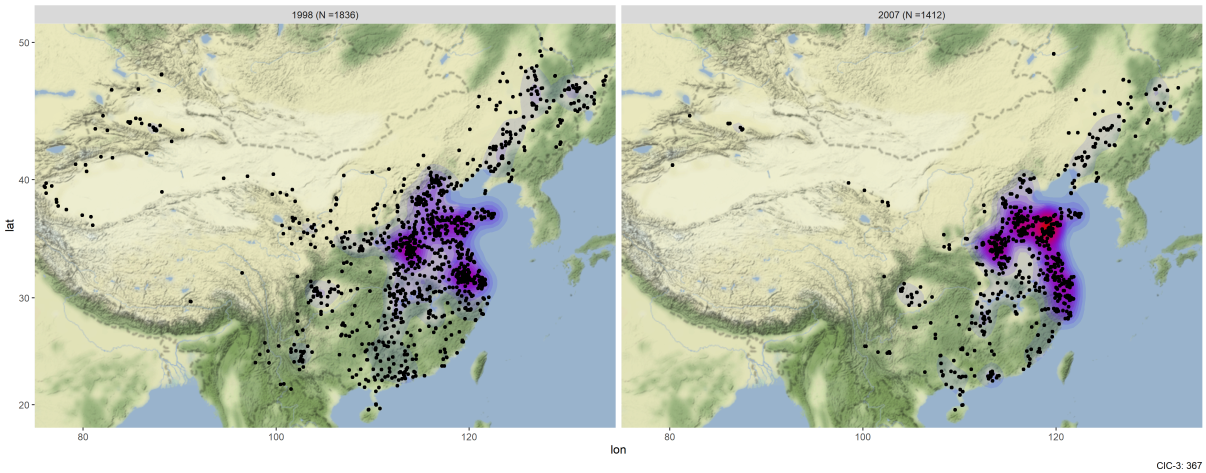


Figure 2: Density of Agricultural Machinery Manufacturing 1998 (left) vs 2007 (right)

In total, there are 161 3-digit industries and 336 prefectures. Whenever an industry is present in a prefecture at a given time we calculate the densities within each ring around the industry-prefecture center in 1998. Rings are defined up to 150km away from the cluster center and can overlap if different industry-prefecture centers are less than 150km apart, which is free from the restrictions of administrative units.

As figure 2 highlights for Agricultural Machinery Manufacturing, we wish to capture the reinforcing role of initially large clusters. To do so we define size quartiles within the industry by prefecture in 1998 to measure the size of the pre-existing cluster.

Table 2.2 provides summary statistics on our final data set which varies by industry, prefecture, ring and year. As expected, employment and firm density is larger in clusters that were initially large in 1998 (4th Quartile). Furthermore, density is decreasing in distance to the cluster center.

A notable feature in table 2.2 is the variation in exporter share amongst the quartiles. As predicted by combining firm sorting into exporting (Melitz 2003) and firm sorting into agglomeration (Gaubert 2018), exporters are disproportionately concentrated in the largest clusters. Bakker et al. (2021) find a similar pattern for the Chinese industrial census in 2004, however they do not distinguish between different industries.

Table 1: Summary of concentric ring data

	Q1:		Q2:		Q3:		Q4:	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Employment Density	.	.	.	.	.	.	.	.
0-10km	45.148	189.812	88.801	249.438	177.974	379.525	737.725	2015.619
10-20km	8.857	68.312	18.658	80.476	42.281	133.949	208.478	570.903
20-30km	6.724	39.344	11.857	40.967	26.734	120.222	110.985	378.095
30-40km	6.776	33.662	11.127	45.863	21.08	81.539	72.317	223.431
40-50km	8.308	39.942	12.474	64.438	19.38	68.061	59.735	192.563
50-150km	8.978	43.278	13.028	51.608	20.456	75.806	45.679	151.498
Firm Density	.	.	.	.	.	.	.	.
0-10km	.273	.407	.323	.538	.496	.886	1.557	3.807
10-20km	.047	.155	.089	.25	.176	.428	.727	1.678
20-30km	.034	.116	.058	.147	.114	.32	.411	1.038
30-40km	.032	.098	.048	.137	.09	.261	.283	.721
40-50km	.034	.107	.05	.167	.085	.247	.225	.562
50-150km	.037	.12	.051	.151	.083	.23	.172	.447
Exporter Share	.	.	.	.	.	.	.	.
0-10km	.133	.321	.208	.373	.26	.379	.319	.358
10-20km	.177	.348	.2	.355	.234	.358	.278	.334
20-30km	.17	.332	.185	.343	.222	.348	.271	.336
30-40km	.181	.337	.182	.33	.215	.338	.261	.33
40-50km	.184	.327	.188	.327	.212	.332	.256	.327
50-150km	.183	.314	.192	.315	.212	.315	.247	.314
Change Employment Density	.	.	.	.	.	.	.	.
0-10km	-.339	195.727	-27.802	216.668	-72.215	347.092	-232.278	1625.439
10-20km	5.291	67.781	4.238	75.304	5.805	114.401	32.505	430.404
20-30km	2.198	33.882	1.589	41.528	5.84	113.886	34.28	307.653
30-40km	2.194	29.223	1.88	39.206	5.51	61.783	20.499	169.712
40-50km	.559	36.864	1.545	51.448	3.514	53.539	16.922	129.119
50-150km	.912	44.37	1.564	38.861	2.953	61.211	11.02	113.177

Density is measured in workers / firms per 10k square meters.

Trade policy uncertainty

Our identification relies on the exogeneity of NTR gaps with respect to Chinese firms' location choices. As outlined in Pierce and Schott (2016), under certain conditions, the US president could grant non-market economies normal trade relations without consulting the Congress. These non-market economies would otherwise face non-NTR tariffs, which the US applied to countries not part of the WTO and which were substantially higher tariffs than those applied to WTO members. China was first granted NTR tariffs in 1980, which could be overturned by congressional disapproval each year. Chinese firms exporting to the US thus faced the risks that they could potentially experience large increases in US tariffs levied on their goods.

This trade policy uncertainty was eliminated when China acceded to the WTO in late 2001, hence representing a considerable reduction in trade uncertainty for Chinese firms. We argue that the gap in NTR and non-NTR tariffs, which was decided in the US not China, and the uncertain annual approval, were exogenous from the perspective of Chinese firms. We exploit differences in NTR gaps across Chinese industries and the establishment of permanent normal trade relations following China's WTO accession to identify the effect of trade liberalization on industrial agglomeration in China.

3 Export and Within-Industry Spatial Concentration

3.1 Between Cluster

Our empirical approach follows Dix-Carneiro and Kovak (2017). We take the differences in our dependent variable relative to 1998 and estimate the model for each ring in each year $t \in [1999, 2007]$ separately. Specifically we estimate

$$Y_{prj}^{t} - Y_{prj}^{1998} = \sum_{q=1}^{4} \beta_{q,t} NTR_j \times Q_{pjq} + X_{prj}^{t} + \mu_P + \gamma_s + \epsilon_{prj}^{t}, \tag{2}$$

where j denotes a 3-digit industry, p denotes a prefecture, r a ring. Q_{pjq} is a quartile dummy for the initial size quartile q, which does not vary over time as the initial size quartile is defined in 1998. X_{prj}^{t} is a vector of controls that includes the difference between t and 1998 in Chinese input and output tariffs. X_{prj}^{t} also contains the initial shares of foreign- and private-owned capital. Foreign-owned capital shares tend to be a good proxy for local policies to boost growth. It further includes a proxy for the share of firms that are in a special economic zone, based on data from Martin and K. Zhang (2021). We also include fixed effects μ_P for province P and γ_s for 2-digit sector s. Standard errors are clustered at the 3-digit industry level to allow for within-industry correlation in error.

The dependent variable is employment density as defined in equation 1, and constructed as changes relative to 1998. The unit is workers per 10,000 square meters. Furthermore, we use the NTR-Gap from 1998, which does not vary over time. Hence, $\beta_{q,t}$ is the cumulative effect of liberalization on outcomes by year t. The NTR gap measures the gap between WTO tariffs and the potential tariffs Chinese firms would have faced if the Congress had overturned NTR tariffs. Hence, the larger the gap, the more substantial the reduction in trade uncertainty that an industry experienced after 2001.

Table 3.1 provides summary statistics on the NTR-Gap as well as the initial size quartiles. The initial size is measured by the employment in an industry-prefecture pair in 1998. Quartiles are defined by the initial size distribution across prefectures in the same industry. Hence the initial size range of the quartiles overlap.

Table 2: Summary of NTR-Gap and initial cluster size

	Mean	SD	Min	Max
NTR Gap	30.52	14.04	0	71.68
Initial Size				
Q1	164.05	201.91	8	2272
Q2	620.3	611.44	50	6104
Q3	1612.14	1540.46	58	14128
Q4	7060.33	10312.2	91	165953

Note: Initial cluster size is measured by the number of workers and varies by industry-prefecture pairs.

