Carnegie Mellon University

Essays on Macroeconomic Policy

Xuege Zhang

A Dissertation submitted to Tepper School of Business
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics

Department of Economics
Tepper School of Business, Carnegie Mellon University
Pittsburgh, PA

Committee Members:

Ali Shourideh

Laurence Ales

Brian Kovak

David Childers

August, 2022

Copyright © Xuege Zhang 2022

ALL RIGHTS RESERVED

Introduction

What is the role of the government in a market economy?

Chapter 1 studies a market failure. I document a fact that in the United States over the past forty years, new business formation has concentrated in large cities. I develop a continuum geographic model where business formations endogenously occur in heterogeneous cities based on ex ante firm productivity. I show that changes in externalities can generate the spatial firm dynamism.

Chapter 2 (joint with Ali Shourideh) studies fiscal policy and reform. We explore the aggregate and distributional effects of tax competition. We develop a theory to analyze how interjurisdictional tax competition affects productivity and inequality by distorting firms' production decisions. Our results suggest that moving toward a coordinated fiscal regime boosts aggregate and local-level productivity. Moreover, tax coordination is more favorable compared to tax competition when wage elasticity is low or when there are more superstar firms.

Chapter 3 studies industrial policy and structural change. I analyze the macroeconomic consequences of place-based industrial policies in the presence of production networks and external economies of scale. I document novel empirical evidence on the spillover effects of a specific place-based industrial policy in China. To rationalize the facts, I develop a quantitative general equilibrium trade model incorporating policies, production networks, trade, and agglomeration externalities. I take the model to Chinese data, designing a new approach that combines sorting and the exact-hat algebra in international trade theories. Results show higher externalities in targeted sectors and aggregate welfare gains through strengthening local sectoral competitiveness.

Chapter 4 (joint with Natasha Che) studies international trade and growth. We design an algo-based export recommendation system using a collaborative filtering method to study the relationship between export structure and growth performance. Our system produces export portfolio for more than 190 economies spanning over 30 years. We find that economies whose export structures are more aligned with the algorithm-recommended export structures achieve better growth performance. Our study provides empirical evidence on export diversification potential and enlights further analyses on comparative advantages.

Disclaimer: The views expressed in this dissertation are those of the author(s) and do not necessarily represent the views of the International Monetary Fund (IMF), its Executive Board, or IMF management.

Acknowledgments

First, I am deeply indebted to my advisors: Ali Shourideh, Laurence Ales, Brian Kovak and David Childers. I have learned so much from them, not only ways to see the world through the lens of economics but also the spirit confronting the boundaries of human knowledge. They have always encouraged my curiosity to explore interesting questions and intuitions throughout my doctoral journey. They inspired me, supported me, and shaped me to be a better thinker.

Second, I am grateful for the fantastic intellectual environment at Tepper School of Business and Carnegie Mellon University (CMU). I thank Andre Mouton, Liyan Shi, Rachel Szymanski, Santiago Tellez, and Shunyuan Zhang for being supportive friends at Tepper. And I thank Rebecca Lessem, Brent Glover, and Ariel Zetlin-Jones for discussions on my early ideas. Laila Lee and Lawrence Rapp provided me with enormous administrative help. I am also influenced by some cool friends (scientists, engineers, artists, and entrepreneurs) at CMU. Their dedication and accomplishment at the frontier of different fields comprise my posterior understanding of the CMU motto *("My heart is in the work")*. I see the future world more positively because of CMU.

Third, I am lucky to have worked with many great colleagues at various institutions (in the public and private sectors, in China and in the U.S.) during the past 12 years. These real-world professional policy practices from different perspectives, especially the economic reforms and policy-making processes that I have been honored to contribute to, allowed me to have a deeper understanding of our fast-changing global economy. Dr. Chen, Dr. Guo, Dr. Liew, Dr. Lin, and Dr. Miao are scholars and policy practitioners who have mentored me and profoundly influenced me from the early stage.

Special thanks to my friends of years/decades: Gus, Iris, Jay, Sisi, Teddy, Xintong, and Yuan for their sincere support.

Last but not least, I want to thank my family for their unconditional love and kindness. This dissertation is dedicated to them.

To my family

Contents

Chapter 1

Spatial Distribution of Business Formation

Abstract

This chapter documents the fact that in the U.S. over the past 40 years, business formation has been concentrated in large cities. I develop a model to explain this concentration through an increase in the extent of production externalities. In particular, this model generates a Pareto distribution for firm entry driven by positive sorting, providing an alternative explanation to a city-size-based distribution. I use the model to show how changes in extent of externalities, size elasticity of urban costs, and the primitive distribution of firm productivity affect the spatial distribution of business formation.

1.1 Introduction

Cities develop with different demographical and technological structures, which are key to understanding regional agglomeration and mobility patterns. When businesses start in a city, they bring job opportunities, technology, and production. This chapter aims to understand the mechanism driving changes in the city-level firm-entry distribution.[1]

This chapter is motivated by a new fact concerning the spatial concentration of firm entry (see Section 1.3 for details). This distributional change is important for several reasons. First, locational choices made by heterogeneous firms are informative in terms of the spatial distribution of productivity, innovation, and resources. Quantifying these effects is useful for policymakers, but doing so is not easy. Second, it has been taken as a stylized fact that larger cities have higher average incomes but also exhibit higher wage gaps, i.e., the urban wage premium (as in Baum-Snow and Pavan, 2011 [1]). Thus, understanding the distribution and its changes can help us understand the sources of income inequality. Third, many policies and incentives affect firms' locational choices.[2]

---

[1] I use *business formation, firm entry,* and *establishment entry* interchangeably in this chapter. Key empirical patterns addressed in this chapter using firm or establishment entries do not vary significantly.

[2] An example of such a policy is a local government subsidy designed to lure firms to locate

Many studies such as Decker et al. (2016) [2] have documented the decline of startups in recent decades. Policy advocates (e.g., Edwards, 2021 [3]) have also sought to remove barriers to new businesses, such as by simplifying the process of obtaining permits and licenses. Analyzing how regulations, either in a centralized economy or a decentralized one, can distort firms' locational choices or optimal city sizes is an important avenue for future investigation.

To explore possible mechanisms behind empirics, I build a model that allows heterogeneous firms to make locational choices driven by production externalities, and derive a spatial equilibrium of firm entry. In this chapter, externalities are defined as extra production benefits in the presence of other firms in the same location. The source and a detailed micro-foundation of variation in production externalities is abstract from this chapter.

The rest of this chapter is organized as follows. Section 1.2 introduces the related literature and contributions of this chapter. Section 1.3 presents facts on business dynamism. Section 1.4 introduces the model and finds its equilibrium. In Section 1.5, I use comparative statics to illustrate how elements of the model drive distributional changes, and I discuss on further suggestive evidence. Section 1.6 concludes.

1.2 Literature Review

This chapter draws and contributes to several branches of literautre. First, this chapter draws from the literature on sorting and matching theory. The spatial sorting mechanism proposed in this chapter is developed from the sorting theory for heterogeneous agents. Positive sorting is originated in matching theory and assignment games (see a survey by Chade et al., 2017 [4]). In the canonical matching model developed by Becker (1973) [5], positive assortative matching comes from output-function increases in a partner's type for a fundamental model based on complementarities in the context of marriage. Recent studies inherit supermodularity including labor sorting (Behrens et al., 2014 [6]) and firm sorting (Gaubert, 2018 [7]). Many previous studies have used a mixture of sorting and selection. Behrens et al. (2014) [6] made a seminal contribution of introducing *ex ante* sorting in location and *ex post* selection in occupation of labor. Baldwin et al. (2006) [8] produced a canonical work that combines the Melitz model with the economic geography model to show that relocating to larger regions is more attractive for productive firms. Previous studies focus less on firm sorting than on labor sorting. Compared to the former ones studing production side, this chapter differs from previous approaches in generating sorting, such as the trade-off between entry cost and price in Nocke (2006) [9], and spatial

their headquarters in designated areas, see Chapter 2 for further discussion. Other examples include public venture-capital funds implemented in Finland, and the technology diffusion and export promotion driven by the Industrial Master Plan in Malaysia. I will explore more place-based, industrial, and export policies will be studied in Chapter 3 and 4.

separation incurring trade costs among locations in Okubo et al. (2010) [10]. This chapter also contributes to the literature on many-to-one sorting. In this chapter, positive sorting, more productive firms sort into large cities, arises (see Section 1.4.2). This is analogous to a many-to-one matching problem, where many firms with similar productivity will optimally choose a single location. As surveyed in Chade et al. (2017) [4], both many-to-one matching and sorting with externalities are still open questions.

Second, this chapter contributes to the literature on firm size distribution. Among the literatures establishing the fact of city size distribution, large cities following the Zipf's Law, and Pareto distributions as in Gabaix (1999) [11], and populated places have been shown to follow log-normal distributions Eeckout et al. (2014) [12]. Ioannides and Skouras (2013) [13] verify the Pareto distribution using multiple different definitions of U.S. cities. City size distributions outside of the U.S. are less unified. Rosen and Resnick (1980) [14] reject Zipf's Law for 36 out of 44 countries. Soo (2005) [15] tests Zipf's Law for 73 countries and finds that the average Pareto exponent is less than 1.

Among the ongoing controversies regarding the exact distributions for firm and city size, two key issues are distributional forms and identifying the mechanisms behind them. There are currently two main streams of thoughts. The first is the random process: Gabaix (1999) [11] presents a canonical model where the Kesten process and Gibrat's law are the underlying mechanisms driving the Pareto distribution. Rossi-Hansberg and Wright (2007) [16] extend this idea, using productivity shocks to explain deviations from Zipf's Law. The second is the recent static approach where Geerolf (2016) [17] introduces a model that generates Pareto distributions for the span of control of managers, and Zipf's law is used for firm size distribution with production functions embedding complementarities. My model is related to that of Geerolf (2016) [17] in that complementarity assumptions are essentially the same, but mine focuses more on sorting firm entries rather than sorting workers. Size distributions derived in this chapter rely on both the primitive distribution and the structure of supply-side optimization.

A side strand of literature related to the size distribution is firm dynamics. The spatial concentration documented in this chapter contributes to the backdrop of a country-level decreasing trend in new business formation. This chapter finds that this decreasing trend holds across cities (see Section 1.3). Papers on firm dynamics have focused mainly on trends of business formation across different ages and sizes at the national level. The most relevant papers are Haltiwanger et al. (2016) [18] and Pugsley and Sahin (2015) [19] who use the Census LBD data set. Rossi-Hansberg and Wright (2007) [16] and Rubinton (2020) [20] examine the connection between establishment scale and firm entry/exit dynamics, focusing on city-level patterns. Other papers consider policy in analyzing firm dynamism. Garicano et al. (2016) [21] estimate the welfare effect of French labor regulation on firm size and productivity distribution. Barwick et al. [22] (2020) and Gu and Jia (2021) [23] study the entry and exit changes responding to Chinese industrial policies and state-owned enterprise (SOE) reforms. Asturias et al. (2017) [24] find that reducing entry barriers to technology adoption for firms generates aggregate growth in Chile and South Korea.

Third, this chapter contributes to studies incorporating supply-side externalities, which have long been discussed in economics. Marshall (1890) [25] proposed three canonical sources of agglomeration externalities: labor (labor-market pooling), goods (input-output linkage), and ideas (knowledge spillover). Since then, externalities have been widely discussed in terms of the above aspects, as well as in terms of endowed amenities (such as natural advantages; see Ellison and Glaeser (1999) [26]). This chapter abstracts from the exact mechanisms generating externalities. Instead, I consider *ex ante* identical locations and embed externalities determining firms' choices in endogenous city distributions. See Section 1.5 for evidence of a change in externalities and a discussion as to the source of this change.

1.3 Facts

This chapter uses the city as the geographical unit of analysis. A city refers to a Metropolitan Statistical Area (MSA), as defined by the Office of Management and Budget (OMB) for federal statistical purposes. According to the OMB, an MSA represents a geographic area consisting of a large population nucleus together with adjacent communities having a high degree of economic and social integration with the nucleus.[3] To understand production agglomeration at the city level, this chapter gathers the total number of establishments, establishment entries and employment from the Census Business Dynamics Statistics (BDS), as well as GDP and population from the BEA, from 1977 to 2014.[4] In the fact section, city size is measured by entries, establishments, population and GDP, and I will compare their distributional changes respectively. In the model section, city size is measured by the number of entries or total number of establishments to understand and highlight the effect of agglomeration on the distribution of business formation. In what follows, I document two facts that motivate the quantitative analysis.

First, the correlation between entry and total number of establishments increases significantly. I use the following set of regressions.

$$Y_{i,t} = \beta_0 + \beta_1 ln(estab)_{i,t} + \beta_2 ln(estab) * \textbf{Year} + \gamma_i + \eta_t + \epsilon_{i,t}$$

where $Y_{i,t}$ denotes $ln(entry)$ and $ln(entry\_pc)$ (per capita) at city-year level. t is from 1977 to 2014, and i denotes cities. Independent variables include to-

[3] The definition used in this chapter is the July 2015 version; for details, see BEA regional section on the MSA list. City and MSA are used interchangeably in the rest of this chapter.

[4] The BEA reports total earnings by metropolitan areas as metro-level GDP.

tal number of establishments and interacting terms. γ_i and η_t are fixed effects. Results (see Table 1.3) show siginificant increases in the correlation between entry and establishment with interacting terms significantly positive. The estimated size elasticities of entry and entry per capita (β_1) are 1.088 and 0.195 respectively. β_2 are both positive and monotonically increasing, indicating the positive correlation increasing over time. OLS plots without fixed effect between $ln\ Entry$, $ln\ Entry\ Per\ Capita$ and $ln\ Establishment$ are shown in Figure 1.13. Figure 1.13.c and 1.13.d show the value of the slopes of fitted line with 95% confidence intervals from 1977 to 2014. The slopes increase significantly in both cases. Accompanying the raising significance is the decline in variation and increase in the explanatory power, reflected by Figure 1.13 with change in standard error and R^2 respectively. Variation of coefficient decreases over time, suggesting a shrinking spread in both cases. R^2 increases from 0 to 0.2 for entry per capita (see solid line in Figure 1.13.f), and this number increases from 0.97 to 0.985 for total entry (see solid line in Figure 1.13.).

Second, firm entry is concentrated in large cities. The concentration is shown by an increase in the share of entry for certain amount of cities. Figure 1.10 shows that establishment and entry have increased disporportionally over time. As indicated in Figure 1.11, the share of GDP and population for the top 20 MSAs remain constant and decrease, but the entry and establishment shares increase significantly.[5] Specifically, in Figure 1.11, the population share is almost flat around 45%, the GDP share remained around 50%, while the establishment entry increased by around 8% (from around 36% to around 44%). Figure 1.12 shows that the cumulative share of establishment entry in top cities (in Figure 1.12, the ten lines from bottom to top aligned represent top 10, 20, 30, ..., 100 MSAs ranked by 2014 nominal GDP) has been increased. The set of the top 10 cities has increased by around 5%, while in the set of the top 100 cities, it has increased by more than 10%.

The concentration is consistent with the nation-wide decline in new business formation documented in Haltiwanger et al. (2014 [2], 2016 [18]) and Pugsley and Sahin (2015) [19]. However, the distribution of this decline across cities is largely unknown. Table 1.1 presents summary statistics for change in entry and establishment. Firm entry in Top 50 MSAs increased in absolute value (from 7137.8 to 8207.3 on average) and share (see the fifth line from the bottom in Figure 1.12) while other MSAs in total decrease in the value (from 640.4 to 552.3 on average) and the share. Total number of establishments and entry both increase in the average and the spread for top cities, while decrease in entry in bottom cities.[6]

---

[5]The rank is based on nominal GDP in each year, which do not change significantly when switching to the real GDP measure. This approach is robust if ranked by population, number of establishment and employment. The same holds true for all ranks mentioned in this chapter later. Note that each share is calculated at the national level.

[6]This fact is robust for entry at firm age 0, which is used to measure the entrepreneur activity in this chapter.

1.4 Model

In this section, I build a simple static model to analyze the effects of changes in the extent of externalities and other key factors on the city size distribution.[7]

1.4.1 Setup

Cities. Assume there is a continuum of cities with mass λ sitting on the interval $[0,\lambda] \equiv \mathscr{I}$.[8] Cities are *ex ante* identical. Both the city size and measure for each size of city are endogenously determined by firms' locational choices. Let $m_i \in \mathbf{R}^+ \cup \{0\}$ denote *ex post city size* and let $f(m_i) \in \mathbf{R}^+ \cup \{0\}$ denote *measure of cities* with size m_i. We index cities by i based on the rank of ex post city size, $i \in \mathscr{I}$, i.e., we build a monotonically increasing function $m_i \equiv m(i) : \mathscr{I} \mapsto \mathbf{R}^+$. Cities differ in externalities, denoted by $\phi(m_i)$, while sizes are endogenously heterogeneous. We assume $\frac{\partial \phi}{\partial m_i} > 0 \; \forall i$, which can be thought of as a production spillover or learning benefit in the presence of other firms in a location.

Firms. Assume there is a unit continuum of firms sitting on the interval $[0,1] \equiv \mathscr{J}$. Firms are exogenously heterogeneous in productivity, drawn from a given distribution $G(a)$ with $supp(G) = [\underline{a}, \infty] \subset \mathbf{R}^+$. We index these firms by j based on the rank of the productivity, $j \in \mathscr{J}$, which can be summarized by a monotonically increasing function $a_j \equiv a(j) : \mathscr{J} \mapsto supp(G)$. Assume each firm hires unit labor as the only input and produce a unit numeraire good. A firm chooses the city size that is endogenously and uniquely determined in equilibrium. Firm a_j's production function $y_{i,j} \equiv y(a_j, m_i)$ at city i is $y_{i,j} = a_j \phi(m_i)$.

Labor. Assume workers are homogeneous and do not make locational choices. Denote $w(m)$ the labor supply that represents the congestion cost increasing in city size.[9] Assume there is a pool of infinite labor in each city. Index labor in each city by k. Denote $w_{R,k}$ as the reservation wage for individual k, capturing heterogeneous value of time for individuals. Assume $w_{R,k}$ i.i.d. drawn from a stationary distribution of reservation wage of the population $F(w)$, with a finite bounded support $\boldsymbol{W}$. Assume $F(\cdot)$ is the same across cities. Each individual participates the labor market only if the equilibrium wage is higher than the reservation wage. Denote the indicator function $\mathbf{1}(w_{R,k})$ as the choice to work

[7]It is as if a steady state considering a firm's maximization objective over time where there is a probability of death in any given period. It can be easily shown that solving for asymmetric entry equilibrium is essentially equivalent to solving for ex post city size conditional on steady state turnover rate. And measure, index, productivity distribution of firms remain the same.

[8]Discrete city case has positive sorting property and similar equilibrium condition. See appendix for illustration and proofs.