We begin our empirical analysis by investigating the effect of trade uncertainty reduction on the within-industry clustering of firms across cities following equation 2. The within-industry clustering is measured as the employment density of a cluster. Without loss of generality, we define a cluster as a 100km circle around the industry-city center as the average radius of a Chinese city is 100km. In the next section, we will show the results with different radii when examing the between-city effect. Figure 3 shows the evolution of the effect of trade liberalization by quartile on employment

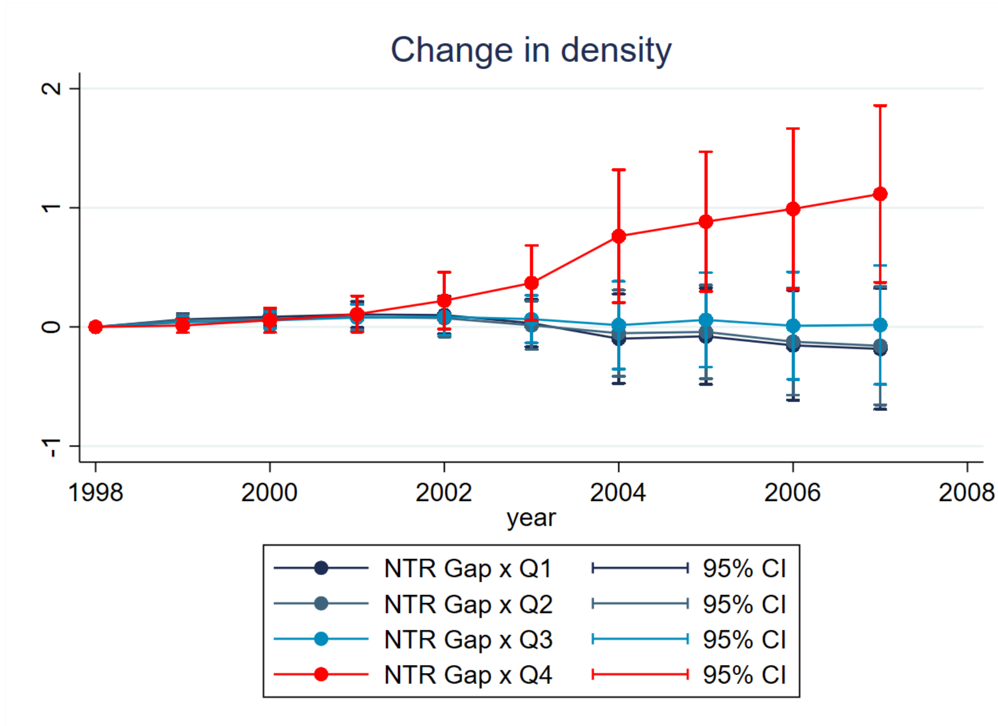


Figure 3: Change in employment density in a 100km around the cluster center by initial size quartile

density. There is no significant change in density associated with NTR Gaps for any quartile before 2001 when China acceded to the WTO. This is in line with expectations, as the reduction in trade uncertainty by the establishment of permanent normal trade relations was tied to the WTO membership of China. This suggests that industries that faced a higher NTR gap were not on a differential trend prior to 2001 compared to industries with a lower gap, irregardless of their initial size quartile.

After 2001, figure 3 shows a sharp increase in employment density associated with a larger reduction in trade uncertainty, however, this only holds for cluster in the fourth initial size quartile. That is, only clusters that were already large in 1998 expanded following trade liberalization. This is in line with expectations that arise by combining Melitz (2003) and Gaubert (2018), as in Bakker et al. (2021). Because of sorting and agglomeration externalities, relatively more productive and exporting firms are disproportionately located in larger clusters. Since these firms benefit the most from better access to foreign markets, it is the largest initial clusters that reinforce themselves, resulting in an increase in the spatial concentration of economic activity and a reinforcing of regional comparative advantages.

3.2 Between and Within Cluster

We now add back the concentric rings to equation 2 to further investigate the effect on increased access to foreign markets within and across clusters.

Just as before, figure 4 displays an increase in employment density for the initially largest clusters only after 2001. Plotting the cummulative effects by 2007 across concentric rings in figure 5 further reveals a reduction in the effect size further away from the initial cluster center[4]. We find the strongest effects at the 10-20km ring which then slowly fade and cease to be significantly different from the other quartiles after 100km.

These effects are large in magnitude. For example, by 2007 a one standard deviation larger NTR Gap lead to a density increase of 21.2% at 10-20km for an average Q4 cluster.

3.3 Decomposition

The overall effects highlighted in figure 5 offer no insights into the margins driving these increases in density. A concern might be that these patterns could be explained by a Melitz (2003) framework alone, without any agglomeration. Exporting firms will be larger even before trade liberalization, potentially contributing to being in the 4th Quartile. It is also these firms that grow the most after liberalization. In this case the results in figure 5 could be explained by within firm changes in employment alone.

When we decompose the effect into the margins of growth, we find that this is not the case. Specifically, we decompose the change in employment density by 2007 into 5 terms:

1. Incumbents firms: Growth of firms that existed before 2001 and at least until 2007
2. Exits: Decline in employment density due to firms that existed before 2001 exiting the sample.
3. Entry: Employment density gains due to new firms entering
4. Switching into the industry: Incumbent firms that switch into the liberalizing industry by 2007

[4] Note that we exclude the first ring from 0 to 10km. Because initial cluster centers are based on weighted average longitude and latitude of all firms in an industry-prefecture combination and because the first ring covers the smallest area, density varies drastically. Indeed, it sometimes includes no firm at all. Because of the large standard deviation, the estimates are extremely imprecise and distort the scaling of our figures.