[9]This is to avoid a coordination failure.

for individual k given an equilibrium wage w_i in city i:

$$\mathbf{1}_i(w_{R,k}) = \begin{cases} 1 & \text{if } w_i \geq w_{R,k} \\ 0 & \text{if } w_i < w_{R,k} \end{cases}$$

Denote L_i as the aggregate labor supply in city i. The labor supply function is as follows.

$$L_i = \int_0^{w_i} \mathbf{1}(x) \cdot dF(x) = F(w_i)$$

1.4.2 Equilibrium

Definition

A spatial equilibrium consists of a set of equilibrium city sizes $\{m_i^*\}_{i\in\mathscr{I}}$, the measure of cities $f^*(m_i)$, and equilibrium wages $\{w_i^*\}_{i\in\mathscr{I}}$ such that:

1. Taking the city size m_i^* and local wage w_i^* as given, firm j's locational choice satisfies the following maximization problem with labor markets cleared locally.

$$\max_i a_j \phi(m_i) - w(m_i), \ \forall j \in \mathscr{J}$$

2. In equilibrium, non-degenerated city sizes solve the following:

$$\exists j \in supp(G) \text{ s.t. } m^*(a_j) \in \{\arg\max_{i\in\mathscr{I}} \ \pi(a_j, m_i; \Theta)\}, \ \forall i \in supp(f)$$

3. The measure of cities satisfies a firm-city matching with continuity and no-bunching conditions met:

$$G(a) = \int_{\underline{m}}^{m^*(a)} mdF(m), \ \forall a \in supp(G)$$

where $F(m) \in \mathscr{C}^2$.

Properties

In what follows, I present properties of the equilibrium. I first impose a functional form on ϕ as in Duranton and Puga (2004) [27]: $\phi(m) = m^\alpha$ for tractability. And I assume labor supply is not perfectly elastic and takes the following functional form (as Gennaioli et al. (2013) [28] and Behrens et al. (2014) [6]) where η is the wage elasticity. Assume this elasticity is identical across cities: $w_i = F^{-1}(m_i) = \gamma m_i^\eta$.

Firm j's problem is to choose a city $i \in \mathscr{I}$ from a given set of potential cities to maximize their profits.[10] With the specified functional form assumptions on $\phi(\cdot)$ and cost $w(\cdot)$, individual maximization problem is: $\max_{m \in \{m_i\}_{i=1}^n} a_j m^\alpha - \gamma m^\eta$. This gives us the first property of positive sorting. When $\eta - \alpha > 0$, $\frac{\partial m}{\partial a_j} > 0 \ \forall a_j \in$ **supp**(G), optimal city size is monotonically increasing in individual productivity.[11]

If all optimal choices are achievable in equilibrium, the above optimization problem immediately indicates: (1) cities are homogeneous in firm type; and (2) all firms make positive profits. The intuition is that when the cost elasticity of city size is small enough, all agents will go to the largest city or do not enter, given that firms have low productivity to even afford a city of any size. I summarize the this property with the following lemma.

Lemma 1. (Positive Sorting) *In equilibrium, given* $\theta = (\alpha, \gamma, \eta, k, \underline{a}, \lambda)$ *and* $\eta > \alpha$, $\frac{\partial m^*}{\partial a_j} > 0$ *and* $\pi(a_j) > 0 \ \forall a_j \in supp(G)$.

The second property is the distribution of total number of firm as a sufficient statistic for distribution of firm entry. The nature of the problem gives the indeterminancy of entry in absence of an exogeneous constant entry rate. The reason is that firms make choices based on the city size. In this model, city size is the only identity of the city, but among cities with the same size, firms are indifferent. Thus, we effectively do not differentiate the pair of cities with their sizes flipped in equilibrium.

Third, there exists possible binding constraints because of the relative measure of cities and firms. We look for an equilibrium that the distribution is continuous on $supp(f)$. Since the relative measure of cities and firms are not necessarily equal, there are cases where the smallest size of cities binds firms' choices at the bottom. To see this, let $\hat{m}(a)$ denote the *break-even city size* for any given productivity a_j. From the optimization problem, we know that $\hat{m}(a_j) = (\frac{a_j}{\gamma})^{\frac{1}{\eta-\alpha}}$. By the setup, the relative measure between cities and firms is λ, $\lambda \in \mathbf{R}^+$. We need to consider the following four cases to derive equilibrium city size distribution.

1. All firms enter without constraints;

2. All firms enter almost without constraint (the *ex post* smallest city happens to be the one optiamlly chosen by the least productive firm);

[10]Firms know which one is potentially larger, i.e., the ex post rank; but they do not know the ex post size. We assume the knowledge of such "potential rank" due to the following reason. In absence of city developer, there could be multiple equilibria in the sense that agents do not know which city is exactly the one with their desired size. In other words, the basline model can pin down a unique city size distribution and a unique set of cutoff productivity, but the set of equilibrium city size is not an ordered set.

[11]To see this, $\pi(a_j) = a_j \phi(m^*) - w(m^*) = a_j^{\frac{\eta}{\eta-\alpha}} \gamma^{\frac{\alpha}{\alpha-\eta}} ((\frac{\alpha}{\gamma})^{\frac{\alpha}{\eta-\alpha}} - (\frac{\alpha}{\eta})^{\frac{\eta}{\eta-\alpha}}) > 0 \Leftrightarrow \eta > \alpha, \forall a_j \in [\underline{a}, \infty)$.

3. All firms enter with constraints (only bunching);

4. Not all firms enter (bunching and staying outside of the market)

Let $\underline{m}$ denote *ex post* the smallest city endogenously determined in equilibrium. For any given set of parameters $\theta = (\alpha, \gamma, \eta, k, \underline{a}, \lambda)$, the endogeneous $\underline{m}$ gives us four sets of equilibrium conditions to characterize the above cases, with their relation to $\hat{m}(\cdot)$ and $m^*(\cdot)$ illustrated in Figure 1.1. Each case corresponds to a relative measure of cities and firms (λ), which is characterized in the analytical solution in next section.

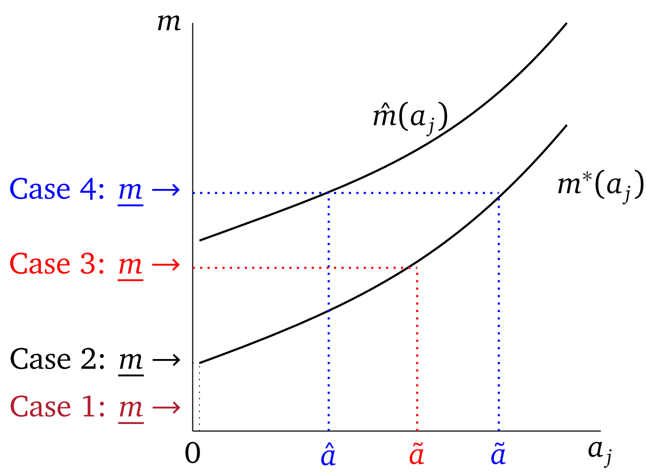


Figure 1.1: Cases when $\underline{m}$ endogenously determined

Equilibrium Firm-City Distribution

For any given set of parameters $\theta = (\alpha, \gamma, \eta, k, \underline{a}, \lambda)$, we obtain a unique city-size distribution satisfying the equilibrium conditions, characterized by the cumulative distribution function $F(m)$ and the productivity cutoffs for exiting $\hat{a}$ and bunching $\tilde{a}$, summarized in the following proposition.

Proposition 1. **(*Ex Post* City Size Distribution)** *For any given set of parameters θ including the extent of externalities α, productivity distribution shape parameter k, scale parameter $\underline{a}$, labor supply coefficient γ, elasticity η, and measure of cities λ (measure of firms normalized to 1), denote $\Theta \equiv (\alpha, \gamma, \eta, k, \underline{a}, \lambda) = \{\Theta \subset \mathbf{R}_+^{\mathbf{card}\Theta}, \eta > \alpha\}$, there exists a unique continuous distribution of ex post city size with or without bunching at the bottom of distribution. When assuming Pareto distribution for firm size distribution, city size distribution follows Pareto, with shape parameter $\xi = k(\eta - \alpha) + 1$.*

Proof. See Appendix 1.7.1. □

The solution of CDF $F^*(m)$ is specified as follows:

- Case 1: When $\Theta \subset \Theta_1 \equiv \{(\alpha,\gamma,\eta,k,\underline{a},\lambda) \subset \mathbf{R}_{+}^{\mathbf{card}\Theta} : \lambda > \bar{\lambda}_2\}$, where $\bar{\lambda}_2 \equiv \frac{k(\eta-\alpha)}{(\frac{\alpha\underline{a}}{\gamma\eta})^{\frac{1}{\eta-\alpha}}(k(\eta-\alpha)+1)}$, there are $\lambda-\beta$ empty cities.

$$F^*(m)=\begin{cases}0, \text{ if } m<\underline{m}\\ 1-\frac{k(\eta-\alpha)(\frac{\underline{a}\alpha}{\gamma\eta})^k}{\beta[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}, \text{ if } m\geq \underline{m}\end{cases}$$

where $\beta=\frac{k(\eta-\alpha)}{[k(\eta-\alpha)+1](\frac{\underline{a}\alpha}{\gamma\eta})^k}$,and $\underline{m}=(\frac{\alpha\underline{a}}{\gamma\eta})^{\frac{1}{\eta-\alpha}}$.

- Case 2: When $\Theta \subset \Theta_2 \equiv \{(\alpha,\gamma,\eta,k,\underline{a},\lambda) \subset \mathbf{R}_{+}^{\mathbf{card}\Theta} : \lambda = \bar{\lambda}_2\}$,

$$F^*(m)=1-\frac{k(\eta-\alpha)(\frac{\underline{a}\alpha}{\gamma\eta})^k}{\lambda[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}.$$

- Case 3: When $\Theta \subset \Theta_3 \equiv \{(\alpha,\gamma,\eta,k,\underline{a},\lambda) \subset \mathbf{R}_{+}^{\mathbf{card}\Theta} : \lambda \in (\bar{\lambda}_1,\bar{\lambda}_2)\}$, where $\bar{\lambda}_1 \equiv \frac{\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1}+(\frac{\eta}{\alpha})^k-1}{(\frac{\underline{a}}{\gamma})^{\frac{1}{\eta-\alpha}}(\frac{\eta}{\alpha})^k}$ and $\bar{\lambda}_2$ as specified above, firms with productivity below $\tilde{a}$ bunch at city with size $\underline{m}$ and we denote $\underline{F}$ as the measure of the bunching size[12].

$$F^*(m)=\begin{cases}0, \text{ if } m<\underline{m}\\ \frac{\underline{F}}{\lambda}, \text{ if } m=\underline{m}\\ \frac{\underline{F}}{\lambda}+\int_{-\infty}^{m}\frac{k(\eta-\alpha)(\frac{\underline{a}\alpha}{\gamma\eta})^k}{x^{k(\eta-\alpha)+2}}dx, \text{ if } m>\underline{m}\end{cases}$$

where $\underline{m}=(\frac{\tilde{a}\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}$, $\underline{F}=\frac{G(\tilde{a})}{\underline{m}}$ s.t.

$$\tilde{a}=\{\tilde{a}\in[\underline{a},\infty):(\frac{\alpha}{\gamma\eta})^{\frac{1}{\alpha-\eta}}\tilde{a}^{\frac{1}{\alpha-\eta}}-\frac{\underline{a}^k\alpha}{\gamma\eta}^{\frac{1}{\alpha-\eta}}\tilde{a}^{\frac{1}{\alpha-\eta}-k}+\beta\underline{a}^{\frac{\theta}{\eta-\alpha}}\tilde{a}^{\frac{\theta}{\alpha-\eta}}-\lambda=0\}.$$

- Case 4: When $\Theta \subset \Theta_4 \equiv \{(\alpha,\gamma,\eta,k,\underline{a},\lambda) \subset \mathbf{R}_{+}^{\mathbf{card}\Theta} : \lambda \leq \bar{\lambda}_1\}$, all firms with productivity under $\hat{a}$ do not enter the market. Firms with producitivty above $\hat{a}$ but below $\tilde{a}$ bunch at city with size $\underline{m}$.

$$F^*(m)=\begin{cases}0, \text{ if } m<\underline{m}\\ \frac{\underline{F}}{\lambda}, \text{ if } m=\underline{m}\\ 1-\frac{k(\eta-\alpha)(\frac{\eta}{\alpha})^k}{\beta[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}, \text{ if } m>\underline{m}\end{cases}$$

[12]In this case, CDF has different form compared to case 4 although it is essentially a special case of case 4. This is because $\tilde{a}$ cannot be solved analytically due to a higher order non-linear equation. But due to the closed-form solution for all variables in case 4, we can write the bound for λ analytically in case 3.

where $\tilde{a} = \{\frac{1}{\lambda}[\frac{a}{(\frac{\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}}(\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1}+(\frac{\eta}{\alpha})^k-1)]\}^{\frac{1}{k+\frac{1}{\eta-\alpha}}}$, $\underline{m} = (\frac{\tilde{a}\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}$, $\underline{F} = \frac{\lambda[(\frac{\eta}{\alpha})^k-1]}{\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1}+(\frac{\eta}{\alpha})^k-1}$, $\beta = \lambda - \underline{F}$.

Proposition 2. (Entry Distribution) *For each ex post city size probability distribution $F^*(m)$, and exogeneous entry/exit rate μ, there exists a unique city-entry steady state probability distribution with CDF $H^*(m)$, which follows Pareto with shape parameter $k(\eta-\alpha)+1$, regardless whether there is bunching or exiting at the bottom of distribution. Specifically, when there is no bunching, the CDF is:*

$$H^*_M(m) = 1 - \frac{\mu^{k(\eta-\alpha)+1}k(\eta-\alpha)(\frac{a\alpha}{\gamma\eta})^k}{\zeta[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}.$$

where $\zeta = \lambda$ in case 1 and $\zeta = \lambda - \underline{F}$ in case 2,3,4 as indicated in Proposition 1.

Proof. See Appendix 1.7.1. □

1.5 Model Implication

In this section, I first confirm that the empirical distribution is consistent with the theoretical result using several statistical approaches. Then, I present some numerical experiments and use the data to identify the change in the extent of externalities that can generate the distributional change of firm entry.

1.5.1 Empirical Distribution

According to Proposition 2, the entry distribution follows Pareto. Using the following robust statistical tests, I show that the empirical city-level establishment entry follows Pareto distribution at the top, with the shape parameter decreasing over time.

1. Pareto regression

I start by following a standard power law OLS regression (Auerbach (1913) [29] and Gabaix (1999) [11]):

$$ln(Rank)_t = \beta_0 + \beta_1 ln(Size)_t + \epsilon_t$$

Measuring city size with population, the Pareto law coefficient is robustly estimated in a range $[0.85, 1.2]$, which can be found in Gabaix (1999) [11] for the U.S. and Soo (2005) [15] for international cases.[13] However, when measuring city size with entry and establishment, Power law (or rank-size) exponent is not

[13]Zipf's Law refers to absolute value of the rank-size coefficient is 1, thus when Pareto coefficient ξ has value 1 it is equivalent to say Zipf's coefficient or Zipf's law exponent. A detailed survey can be found in Gabaix (2003).

close to 1 and it increases significantly in both cases. Figure 1.14 shows the log-log plot of rank versus entry. The flatter slope in the power law plot corresponds to the decrease in the shape parameter of Pareto distribution; see Proposition 3 below. Figure 1.16 compares power law coefficients for entry, establishment, GDP and population. There are two main observations. First, there is a significant linear relation between the rank and the size (although different from −1 over time). Second, contrary to the city size measured by GDP and population, there is a significant increase in Zipf's coefficient for entry and establishment. In the data, we observe an increase in the Zipf's coefficient (or decrease in absolute value of Zipf's coefficient), i.e., a decrease in the shape parameter of Pareto distribution in the continuous case.[14]

Proposition 3. *Assuming a random variable X follows Pareto distribution with a shape parameter a and its discrete realization has a Zipf's coefficient b (power law), then an increase in Zipf's coefficient b is equivalent in a decrease in a.*

Proof. See Appendix 1.7.1. □

2. Mean excess function and QQ-plot. Denote X a random variable and x_F as the right endpoint. I plot the following mean excess function over threshold value u: $e(u) = \mathbf{E}(X - u|X > u), 0 \leq u < x_F$; see results in Figure 1.17. Heavy-tailed distributions typically have mean excess function diverging (to infinity in the extreme) (Embrechts et al., 1997 [30]), which is consistent with the plot for entry. QQ-plot is often used to refer extreme value family with linearity of the fitted line. As instructed by Embrechts (1997) [30], I fit the distribution with MLE and have the estimation plugged in QQ-plot as the standard inspection; see results in Figure 1.18.

3. Pickands estimator (Pickands III, 1975). Pickands estimator ($\hat{\xi}^{(P)}_{k,n} = \frac{1}{ln2} ln \frac{X^n_k - X^n_{2k}}{X^n_{2k} - X^n_{4k}}$) does not show a significant change in the shape parameter. See Figure 1.19.

4. Hill estimator (Hill, 1975). Hill estimator ($\hat{\xi}^{(H)}_{k,n} = \left(\frac{1}{k}\sum_{j=1}^{k} lnX_{j,n} - lnX_{k,n}\right)^{-1}$) shows the range for Pareto in terms of order statistics, which indicates distributional changes over time. (See Figure 1.20) Clauset et al. (2009) [31] design an algorithm for an estimator[15] combining maximum-likelihood and Kolmogorov-Smirnov statistics to test the Power distribtion in empirical data, which is essentially equivalent to Hill estimator.[16] In this chapter, I use the well-developed algorithm in Clauset et al. (2009) [31] to estimate the shape parameter. The change is around 0.08 (from 1.09 to 1.01 from 1977 to 2014), close to the associated change from the truncated Power law.

[14]An alternative way is to use Benford's Law for Pareto distribution. See https://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/ for their equivalence.

[15]The estimator they use is $\hat{\kappa} = 1 + n\left(\sum_{i=1}^{n} ln \frac{X_i}{X_{min}}\right)^{-1}$, where $\kappa = \frac{1}{\xi}$.

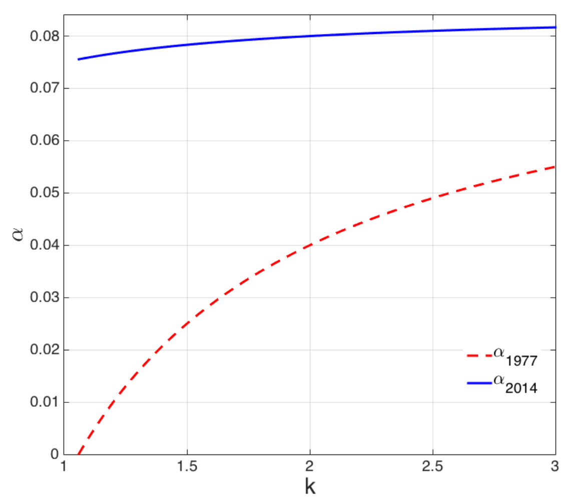
[16]I also test this empirically and the result does not change significantly.

5. Generalized Pareto Distribution MLE (Smith, 1990) with likelihood function is $l((\xi,\beta);\mathbf{X}^{(n)}) = -nln\beta - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n} ln\left(1+\frac{\xi}{\beta}X_i\right)$ where $\mathbf{X}^{(n)} \equiv (X_i^{(n)})_{i=1}^{n}$. Both the Generalized Pareto Distribution (GPD) MLE and the Hill estimator show significant increases in ξ for the order statistics after the top 100 (see Figure 1.20) and for the top 200, 100, 50 and 20 entry cities (see Table 1.4 and 1.5). The results indicate a significant fatter tail. However, according to Table 1.4 and Table 1.5, the estimated shape parameters are sensitive when changing the order statistics. For robustness, I adopt a smaller change as the conservative result.

1.5.2 Quantitative Analysis

I first use the empirical distribution to calibrate values in Proposition 2. Figure 1.2 shows the change in the extent of externalities (α) when assuming productivity distribution (starting from $k = 1.05$ as Luttmer (2007) [32]) and wage elasticity ($\eta = 0.082$ as Behrens et al. (2014) [6]). The magnitudes are consistent with current literature such as Acemoglu and Angrist (2000) [33], Moretti (2004) [34] and Behrens et al. (2014) [6].[17]

Figure 1.2: Relative Change in Externalities (α) Corresponding to Productivity Parameter (k)



As stated in Proposition 2 in Section 1.4.2, the distribution in the range where firms choose their optimal sizes is: $H_M^*(m) = 1 - \frac{\mu^{k(\eta-\alpha)+1}k(\eta-\alpha)(\frac{\underline{a}\alpha}{\gamma\eta})^k}{\zeta[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}$. Thus we can summarize the comparative statics of concentration with the following lemma.

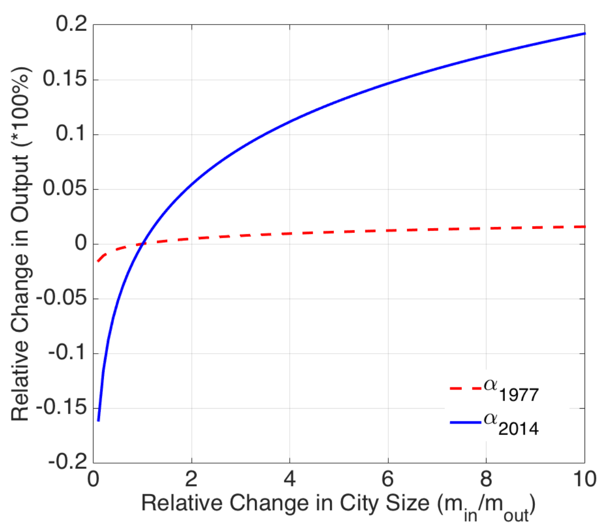
Lemma 2. (Concentration) *An increase in the extent of externalities induces a fatter tail in firm entry distribution, i.e.,* $\frac{\partial\theta}{\partial\alpha} < 0, \ \forall k > 0$.

[17] Their estimates vary from 0.03 to 0.15.

The intuition is that when α increases, firms are driven by a disporportional increase in their benefits from city sizes. Thus, they can afford higher costs to set up business in large cities. Therefore, large cities become disporportionally larger, i.e., a concentration occurs at the aggregate level.

To understand the change in magnitiude, I compute the change in relative output corresponding to the change in externalities. k is taken to be 1.05 as in Luttmer (2007) [32]. As shown in Figure 1.3, for a firm with productivity $a_j \in supp(G)$, moving to a city twice as large as before will incur a 1% net increase in output with the extent of externalities as the value in 1977. However, this number will be 5% with the extent of externalities as the value in 2014. Similarly, moving to a city, for instance, 10 times as large as before, the net increase in outputs would be around 2% in 1977 but will be around 19.5% in 2014.

Figure 1.3: Relative Change in Output Corresponding to Change in Externalities (α)



Although the identified magnitudes are within a reasonable range, few existing studies have identified the change in the extent of externalities over time. And it is difficult to reach an conclusion whether this change (either in α or in output) is economically significant at this point. For further studies, it is worthwhile exploring how this concentration of firm entry shapes the distribution of productivity and income inequality.

Because there is no direct empirical evidence indicating an increase in the extent of externalities, an intuitive source of such change is proposed. In my model setup, the production externality is defined as the extra benifit from being in a larger pool of firms around, for example, a specialization in a production network (either firms producing similar products, or upstream/downstream firms). With a technological improvements, firms and industries are better off concentrating on smaller ranges of tasks, i.e., specializing in a narrower types of techologies (or in a narrower skill-level conditional on the same type of technology). However, a specialization does not mean that production becomes more isolated, or less diversified.[18] Instead, firms can increase their dependence on each other. In other words, technologies can be seperated into different layers of firms in terms of their productivity (ability to adapt to such technologies), but their overall network can be made tighter to ensure professional and matched supports. This setup is not equivalent to the Marshallian externalities in terms of input-output linkages as in Marshall (1890) [25], which is conditional on any technology or productivity level. Rather, I focus on the increase in benefits when technology improves. Such improvement can be any productivity increase, such as that coming from innovation and automation, or better business environment, or improved learning from peers. An example of the latter is that firms become better aware of their own and others' (dis)advantages, leading them to rely more on a direct searches. In this case, the efficiency is reflected in the ability to determine what they need to learn and whom they need to learn it from.

According to the Proposition 2, concentration can also come from changes in the productivity distribution (k) and/or the wage elasticity (η). Because the bottom distribution is not robust, we lack moments with which to identify the entry data directly. To address the issue, there is evidence suggesting a change in k. Specifically, Axtell (2001) [35] identifies the firm size distribution with Census and Compustat data, and finds that shape parameters are stable over time. Moreoever, even if there are evidence indicated that the wage elasticity and firm size distribution are invariant over time, a benchmark is needed to assess the change in α. One way to find such a level is to use firms' profits data accompanied by detailed information on location and firm-specific characteristics. Some studies are going to this direction, such as the identification of human capital externalities as in Moretti (2004) [34], Gennaioli et al. (2013) [28] and Behrens et al. (2014) [6]. That being said, the change in these estimates over time is still ambiguous and worthwhile for further studies.

1.6 Conclusion

U.S. cities differ in their abilities to attract firms and labor. Business formation is concentrated in large cities, accompanied by an overall decline in entries across all cities. This chapter establishes facts regarding the increasing correlations between entry, entry per capita, and city size, as well as firm-entry concentration from 1977 to 2014. Furthermore, the empirical city size distribution follows a Pareto distribution at the top, with a shape parameter that significantly decreases over time when measuring the size with entries or total number of establishments.

In this chapter, I build a heterogeneous firms model to explain this concen-

[18] This is also relevant to the discussion in Chapter 4 at detailed product level.

tration pattern. In equilibrium, firms are sorted to different city sizes. They trade off between complementarity of size and producitivity, and local fixed cost. When the extent of externalities increases, the city size distribution has a fatter tail, indicating the disproportionate pull of larger cities.

This chapter provides a means with which to understand several potential mechanisms in driving the changing geography of firm dynamics. Many interesting questions have arisen; this chapter is a first step towards deeper policy analysis. Relevant and important questions include the economic significance of the concentration and other drivers. In reality, firms' locational choices are driven not only by the scale of the city size but also by variation in entry costs, labor constraints, prices of goods and so forth. To consider other major factors affecting firms' locational choice and spatial distribution of business formation requires a more quantitative model with relevant elements and appropriate firm-level data.

In the remainder of this dissertation, I will address other aspects that require study on production in space and some macroeconomic policies on impact. One relevant aspect is the drivers behind the spatial changes. Structural changes of spatial economic activities involve not just size but also structure.[19] Data suggests that firm entry is concentrated in large cities. However, we also see a large amount of heterogeneity within city groups. Figure 1.22 shows the divergence in firm entry between Dallas and Detroit from 1977 to 2014. Back in 1977, these two cities were similar in terms their relative total establishments and entries: both accounting for 1.5% to 1.7% of the total national entries. However, both the numer of entries and the number of establishments have increased significanly in Dallas, while in Detroit they have decreased over time. In 2014, the disparity in total establishments had increased from 0.1% to nearly 1%, and the disparity in total entries had increased to around 1.2%. Given that they are both among the 20 most populous U.S. cities, the reason for the divergence could be structural changes or location-based policies. This chapter cannot explain such divergences among large cities, but they could be important to understanding inequality in the U.S.[20]

Dallas today has a sprawling startup community. If we anatomize cities with respect to productivity dynamics, we find a set of cities, including San Francisco, Seattle, and Boston, that feature a significant number of entrepreneurs and technological innovations. Cities such as Cleveland and Detroit have seen not only decreasing firm entry but also decreasing production and populations. This divergence is documented in Moretti (2012) [37].

Another factor to consider is regional policies that shape firms' locational decisions and thus change inequality and overall productivity. Texas has no state or local income tax, and in 2014 *CEO Magazine* ranked Texas the number one

[19] I study this idea further in Chapter 3.

[20] Looking at granular patterns reveals a tension between aggregate and local product markets, as documented in Rossi-Hansberg et al. (2021) [36].

state for attracting businesses. The Dallas Regional Chamber also supports many large-scale start-up activities to bolster the business-friendly environment. In addition, special regulations at the subcentral-government level, including tax cuts, rebates, and promises of government services, induce new businesses to locate in specific regions. Tax competition among states or cities can also significantly drive firms' locational decisions (Ruben, 2014 [38]). These points lead to the discussion in Chapter 2.

1.7 Appendix

Table 1.1: Summary Statistics for GDP, Population, Establishment and Entry in 1977 and 2014

	1977				2014			
	Mean	25%	75%	S.D.	Mean	25%	75%	S.D.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP in Top 50	1.92e+07	7871205	1.94e+07	2.42e+07	1.82e+08	7.71e+07	1.92e+08	2.01e+08
GDP in Others	1423646	646446	1897272	1115830	1.27e+07	5284769	1.71e+07	1.03e+07
Population in Top 50	2308092	1015639	2476457	2714391	3491294	1609533	4438715	3366137
Population in Others	203932.8	99335	269295	151502.7	310132.3	143428	414926	235510.6
Entry in Top 50	7137.8	3024	7031	8137.5	8207.3	3118	9242	9179.3
Entry in Others	640.4	307	823	479.2	552.3	224	756	464.3
Establishment in Top 50	42904.7	18204	42714	54540.9	75457.9	34489	86840	79130.7
Establishment in Others	3720.6	1812	4967	2787.8	6081.0	2706	8044	4662.8

Notes: There are 366 observations each year in BDS. Top 50, all MSA after 50 and bottom 50 categorized in the table are based on rank of 2014 nominal GDP. Ranking by nominal GDP in 1977 and 2014 in respective years do not change their share significantly since the rank are almost stable in top 50 cities, especially when only accumulative statistics are required. 25% and 75% represent the first and third quartiles. S.D. stands for standard deviation. Refer to Section 1.3 for further details.

1.7.1 Main Proofs

Proposition 1. (*Ex Post* City Size Distribution) *For any given set of parameters θ, there exists a unique continuous distribution of ex post city size with or without bunching at the bottom of distribution. When assuming Pareto distribution for firm size distribution, city size distribution follows Pareto, with shape parameter $\xi = k(\eta - \alpha) + 1$.*

Proof. Consider cases where firms are constrained and unconstrained by endogeneously determined smallest city. When firms are choosing their optimal size in equilibrium: $\underline{m} \leq m^*(\underline{a})$. Equivalently, city size distribution ($F(m_i)$) has no mass point, i.e., it is absolute continuous with respect to Lebesgue measure. Equilibrium is $F(m)$ (or $f(m)$), $\underline{m}$ such that:

$$\begin{cases} m^*(a_j) = (\frac{a_j \alpha}{\gamma \eta})^{\frac{1}{\eta - \alpha}} \\ G(a) = \int_{\underline{m}}^{m^*(a)} m dF(m) \\ \int_{-\infty}^{\infty} dF(m) = \lambda \\ \underline{m} = m^*(\underline{a}) \end{cases} \tag{1.1}$$

Absolute continuity assumption and monotonicity of the mapping $m^*(a_j)$ allows us to write the add-up condition for measure of firms. We count total measure of firms under productivity a, $G(a)$, by integrating city size m (measured by number of firms choosing this city) from the smallest $\underline{m}$ all the way up to the city size $m^*(a)$ chosen by firm with productivity a. Note we only focus on cases where there is no bunching in the middle of the distribution, as indicated in data. In other words, $F(m)$ is continuous on $[\underline{m}, \infty)$, where $\underline{m}$ denotes the lower bound of city size, i.e., *ex post* smallest city in equilibrium. Thus, there exists a continuous density $f(m)$ on $[\underline{m},\infty)$ such that $G(a)=\int_{\underline{m}}^{m^*(a)} mdF(m)$. As we will see below, even with the bunching and exiting, the distribution above smallest ex post city $\underline{m}$ will not change. This is because the optimal choice of firms that are not bound by the smallest city size will remain same size choices. So the bunching and exiting only change distribution at the bottom.

Because of the potential binding constraint of endogeneous lower bound, we need to consider four cases mentioned in Section 1.4.2 and Figure 1.1.

1. **(Case 1 and Case 2)** When all firms enter and ex post smallest city size is not binding firms' decision. First consider the case where ex post smallest city size is exactly the unconstrained optimal choice of the firm with the lowest productivity, i.e., $\underline{m}=m^*(\underline{a})$. In equilibrium,

$$\begin{cases} m^*(a_j)=(\frac{a_j\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}} \text{ (optimal choice)} \\ G(a)=\int_{\underline{m}}^{m^*(a)} mdF(m) \text{ (measure of firms)} \\ \int_{-\infty}^{\infty} dF(m)=\lambda \text{ (measure of cities)} \\ \underline{m}=m^*(\underline{a}) \end{cases}$$

In this case, $\int_{-\infty}^{\infty} dF(m)=\int_{m^*(\underline{a})}^{\infty} f(m)dm,\ \forall m\in\mathbf{R}$. Solving the ODE of the measure of firms specified above gives $f(m^*(a))=\frac{k\underline{a}^k(\eta-\alpha)(\frac{\alpha}{\gamma\eta})^k}{m^{k(\eta-\alpha)+2}}$. For simplicity afterwards, denote $\beta\equiv\frac{k(\eta-\alpha)}{\underline{m}(k(\eta-\alpha)+1)}, \theta\equiv k(\eta-\alpha)+1$. When $\underline{m}=m^*(\underline{a})$, with some algebra, we obtain the density for equilibrium city size (before normalizing mass to 1):

$$f(m^*(a))=\beta\frac{\theta\underline{m}^{\theta}}{m^{\theta+1}}$$

$$\text{with } \int_{-\infty}^{\infty} f(m)dm=\lambda \Rightarrow \lambda=\frac{k(\eta-\alpha)}{(\frac{\alpha\underline{a}}{\gamma\eta})^{\frac{1}{\eta-\alpha}}(k(\eta-\alpha)+1)}\equiv\lambda_1$$

When the smallest city size is not binding the optimal choice of the least productive firm, i.e., $\underline{m}<m^*(\underline{a})$. It must be that $\underline{m}=0$ since no firms in $supp(G)$ are willing to enter any city smaller than $m^*(\underline{a})$. In this case, $\lambda>\frac{k(\eta-\alpha)}{(\frac{\alpha\underline{a}}{\gamma\eta})^{\frac{1}{\eta-\alpha}}(k(\eta-\alpha)+1)}$, empty cities have total measure: $\lambda-\frac{k(\eta-\alpha)}{(\frac{\alpha\underline{a}}{\gamma\eta})^{\frac{1}{\eta-\alpha}}(k(\eta-\alpha)+1)}$.

2. **(Case 3)** When there is only bunching: $\underline{m}\in(m^*(\underline{a}),m^*(\tilde{a}))$. $\tilde{a}$ denotes the cutoff productivity where firms lower than $\tilde{a}$ will bunch at city with size $\underline{m}$. Let $\underline{F}$ denote the value of the mass point, i.e., the measure of city with size $\underline{m}$. Note all firms are still making positive profits. i.e., no firms are out, although some are not going to their theoretically optimal size m^*.

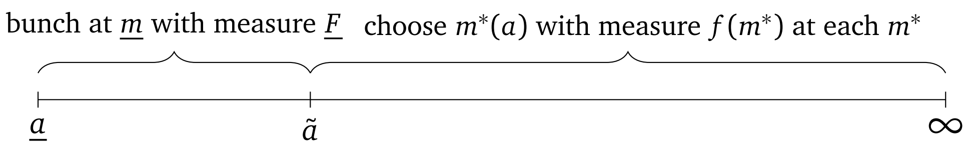


Figure 1.4: When all firms enter and bunching exists

Equilibrium is $F(m)$ (or $f(m)$), $\underline{m}, \tilde{a}, \underline{F}$ s.t.:

$$\begin{cases} m^*(a_j) = (\frac{a_j \alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}} \\ G(a) = G(\tilde{a}) + \int_{\underline{m}}^{m^*(a)} m f(m) dm,\ (a > \tilde{a}) \text{ (measure of firms)} \\ \underline{F} + \int_{\underline{m}}^{\infty} f(m) dm = \lambda \text{ (measure of cities)} \\ G(\tilde{a}) = \underline{m} \cdot \underline{F} \\ \underline{m} = m^*(\tilde{a}) \end{cases}$$

When all firms enter but constrained by the ex post smallest size, i.e. $\underline{m} \in (m^*(a), \hat{m}(a))$.

$$\underline{F} + \int_{-\infty}^{\infty} dF(m) = \lambda$$

$$\Rightarrow \frac{1 - (\frac{\underline{a}}{\tilde{a}})^k}{(\frac{\alpha\tilde{a}}{\gamma\eta})^{\frac{1}{\eta-\alpha}}} + \frac{k(\eta-\alpha)\underline{a}^k}{(k(\eta-\alpha)+1)(\frac{\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}} \tilde{a}^{k+\frac{1}{\eta-\alpha}}} = \lambda$$

The above is a non-linear equation of $\tilde{a}$ which cannot be solved analytically. However, we can solve case 4 analytically, thus all firms enter with bunching when

$$\lambda \in (\frac{\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1} + (\frac{\eta}{\alpha})^k - 1}{(\frac{\underline{a}}{\gamma})^{\frac{1}{\eta-\alpha}} (\frac{\eta}{\alpha})^k}, \frac{k(\eta-\alpha)}{(\frac{\alpha\underline{a}}{\gamma\eta})^{\frac{1}{\eta-\alpha}}(k(\eta-\alpha)+1)})$$

3. **(Case 4)** When there are both bunching and exiting: $\underline{m} \in (m^*(\hat{a}), m^*(\tilde{a}))$. $\hat{a}$ denotes the cutoff productivity of firms breaking even at city size, i.e., $\{\hat{a}\} := \{a_j : a_j \underline{m}^\alpha - \gamma \underline{m}^\eta = 0\}$. $\tilde{a}$ denotes the cutoff productivity of firms with optimal choice exactly equals the size of the smallest city in equilibrium, i.e., $\{\tilde{a}\} := \{a_j : m^*(a_j) = \underline{m}\}$. Thus, firms productivity within the rage $(\hat{a}, \tilde{a})$ will bunch at city with size $\underline{m}$.