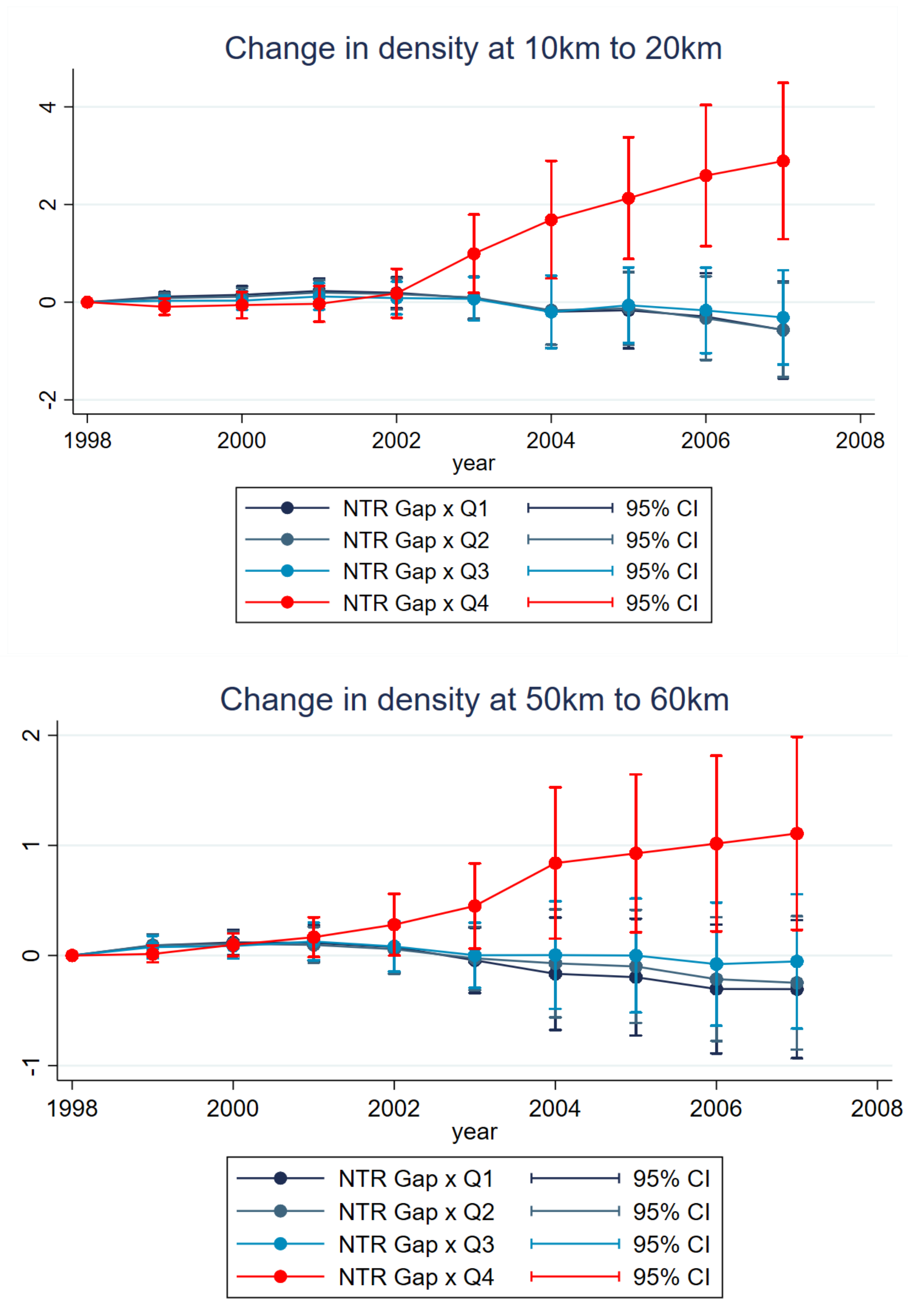


Figure 4: Change in employment density at different distances

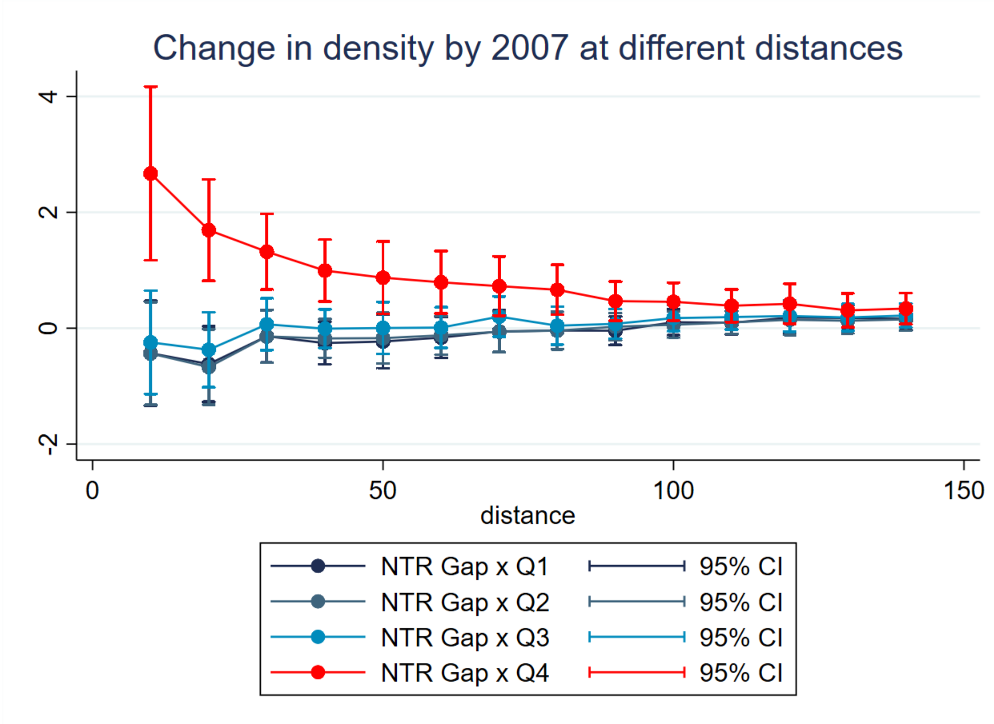


Figure 5: Cumulative change in employment density by 2007

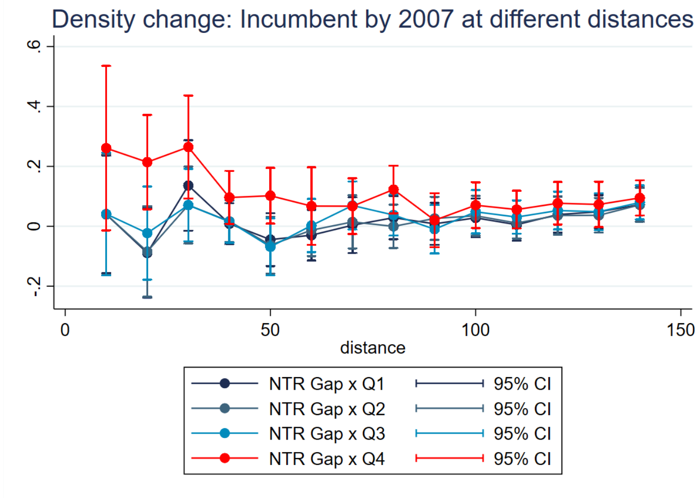
5. Switching out of the industry: Incumbent firms that leave the liberalizing industry by 2007

This decomposition is exact except for firms relocating between the rings. As this rarely happens and does not contribute to the observed effects we omit this term.

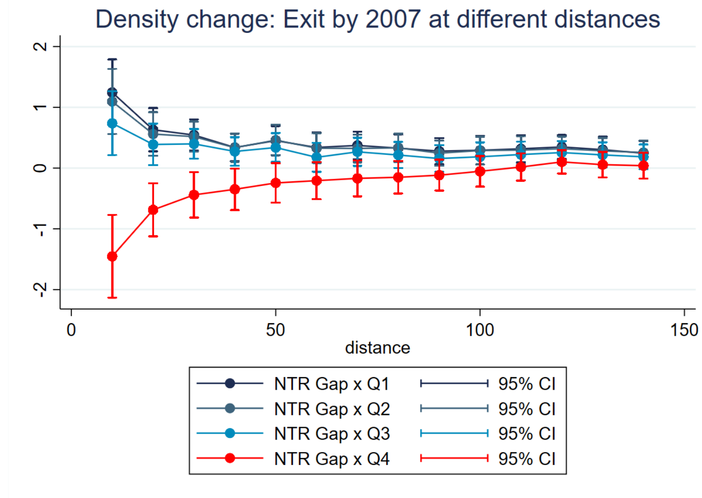
Figure 6 reveals that the positive effect of trade liberalization on employment density in the fourth quartile is predominately driven by firms entering. The positive effect on concentration that we observe appears to be driven mostly by the extensive margin of agglomeration, where new firms enter in booming clusters.

Surprisingly, the initial stock of firms does not benefit. They either exit, reducing employment density, or if we only focus on the survivors to not appear to grow. This is at odds with predictions from a standard Melitz type model, where productive incumbent firms should expand following an increased access to foreign markets.

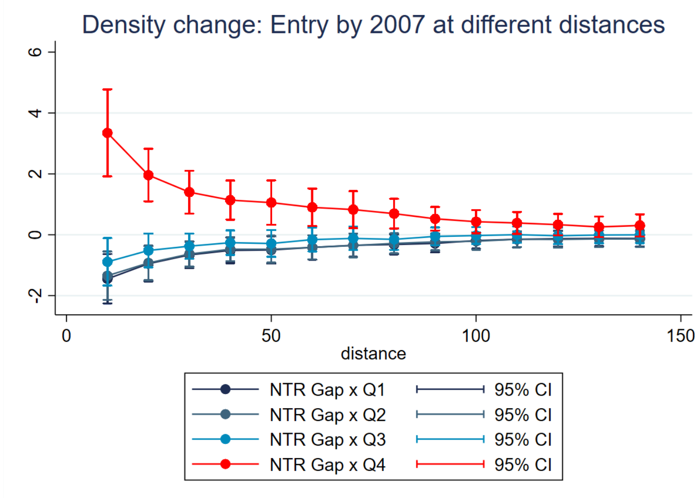
Additionally, incumbent local firms tend to switch their industry to join or leave the liberalizing industry. On net this also contributes to the positive effect on employment density, as more firms switch into the liberalizing industry than out of it. All in all, the results indicate that initially large clusters of an industry experiencing stronger trade liberalization tend to become more dynamic.



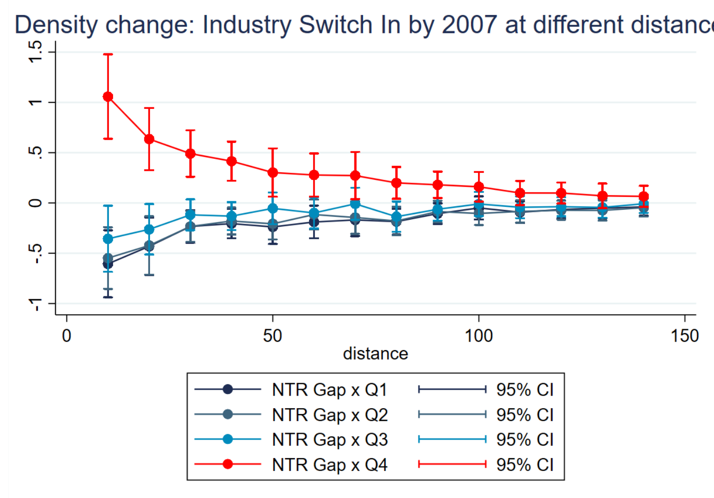
(a) Incumbents



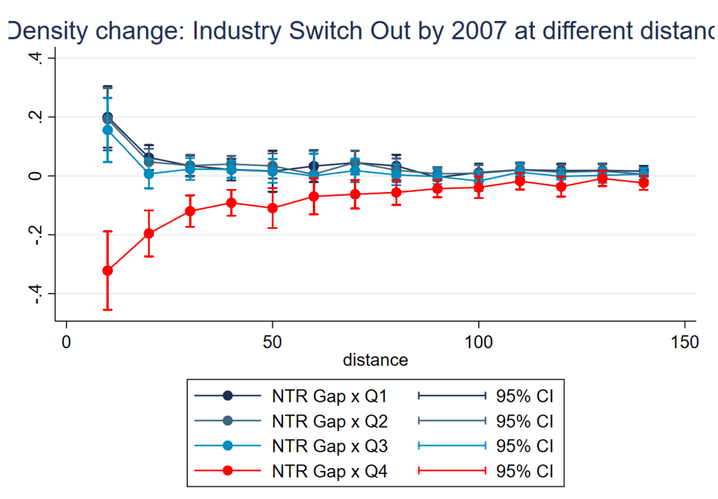
(b) Exits



(c) Entry

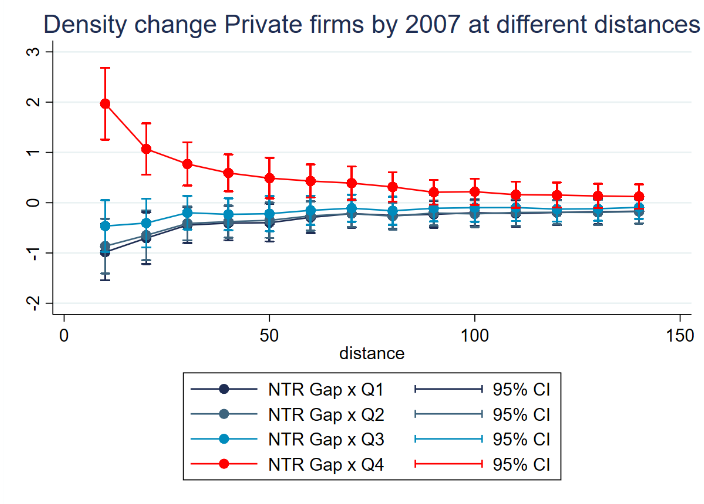


(d) Industry in Switching

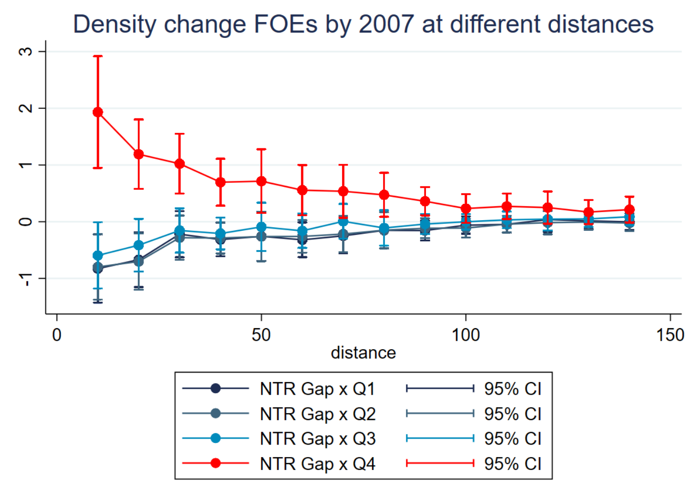


(e) Industry out switching

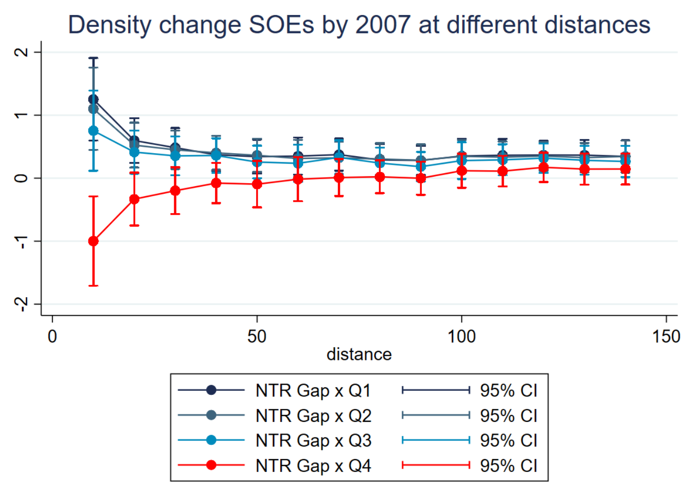
Figure 6: Decomposed cumulative change in employment density by 2007 by the margin of growth



(a) Private



(b) Foreign



(c) State and Collectivley owned

Figure 7: Decomposed cumulative change in employment density by 2007, by firm ownership

The adverse effects for incumbent local firms might also reflect the extensive privatization of the Chinese economy at the time. The removal of barriers to entry could have allowed private firms to enter in booming clusters and hence increasing the local competition for incumbents. Indeed, as figure 7 shows, the positive employment density effects we document so to exclusively come from private and foreign-owned firms, whereas state and collectively owned enterprises contract. Contrastingly, in initially smaller clusters where there are no substantial entry effects, SOEs and collectively owned enterprises benefit from increased access to foreign markets.

3.4 Robustness

Though NTR-Gaps are arguably exogenous from the perspective of Chinese firms, there are still some potential confounders that remain.

Though equation 2 already accounts for the presence of FDI and special economic zones, other local factors that influence density and correlate with NTR gaps may remain. For instance, local governments may invest in infrastructure to reduce internal trade cost as a response to local growth due to trade liberalization. Assuming that this further boosts local growth, for example by counteracting dispersion forces, we may be overstating the reinforcing local effects due to better access to foreign markets.

Another concern may be related to local spillover effects as highlighted in figure **??**. If local spillovers exist and are positive we may conflate them with the direct effect of a reduction in trade uncertainty and therefore overstate the direct effect.

To address these concerns we extend the model in equation 2 to further capture the change in road length of each prefecture relative to 1998, the establishment of high-speed railway connections, and the employment weighted average prefecture NTR-Gap. Figure 8 displays cumulative employment (top) and firm (bottom) density effects while including these additional controls. The results remain virtually unchanged, indicating that our results are robust to these concerns.

Our empirical approach relies on separately estimating equation 2 for each year to obtain cumulative effects by each year. A disadvantage of this approach is that is does not allow us to include a more flexible set of fixed effects that can account for trends over time. Since density around the fourth quartile cluster centers is larger than for lower quartiles by definition, any subsequent changes in absolute density are likely also to be larger than for lower quartiles. This is a potential concern, especially if larger prefectures are on a different growth path than smaller prefectures and larger prefectures are likely to contain more Q4 clusters.

We address these concerns by modifying our empirical approach to a regression that pools all years. Specifically, we estimate

$$Y_{prjt} - Y_{prj}^{1998} = \sum_{q=1}^{4} \beta_q NTR_j \times Q_{pjq} \times Post_t + X_{prjt} + \mu_{pt} + \gamma_j + \epsilon_{prjt} \tag{3}$$

Where again j denotes a 3-digit industry, p denotes a prefecture, r a ring. Q_{pjq} is a quartile dummy

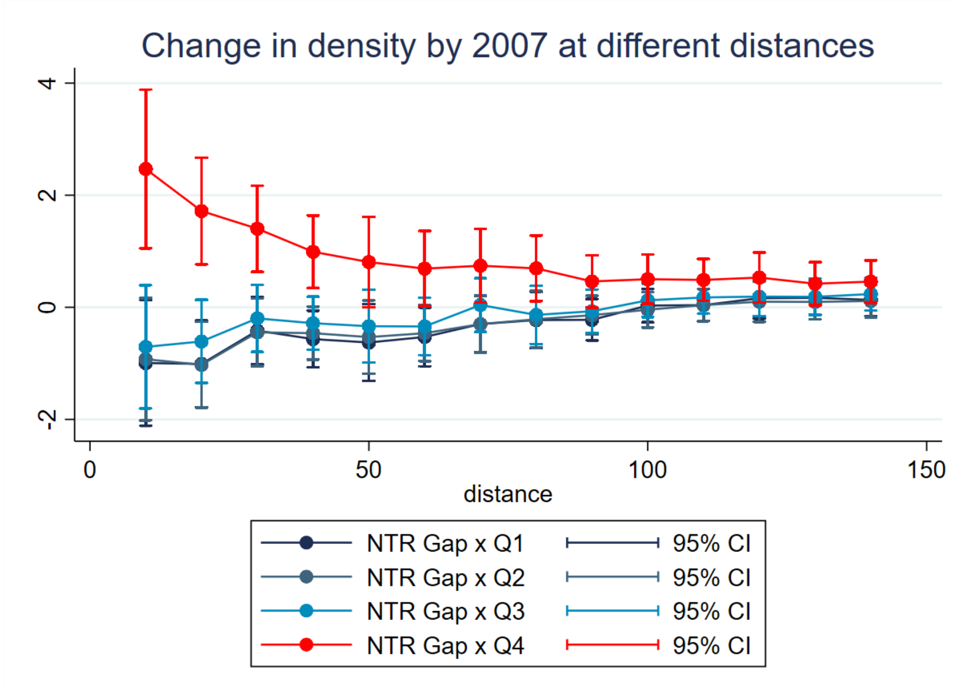


Figure 8: Cumulative density effects by 2007 with additional controls

for the initial size quartile q. $Post_t$ is a binary variable taking one after China's accession to the WTO in 2001. X_{prjt} is a vector of controls that includes the difference between t and 1998 in Chinese input and output tariffs, private and foreign ownership share in 1998, and SEZ shares in t. The time variation now allows us to include a more flexible set of fixed effects. Hence, μ_{pt} controls for prefecture time trends and γ_j controls for time invariant differences between 3-digit industries. Additional controls such as the change in a prefectures road length, high-speed railway connections, and weighted average prefecture NTR gaps are absorbed by μ_{pt}.

The dependent variable remains the total change in employment density at different rings r for each cluster. Yet larger total changes in Q4 may appear larger than for lower quartiles simply because Q4 clusters have a higher density to begin with. Therefore we also take changes in inverse-hyperbolic sine transformed employment density to capture relative effects. Note however, that relative effects reveal very little about changes in the spatial concentration of economic activity, with which we are concerned. For example, a 10% increase in the density of a Q1 ring is small relative to a 1% increase in the density of a Q4 ring. Even if in relative terms, the density of the smallest clusters increase more than the density of the largest clusters, the overall spatial concentration of economic activity is still likely to increase.

Table 3.4 displays results for the pooled regression as specified in equation 3. The results still display the distinct pattern of reinforcing initially large clusters. Even in relative terms and after controlling for prefecture time trends, Q4 clusters' density grow substantially more than smaller clusters as a response to increased access to foreign markets. Note that these effects are averages of the effects across rings and years post 2001 and therefore not directly comparable to the effects based on equation 2.

Table 3: Pooled Regression

	(1)	(2)
	Change in density	Growth in density (ihs)
NTRgap × Q1 × WTO	0.151	0.0777^{**}
	(0.106)	(0.0374)
NTRgap × Q2 × WTO	0.170^{*}	0.0807^{**}
	(0.103)	(0.0376)
NTRgap × Q3 × WTO	0.250^{**}	0.104^{**}
	(0.121)	(0.0427)
NTRgap × Q4 × WTO	0.715^{***}	0.178^{***}
	(0.242)	(0.0578)
City x Year FE	Yes	Yes
Industry FE	Yes	Yes
N	2459786	2459786

Column 1 shows total changes in density relative to 1998. Density is measured in workers per 10k square meters. Column 2 shows changes in inverse-hyperbolic sine transformed density (×100). Standard errors (in parentheses) are clustered at the 3-digit industry level. For consistency with the previous figures the 0-10km ring is excluded.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4 Export and Co-agglomeration

The focus so far has been on the direct effect of trade liberalization on within industry spatial concentration. However, industries do not operate in isolation, but rather are interdependent. Ellison, Glaeser, and Kerr (2010) already showed that industries that are economically close to one another tend to co-agglomerate. These co-agglomeration patterns are strongest for firms linked by input-output relations (Ellison, Glaeser, and Kerr 2010). Helm (2020) too finds that trade shocks to one industry spill over locally to other industries, but only for industries connected by labor linkages.

To address whether increased foreign market access also reinforces co-agglomeration patterns, we modify equation 2 to measure the density change in related industries rather than the liberalizing industry.

$$\sum_{k\neq j}\omega_{jk}(Y_{pk}^{rt}-Y_{pk}^{r1998})=\sum_{q=1}^{4}\beta_q^{rt}NTR_j\times Q_{pjq}+X_{pj}^{rt}+\mu_P+\gamma_s+\epsilon_{pj}^{rt} \tag{4}$$

Where ω captures the related industries through three linkages: input, output and labor pooling. Input linkages are captured by input shares taken from Chinese IO tables from 2002, where ω measures the cost expenditure share of the downstream industry j spent on the upstream industry k. Similarly, output linkages are measured by the output shares, where ω measures the share of the upstream industry j's output sold to the downstream industry k. To measure the extent to which industries use a similar labor pool we calculate bilateral occupation correlations between industries. To do so we use the Chinese Population Census from 2000, which reports industry and occupation of employment on a detailed level. We then calculate the correlation in the occupation share of each industry pair to measure labor pooling. In the few cases that the correlation is negative, we replace it with zero to avoid negative employment.