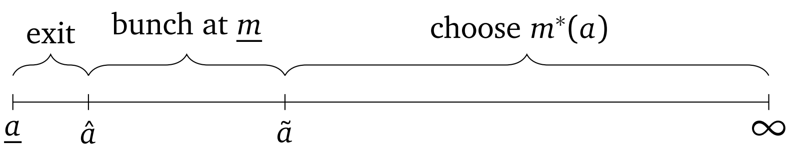


Figure 1.5: When both bunching and exiting exist

Equilibrium is $F(m)$ (or $f(m)$), $\underline{m}, \tilde{a}, \hat{a}, \underline{F}$ s.t.:

$$\begin{cases} m^*(a_j) = (\frac{a_j \alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}} \\ G(a) = G(\tilde{a}) + \int_{\underline{m}}^{m^*(a)} m dF(m), \ (a > \tilde{a}) \text{ (measure of firms)} \\ \underline{F} + \int_{\underline{m}}^{\infty} dF(m) = \lambda \text{ (measure of cities)} \\ G(\tilde{a}) - G(\hat{a}) = \underline{m} \cdot \underline{F} \\ \underline{m} = m^*(\tilde{a}) \\ \hat{a}\underline{m}^{\alpha} - \gamma \underline{m}^{\eta} = 0 \text{ (} \hat{a} \text{ breaks even at } \underline{m}) \end{cases}$$

Thus we obtain $f(m) = \beta \frac{\theta \underline{m}^{\theta}}{m^{\theta+1}}$ on $(\underline{m}, \infty)$, as well as the cutoff and mass point in the bottom of the distribution:

$$\underline{F} = \frac{\underline{a}^k((\frac{\eta}{\alpha})^k - 1)}{(\frac{\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}} \tilde{a}^{-k-\frac{1}{\eta-\alpha}}$$

$$\Rightarrow \tilde{a} = \{\frac{1}{\lambda}[\frac{\underline{a}}{(\frac{\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}}(\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1} + (\frac{\eta}{\alpha})^k - 1)]\}^{\frac{1}{k+\frac{1}{\eta-\alpha}}}$$

When there are firms exiting and bunching, i.e., $\underline{m} \in (\hat{m}(a), \infty)$.

$$\underline{F} + \int_{\underline{m}}^{\infty} f(m)dm = [\frac{\underline{a}^k k(\eta-\alpha)(\frac{\gamma\eta}{\alpha})^{\frac{1}{\eta-\alpha}}}{k(\eta-\alpha)+1} + \underline{a}^k[(\frac{\eta}{\alpha})^k - 1](\frac{\gamma\eta}{\alpha})^{\frac{1}{\eta-\alpha}}]\tilde{a}^{-k-\frac{1}{\eta-\alpha}} = \lambda$$

$$\Rightarrow m^*(\tilde{a}) > (\frac{\underline{a}}{\gamma})^{\frac{1}{\eta-\alpha}} \Leftrightarrow \lambda < \frac{\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1} + (\frac{\eta}{\alpha})^k - 1}{(\frac{\underline{a}}{\gamma})^{\frac{1}{\eta-\alpha}}(\frac{\eta}{\alpha})^k} \equiv \lambda_2$$

At the end, we check whether the solution set is empty. Note all terms are positive in λ_1 and λ_2, thus $\lambda_1, \lambda_2 > 0$. In addition, $\Theta = \{(\alpha, \gamma, \eta, k, \underline{a}, \lambda) : \lambda_2 - \lambda_1 > 0\} \neq \emptyset \ \forall(\alpha, \gamma, \eta, k, \underline{a}, \lambda) \in \mathbf{R}_+^6$ and $\eta > \alpha$.

Therefore, a unique solution exists for any give set of parameters. And we obtain the the distribution for *ex post* city size $F^*(m)$ in equilibrium in the main content. And we obtain the Pareto distribtuion with shape parameter $\xi \equiv k(\eta - \alpha) + 1$. □

Proposition 2. (Entry Distribution) *For each ex post city size probability distribution $F^*(m)$, and exogeneous entry/exit rate μ, there exists a unique city-*

entry steady state probability distribution has CDF $H^*(m)$, which follows Pareto with shape parameter $k(\eta-\alpha)+1$, regardless whether there is bunching or exiting at the bottom of distribution. Specifically, when there is no bunching, the CDF is:

$$H^*_M(m) = 1 - \frac{\mu^{k(\eta-\alpha)+1}k(\eta-\alpha)(\frac{a\alpha}{\gamma\eta})^k}{\zeta[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}$$

where $\zeta = \lambda$ in case 1 and $\zeta = \lambda - \underline{F}$ in case 2,3,4 as indicated in Proposition 1.

Proof. Consider random variable $X \sim F_X(x)$ with CDF F_X and randome variable $M = \frac{X}{\mu} \equiv h(X)$, where μ is a constant entry/exit rate in steady state.

Denote χ as sample space of X, by transformation of random variable:

$$\begin{aligned} F_M(m) &= \mathbf{P}_M(M \leq m) \\ &= \mathbf{P}_M(h(X) \leq m) \\ &= \mathbf{P}_X(x \in \chi : g(X) \leq m) \\ &= \int_{\{x\in\chi:g(X)\leq m\}} f^*_X(x')dx' \\ &= 1 - \frac{k(\eta-\alpha)(\frac{a\alpha}{\gamma\eta})^k}{\zeta[k(\eta-\alpha)+1](\frac{m}{\mu})^{k(\eta-\alpha)+1}} \\ f^*_m(M) = F'_M(m) &= \frac{\mu^{k(\eta-\alpha)+1}k(\eta-\alpha)(\frac{a\alpha}{\gamma\eta})^k}{\zeta^2[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+2}} \end{aligned}$$

where $\zeta = \lambda$ in case 1 and $\zeta = \lambda - \underline{F}$ in case 2,3,4 as indicated in Proposition 1.

Note that the baseline model is without initial size, equivalent to the case where there is initial size but with entry/exit rate $\lambda = 1$. The adding-up condition needs to change when $\lambda \neq 1$ since the total measure of firms only entrants, not incumbents. In other words, to solve for entry, we need to change equilibrium condition to capture the measure of entrants. However, this gives the same results without taking into account of entrants and conduct ex post random variable transformation.

When $\lambda = 1$, for $x \in [\underline{m}, m^*] \equiv \{x : F(x) \in \mathscr{C}^2\}$ where x is variable for *ex post* city size: $G(a) = \int_{-\infty}^{m^*(a)} x dF(x)$. When $\lambda \in (0,1)$, for $x \in [\underline{m}, m^*_e] \equiv \{x : F(x) \in \mathscr{C}^2\}$ where x is variable for entry: $G(a) = \int_{-\infty}^{m^*_e(a)} x dF(x)$.

Therefore the distribution for entry $H^*(m)$ in equilibrium is:

- Case 1:

$$H^*(m) = \begin{cases} 0, \text{ if } m < \underline{m} \\ 1 - \frac{\mu^{k(\eta-\alpha)+1}k(\eta-\alpha)(\frac{a\alpha}{\gamma\eta})^k}{\beta[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}, \text{ if } m \geq \underline{m} \end{cases}$$

  where $\beta = \frac{k(\eta-\alpha)}{[k(\eta-\alpha)+1](\frac{a\alpha}{\gamma\eta})^k}$, $\underline{m} = (\frac{\alpha a}{\gamma\eta})^{\frac{1}{\eta-\alpha}}$

- Case 2: $H^*(m) = 1 - \frac{\mu^{k(\eta-\alpha)+1}k(\eta-\alpha)(\frac{\underline{a}\alpha}{\gamma\eta})^k}{\lambda[k(\eta-\alpha)+1]m^{k(\eta-\alpha)+1}}$.

- Case 3: firms with productivty below $\tilde{a}$ bunch at city with size $\underline{m}$.[21]

$$H^*(m) = \begin{cases} 0, \text{ if } m < \underline{m} \\ \frac{\underline{F}}{\lambda}, \text{ if } m = \underline{m} \\ \frac{\underline{F}}{\lambda} + \int_{\underline{m}}^{m} \frac{\mu^{k(\eta-\alpha)+1}k(\eta-\alpha)(\frac{\underline{a}\alpha}{\gamma\eta})^k}{x^{k(\eta-\alpha)+2}} dx, \text{ if } m > \underline{m} \end{cases}$$

where $\underline{m} = (\frac{\tilde{a}\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}$, $\underline{F} = \frac{G(\tilde{a})}{\underline{m}}$ s.t.

$$\tilde{a} = \{\tilde{a} \in [\underline{a}, \infty) : (\frac{\alpha}{\gamma\eta})^{\frac{1}{\alpha-\eta}} \tilde{a}^{\frac{1}{\alpha-\eta}} - \frac{\underline{a}^k \alpha^{\frac{1}{\alpha-\eta}}}{\gamma\eta} \tilde{a}^{\frac{1}{\alpha-\eta}-k} + \beta \underline{a}^{\frac{\theta}{\eta-\alpha}} \tilde{a}^{\frac{\theta}{\alpha-\eta}} - \lambda = 0\}$$

- Case 4: When $\Theta \subset \Theta_4 \equiv \{(\alpha, \gamma, \eta, k, \underline{a}, \lambda) \subset \mathbf{R}_+^{\mathbf{card}\Theta} : \lambda_e = \frac{\lambda}{\mu} < \chi_2\}$, where $\chi_2 \equiv \frac{\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1} + (\frac{\eta}{\alpha})^k - 1}{\mu(\frac{\underline{a}}{\gamma})^{\frac{1}{\eta-\alpha}}(\frac{\eta}{\alpha})^k}$, all firms with productivity under $\hat{a}$ do not enter the market. Firms with producitivty above $\hat{a}$ but below $\tilde{a}$ bunch at city with size $\underline{m}$. where $\tilde{a} = \{\frac{1}{\lambda}[\frac{\underline{a}}{(\frac{\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}}(\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1} + (\frac{\eta}{\alpha})^k - 1)]\}^{\frac{1}{k+\frac{1}{\eta-\alpha}}}$, $\underline{m} = (\frac{\tilde{a}\alpha}{\gamma\eta})^{\frac{1}{\eta-\alpha}}$, $\underline{F} = \frac{\lambda[(\frac{\eta}{\alpha})^k - 1]}{\frac{k(\eta-\alpha)}{k(\eta-\alpha)+1} + (\frac{\eta}{\alpha})^k - 1}$, $\beta = \lambda - \underline{F}$. □

Power Law Coefficient and Shape Parameter of Pareto Distribution

We refer Power Law coefficient to the size-rank (ln scale) coefficient and Zipf's coefficient a special case of Power Law coefficient with shape parameter 1. Note that Power Law coefficient is between discrete order statistics (rank) and a certain level, whereas Pareto distribution is between continuous cumulative probability (below or above a certain threshold) and the level.

Proposition 3 *Assuming a random variable follows Pareto distribution with shape parameter a and its discrete realization has Power Law coefficient b, then increase in rank-size coefficient absolute value of b is equivalent in decrease in a.*

Proof. Denote a set of discrete realization of random variable M with $X^{(n)} := \{M_i\}_{i=1}^n$, where i index the rank of M. By setup,

$$ln(i) = ln(c) - bln(m) \Rightarrow i \propto m^{-b},\ m \propto i^{-\frac{1}{b}}$$

Denote $m(i) := M_i$. By Power Law definition: $\mathbf{E}(M_i) = m(i) = ci^{-\frac{1}{b}} : \mathbf{N} \mapsto \mathbf{R}^+$

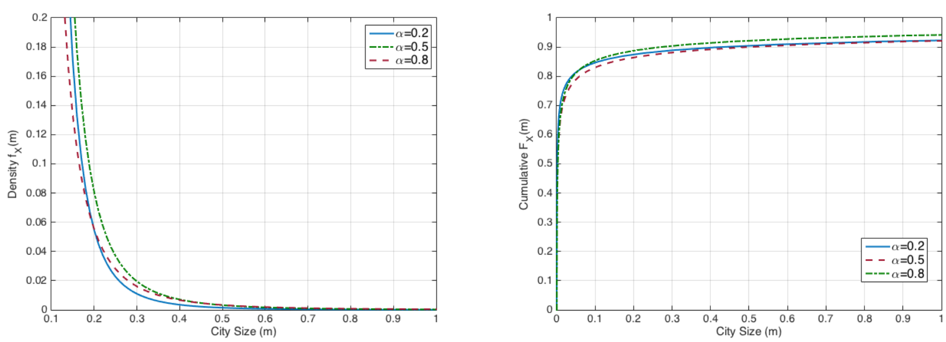
[21]In this case, CDF cannot be written in the form only with exogeneous parameters. This is because $\tilde{a}$ cannot be solved analytically due to a higher order non-linear equation thus lower bound of integral $\underline{m}$ cannot be expanded. But due to the closed-form solution for all variables in case 4, we can write the bound for λ analytically.

for some constant c. Thus $\mathbf{P}(M \geq m(i)) = \mathbf{P}(M \geq ci^{-\frac{1}{b}}) = \tilde{c}i$ for some constant $\tilde{c}$. By change of variable, we have $\mathbf{P}(M \geq n) \propto n^{-b}$, which is cumulative distribution of M and $\mathbf{P}(M = n) \propto n^{-1-b} \equiv n^{-a}$ where $a = 1 + b$, where b is Power Law coefficient and a the shape parameter of Pareto distribution. Therefore, a decrease in b (i.e., an increase in Power Law coefficient) is equivalent to increase in shape parameter of Pareto distribution. □

1.7.2 Comparative Statics for Other Distribution Properties

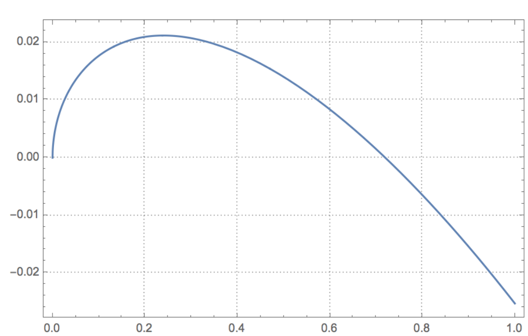
First, I use numerical experiments to illustrate comparative statics when there is both bunching and existing. Due to the lack of robustness of the Pareto distribution in bottom of empirical data, we only present an illustrative numerical experiment in this section to see how externalities drives the distribution at the bottom. $\eta = 1.5$ is chosen as in Saiz (2010) [39]. Other parameters are chosen as: $k = 1.1, \underline{a} = 0.1, \gamma = 1, \lambda = 0.1$. For each case we vary α on $[0, 1]$. We first illustrate the PDF and CDF when $\alpha = 0.2, 0.5, 0.8$.

Figure 1.6: Change in PDF (Left) and CDF (Right) when Increasing α



Note that the shape parameters and scale parameters are changing at the same time. Fatter tail is accompanied by a change in the bottom of distribution. Specifically, the shape parameter decreases monotonically while the scale parameter does not (see Figure 1.7).

Figure 1.7: Change in Scale Parameter when Increasing α

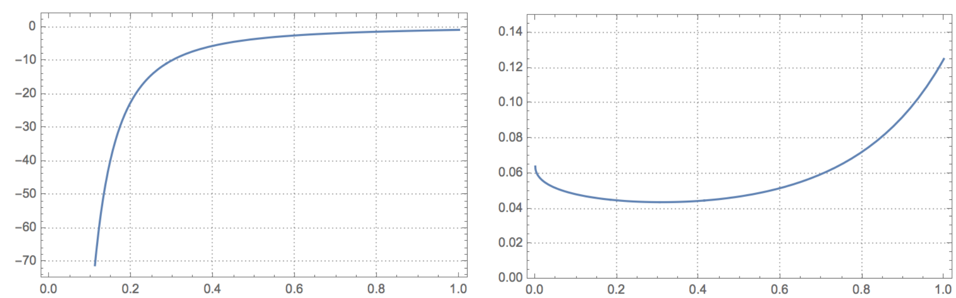


To see this,

$$f(m) = \frac{k(\eta-\alpha)(\frac{\underline{a}\alpha}{\gamma\eta})^k}{m^{k(\eta-\alpha)+2}}$$
$$\Rightarrow \frac{\partial\theta}{\partial\alpha} = -k < 0$$
$$\Rightarrow \frac{\partial\beta}{\partial\alpha} = \frac{\underline{a}k^2(\eta-\alpha)\left(\frac{\alpha\underline{a}}{\gamma\eta}\right)^{k-1}}{\gamma\eta} - k\left(\frac{\alpha\underline{a}}{\gamma\eta}\right)^k$$

Effects on Cutoffs. Figure 1.8 shows when increasing α, firms below and close to the bunching cutoffs are entering the market, cutoff for bunching decreases and cutoff for exiting increases. This is also reflected by the fatter tail of the city size distribution. On the other hand, the exting cutoff increases, thus the mass of firms exiting the market also increases. This corresponds to the change of scale parameter of the Pareto distribution.

Figure 1.8: Left: Effects on bunching cutoff ($\frac{\partial\tilde{a}}{\partial\alpha}$) Right: Effects on exiting cutoff ($\frac{\partial\hat{a}}{\partial\alpha}$)



Effects on Bunching. Figure 1.9 shows that increasing externalities induces a larger size of bunching city while we see its measure decreasing. The decreasing measure corresponds to a decrease in difference of the above two cutoffs and thus the decrease in mass of firms choosing to accept to bunch as opposed to either exit or enter their unconstrained optimal size (if achievable).

Figure 1.9: Left: Effects on the bunching size ($\frac{\partial\underline{m}}{\partial\alpha}$) Right: Effects on the mass of bunching ($\frac{\partial\underline{F}}{\partial\alpha}$)

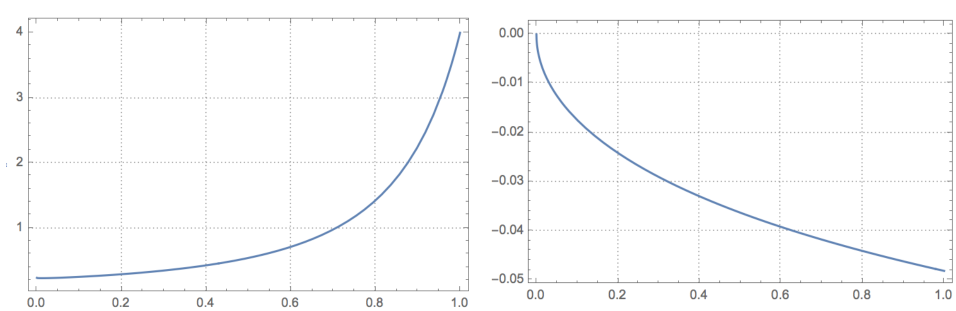
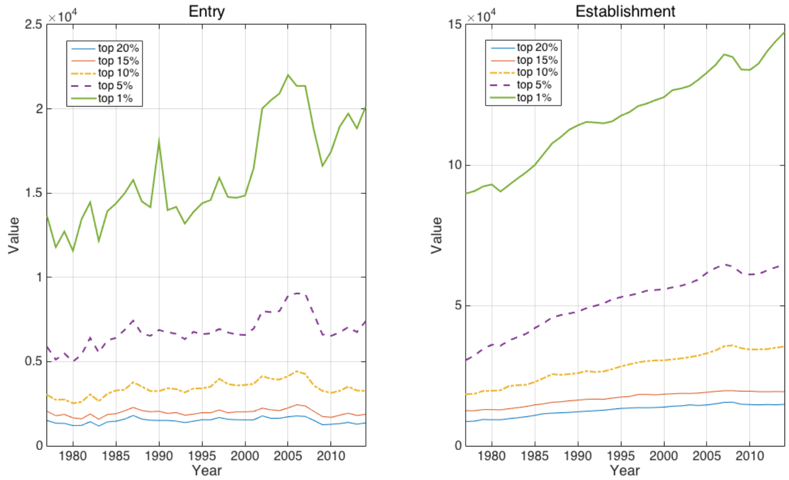


Figure 1.10: Change of Percentile for Entry and Establishment (1977-2014)



Notes: Plots show how value of top 1%, 5%, 10%, 15% and 20% percentile (from the top curve the the bottom curve) in entry (left) and establishment (right) change from 1977 to 2014. Entry and establishment have increased disporportionally, especially for top 1% and top 5% MSAs. Refer to Section 1.3 for further details. Source: BDS.