The measure for the cummilative change in co-agglomeration $\sum_{k\neq j}\omega_{jk}(Y_{pk}^{rt}-Y_{pk}^{r1998})$ is the weighted total employment density in a given ring, where the weight ω reflects different linkages between the liberalizing industry j and all other local industries $k\neq j$.

The specification otherwise remains the same as in equation 2, so that we estimate the effect of industry j's reduction in trade uncertainty on local density changes of other industries related to industry j. The results are displayed in figure 9. We find positive spillover effects on local firms upstream and downstream of industry j facing trade liberalization, but only for clusters of j that were initially large. This implies that trade liberalization also induces co-agglomeration, but only for the clusters that were large enough to take advantage of the increased access to foreign markets. This is also in line with Ellison, Glaeser, and Kerr (2010), who find that input-output linkages matter the most for the determination of co-agglomeration patterns.

In contrast, we find negative spillovers for industries using a similar labour force. This implies a competition effect between industries using a similar labour force. For labour to be scarce locally we need imperfect labor mobility between regions, which might be the case in China due to the Hukou policy which introduces mobility frictions. This is consistent with the findings of Curuk and Vannoorenberghe (2017) for the U.S., whose results imply that specific skill complementaries and knowledge transfusion through worker flows may not outweigh the effects of competition for the same pool of labor.

5 Theoretical Framework

In this section, we lay out a spatial Melitz model to understand the patterns we document in this paper. The model can be viewed as a direct extension of Melitz (2003) to an environment with multiple regions within a country. Regions can be interpreted as cities, but can also be interpreted as "circles" of a city in the context of our empirics. To ease the presentation we will always refer to "cities" for these heterogeneous exporting regions. The model can also be viewed as a simplified version of Bakker et al. (2021), where we impose a particular spatial structure to both capture the salient feature of China's regional development experience and obtain clean analytical results that can be contrasted with data patterns.

5.1 Environment

We specifically consider an "Rural China–Chinese cities–U.S." spatial structure: one rural area of China, N exogenously fixed Chinese cities, and one foreign country (US). Rural China supplies labor to firms in cities, which then produce to fulfill the demand of US consumers. We focus on the problem of firms located in Chinese cities, that aim to export to the US.

We normalize the total mass of workers to 1. Workers are initially exogenously distributed in the rural area and among cities. Specifically, the share of workers located in cities is η, and in the rural area is $1 - \eta$. Cities potentially differ in the initial size of the labor force. Denote city j endows ηe_j units of labor. Firms in city j will produce using the labor endowment ηe_j and labor migrated from the rural area. We abstract from migration between cities. This structure well mimics the development experience of China after joining WTO, where the growth engine of cities is the workers migrating from the rural area.

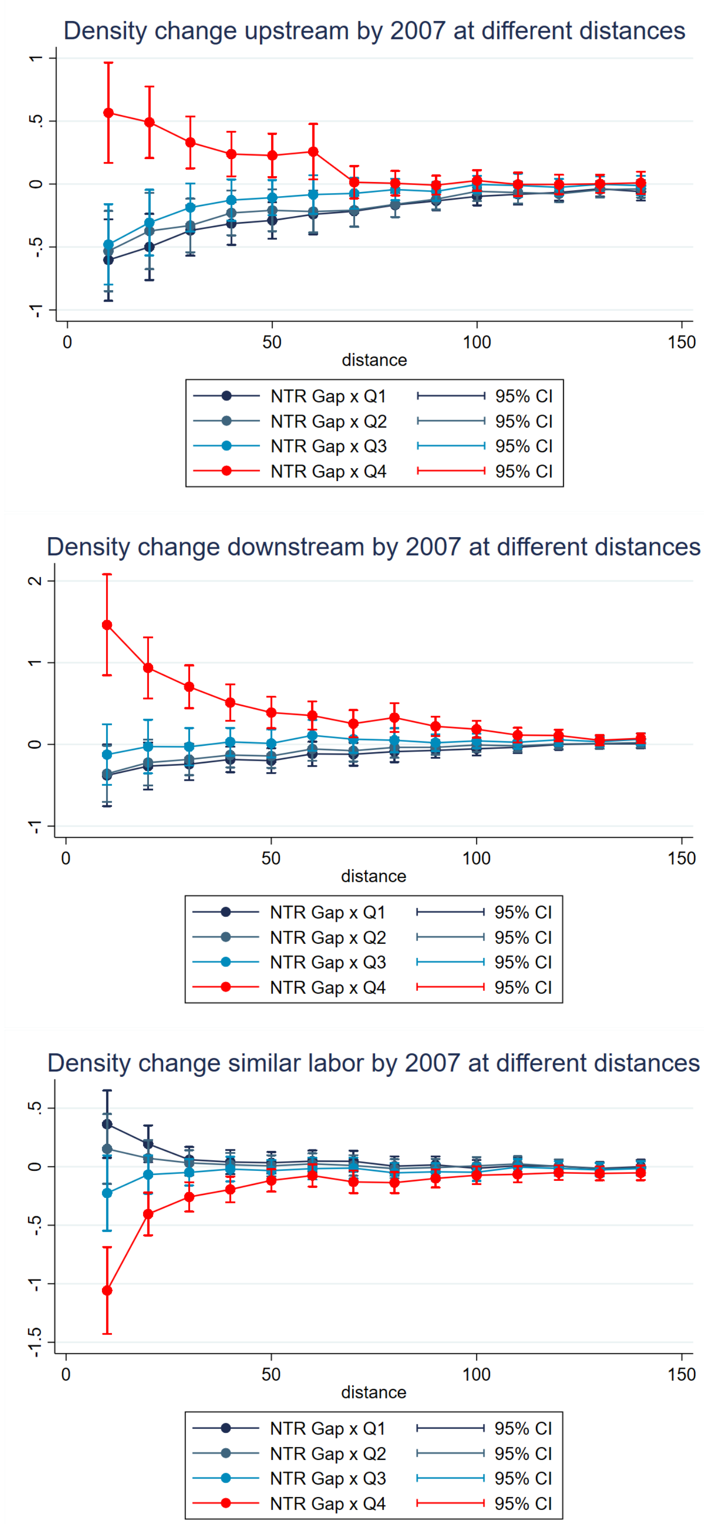


Figure 9: Cumulative co-agglomeration effects by 2007

Rural workers receive an exogenous value b if staying in the rural area. They can also choose to move to one of the N cities. If the worker moves to city j, he will get wage w_j instead of b. We allow for preference shocks for each city of choice to smooth the migration. So workers in the rural area face the following problem:

$$\max\{b+\epsilon_0, w_1+\epsilon_1, \ldots, w_j+\epsilon_j, \ldots, w_N+\epsilon_N\}$$

where ϵ_j is the Gumbel distributed preference shock with location parameter 0 and scale parameter α. The probability of moving to city j is thus $\frac{1}{e^{\alpha(b-w_j)}+\sum_{i=1}^{N} e^{\alpha(w_i-w_j)}}$. Therefore, combining local workers and migrating workers, the total labor supply to city j is

$$L_j^d(\mathbf{w}) = \eta e_j + \frac{1-\eta}{e^{\alpha(b-w_j)}+\sum_{l=k}^{N} e^{\alpha(w_k-w_j)}} \tag{5}$$

where $\mathbf{w}$ collects the wages of all cities: $\mathbf{w}=\{w_1,\ldots,w_N\}$.

Foreign demand. Foreigners' utility function is standard CES with substitution of elasticity σ. Therefore the foreigners' problem is

$$\begin{aligned} \max_{x_i} U &= \left(\int_{i\in\Omega} x_i^{\frac{\sigma-1}{\sigma}} \mathrm{d}i\right)^{\frac{\sigma}{\sigma-1}} \\ \text{s.t. } & \int_{i\in\Omega} p_i x_i \mathrm{d}i = \mathbb{W} \end{aligned}$$

where Ω is the universe of varieties, p_i and x_i are prices and consumption of variety i, and $\mathbb{W}$ is the total expenditure of US consumers on Chinese exports. For the purpose of this paper, we focus on Chinese exporters and treat $\mathbb{W}$ as exogenous. The foreign demand faced by individual firm/variety i is thus

$$\begin{aligned} x_i &= \left(\frac{p_i}{\mathbb{P}}\right)^{-\sigma} \frac{\mathbb{W}}{\mathbb{P}} \\ \text{s.t. } \mathbb{P} &= \left(\int_{i\in\Omega} p_i^{1-\sigma} \mathrm{d}i\right)^{\frac{1}{1-\sigma}} \end{aligned}$$

where $\mathbb{P}$ is the standard aggregate price index. This formulation states that firms in different cities face the same demand curve. How much a firm can sell ultimately depends on its ability to charge low prices.

Firm problem. Firms use labor only and (for now) produce for foreigners only. Firms engage in monopolistic competition. The labor cost of producing x units of output for an operating firm is given by

$$l(\phi) = f + \frac{x}{\phi} \tag{6}$$

where f is the fixed cost of production in terms of labor, and ϕ is the firm-specific productivity. For a given wage level w_j, the pricing problem of a firm with productivity ϕ in city j yields

$$p_j(\phi) = \frac{w_j}{\rho(1-\tau)\phi} \tag{7}$$

where $\rho = \frac{\sigma-1}{\sigma}$. The output, revenue, and firm profit corresponding to $p_j(\phi)$ is thus

$$\begin{aligned}
x_j(\phi) &= \left(\frac{w_j}{\rho(1-\tau)\phi}\right)^{-\sigma} \frac{\mathbb{W}}{\mathbb{P}^{1-\sigma}} \\
r_j(\phi) &= p_j(\phi)x_j(\phi) = \mathbb{W}\left(\frac{\mathbb{P}\rho(1-\tau)\phi}{w_j}\right)^{\sigma-1} \\
\pi_j(\phi) &= \frac{r_j(\phi)}{\sigma} - f_j w_j
\end{aligned}$$

Firm entry and aggregation. In each city, there is a large (unbounded) pool of prospective entrants into the exporting business. To enter, firms have to make an initial investment modeled as a fixed entry cost F. Firms then draw their initial productivity from a city-specific distribution with CDF $G_j(\phi)$ and PDF $g_j(\phi)$. Conditional on the productivity drawn, firms decided whether to stay and produce/export or to exit and not produce. If the firm does produce, it then faces a constant probability δ in every period of a bad shock that would force it to exit.

Given that ϕ once drawn is fixed for a firm, there will be a cutoff productivity ϕ_j^* for each city j, such that entrants will find unprofitable below this cutoff once entering. So the observed firm distribution of city j, $\mu_j(\phi)$, will be a truncated distribution on G_j:

$$\mu_j(\phi) = \begin{cases} \frac{g_j(\phi)}{1-G_j(\phi_j^*)}, & \text{if } \phi \geq \phi_j^* \\ 0, & \text{o.w.} \end{cases}$$

where $1 - G_j(\phi_j^*)$ is the probability of successful entry in city j. As is standard in the Melitz model, denote $\tilde{\phi}_j$ be the observed average productivity defined as follows

$$\tilde{\phi}_j(\phi_j^*) = \left[\frac{1}{1 - G_j(\phi_j^*)} \int_{\phi_j^*}^{\infty} \phi^{\sigma-1} g_j(\phi) \mathrm{d}\phi\right]^{\frac{1}{\sigma-1}} \tag{8}$$

Revenue and profit of a firm with productivity $\tilde{\phi}_j$ would be as follows

$$\bar{r}_j := r_j(\tilde{\phi}_j) = \left(\frac{\tilde{\phi}_j(\phi_j^*)}{\phi_j^*}\right)^{\sigma-1} r_j(\phi_j^*)$$

$$\bar{\pi}_j := \pi_j(\tilde{\phi}_j) = \frac{r_j(\tilde{\phi}_j)}{\sigma} - fw_j = \left(\frac{\tilde{\phi}_j(\phi_j^*)}{\phi_j^*}\right)^{\sigma-1} \frac{r_j(\phi_j^*)}{\sigma} - fw_j$$

Using the definition of cutoff productivity of each city $\pi_j(\phi_j^*) = 0$ yields

$$\bar{\pi}_j = \left[\left(\frac{\tilde{\phi}_j(\phi_j^*)}{\phi_j^*}\right)^{\sigma-1} - 1\right] fw_j \tag{9}$$

Also free entry in each city j yields

$$\left[1 - G_j(\phi_j^*)\right] \frac{\bar{\pi}_j}{\delta} - F = 0 \tag{10}$$

Combining equation (9) and (10) we obtain the following equation

$$\underbrace{\frac{\delta F}{1 - G_j(\phi_j^*)}}_{\text{city } j \text{ average profit implied by free entry}} = \underbrace{\left[\left(\frac{\tilde{\phi}_j(\phi_j^*)}{\phi_j^*}\right)^{\sigma-1} - 1\right] fw_j}_{\text{city } j \text{ average profit implied by zero profit at } \phi_j^*} \tag{11}$$

This equation indicates that in each city j, productivity cutoff ϕ_j^* should equalize the average profit implied by the free entry condition and the average profit implied by cutoff productivity leading to exactly zero profits. Given wage w_j, this condition yields for each city the cutoff productivity ϕ_j^*, which summarizes all the firm-level information for city j. Given ϕ_j^*, other firm-level objects, such as $\bar{r}_j, \bar{\pi}_j, \tilde{\phi}_j$, and firm distribution μ_j, can all be derived accordingly.

5.2 Equilibrium

What is left to be determined in the equilibrium are wage w_j and mass of firms M_j for each city. First note that the aggregate price index is

$$\begin{aligned}\mathbb{P} &= \left[\sum_{j=1}^{N} \int_0^\infty p_j(\phi)^{1-\sigma} M_j \mu_j(\phi) \mathrm{d}\phi\right]^{\frac{1}{1-\sigma}} \\ &= \left[\sum_{j=1}^{N} M_j P_j^{1-\sigma}\right]^{\frac{1}{1-\sigma}}\end{aligned}$$

where $P_j = \left[\int_0^\infty p_j(\phi)^{1-\sigma} \mu_j(\phi) \mathrm{d}\phi\right]^{\frac{1}{1-\sigma}} = p_j(\tilde{\phi}_j)$ is the city-specific price index, which is a function of the wage level w_j.

Goods Market Clearing. For each city j, total production must equal total demand. Total demand is city specific because the pricing of the firm depends on both the firm-level productivity ϕ and the city-level wage rate w_j. We can express the demand faced by the firm with productivity draw $\tilde{\phi}_j$ as follows

$$x_j(\tilde{\phi}_j) = \left(\frac{p_j(\tilde{\phi}_j)}{\mathbf{P}}\right)^{-\sigma} \frac{\mathbb{W}}{\mathbf{P}}$$

A bit of algebra yields the following equation on firm mass M_j

$$M_j = \frac{\mathbb{W}}{\bar{r}_j \sum_{i=1}^{N} \frac{M_i}{M_j} \left(\frac{P_i}{P_j}\right)^{1-\sigma}} \tag{12}$$

where $\bar{r}_j$ is the revenue of a firm in city j with productivity draw $\tilde{\phi}_j$. Given w_j, this is a system with N equations and N unknowns $\{M_j\}_{j=1}^{N}$. If symmetry is imposed such that $G_j = G$ for all city

j, then the system is reduced to

$$MN = \frac{\mathbb{W}}{\bar{r}}$$

which becomes an accounting identity that says, under symmetry, the number of firms in the economy equals the total expenditure divided by the firms' average revenue.

Labor Market Clearing. Finally, the wage is determined by equating the labor supply with the labor demand in each city or location. Labor demand in each city consists of two parts: labor used for entering and labor used for production. Labor used for entering is

$$L_j^e(\mathbf{w}) = F \frac{\delta M_j}{1 - G_j(\phi_j^*)} \tag{13}$$

where in the last term we impose the steady-state condition for the law of motion of firm mass. Labor used for production is

$$L_j^p(\mathbf{w}) = M_j \left(f_j + \frac{x_j(\tilde{\phi}_j)}{\tilde{\phi}_j} \right) \tag{14}$$

Then the labor market clearing condition states

$$L_j^e(\mathbf{w}) + L_j^p(\mathbf{w}) = L_j^s(\mathbf{w}) \tag{15}$$

where $L_j^s(\mathbf{w})$ is the city j labor supply from equation (5). This is also a system with N equations and N unknowns. We can thus express the equilibrium a bit more formally as follows.

Spatial-Melitz equilibrium. A spatial-Melitz equilibrium consists of, for each city $j \in \{1, 2, \ldots, N\}$, the wage w_j, mass of firms M_j, and production/export cutoff ϕ_j^*, such that for a given set of city-specific productivity distribution $\{G_j(\phi)\}_{j=1}^N$, the following conditions hold

1. Productivity cutoff ϕ_j^* satisfies the free entry condition for each city j according to equation (11).

2. Mass of firms M_j clears the goods market for each city j according to equation (12).

3. Wage w_j clears the labor market for each city j according to equation (15).

5.3 Analysis

Following Helpman et al. (2004) and Bakker et al. (2021), we postulate that productivity draws are Pareto distributed. Specifically, we make the following assumption

Assumption 1 *City-specific productivity distribution has the CDF $G_j(\phi)$ and corresponding PDF $g_j(\phi)$ as follows*

$$G_j(\phi) = 1 - \left(\frac{\ell_j}{\phi}\right)^k$$
$$g_j(\phi) = k\ell_j^k \phi^{-k-1}$$

where $\ell_j > 0$ is the city-specific productivity lower bound, and k is the common shape parameter.

Essentially, we postulate that cities are different in terms of the upper Pareto tail. Cities with higher ℓ_j feature a thicker upper tail. Given the assumption above, we first show how the city-specific productivity cutoff ϕ_j^* (and corresponding average observed productivity $\tilde{\phi}_j$) depends on the wage level w_j.

Proposition 1 *Given Assumption 1, at a given wage level w_j for city j, the city-specific productivity cutoff ϕ_j^*, and corresponding average observed productivity $\tilde{\phi}_j$, are given as follows*

$$\phi_j^* = \max\left\{\ell_j \left[\frac{\sigma - 1}{k + 1 - \sigma}\frac{fw_j}{\delta F}\right]^{\frac{1}{k}}, \ell_j\right\} \tag{16}$$
$$\tilde{\phi}_j = \left[\frac{k}{k+1-\sigma}\right]^{\frac{1}{\sigma-1}} \phi_j^* \tag{17}$$

Proposition 1 describes the dependence of cutoff productivity on both wages and city-specific productivity distributions. First, the cutoff productivity is (weakly) increasing in the wage level. This is because higher wages make entry less profitable, thus requiring higher productivity to make entry break even. Second, the cutoff productivity is (strictly) increasing in ℓ_j. Since the entry cost is assumed to be the same across cities, the free entry thus implies more competition in cities with a thinker upper tail. Finally, as in the standard Melitz model, the average observed productivity $\tilde{\phi}_j$ is proportional to the cutoff productivity, which is specific to the Pareto distributed productivity assumption.