Table 1.2: Summary Statistics for Entry Per Capita and Entry Per Employee in 2014

	ln Entry Per Capita				ln Entry Per Employee			
	Mean	25%	75%	S.D.	Mean	25%	75%	S.D.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MSA in Total	-6.33	-6.53	-6.13	0.29	-5.27	-5.49	-5.06	0.30
Top 50 MSA	-6.12	-6.27	-5.97	0.18	-5.23	-5.36	-5.04	0.21
All MSA After 50	-6.37	-6.56	-6.17	0.30	-5.30	-5.50	-5.07	0.31
Bottom 50 MSA	-6.40	-6.62	-6.21	0.29	-5.38	-5.56	-5.20	0.31

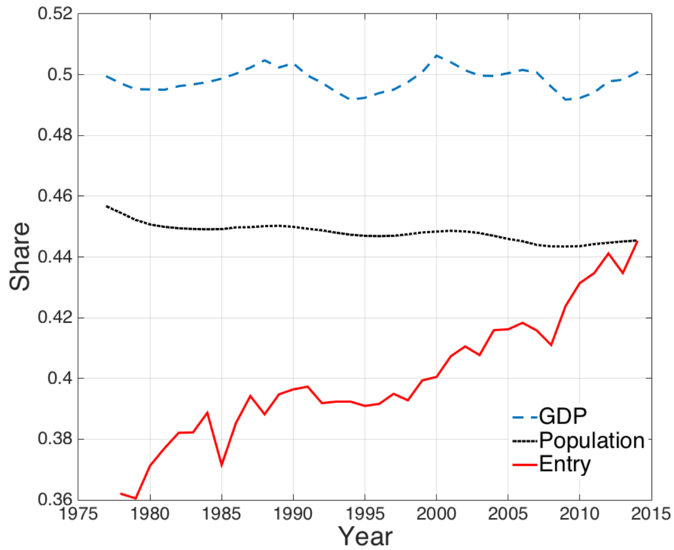
Notes: There are 357 and 366 observations in entry per capita and entry per employee respectively due to the nuance difference in counting MSA in BEA and BDS. Top 50, all MSA after 50 and bottom 50 categorized in the table are based on rank of 2014 nominal GDP. 25% and 75% represent the first and third quartiles. S.D. stands for standard deviation. Refer to Section 1.3 for further details.

Table 1.3: Robust Regression Result for Entry and Entry Per Capita Versus City Size

Variable	ln Entry	ln Entry Per Capita
	(1)	(2)
ln Establishment	1.088***	0.195***
	(0.0078)	(0.0105)
ln Establishment*Year	pos.& signif.	pos.& signif.
Constant	-2.455***	-7.160***
	(0.0643)	(0.0863)
City Fixed Effect	Yes	Yes
Year Fixed Effect	Yes	Yes
Observations	13,908	13,566
R^2	0.994	0.88

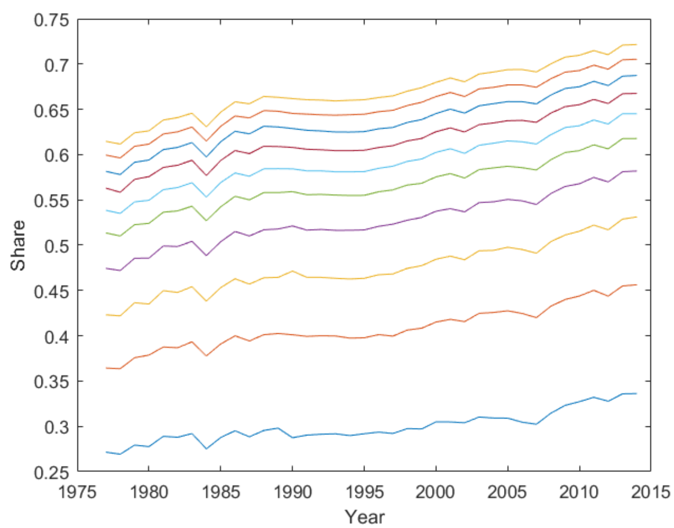
Notes: Using entry per capita and entry per employee do not change the significance of the increase in coefficient of the interacting term. There are 357 and 366 observations in entry per capita and entry per employee respectively due to the nuance difference in counting MSA in BEA and BDS. Standard errors in parentheses: *** p<0.01, ** p<0.05, *p<0.1. Refer to Section 1.3 for further details.

Figure 1.11: Share of GDP, Population, Entry and Establishment in Top 20 MSAs



Notes: Share of establishment entry in top 20 MSAs have increased significantly over time while share of GDP and population remain constant or decrease. This pattern also applies to top 10, 30, ..., 100 MSAs. See a specific illustration of entry in Figure 1.12. The rank is based on nominal GDP in each year. Each share is calculated over national level. Refer to Section 1.3 for further details. Source: BEA, BDS.

Figure 1.12: Share of Establishment Entry in Top Ranked MSAs in the U.S. (1977-2014)



Notes: The bottom to top lines (ten lines in total) stand for the share of establishment entry in top 10, 20, 30, ... 100 MSAs, ranked by 2014 nominal GDP. Ranking by nominal GDP in each year does not change the pattern significantly. Refer to Section 1.3 for further details.

Table 1.4: GPD MLE for Shape and Scale Parameters

Groups	All		Top 300		Top 200	
	1977	2014	1977	2014	1977	2014
	(1)	(2)	(3)	(4)	(5)	(6)
$\underline{m}$	52	73	226	145	446	368
$\underline{\hat{m}}$	533.19	593.55	540.66	591.78	499.42	482.19
(s.e.)	(66.17)	(81.24)	(64.80)	(78.97)	(52.60)	(58.46)
$\hat{\xi}$	0.82	0.92	0.80	0.90	0.76	0.92
(s.e.)	(0.12)	(0.13)	(0.11)	(0.13)	(0.09)	(0.10)
Rate	78.82%	67.61%	80.99%	70.49%	96.88%	84.56%
Convergence	Yes	Yes	Yes	Yes	Yes	Yes

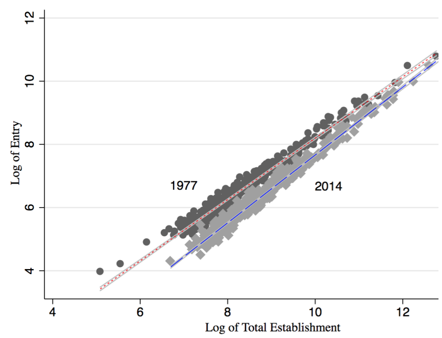
Notes: $\underline{m}$ is the trucated threshold value in data and $\underline{\hat{m}}$ is the fitted threshold for Generalized Pareto Distribution (GPD). "Rate" is the proportion of data points above the threshold. The last row reports successful convergence.

Table 1.5: GPD MLE for Shape and Scale Parameters (continued)

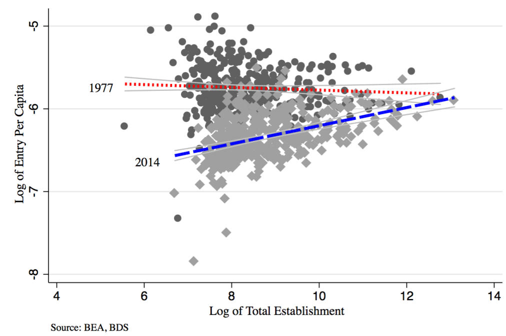
Groups	Top 100		Top 50		Top 20	
	1977	2014	1977	2014	1977	2014
	(7)	(8)	(9)	(10)	(11)	(12)
$\underline{m}$	1099	1048	2205	2091	5759	7123
$\underline{\hat{m}}$	690.24	487.50	778.32	596.75	836.04	661.29
(s.e.)	(62.89)	(50.69)	(67.99)	(57.28)	(71.44)	(61.37)
$\hat{\xi}$	0.65	0.82	0.50	0.70	0.46	0.64
(s.e.)	(0.08)	(0.10)	(0.07)	(0.09)	(0.07)	(0.08)
Rate	100%	100%	100%	100%	100%	100%
Convergence	Yes	Yes	Yes	Yes	Yes	Yes

Notes: $\underline{m}$ is the trucated threshold value in data and $\underline{\hat{m}}$ is the fitted threshold for Generalized Pareto distribution. "Rate" is the proportion of data points above the threshold. The last row reports successful convergence.

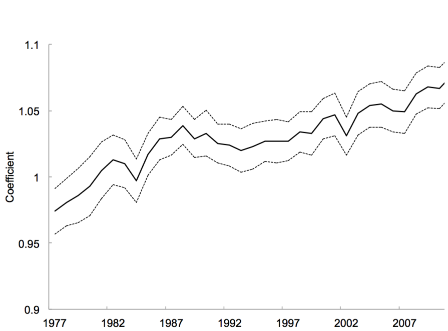
Figure 1.13: Entry and City Size in all MSAs (1977-2014)



a. Entry V.S. Establishment

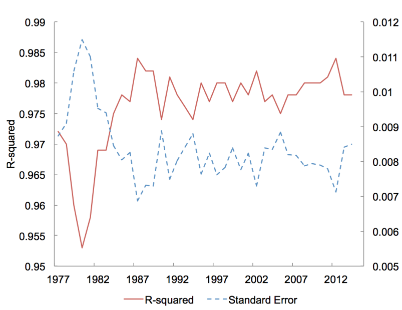


b. Entry Per Capita V.S. Establishment

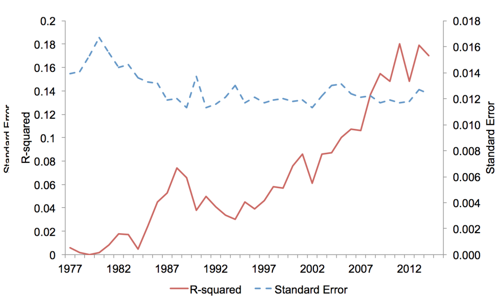


c. Coefficient (95% CI) in Figure 1.13.a

d. Coefficient (95% CI) in Figure 1.13.b



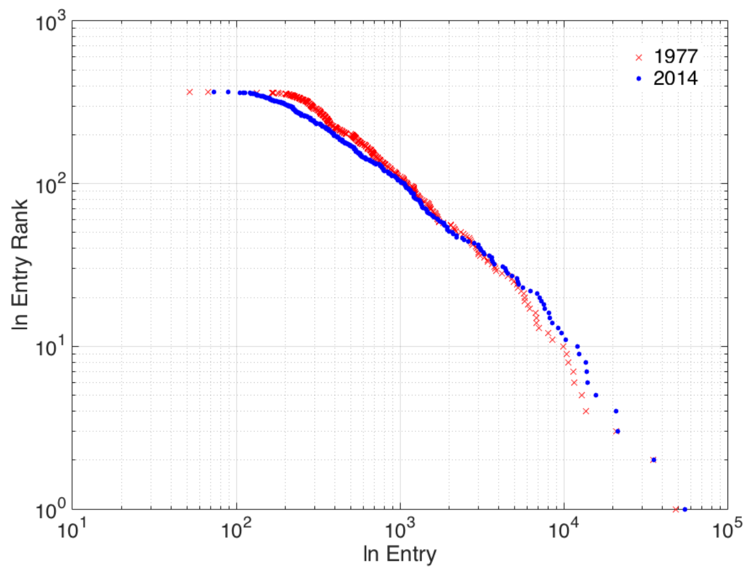
e. R^2 and S.E. of Coefficient in Figure 1.13.a



f. R^2 and S.E. of Coefficient in Figure 1.13.b

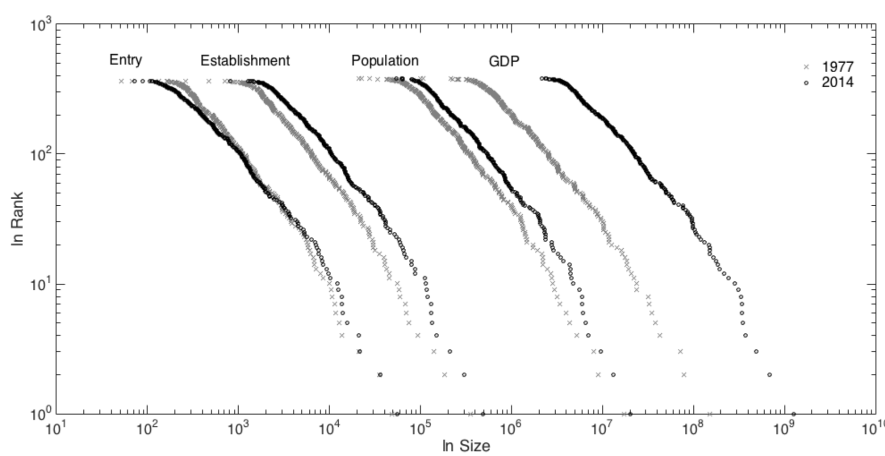
Notes: Figure 1.13.a and Figure 1.13.b are scatter plots for only 2 years: 1977 and 2014. Figure 1.13.c-f show the change of coefficient from 1977 to 2014, in total 38 years. "S.E." stands for standard errors.

Figure 1.14: Comparison of Zipf's Law for Entry, Establishment, Population and GDP in all MSAs in the U.S. (1977-2014)



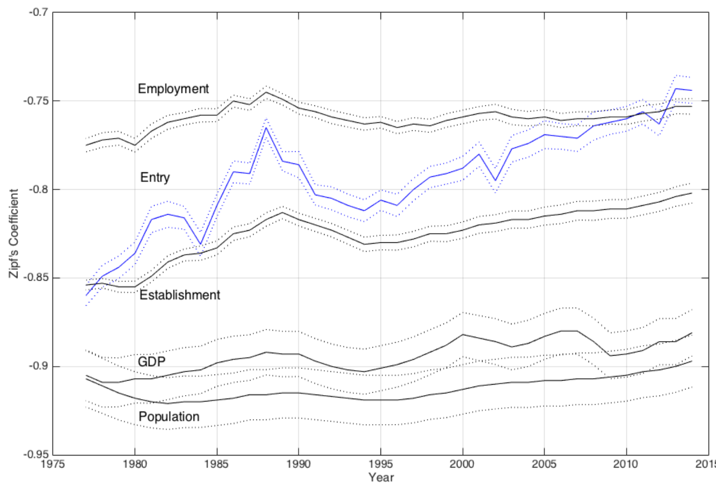
Notes: Refer to Section 1.5.1 for further details. x-axis is the natural log of establishment entry, y-axis is log of rank of establishment entry.

Figure 1.15: Comparison of Zipf's Law for Entry, Establishment, Population and GDP in all MSAs in the U.S. (1977-2014)



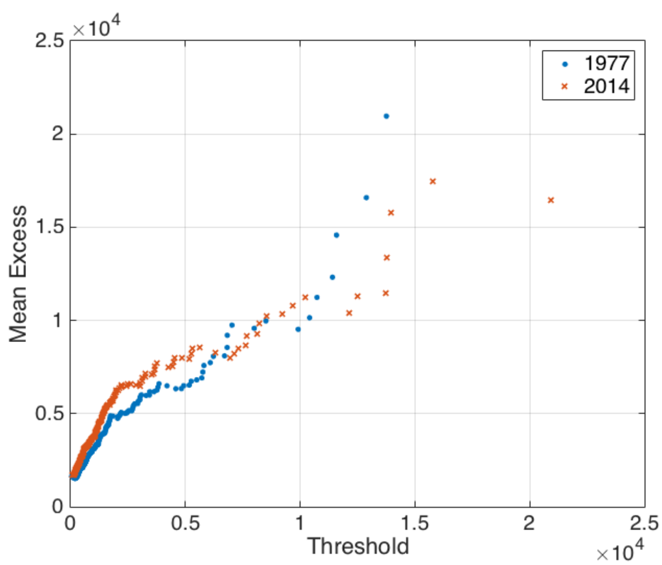
Notes: Refer to Section 1.5.1 for further details. x-axis is the natural log of city size measured by total number of entry, establishment, population and GDP.

Figure 1.16: Zipf's Coefficients (with 95% Confidence Interval) for Establishment Entry, Total Number of Establishments GDP and Population (1977-2014)



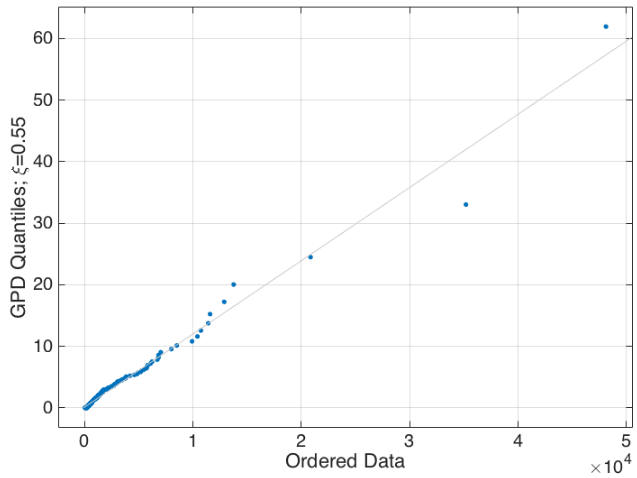
Notes: This graph shows entry rank versus entry size in log-log scale. Statistically significant linear relationship indicates a Pareto distribution. Regression run in plot: $ln(Rank)_t = \beta_0 + \beta_1 ln(Size)_t + \epsilon_t$, where *size* is measured by level of entry in each city. Over time, the slope becomes less steeper with intercept decreases. Zipf's coefficients associated with trucated data adapted for MLE Pareto estimation have similar pattern. Refer to Section 1.5.1 for further details.