The second proposition we establish is that cities with thicker productivity upper tail will have higher wages if initially, all workers are in the rural area, i.e., $\eta = 0$. $\eta = 0$ is required as we do not allow migration decisions of city workers. Nevertheless, this does not affect our main proposition on the effect of trade cost reduction.

Proposition 2 *Given Assumption 1 and $\eta = 0$, city-specific wages w_j will be weakly increasing in ℓ_j.*

Now we introduce a second assumption linking ℓ_j to initial city size e_j. The idea is that, potentially because of sorting and learning effects, cities with a larger population will have a higher lower bound of productivity.

Assumption 2 *City-specific productivity draws lower bound ℓ_j increases with initial city size e_j.*

Given this additional assumption, the model yields a clear prediction that a decrease in common trade cost (or tariff) τ will lead to divergence between cities in terms of employment.

Proposition 3 *Given Assumption 1 and 2. When $N = 2$ and $e_1 > e_2$, then $L_1 > L_2$. Additionally, when τ decreases, $\frac{L_1}{L_2}$ increases.*

6 Conclusion

The widening unequal spatial concentration of economic activities has become a central topic for policy makers and academics. In this paper, we empirically investigate an implication that arises by combining new developments on agglomeration theory (Gaubert 2018) which a Melitz (2003) type framework as in Bakker et al. (2021) and Bakker (2021). Increasing access to foreign markets increases the demand for manufactured goods, which leads to spatial concentration.

Using geo-coded data on Chinese manufacturing firms between 1998 and 2007 and a large reduction in trade uncertainty for Chinese exporters in 2001, we show that initially large agglomeration clusters reinforce themselves as predicted by theory. We measure agglomeration along two dimensions, "thickness" and "closeness". Thickness captures the local density of employment, and hence the intensive and extensive margins of agglomeration. Closeness measures the local density of firms, and hence only the extensive margin. Trade liberalization increases thickness and closeness, but only for local industry clusters that were already large before liberalization took place.

These effects are large but attenuate as one moves away from the initial cluster center. For example, by 2007 a one standard deviation larger reduction in trade uncertainty lead to an employment density increase of 21.2% at 10-20km for an average cluster in the largest initial size quartile. We find these reinforcing agglomeration effects until 100km away from the initial center, after which they cease to be significant.

Decomposing the changes in economic density yields that they are primarily driven by firms entering the sample and their subsequent employment growth. Firms that were originally around before liberalization decline in face of the increased competition from new entrants. Some of the original firms also switch their industry to join the local industry experiencing greater trade liberalization.

We further explore local spillover effects between related industries. Trade liberalization propagates to industries upstream and downstream in the supply chain, hence inducing co-agglomeration. In contrast, industries using a similar labour pool seem to suffer from other local industries expanding due to increased access to foreign markets.

These results offer new insights into the relation between trade and agglomeration, and how trade liberalization shapes the spatial distribution of economic activity. We show that the effect of trade uncertainty reductions has heterogeneous effects within sectors, depending on the strength of local agglomerations. Specifically, better access to foreign markets increases the spatial concentration of economic activity because it only leads to expansions of agglomeration clusters that were large initially.

References

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price (2016). "Import Competition and the Great US Employment Sag of the 2000s". In: Journal of Labor Economics 34.S1, pp. 141–198.

Alfaro, Laura and Maggie Chen (2014). "The global agglomeration of multinational firms". In: Journal of International Economics 94.2, pp. 263–276.

Alfaro, Laura, Maggie Chen, and Harald Fadinger (2019). Spatial Agglomeration and Superstar Firms: Firm-level I Harvard Business School.

Alfaro, Laura and Maggie X Chen (2018). "Selection and market reallocation: Productivity gains from multinational production". In: American Economic Journal: Economic Policy 10.2, pp. 1–38.

Autor, David H., David Dorn, and Gordon H. Hanson (2013). "The China Syndrome: Local Labor Market Effects of Import Competition in the United States". In: American Economic Review 103.6, pp. 2121–68.

Bakker, Jan David (2021). Trade and Agglomeration: Theory and Evidence from France. Working Paper.

Bakker, Jan David, Alvaro Garcia Marin, Andrei V Potlogea, Nico Voigtländer, and Yang Yang (2021). Cities, Productivity, and Trade. Working Paper.

Baldwin, Richard E. and Toshihiro Okubo (2005). "Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting". In: Journal of Economic Geography 6.3, pp. 323–346.

Balsvik, Ragnhild, Sissel Jensen, and Kjell G Salvanes (2015). "Made in China, sold in Norway: Local labor market effects of an import shock". In: Journal of Public Economics 127.C, pp. 137–144.

Bloom, Nicholas, Andre Kurmann, Kyle Handley, and Philip Luck (2019). The Impact of Chinese Trade on U.S. Em 2019 Meeting Papers 1433. Society for Economic Dynamics.

Brandt, Loren, Johannes Van Biesebroeck, Luhang Wang, and Yifan Zhang (2017). "WTO Accession and Performance of Chinese Manufacturing Firms". In: American Economic Review 107.9, pp. 2784–2820.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang (2012). "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing". In: Journal of Development Econo 97.2, pp. 339–351.

— (2014). "Challenges of working with the Chinese NBS firm-level data". In: China Economic Review 30.C, pp. 339–352.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sebastien Roux (2012). "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection". In: Econometrica 80.6, pp. 2543–2594.

Coşar, A. Kerem and Pablo Fajgelbaum (2016). "Internal Geography, International Trade, and Regional Specialization". In: American Economic Journal: Microeconomics 8.1, pp. 24–56.

Curuk, Malik and Gonzague Vannoorenberghe (2017). "Inter-sectoral labor reallocation in the short run: The role of occupational similarity". In: Journal of International Economics 108, pp. 20–36.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum (2014). "The Rise of the East and the Far East: German Labor Markets and Trade Intergration". In: Journal of the European Economic Association 12.6, pp. 1643–1675.

Dix-Carneiro, Rafael and Brian K. Kovak (2017). "Trade Liberalization and Regional Dynamics". In: American Economic Review 107.10, pp. 2908–46.

Donoso, Vicente, Victor Martin, and Asier Minondo (2015). "Do Differences in the Exposure to Chinese Imports Lead to Differences in Local Labour Market Outcomes? An Analysis for Spanish Provinces". In: Regional Studies 49.10, pp. 1746–1764.

Duranton, Gilles and Henry G. Overman (2005). "Testing for Localization Using Micro-Geographic Data". In: The Review of Economic Studies 72.4, pp. 1077–1106.

Duranton, Gilles and Diego Puga (2004). "Micro-Foundations of Urban Agglomeration Economies". In: Cities and Geography. Ed. by Henderson, J. Vernon and Thisse, Jacques-Francois. Vol. 4. Handbook of Regional and Urban Economics. Elsevier, pp. 2063–2117.

Ellison, Glenn and Edward L. Glaeser (1997). "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach". In: Journal of Political Economy 105.5, pp. 889–927.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr (2010). "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns". In: American Economic Review 100.3, pp. 1195–1213.

Fajgelbaum, Pablo and Stephen J. Redding (2022). "Trade, Structural Transformation, and Development: Evidence from Argentina 1869–1914". In: Journal of Political Economy 130.5, pp. 1249–1318.

Gathmann, Christina, Ines Helm, and Uta Schoenberg (2020). "Spillover Effects of Mass Layoffs". In: Journal of the European Economic Association 18.1, pp. 427–468.

Gaubert, Cecile (2018). "Firm Sorting and Agglomeration". In: American Economic Review 108.11, pp. 3117–53.

Greenaway, Sir David and Richard Kneller (2008). "Exporting, productivity and agglomeration". In: European Economic Review 52.5, pp. 919–939.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings". In: Journal of Political Economy 118.3, pp. 536–598.

Handley, Kyle and Nuno Limão (2017). "Policy uncertainty, trade, and welfare: Theory and evidence for China and the United States". In: American Economic Review 107.9, pp. 2731–83.

Helm, Ines (2020). "National Industry Trade Shocks, Local Labour Markets, and Agglomeration Spillovers". In: Review of Economic Studies 87.3, pp. 1399–1431.

Helpman, Elhanan, Marc J Melitz, and Stephen R Yeaple (2004). "Export versus FDI with heterogeneous firms". In: American economic review 94.1, pp. 300–316.

Koenig, Pamina (2009). "Agglomeration and the export decisions of French firms". In: Journal of Urban Economics 66.3, pp. 186–195.

Koenig, Pamina, Florian Mayneris, and Sandra Poncet (2010). "Local export spillovers in France". In: European Economic Review 54.4, pp. 622–641.

Krugman, Paul (1991). "Increasing Returns and Economic Geography". In: Journal of Political Economy 99.3, pp. 483–99.

Lu, Yi, Jin Wang, and Lianming Zhu (2019). "Place-based policies, creation, and agglomeration economies: Evidence from China's economic zone program". In: American Economic Journal: Economic Policy 11.3, pp. 325–60.

Marshall, Alfred (1890). The Principles of Economics. McMaster University Archive for the History of Economic Thought.

Martin, Leslie A. and Katie Zhang (2021). "Changing Population Exposure to Pollution in China's Special Economic Zones". In: AEA Papers and Proceedings 111, pp. 406–09.

Melitz, Marc J. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity". In: Econometrica 71.6, pp. 1695–1725.

Nocke, Volker (2006). "A Gap for Me: Entrepreneurs and Entry". In: Journal of the European Economic Associatio
4.5, pp. 929–956.

Pierce, Justin R. and Peter K. Schott (2016). “The Surprisingly Swift Decline of US Manufacturing Employment”. In: American Economic Review 106.7, pp. 1632–62.

Redding, Stephen J. (2020). Trade and Geography. NBER Working Papers 27821. National Bureau of Economic Research, Inc.

Redding, Stephen J. and Esteban Rossi-Hansberg (2017). “Quantitative Spatial Economics”. In: Annual Review of Economics 9.1, pp. 21–58.