Figure 1.17: Mean Excess Function for Entry in 1977 and 2014



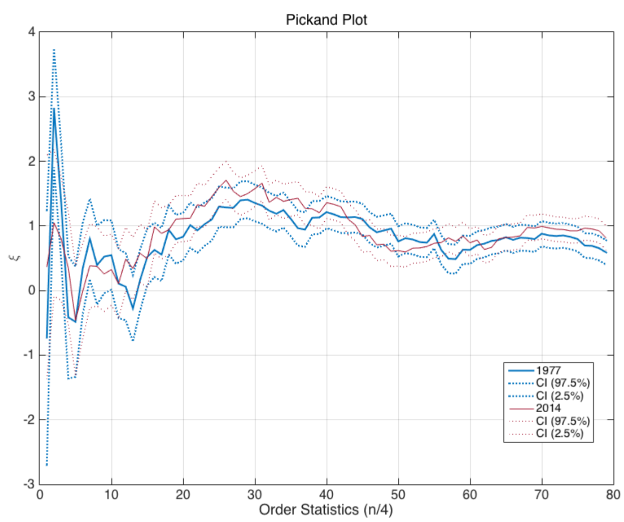
Notes: We plot the following mean excess function ($e(u) = \mathbf{E}(X - u|X > u), 0 \leq u < x_F$) over threshold value u. I.e., the above plot consists $\{(X_i^n, e_n(X_i^n)) : i = 1, ..., n\}$. As in Embrenchts (1997), mean excess function is used only as a graphical method to distinguish light- and heavy-tailed models.

Figure 1.18: QQ-plot for Entry in 1977



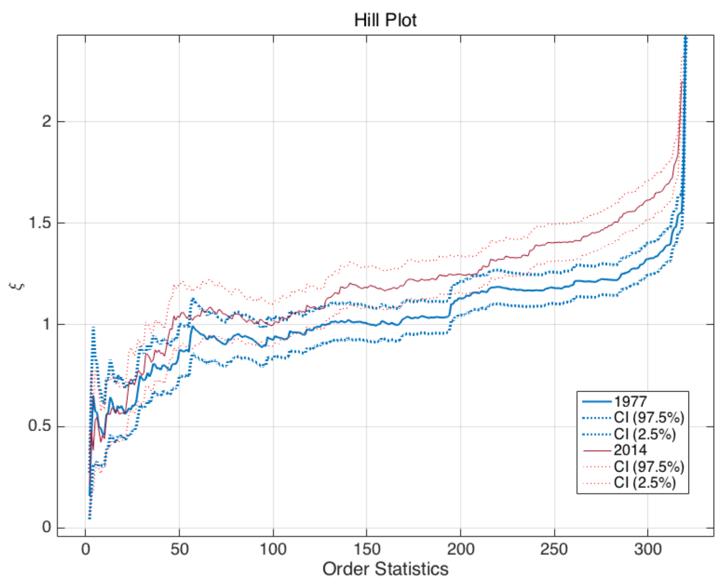
Notes: QQ-plot with ξ indicated using Pareto coefficient in GPD MLE. Refer to Section 1.3 for more details.

Figure 1.19: Pickand's Estimator with 95% CI for Entry in 1977 and 2014



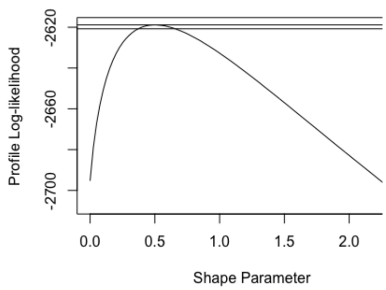
Notes: Pickand's estimator shows stability for shape parameter after order statistics around 20. Due to the derivation of this estimator (for formula refer Section 1.3) this order statistics represents around 80th city. However, estimation does not display significant change from 1977 to 2014.

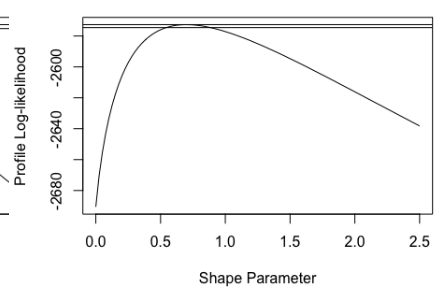
Figure 1.20: Hill's Estimator with 95% CI for Entry in 1977 and 2014



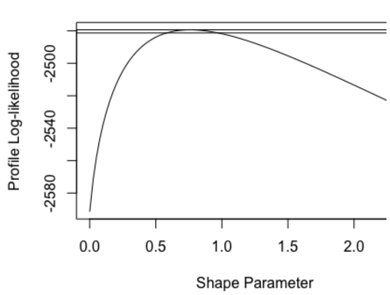
Notes: Hill's estimator shows stability for shape parameter after order statistics around 50. However, several cities at the top is also noisy. 95% confidence interval show significant increase in shape parameter from 1977 to 2014. Note that this shape parameter is inverse of the shape parameter we have in out model due to the original derivation of the formula (see Embrechts (1997) Chapter 6.4.2 for details). For this reason we omit some of the data points when we do calibration.

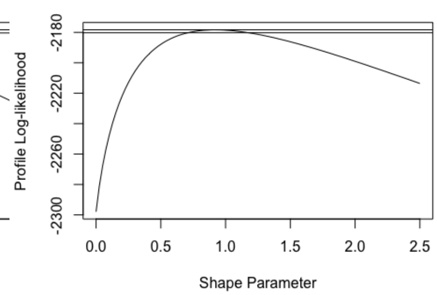
Figure 1.21: Pareto MLE by Different Order Statistics Groups





a. MLE for Top 50 MSAs in 1977 b. MLE for Top 50 MSAs in 2014

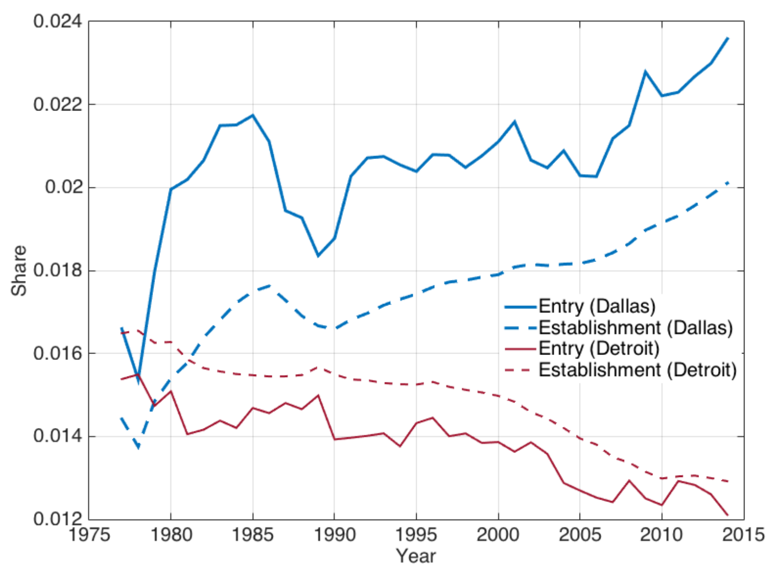




c. MLE for Top 200 MSAs in 1977d. MLE for Top 200 MSAs in 2014

Notes: This is the Pareto fitted shape parameter log-likelihood plot. We sort MSA by number of entry in 1977 and 2014 respectively. I.e., the number of MSA fitted in the model is exactly the trucated order statistics. Then, we fit MLE with top 20, 50, 100, 200 and 300 MSAs. Here we just show Top 50 and Top 200. Full set of MLE with standard errors are presented in Table 1.4.

Figure 1.22: Share of Entry and Establishment in Dallas and Detroit (1977-2014)



Notes: This plot compares the share of entry and establishment in Dallas and Detroit from 1977 to 2014. Two similar cities in terms of entry and establishment diverge over time. Source: BDS.

Chapter 2

Macroeconomic Effects of Tax Competition

(joint with Ali Shourideh)

Abstract

What are the aggregate and distributional effects of spatial corporate tax competition? To answer this question, we develop a theory to analyze how the existence of the interjurisdictional tax competition affects productivity and inequalities. In a tractable model, we characterize city tax and productivity distribution in both coordinated and uncoordinated fiscal policy regimes. We show that the existence of tax competition reduces both aggregate and city-level total factor productivity (TFP). And the size inequality can be driven by shocks in wage elasticity, the heterogeneity of firm productivity, or the city TFP distribution at the aggregate level. Our results indicate that moving from tax competition to tax coordination is more favorable when the wage elasticity is low or when there are more superstar firms. Moreover, our framework sheds light on the distributional effects of tax competition in a constrained fiscal policy regime.

2.1 Introduction

Tax competition has quickly become a major subject of both national and global policy debates, chiefly for two reasons. First, there are numerous ongoing "bidding wars" among local governments within countries as well as among nations around the world. A recent example is the bidding war for Amazon's second U.S. headquarters (HQ2). To secure the HQ2 campus, state and local governments offered huge tax-incentive packages in their proposals.[1] Another well-known case is the "border war" for Kansas City between Missouri and Kansas. Gov-

[1] In the first round, 238 cities bid for HQ2 across the U.S. Among 20 finalists, Maryland proposed an $8.5 billion fiscal incentive package to lure Amazon to locate HQ2 in the Montgomery County. This proposal, released in April 2018, set a new U.S. record for the highest bid. As a result, Amazon splits HQ2, half going to Long Island City, in New York City, (with a $1.5 billion incentive package and an extra $1.3 billion tax break) and half going to the Crystal City neighborhood in Arlington, Virginia (with a $573 million incentive package). Amazon proposed to

ernments on both sides of the state border have offered tax breaks to private companies. And many jobs and plants have been relocated because of changes in policy differentials across the border. These cases are just two among many examples of interjurisdictional competitions. Figure 2.20 presents more cases in which state or local governments have attempted to convince specific firms to relocate or to build new factories in their localities with tax incentive packages valued at more than $50 million.[2] We highlight the top five records in Figure 2.20. The earliest megadeal recorded in the data set was in 1976; the number of deals in recent years has grown significantly.

Second, tax competition is relevant to the design and implementation of an economy's fiscal policies, as well as the global taxation regime. This issue is especially critical in light of globalization (Autor et al., 2013 [40], Autor et al., 2013 [41]) and automation (Acemoglu and Restrepo, 2017 [42]), which have driven up job losses in certain regions and sectors. Furthermore, policy advocates have supported people in moving to areas and sectors with abundant opportunities, or supported firms in moving to localities to create more jobs. Local fiscal policies provide for infrastructure, job training, and corporate tax cuts to economically support people, entrepreneurship, and firms' R&D (Chen et al., 2021 [43]).[3]

Should the central government allow local governments to compete? If these supports and the actions to lure labor or firms are reasonable, how should governments implement these policy incentives? Are these policies optimal under a decentralized regime? Are big cities operating inefficiently due to tax competition? To answer these questions, we need a theoretical framework to determine the potential cost of the decentralized fiscal policy regime and to analyze the relevant mechanisms.

This chapter explores the macroeconomic consequences of the existence of tax competition in which local governments aim to attract heterogeneous firms. Specifically, we develop a theoretical framework that incorporates some key elements that previous models cannot speak to. These elements are based on stylized facts regarding the "urban premium" and sorting in space. The two main drivers of the urban premium are sorting and agglomeration externalities (Combes et al. (2012) [45]; Behrens et al. (2014) [6]; Gaubert (2018) [7]). However, none of these papers endogenize spatial policy. Ossa's (2015) [46] research is one of few general equilibrium settings embedding endogenous policy, but it is abstract from measuring distributional distortions. In this chapter, we consider a tax competition in combination with spatial firm sorting. This is be-

---

create 25,000 jobs at each location. Souce: Vox. https://www.vox.com/the-goods/2018/11/5/18066814/amazon-hq2-locations-selected.

[2] These are named as "megadeals" in the data set provided by Good Jobs First and the *New York Times*.

[3] See a survey by Keen and Konrad (2014) [44] for more detail on practices related to tax competitions, both in the U.S. and around the world.

cause tax competition changes the productivity distribution across locations by distorting firms' locational choices. The theory proposed in this chapter enables us to analyze how a tax competition amplifies the sorting effects of heterogeneous agents in space, and thus affects spatial distribution of productivity and income inequality. We solve for a Nash equilibrium policy in an uncooperative game in comparison to the cooperative regime.

Our main contribution is to bring together the tax competition and the spatial general equilibrium models. We develop a framework to characterize a distributional response to endogenous location-based policy (under both uncoordinated and coordinated regimes) with heterogeneous firms' locational choices. Our model features the positive sorting and a "race to the bottom," as documented in the standard tax-competition literature. And our framework facilitates the understanding on fiscal externalities in a competitive policy regime.

Our theoretical framework produces three main results. First, the existence of tax competition reduces both aggregate TFP and city-level TFP, whereby big cities are affected more than small cities in terms of the level of size and average productivity. Second, in tax competition, size inequality is high when wage elasticity is low, which is the case when there are more superstar firms, or when there are fewer superstar cities in the economy. Third, the cost of tax competition decreases when elasticity is high and firm productivity distribution has a thinner tail. And firms' locational choices at the top remain the same when switching tax regimes. Tax competition triggers a race to the bottom, allowing many low-productivity firms to enter the market and making the whole economy worse off. The optimal policy is achieved when benefits from the total behavioral effect are extended further, to the point when cross-behavioral effects propagated in general equilibrium are internalized.

The rest of this chapter is structured as follows. Section 2.2 connects the paper to the related literature. Section 2.3 presents relevant facts that motivate the model. Section 2.4 introduces the model, and discusses equilibrium together with new elasticities. Section 2.5 describes key properties in cooperative and uncooperative regimes. Section 2.6 presents comparative statics within and across regimes to shed light on the aggregate effects of tax competition. Section 2.7 concludes.

2.2 Literature

Although studies on tax competition have existed for decades, there is no consensus on the aggregate consequences of fiscal regimes. Economists debate whether the interjurisdictional tax competition within a country drives the economy away from the social optimum. The primary concerns include the inefficient provision of public goods and the ambiguous effect on inequalities. Few theoretical tax competition studies explicitly analyze the distributional impact in general equilibrium, while many empirical works are limited to estimate signs of local policy response functions.

This chapter contributes to several strands of literature. First, it relates to a large body of research in public finance and tax competition. Policymakers discuss tax competition and tax harmonization to address fiscal challenges to improve business environments and innovation incentives and to reduce inequalities. For example, in the European Union and the United States, debates have taken place regarding future tax reforms on the value added tax (Devereux and Loretz, 2012 [47]). Recent studies discuss policies designed to harmonize global or regional corporate tax rates (such as Hines, 2022 [48]). In October 2021, 137 member jurisdictions of the OECD/G20 Inclusive Framework on Base Erosion and Profit Shifting (BEPS) reached an international agreement on a minimum effective corporate tax rate of 15% to diminish profit shifting and limit international corporate tax competition.[4]

Traditional tax-competition literature discusses whether the strategic competition between localities induces a race to the bottom, or whether tax coordination across locations is more desirable. However, answers to these questions have generally come from an empirical standpoint. Moreover, with a few recent quantitative exceptions, firm heterogeneity has been neglected in many tax-competition studies. Current theories explain asymmetries in tax rates among competing jurisdictions when each jurisdiction chooses a uniform tax rate within its boundaries (such as Wilson (1987) [51] and Ottaviano and van Ypersele (2005) [52]). Approaches to solving this problem vary substantially. Kanbur and Kenn (1991) [53] develop a model with a single good in a partial-equilibrium setup where governments are revenue maximizers. Nielsen (2001) [54] extends the scope of Kanbur and Kenn's (1991) [53] work to welfare-maximizing governments, but this paper relies on an additively separable relationship between consumer surplus and revenue. Haufler and Stahler (2009) [55] studied tax competition in relation to heterogeneous firms' production under partial equilibrium. In this paper, the concept of tax competition is consistent with that of Wilson and Wildasin (2004) [56], in that it is defined as the process of uncooperative setting of tax rates to attract mobile tax bases.

Second, our framework adds to an extensive literature studying spatial economics and international trade. The spatial-economics literature with international trade models is relevant to our approach in that it incorporates microfoundations for firms' incentives to locate and produce in space. However, current studies fall short of comparing noncooperative and cooperative regimes in a spatial environment. The most relevant study is by Ossa (2015) [46], who considers a tax break competition between U.S. states using aggregate data. However, Ossa's (2015) [46] work constitutes more of a quantitative analysis that focuses less on distributional wedges and the mechanisms behind. Other relevant papers

[4]From OECD (2021) [49]. Also see a relevant theoretical work by Hebous and Keen (2021) [50].

discuss local policies, focusing either on sorting, as Gaubert (2018) [7] does, on optimal spatial policies, as do Fajgelbaum and Gaubert (2020) [57], or on spatial misallocation, as do Serrato and Zidar (2016) [58] and Fajgelbaum et al. (2019) [59]. However, they stop short of considering strategic interactions.

Urban papers, on the other hand, primarily analyze the subject from an empirical perspective and focus more on partial equilibrium. We inherit standard microfoundations when modelling the urban congestion. On top of that, we incorporate competitive and coordinated policy regimes in firm sorting as a source of agglomeration forces. By endogenizing entry margins and taxes, we analyze agents' decisions under a general equilibrium framework to address productivity distribution and inequality. In our paper, in a departure from current studies, the fiscal externality induced by the competitive tax rates incorporates general equilibrium effects in the spatial sorting. We include conventional behavioral effects as in Laffer curve analysis. However, the conventional fiscal competition theories are not enough to explain the propagation channel in different fiscal regimes. Using a heterogeneous-agent continuum-location model, we tractably analyze how aggregates are affected by tax competition and its social costs.

2.3 Motivating Facts

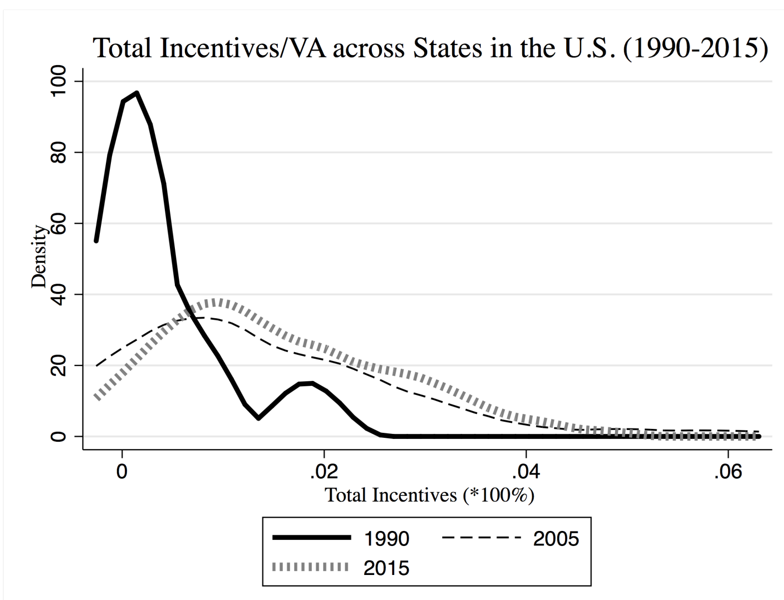
In this section, we describe relevant suggestive evidence from literature that motivates our framework, and present two sets of new facts on local fiscal incentives in the U.S.