Redding, Stephen J. and Daniel M. Sturm (2008). “The Costs of Remoteness: Evidence from German Division and Reunification”. In: American Economic Review 98.5, pp. 1766–97.

A Appendix: Tables

Table A1: Summary of firm data

	Number of Firms	Value Added	Output	Employment	Export	Net value of fixed assets
1998	165110	1.94	6.77	61.96	1.08	4.41
1999	162029	2.16	7.27	58.05	1.15	4.73
2000	162884	2.54	8.57	55.59	1.46	5.18
2001	171243	2.83	9.54	54.41	1.62	5.54
2002	181552	3.3	11.08	55.21	2.01	5.95
2003	196213	4.2	14.23	57.48	2.69	6.61
2004	278984	5.72	20.17	66.16	4.05	7.95
2005	271830	7.22	25.16	68.96	4.77	8.95
2006	301957	9.11	31.65	73.58	6.05	10.57
2007	336763	11.7	40.5	78.74	7.33	12.34

Note: This table shows aggregates of the underlying firm level data by year. This includes all firms in the sample, including non-manufacturing industrial firms. Employment is in million workers. Value added, output, exports and net value of fixed assests are in trillion RMB. All values are in nominal terms.

B Appendix: Figures

C Appendix: Model Derivations

For a city with a higher ℓ_j, the lower bound of the productivity draw is higher. The observed distribution for city j is then

$$\mu_j(\phi) = \frac{g_j(\phi)}{1 - G_j(\phi_j^*)} = k{\phi_j^*}^k \phi^{-k-1}$$

$$G_j^{\text{obs}}(\phi) = \frac{G_j(\phi)}{1 - G_j(\phi_j^*)} = \frac{1 - \left(\frac{\ell_j}{\phi}\right)^k}{\left(\frac{\ell_j}{\phi_j^*}\right)^k}$$

Note that, compared to $g_j(\phi)$, $\mu_j(\phi)$ is shape preserved with however a shifted support from $[\ell_j, \infty)$ to $[\phi_j^*, \infty)$. It follows that the ex-post average productivity $\tilde{\phi}_j$ is

$$\begin{aligned}
\tilde{\phi}_j &= \left[\frac{1}{1 - G_j(\phi_j^*)} \int_{\phi_j^*}^{\infty} \phi^{\sigma-1} g_j(\phi) \mathrm{d}\phi\right]^{\frac{1}{\sigma-1}} \\
&= \left[\frac{k\ell_j^k}{1 - G_j(\phi_j^*)} \frac{{\phi_j^*}^{\sigma-k-1}}{k+1-\sigma}\right]^{\frac{1}{\sigma-1}}, \text{ if } \phi_j^* \geq \ell_j \\
&= \begin{cases} \left[\frac{k}{k+1-\sigma}\right]^{\frac{1}{\sigma-1}} \phi_j^*, & \text{if } \phi_j^* \geq \ell_j \\ \left[\frac{k}{k+1-\sigma}\right]^{\frac{1}{\sigma-1}} \ell_j, & \text{if } \phi_j^* < \ell_j \end{cases}
\end{aligned}$$

which indicates that, under the Pareto distribution, if the cutoff is meaningful ($\phi_j^* \geq \ell_j$), then the ex-post average productivity grows linearly with the productivity cutoff. This property is useful. What follows is that the average profit, $\bar{\pi}_j$, does not depend on the productivity cutoff but on wage level only

$$\bar{\pi}_j = \left[\left(\frac{\tilde{\phi}_j(\phi_j^*)}{\phi_j^*} \right)^{\sigma-1} - 1 \right] f_j w_j = \frac{\sigma - 1}{k + 1 - \sigma} f_j w_j$$

Free entry condition then yields the probability of successful entry, $1 - G_j(\phi_j^*)$, and the productivity cutoff, ϕ_j^*, to be the following

$$\begin{aligned} 1 - G_j(\phi_j^*) &= \min \left\{ \frac{k + 1 - \sigma}{\sigma - 1} \frac{\delta F_j}{f_j w_j}, 1 \right\} \\ \phi_j^* &= \max \left\{ \ell_j \left[\frac{\sigma - 1}{k + 1 - \sigma} \frac{f_j w_j}{\delta F_j} \right]^{\frac{1}{k}}, \ell_j \right\} \end{aligned}$$

The intuition up until now is as follows. Let's say ℓ_j is increasing in e_j. Suppose wages are all the same across cities. Then because of the property of the Pareto distribution, the average profit GIVEN successful entry and the probability of successful entry are both identical across cities. What's different is that entrants draw higher productivity in big cities, and the cutoff productivity is also higher in big cities. Thus, the price level is lower and the firm size is larger in big cities.

Given $\phi_j^*(w_j)$, one can then solve the $2N$ market clearing equations of N product markets and N labor markets. As is shown above, product markets are characterized by

$$\begin{aligned} x_j(\tilde{\phi}_j) &= \left(\frac{p_j(\tilde{\phi}_j)}{\mathbf{P}} \right)^{-\sigma} \frac{\mathbb{W}}{\mathbf{P}} \\ \Longrightarrow \bar{r}_j &= P_j^{1-\sigma} \frac{\mathbb{W}}{\sum_{i=1}^N M_i P_i^{1-\sigma}} \\ \Longrightarrow \left(\frac{\tilde{\phi}_j(\phi_j^*)}{\phi_j^*} \right)^{\sigma-1} r_j(\phi_j^*) &= \frac{k}{k + 1 - \sigma} \sigma f_j w_j = P_j^{1-\sigma} \frac{\mathbb{W}}{\sum_{i=1}^N M_i P_i^{1-\sigma}} \\ \Longrightarrow \sum_{i=1}^N M_i \left(\frac{P_i}{P_j} \right)^{1-\sigma} &= \frac{k + 1 - \sigma}{k\sigma} \frac{\mathbb{W}}{f_j w_j} \\ \Longrightarrow \frac{\mathbb{W}}{\bar{r}_j} &= M_j \sum_{i=1}^N \frac{M_i}{M_j} \left(\frac{P_i}{P_j} \right)^{1-\sigma} \end{aligned}$$

where city-specific price ratio $\frac{P_i}{P_j}$ is given by

$$\frac{P_i}{P_j} = \frac{p_i(\tilde{\phi}_i)}{p_j(\tilde{\phi}_j)} = \frac{w_i}{w_j} \frac{\tilde{\phi}_j}{\tilde{\phi}_i} = \frac{w_i}{w_j} \frac{\phi_j^*}{\phi_i^*}$$

The intuition: cities with a lower price have a larger mass of firms. So, holding wages constant, a city with a higher ℓ_j will have a higher cutoff productivity ϕ_j^* and a lower price level P_j, resulting in more firms/varieties M_j.

A.3 Two asymmetric cities: $N = 2$

When there are only two cities, i.e., $N = 2$, we are able to derive some analytic results. Note that goods market clearing and $N = 2$ imply

$$M_1 \bar{r}_1 = M_2 \bar{r}_2$$

so we have

$$\begin{aligned}
\sum_{i=1}^{N} \frac{M_i}{M_j} \left(\frac{P_i}{P_j}\right)^{1-\sigma} &= \sum_{i=1}^{N} \frac{\bar{r}_j}{\bar{r}_i} \left(\frac{P_i}{P_j}\right)^{1-\sigma} \\
&= \sum_{i=1}^{N} \frac{f_j w_j}{f_i w_i} \left(\frac{w_i \phi_j^*}{w_j \phi_i^*}\right)^{1-\sigma} \\
&= \sum_{i=1}^{N} \frac{f_j w_j^\sigma \phi_j^{*1-\sigma}}{f_i w_i^\sigma \phi_i^{*1-\sigma}} \\
&:= R_j(\mathbf{w})
\end{aligned}$$

which is a function of wages only. It follows that M_j can be expressed as

$$M_j(\mathbf{w}) = \frac{\mathbb{W}}{\bar{r}_j(w_j) R_j(\mathbf{w})}$$

where $\bar{r}_j = \frac{k\sigma}{k+1-\sigma} f_j w_j$. It follows that both portions of the labor demand can be written as a function of wages as well

$$\begin{aligned}
L_j^e(\mathbf{w}) &= \frac{\delta F_j M_j}{1 - G_j(\phi_j^*)} \\
L_j^p(\mathbf{w}) &= M_j \left(f_j + \tilde{\phi}_j^{\sigma-1} \left[\frac{\rho(1-\tau)}{w_j}\right]^\sigma \frac{\mathbb{W}}{\mathbb{P}^{1-\sigma}}\right)
\end{aligned}$$

And the total labor demand is given by

$$L_j^d(\mathbf{w}) = \underbrace{M_j(\mathbf{w})}_{\text{mass of firms}} \left(\underbrace{\frac{\delta F_j}{1 - G_j(\phi_j^*)}}_{\text{labor demand for entry}} + \underbrace{f_j}_{\text{fixed cost}} + \underbrace{\frac{k}{k+1-\sigma} \phi_j^{*\sigma-1} \left[\frac{\rho(1-\tau)}{w_j}\right]^\sigma \frac{\mathbb{W}}{\mathbb{P}^{1-\sigma}}}_{\text{labor demand for variable production}} \right)$$

where

$$
\begin{aligned}
1 - G_j(\phi_j^*) &= \min\left\{\frac{k+1-\sigma}{\sigma-1}\frac{\delta F_j}{f_j w_j}, 1\right\} \\
\phi_j^* &= \max\left\{\ell_j \left[\frac{\sigma-1}{k+1-\sigma}\frac{f_j w_j}{\delta F_j}\right]^{\frac{1}{k}}, \ell_j\right\} \\
\mathbb{P} &= \left[\sum_{j=1}^{N} M_j P_j^{1-\sigma}\right]^{\frac{1}{1-\sigma}}
\end{aligned}
$$