Current studies have documented the existence of tax competitions and the impact of tax changes on firm mobility and production decisions. Chirinko and Wilson (2017) [60] demonstrate the existence of tax competition across locations in the U.S. Devereux and Loretz (2013) [47] analyzes both domestic and international cases. There are also studies discussing exogenous shocks affecting the tax competition. For example, Baldwin and Krugman (2004) [8] argues that greater economic integration can lead to a "race to the top" rather than a "race to the bottom". Regarding the impact of tax changes on firm's choices, Giroud and Joshua (2019) [61] and Serrato and Zidar (2016) [58] documented that C-corporations reduced their activity when states increase corporate tax rates. Moretti and Wilson (2017) [62] estimated mobility on the worker side responding to tax changes. In terms of estimating the effect of mobile firms on local TFP, Greenston et al. (2010) [63] uses data from historical magazine *Site Selection* to test the agglomeration spillover of a winning location when competing for firm's locational choice. TFP of incumbent plants and location-wise TFP in winning counties are higher than that in losing counties. They also confirm that productivity gains generated by the spillover are reflected in higher local factor prices. In a more recent work, Nallareddy et al. (2021) [64] documents an increase in income inequality associated with corporate tax cuts in the U.S.

On top of current literature, we explore some new facts to motivate our theoretical analysis. First, we find a significant amount of public expenditure involved in tax incentives, which have increased (relative to the economic scale) over time. Our principal source of data is the Panel Database on Incentives and Taxes (PDIT) provided by the UpJohn Institute. This dataset was released in 2017, and it includes 33 states that account for over 90% GDP. PDIT recategorizes NAICS into 45 industries indexed by a new 2–3-digit industry code. For each state, we observe both state business taxes and state incentives from 1990 to 2015. State business taxes include: business property taxes (36%), sales taxes on business inputs (21%), state taxes on corporate income, and state gross receipts taxes. Incentives include property tax abatements, customized job training subsidies, investment tax credits (ITCs), job creation tax credits (JCTCs), and R&D tax credits.[5]

Based on PDIT, we find a significant increase of fiscal incentives on business activities among localities in the U.S. Specifically, the share of total incentives out of the total tax revenue has increased from less than 10% to 30% from 1990 to 2015 across U.S. states (see Figure 2.22). Looking into the distributional changes, we find a significant increase in total incentives relative to state value-added, reflected by the right-shifted density in Figure 2.1 over time. Moreover, we also find a decrease in local (state-level) tax revenue across locations, reflected by a left-shifted density in Figure 2.2.[6] In addition to the mean shifting, the variance of tax revenue decreases while the variance of incentive increases.

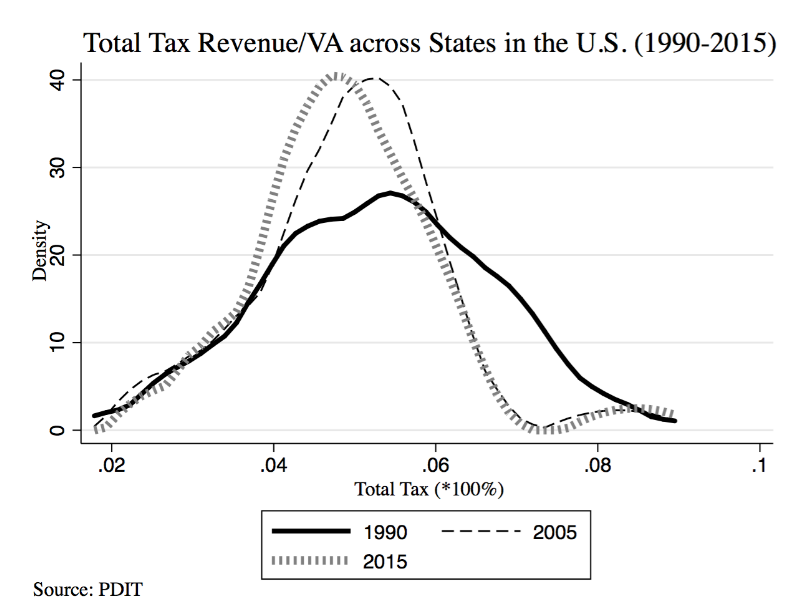
Figure 2.1: Distribution of Incentive Shifts to the Right



[5]States with the highest incentives are New Mexico, Tennessee, New York, Louisiana, and South Carolina. See Bartik (2017) [65] for a more systemetic introduction.

[6]In the Appendix 2.8.11 Figure 2.24, we also show the left-shifted density in the corporate tax income over time.

Figure 2.2: Distribution of Tax Revenue Shifts to the Left



In addition, we test whether changes in tax incentives are closely correlated with changes in international trade flows between the U.S. and its partners. We conduct this test because the international trade volume has increased significantly since the 1990s, which can be considered a driver of the locational decision of multi-national corporations (as in Helpman, 1984 [66] and Wang, 2021 [67]). This is especially relevant after the China's accession to WTO in early 2000s and the intensive trade flows between the U.S. and China.

To do this, we combine state-level taxes and tax incentives data in PDIT and the state-level trade data in Census USA Trade.[7] Since the data set only includes 2014 information, we match this with PDIT data in 2014. We investigate the relationship between sectoral trade flows and incentives using the following specification:

$$y_{i,k} = \beta_0 + \beta_1 x_{i,k} + \gamma_i + \delta_k + \epsilon_{i,k}$$

where $x_{i,k}$ indicates policy measures (incentives, tax revenue, and the incentive shares out of revenue) calculated from PDIT. We index i for location, k for sector. $y_{i,k} \equiv \frac{y_{i,k}}{\sum_{i,k} y_{i,k}}, \frac{y_{i,k}}{\sum_i y_{i,k}}, \frac{y_{i,k}}{\sum_k y_{i,k}}$ are import and export shares calculated from Census. Results are in Table 2.1 and 2.2. We find that local fiscal incentives have insignificant relationship with international trade in the U.S. As regression results indicate, higher incentives do not have a significant correlation with import and export. The export has a few more significant results, but not much compared to other variables. These facts help motivate our model of interjurisditional competitions within a country.

[7]For our purpose, NAICS codes are matched to PDIT from Census USA Trade. See https://usatrade.census.gov.

2.4 Model

2.4.1 Setup

Cities. There is a continuum of cities.[8] Cities are heterogeneous in *ex ante TFP* b, which is drawn from a distribution $G_b(b)$ with $supp(G_b) = \mathscr{B} \subset \mathbf{R}^+$. Each city has *ex post productivity* as the average productivity of firms that choose this city in equilibrium. *Ex post TFP* is a function of *ex post productivity* and *ex ante TFP*, which will be clarified in the firm's setup. We index each city by its inverse CDF of the TFP distribution: $i = G^{-1}(b)$ to faciliate the elasticity analysis in Section 2.4.3. Cities are ex ante empty. Denote $L(b)$ the *ex post city size* in city b determined in equilibrium. Assume there is no agglomeration externalities in each location.[9] Production occurs in cities that are not empty ex post. Production function will be clarified later in the firm's setup.

Each city features an endogenous congestion cost which is identical for all firms choosing to locate in this city. We denote the congestion as $w(L(b))$. There are many ways to model it in the urban literature, such as commuting costs, local land rent, or local wage.[10] In this paper, we follow Behrens and Robert-Nicoud (2015) [68] by assuming an exogenous local wage function:

$$\frac{\partial w(L(b))}{\partial L(b)} > 0$$

Thus far, we have not specified any functional form for $w(L)$. We will assume specific forms in Section 2.4.4 when solving for the equilibrium.

Governments. Assume there is one local government in each city that shares the same index with the city. We assume the government at city with TFP b levies a local lump sum tax $T(b)$ on the profit of a firm that chooses this location in equilibrium. This tax is place-based instead of firm-based.[11] A government's objective is to maximize local tax revenue, and end up to operate at the peak of the Laffer curve. In the tax competition literature, it belongs to a convention of how governments behave (e.g., as in Keen and Kotsogiannis (2004) [69]), often broadly referred as the "Rawlsian" or "Leviathan" governments. This type of government maximizes revenue to rebate people without income, or they invest revenue to provide public goods. This assumption might be different from what occurs in the real world. We model Laffer government to capture one aspect of

---

[8] We will also show the discrete setup to understand policy responses more intuitively.

[9] In the appendix, we relax this assumption and show key characterizations under different policy regimes.

[10] As a part of the "urban premium", it has been well documented in the urban literature that larger cities have higher wages. In this paper, we model congestion with wage. Other microfoundations are also possible, such as land.

[11] This is not completely realistic, but it captures many cases where local economic development tax credits are offered. PDIT documents place-based incentives.

the trade-offs on the government side such that the social welfare increases in revenue. When moving away from maximizing revenue, we would effectively obtain a different Laffer curvature or be constrained in a segment of the Laffer curve. How governments deal with tax revenue ex post is also irrelevant in this paper and we therefore do not offer an opinion on it.

Policy Regimes. In this paper, policy regimes refer to laissez-faire, tax competition and tax coordination. As in Wilson and Wildasin (2004) [56], tax competition is defined as a process of the uncooperative setting of tax rates to attract mobile tax bases. Specifically, we are interested in the Nash equilibrium tax, i.e., government b sets tax $T(b)$ taking government $-b$'s strategy as given. Tax coordination stands for a regime where there is a central government that sets a menu of tax rates across locations and maximize total revenue. We will compare the uncooperative equilibrium with that in coordination. Laissez-faire stands for the case where there is no government, and thus no tax levied.

Firms. There is a mass λ firms in the economy. Firms are heterogeneous in productivity denoted by a, which is drawn from an exogenous cumulative distribution $G_a(a)$, with $supp(G_a) \equiv \mathscr{A} \subset \mathbf{R}^+$. Assume $G_a(\cdot)$ is absolutely continuous. Each firm hires a unit labor and produces a unit numeraire good. Assume there is no sectoral heterogeneity. The total size of a city with TFP level b is characterized by the measure of firms in equilibrium:

$$L(b)g_b(b) = \int_{\mathscr{A}} \mathbf{1}(i^*(a) = b)dG_a(a), \ \forall b \in \mathscr{B}$$

where $\mathbf{1}(i^*(a) = b) = 1$ if firm with productivity a optimally chooses city b in equilibrium; and 0 otherwise. Firm a's payoff in location b is thus $\pi(a, b) = \phi(a, b) - w(L(b)) - T(b)$. We follow Duranton (2004) [27] and assume ϕ is supermodular in size and TFP ($\phi_{ab} > 0$). Note that the congestion cost is identical for all firms that choose to locate in the same city. Thus, there exists positive sorting in firms' productivity and in city size.[12] Each firm makes decisions taking governments' policies as given.[13]

Workers. There is a continuum of workers that are identical and fixed in each city. Each worker is hired at local wage rate $w(L(b))$. After firms choose locations, equal number of new jobs are immediately created locally. It can be shown that this setup induces a general supply curve such that local aggregate

[12] To see this, denote ex post smaller city with L_i.

$$\pi_i(a) > \pi_j(a) \Leftrightarrow a < \frac{w(L_j) - w(L_i) + T_j - T_i}{b_j - b_i}$$

The sorting effect requires $b_i \neq b_j$, since otherwise all firms prefer smaller city to pay lower cost and a coordination failure occurs.

[13] We solve the model backwards: first, we consider each firm's location choice problem (*"location game"* hereafter). Then, we investigate strategic interactions between the governments (*"policy game"* hereafter).

labor supply is equal to city size in equilibrium ($L^*(b)$). In Section 2.4.4, we assume labor supply is not perfectly elastic and takes a parametric form that is standard in the literature.

2.4.2 Equilibrium

Given the above structure, a spatial general equilibrium in tax competition is the set of size, tax and productivity distributions $\{L^*(b), T^*(b), a^*(b)\}$ such that: (1) Taking size and tax as given, firm with productivity $a \in \mathscr{A}$ choose a city $b \in \mathscr{B}$ to maximize profits:

$$\max_{b \in \mathscr{B}} \{\phi(a,b) - w(L(b)) - T(b), 0\}, \ \forall a \in \mathscr{A}$$

(2) Taking strategies in other locations ($-b \in \mathscr{B} \setminus \{b\}$) as given, local government b chooses a level of lump sum tax $T(b)$ to maximize the local tax revenue:

$$\max_{T(b)} L(T(b); b) T(b; T(-b)), \ \forall b \in \mathscr{B}$$

(3) Local labor market-clearing condition is implicitly imposed in $w(L(b))$. Location size $L(b)$ and productivity distribution $a(b)$ is determined in general equilibrium upon firms and governments make choices:

$$\int_b^{\infty} L(b) dG_b(b) = \lambda(1 - G_a(a(b))), \ \forall b \in \mathscr{B}$$

In tax coordination, a firm's problem and market-clearing condition remain the same, whereas a government's problem changes. The central government maximizes total tax revenue across all cities through the selection of a menu of taxes, i.e.,

$$\max_{\{T(b)\}_{b \in \mathscr{B}}} \int_{\mathscr{B}} L(T(b), b) T(b) dG_b(b)$$

Next, we highlight the key difference between our model and current theories. We bring together conventional behavioral effects and fiscal externalities in a general equilibrium, captured by new endogenous elasticities across locations in both discrete and continuum setup. Consistent with current literature, two major forces drive the competitive policy equilibrium: conventional behavioral effects and fiscal externalities. "Conventional" is in the sense that when fixing an arbitrary city, the size of this city decreases when a local tax increases. This is captured by $L(T(b); b)$. Furthermore, fiscal externalities operate on top of it as the regime distortion. When choosing the optimal tax, local governments do not take into account their distributional effects. How the conventional behavioral effects differ across locations (we refer to this as the "cross-behavioral effect", or "cross-elasticity") as general equilibrium outcomes have not been analyzed before.[14] To understand macroeconomic effects, we will analyze how they interact with other factors under different policy regimes as endogenous cross-behavioral effects depend on both productivity distribution and policy regimes. In the following section, we will show equilibrium elasticities with discrete- and continuum-location cases to provide tractable analyses.

2.4.3 Elasticity in Heterogeneous Locations with Endogenous Location Decision

The *cross-elasticity* summarizes a change in tax in any location transmitted to all other locations through general equilibrium effects. We highlight key components and the policy response in two environments respectively.[15]

Discrete Setup

Locations. There are $n \in \mathbf{N}$ cities sitting on interval $[0,1]$. A city sitting at the i^{th} node from 0 can be indexed by $\frac{i}{n}$, $i \in \{1, 2, ..., n\}$. Denote the set of city by $\mathscr{I} \equiv \{\frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\}$. We normalize the original index i by n so that we can then represent the cumulative portion of all cities, which is also useful when extending to the continuum case. The heterogeneous TFP setup is as we defined in main model setup in Section 2.4.1. In discrete case, the exogenous TFP is denoted by b_i at each $i \in \mathscr{I}$. Denote $\mathscr{B} \equiv \{b_i\}_{i \in \mathscr{I}}$ as the set of exogenous TFP across locations. Without loss of generality, assume $b_{\frac{1}{n}} < b_{\frac{2}{n}} < ... < b_1$, i.e., ordering cities with their exogenous TFP from the lowest to the highest. Equivalently, there is a bijection between city orders and TFP: $b(i) : \mathscr{I} \mapsto \mathscr{B}$, where $b_i < b_j$, $\forall i < j \in \mathscr{I}$.

Firms. There is a mass Λ_n firms.[16] A firm's payoff is as above in Section 2.4.1. Denote $L_{\frac{i}{n}}$ as total measure of firms in city i in equilibrium. Firm a's payoff in location i is the following:

$$\pi(a, b_{\frac{i}{n}}) = \phi(a, b_{\frac{i}{n}}) - w(L_{\frac{i}{n}}) - T(b_{\frac{i}{n}})$$

Positive sorting property still holds as a result of supermodularity between local TFP and firm productivity, in addition to the increasing wage function. This

---

[14]Note that *elasticity* mainly refers to a behavioral effect throughout this chapter unless specified, i.e., how size in an arbitrary city changes in relation to tax change in a specific city.

[15]Since general setups are the same as Section 2.4.1, we only state what are different from that in continuum case as follows. In the Appendix, we also show a simple baseline model with discrete locations and uniform productivity and TFP distribution where we can solve cross-elasticity analytically.

[16]The subscript n denotes the total mass corresponding to a given number of cities in the economy. Otherwise, as we increase number of cities, the size of each city will effectively go to zero.

means that more productive firms will sort themselves into ex post larger locations. And the intuition is that all firms would want to go to ex ante productive cities, but only the most productive firms can bear high cost in ex post large cities.

The difference between continuum and discrete cases is that, in the discrete case, there is no perfect positive sorting. What pins down equilibrium productivity distribution will be a set of productivity cutoffs (denoted by $\{a_{\frac{i}{n}}\}_{i=1,\ldots,n}$). We still see more productive firms end up being in ex post larger locations, but around other productive ones, i.e., locations are not ex post homogeneous vis-à-vis the firm type. This is consistent with the result in Behrens and Robert-Nicoud (2015) [68], whereas they do not include either policy or policy regime(s). As a part of equilibrium results, discreteness rules out zero tax equilibrium across locations.[17] This will be discussed in detail in the entry margin section.

With firms maximizing profits and governments maximizing tax revenue, an equilibrium in discrete locations consists of a sequence of taxes, city sizes and productivity cutoffs $\{T_{\frac{i}{n}}, L_{\frac{i}{n}}, a_{\frac{i}{n}}\}_{i=1,\ldots,n,\ n\in\mathbf{N}}$ such that

$$\phi(a_\Delta, b_\Delta) - w(L_\Delta) - T_\Delta = 0 \tag{2.1}$$

$$\phi(a_p, b_p) - w(L_p) - T_p = \phi(a_p, b_{p-\Delta}) - w(L_{p-\Delta}) - T_{p-\Delta},\ \forall p \in \mathscr{I} \setminus \{\Delta, 1\} \tag{2.2}$$

$$\Lambda_n[G(a_{p+\Delta}) - G(a_p)] = L_p,\ \forall r \in \mathscr{I} \setminus \{1\} \tag{2.3}$$

$$\Lambda_n[1 - G(a_1)] = L_1 \tag{2.4}$$

where $\Delta \equiv \frac{1}{n}$ the index of the least productive city. We denote $p, r \in \mathscr{I}$ as some generic cities, which is the inverse cdf of each city's ex ante TFP. Equation (2.1) and (2.2) characterize indifferent productivity at entry margin and between any two arbitrary nearby (in terms of ex ante TFP) cities. Equations (2.3) and (2.4) characterize the adding-up feature: size of city with index $\frac{i}{n}$ equal to the measure of firms between two associated productivity cutoffs.

Consider a tax purterbation in city r. In equilibrium, elasticity of size and

[17] To see this, suppose the equilibrium is zero tax in all cities, then any city with positive mass of firms, which must be heterogeneous, has the motivation to deviate by levying ϵ tax without lossing all of firms.

productivity cutoffs to a unit increase in T_r is characterized by:

$$\phi_a(a_\Delta, b_\Delta)\frac{\partial a_\Delta}{\partial T_r} - w'(L_\Delta)\frac{\partial L_\Delta}{\partial T_r} - \mathbf{1}[\Delta = r] = 0 \tag{2.5}$$

$$\phi_a(a_p, b_p)\frac{\partial a_p}{\partial T_r} - w'(L_p)\frac{\partial L_p}{\partial T_r} - \mathbf{1}[r = p] = \phi_a(a_p, b_{p-\Delta})\frac{\partial a_p}{\partial T_r} - w'(L_{p-\Delta})\frac{\partial L_{p-\Delta}}{\partial T_r} - \mathbf{1}[r = p+\Delta] \tag{2.6}$$

$$\Lambda_n\left[g(a_{p+\Delta})\frac{\partial a_{p+\Delta}}{\partial T_r} - g(a_p)\frac{\partial a_p}{\partial T_r}\right] = \frac{\partial L_p}{\partial T_r} \tag{2.7}$$

$$-\Lambda_n g(a_1)\frac{\partial a_1}{\partial T_r} = \frac{\partial L_1}{\partial T_r} \tag{2.8}$$

With conditions (2.5)-(2.8), we immediately obtain how one perturbation is accumulated and transmitted to other cities, shown in the proof of proposition in appendix.

From Discrete to Continuum

Proposition 1. If elasticity satisfies conditions (2.1)- (2.4) in discrete case, then cross-elasticity $\{\frac{dL_p}{dT_r}\}_{p,r\in\mathscr{I}}$ satisfies the following equations in the limit:

$$-\int_p^1 \frac{dL_{\hat{p}}}{dT_r}d\hat{p}\frac{\phi_{ab}(a_p, b_p)\, b_p'}{\lambda g(a_p)} - w''(L_p)L_p'\frac{dL_p}{dT_r} - w'(L_p)\frac{d}{dp}\frac{dL_p}{dT_r} = 0, \forall p > r$$

$$-\left[-\frac{1}{w'(L_r)} + \int_p^1 \frac{dL_{\hat{p}}}{dT_r}d\hat{p}\right]\frac{\phi_{ab}(a_p, b_p)\, b_p'}{\lambda g(a_p)} - w''(L_p)L_p'\frac{dL_p}{dT_r} - w'(L_p)\frac{d}{dp}\frac{dL_p}{dT_r} = 0, \ \forall p < r$$

$$-\frac{\phi_a(a_0, b_0)}{\lambda g(a_0)}\left[\int_0^1 \frac{dL_{\hat{p}}}{dT_r}d\hat{p} - \frac{1}{w'(L_r)}\right] - w'(L_0)\frac{dL_0}{dT_r} = 0$$

Proof. See Appendix 2.8.1. □

Note that Proposition 1 features an important equilibrium proporty of cross-elasiticity. Fix a location p, and consider a tax perturbation in any other location r, elasticity below and above this perturbation city is not symmetric. To derive the distribution in a more intuitive and economically meaningful way, we will change the measure from p, r to TFP b for elasticity from now on. With a little abuse use of notation, we denote cross-elasticity of an arbitrary city with TFP b with respect to tax perturbation city with TFP p ($b \neq p \in \mathscr{B}$) to be $f(b, p)$ where

$$f(b,p) = \begin{cases} f^L(b,p) \equiv \frac{dL(b)}{dT(p)}, \ \forall b < p \\ f^U(b,p) \equiv \frac{dL(b)}{dT(p)}, \ \forall b > p \end{cases}$$

Further, define

$$F^L(b,p) := -\int_b^\infty f^L(u,p)dG(u)$$

Lemma 1. *(Continuum elasticities)* Given productivity distribution $G_a(b)$ and TFP distribution $G_b(b)$, cross-elasticity $\{f^L(b,p), f^U(b,p)\}$ are characterized by the following partial integro-differential equations $E1, E2$ with a system of boundary conditions:

$$-\int_b^\infty f^U(u,p)dG_b(u)\Psi(b) - \mathscr{W}(b)f^U(b,p) - w'(L(b))f_b^U(b,p) = 0, \forall p > b$$
$$-\left[-\frac{1}{w'(L(p))} + \mathscr{L}^L(b)\right]\Psi(b) - \mathscr{W}(b)f^L(b,p) - w'(L(b))f_b^L(b,p) = 0, \forall p < b$$

where $\mathscr{L}^L(b) \equiv \int_b^\infty f^L(u,p)dG_b(u)$, $\Psi(b) \equiv \frac{\phi_{ab}}{\lambda g_a(a(b))}$, $\mathscr{W}(b) \equiv w''(L(b))L'(b)$, and

$$-\frac{\underline{b}}{\lambda\alpha_c a(\underline{b})^{\alpha_s}}\left(\mathscr{L}^L(\underline{b}) - \frac{1}{\kappa\sigma L(p)^{\sigma-1}}\right) - \kappa\sigma L(\underline{b})^{\sigma-1}\frac{dL(\underline{b})}{dT(p)} = 0, \forall p \in \mathscr{B}$$
$$F^U(\infty,p) = 0, \forall p \in \mathscr{B}$$
$$F^L(\underline{b},\underline{b}) + \frac{1}{\kappa\sigma L(\underline{b})^{\sigma-1}} = 0$$
$$F^U(b,b) - F^L(b,b) = 0, \forall b \in \mathscr{B}$$
$$f^U(b,b) - f^L(b,b) = 0, \forall b \in \mathscr{B}$$

Proof. See Appendix 2.8.1. □

In the appendix, we extend our continuum model with externalities and alternative government roles to compare key optimality conditions.

2.4.4 Equilibrium in Comparison

We use the following assumptions on the exogeneous productivity distribution, profit and local costs to analyze equilibrium. These assumptions facilitate tractability in the coordinated equilibrium, as will be clear in following propositions. We provide some extensions in the appendix.

Denote $\Theta \equiv \{\alpha_c, \alpha_s, \beta_c, \beta_s, \kappa, \lambda, \sigma\}$ the set of primitives.

Assumption 1. Assume the location TFP b and a firm' productivity a have the following distribution: $a \sim G_a = Pareto(\left(\frac{\alpha_c}{-1-\alpha_s}\right)^{\frac{1}{-1-\alpha_s}}, -1-\alpha_s)$, $b \sim G_b = Pareto(\left(\frac{\beta_c}{-1-\beta_s}\right)^{\frac{1}{-1-\beta_s}}, -1-\beta_s)$. Denote $a(b) \in [\underline{a}, \infty) \equiv \mathscr{A} = supp(G_a), b \neq$

$p \in [\underline{b}, \infty) \equiv \mathscr{B} = supp(G_b)$. For simplicity, shape parameters are denoted by $k_a \equiv \left(\frac{\alpha_c}{-1-\alpha_s}\right)^{\frac{1}{-1-\alpha_s}}$ and $k_b \equiv \left(\frac{\beta_c}{-1-\beta_s}\right)^{\frac{1}{-1-\beta_s}}$.A firm's profit features supermodularity between local TFP and own productivity: $\phi(a,b) = ab$. At last, the wage function has the following form (as Behrens et al. (2014) [6] and Gennaioli et al. (2013) [28]) where σ is the wage elasticity: $w(L) = \kappa L^{\sigma}$.

2.4.5 Equilibrium under Laissez-Faire

First, consider equilibrium size and productivity in laissez-faire, i.e., no tax levied by any government across locations.[18]

Laissez-faire economy. Equilibrium in laissez-faire is a productivity distribution and a city size distribution $\{a^*(b), L^*(b)\}$ such that taking city TFP b, city size $L(b)$ and local wage $w(L)$ as given, a firm with productivity a makes locational choice b to maximize profits: $\max_{b\in\mathscr{B}}\ ab - w(L(b)),\ \forall a \in \mathscr{A}$. And upon the firm's choice being made, local labor markets clear: $\int_b^{\infty} L(b)dG_b(b) = \lambda(1 - G(a(b))),\ \forall b \in \mathscr{B}$.

Lemma 2. Equilibrium allocations in laissez-faire have power form solution $L(b) = c_L^* b^{S_L^*}, a(b) = c_A^* b^{s_A^*}$, where equilibrium allocations on size and productivity have the following coefficients and exponents: $s_L^* = \frac{\alpha_s+\beta_s+2}{\alpha_s\sigma+\sigma-1}$, $s_A^* = \frac{\beta_s\sigma+\sigma+1}{\alpha_s\sigma+\sigma-1}$, and

$$c^*_{L,Laissez-faire} = \left(\frac{\alpha_c\lambda(\sigma s_L^* - 1)(\kappa\sigma s_L^*)^{\alpha_s+1}}{\beta_c}\right)^{\frac{1}{1-(\alpha_s+1)\sigma}} \equiv C_L^{LF}$$

$$c^*_{A,Laissez-faire} = \kappa\sigma s_L^*\left(\left(\frac{\alpha_c\lambda(\sigma s_L^* - 1)(\kappa\sigma s_L^*)^{\alpha_s+1}}{\beta_c}\right)^{\frac{1}{1-(\alpha_s+1)\sigma}}\right)^{\sigma} \equiv C_A^{LF}$$

Proof. See appendix. □

[18] A world with harmonized tax, an equal positive tax across locations, is a generalized version of this. As shown in the proof, tax harmonization will only change the entry margin, but not change the equilibrium allocation distribution at the top. For now we just focus on zero tax case to show the general intuition to compare different regimes. We will go back to tax harmonization in comparative statics in Section 2.6.

2.4.6 Equilibrium under Tax Competition

Given the equilibrium in uncooperative game equilibrium defined in Section 2.4.2, equilibrium conditions can be summarized as:

$$-\frac{T(b)}{w'(L(b))} + L(b) = 0, \forall b \in \mathscr{B} \tag{2.9}$$

$$a(b) - w'(L(b))L'(b) - T'(b) = 0, \forall b \in \mathscr{B} \tag{2.10}$$

$$L(b)g_b(b) - \lambda g_a(a(b))a'(b) = 0, \forall b \in \mathscr{B} \tag{2.11}$$

where the location game of firms satisfies:

$$\pi(a(\underline{b}), \underline{b}) = 0$$

$$\int_{\underline{b}}^{\bar{b}} L(b)dF(b) - \lambda\left(1 - G(a(\underline{b}))\right) = 0$$

$$\lim_{b\to\infty} a(b) = \infty$$

In addition, elasticities satisfy the system of integro-differential equations with a set of boundary value conditions in Lemma 1.

Before proceeding the equilibrium, we highlight a rationale behind local optimal taxes. As the policy condition indicates, in tax competition, local government will levy a tax proportional to local cost, i.e.,

$$T(b) \propto w'(L(b))L(b) \propto w(L)$$

The intuition is that local government maximizes the local revenue without taking into account their policy effect on other locations. Consider one unit increase in local tax $T(b)$, mechanical effect will just be $L(b)$ holding sizes fixed. At the same time, a local government knows that there will be firms leaving its city, which is determined by self-elasticity $\frac{1}{w'(L(b))}$. The only behavioral effect that matters for the local government comes from its own city. As will be shown in later sections, the cross-elasticities, which are essentially irrelevant for local governments, are key terms in the coordination equilibrium.

We summarize the uncooperative equilibrium with the following proposition.

Proposition 2. *(Equilibrium allocation in tax competition)* In a tax competition, the equilibrium size, productivity and tax are in the form of $L(b) =$

$c_L^* b^{S_L^*}$, $T(b) = c_T^* b^{s_T^*}$, $a(b) = c_A^* b^{s_A^*}$, where

$$\text{Size coefficient: } c_L^* = \left(\frac{\alpha_c \lambda (\sigma s_L - 1)(\kappa\sigma(\sigma+1)s_L)^{\alpha_s+1}}{\beta_c} \right)^{\frac{1}{1-(\alpha_s+1)\sigma}}$$

$$\text{Size shape: } s_L^* = \frac{\alpha_s + \beta_s + 2}{\alpha_s\sigma + \sigma - 1}$$

$$\text{Productivity coefficient: } c_A^* = \kappa\sigma(\sigma+1)s_L^* \left(\underbrace{\left(\frac{\alpha_c \lambda (\sigma s_L^* - 1)(\kappa\sigma(\sigma+1)s_L^*)^{\alpha_s+1}}{\beta_c} \right)^{\frac{1}{1-(\alpha_s+1)\sigma}}}_{c_L^*} \right)^{\sigma}$$

$$= (\sigma+1)^{\frac{1}{1-(\alpha_s+1)\sigma}} C_A^{LF} > 0$$

$$\text{Productivity shape: } s_A^* = \frac{\beta_s\sigma + \sigma + 1}{\alpha_s\sigma + \sigma - 1}$$

$$\text{Tax coefficient: } c_T^* = \kappa\sigma \left(\underbrace{\left(\frac{\alpha_c \lambda (\sigma s_L^* - 1)(\kappa\sigma(\sigma+1)s_L^*)^{\alpha_s+1}}{\beta_c} \right)^{\frac{1}{1-(\alpha_s+1)\sigma}}}_{c_L^*} \right)^{\sigma}$$

$$\text{Tax shape: } s_T^* = \frac{\alpha_s\sigma + \beta_s\sigma + 2\sigma}{\alpha_s\sigma + \sigma - 1}$$

where C_L^{LF}, C_A^{LF} are functions of fundamentals specified in Lemma 2.

Proof. See appendix. □

As Proposition 2 shows, size and productivity coefficients are proportional to that in Laissez-faire. In particular, size coefficient c_L^* dampens the laissez-faire constant C_L^{LF} while the productivity coefficient c_A^* is accelerated. This suggests that the city size decreases overall whereas the average productivity increases in these shrunk cities. The reason is that the cost is driven up in these locations by the tax. We will compare the uncooperative equilibium distribution with the centralized regime once we establish the coordinated results in the next section.

2.4.7 Equilibrium under Tax Coordination

Under tax coordination, location games (2.10 and 2.10) and elasticity conditions (as $E1$ and $E2$ in the appendix) remain the same, but policy conditions change to the following:

$$\underbrace{\int_{\underline{b}}^{p} \chi^L(u,p)du + \int_{p}^{\infty} \chi^U(u,p)du}_{\text{cross-behavioral effect}} - \underbrace{\frac{T(p)}{\kappa\sigma L(p)^{\sigma-1}}}_{\text{self-behavioral effect}} + \underbrace{L(p)}_{\text{mechanical effect}} = 0, \forall p \in \mathscr{B} \tag{2.12}$$

where $\chi^L(u,p) \equiv T(u)f^L(u,p)g_b(u)$ and $\chi^U(u,p) \equiv T(u)f^U(u,p)g_b(u)$, with integration range below and above p respectively. The difference between $P1$ and $P2$ are two integrals in the first two terms in $P2$, i.e., the "cross behavioral effects". They represent how a tax perturbation in location p will incur an inflow to all other locations. In this fiscal regime, the central government tradeoff between the gain with a higher tax in any arbitrary location and the *potential* loss from the inter-regional flow outcome triggered by such tax perturbation. We will illustrate the below p and above p distinction in the next section.

Different from policy optimality condition in competition, the coordinated equilibrium also consists elasticity functions. These are characterized by a system of partial integro-differential equations associated with a system of boundary value conditions in both Fredholm and Volterra types. And we find analytical solution for cooperative equilibrium, as summarized in the following proposition.

Proposition 3. *(Equilibrium allocation in coordination)* Equilibrium allocation features Pareto distribution. Specifically, size, productivity feature same shape paratmers as in competition, while scale parameters are as functions of ρ.

$$\text{Size: } c_L^*(\rho) = \left(\frac{\alpha_c\lambda(\sigma s_L - 1)(\kappa\sigma(\rho\sigma+1)s_L)^{\alpha_s+1}}{\beta_c}\right)^{\frac{1}{1-(\alpha_s+1)\sigma}} = (\rho\sigma+1)^{\frac{\alpha_s+1}{1-(\alpha_s+1)\sigma}} C_L^{LF}$$

$$s_L^* = \frac{\alpha_s + \beta_s + 2}{\alpha_s\sigma + \sigma - 1}$$

$$\text{Productivity } c_A^*(\rho) = \kappa\sigma(\rho\sigma+1)s_L^* c_L^*(\rho)^{\sigma} = (\rho\sigma+1)^{\frac{1}{1-(\alpha_s+1)\sigma}} C_A^{LF}$$

$$s_A^* = \frac{\beta_s\sigma + \sigma + 1}{\alpha_s\sigma + \sigma - 1}$$

$$\text{Tax: } c_T^*(\rho) = \rho\kappa\sigma c_L^*(\rho)^{\sigma}$$

$$s_T^* = \frac{\alpha_s\sigma + \beta_s\sigma + 2\sigma}{\alpha_s\sigma + \sigma - 1}$$

where $\rho^* = \frac{\alpha_s\sigma+\sigma-1}{(\alpha_s+2)\sigma}$ and $t^* = \frac{\frac{c_T r_1 \underline{b}^{s_T+r_1}}{s_T+r_1} + \frac{c_T r_2 \underline{b}^{s_T+r_2}}{s_T+r_2}}{\underline{b}^{r_1} + \frac{\underline{b}^{r_2}}{\zeta}}$. Elasticities $\{f^U(b,p), f^L(b,p)\}$

are the following:

$$f^U(b,p) = \frac{r_2 c_L^{1-\sigma} b^{-\beta_s + r_2 - 1} (r_2 p^{r_2} + r_1 \zeta p^{r_1}) p^{-r_1 - r_2 - \sigma s_L + s_L}}{\beta_c \kappa \sigma \zeta (r_2 - r_1)}, \ \forall p > b \in \mathscr{B}$$
$$f^L(b,p) = \frac{r_2 b^{-\beta_s - 1} c_L^{1-\sigma} (r_2 b^{r_2} + r_1 \zeta b^{r_1}) p^{-r_1 - \sigma s_L + s_L}}{\beta_c \kappa \sigma \zeta (r_2 - r_1)}, \ \forall p < b \in \mathscr{B}$$

and $r_1 = \frac{\theta_2 - \theta_1 + \sqrt{(\theta_1 - \theta_2)^2 + 4\theta_2}}{2\theta_2}$ and $r_2 = \frac{\theta_2 - \theta_1 - \sqrt{(\theta_1 - \theta_2)^2 + 4\theta_2}}{2\theta_2}$ where $\theta_1 \equiv \frac{\lambda \kappa \sigma \alpha_c c_A^{\alpha_s} c_L^{\sigma-1}(s_L(\sigma-1) - \beta_s)}{\beta_c}$, $\theta_2 \equiv \frac{\lambda \kappa \sigma \alpha_c c_A^{\alpha_s} c_L^{\sigma-1}}{\beta_c}$, $\zeta \equiv -\left(\frac{\frac{B}{A} \frac{r_2}{\beta_c} \underline{b}^{r_2 - \beta_s - 1} - \underline{b}^{r_2}}{\frac{B}{A} \frac{r_1}{\beta_c} \underline{b}^{r_1 - \beta_s - 1} - \underline{b}^{r_1}} \right)$, $B \equiv \kappa \sigma c_L^{\sigma-1} \underline{b}^{s_L(\sigma-1)}$ and $A \equiv \frac{\underline{b}}{\lambda \alpha_c c_A^{\alpha_s} \underline{b}^{s_A \alpha_s}}$. Note that c_L^*, c_A^*, c_T^* are different that in competition.

Proof. See appendix. □

Proposition 3 indicates that size and productivity distribution have coefficients proportional to that in laissez-faire. We will explore properties in coordination in the next section.

2.4.8 Elasticity

In this section, we discuss properties to understand the general equilibrium effects, which facilitates us to derive optimal policies in the coordinated regime in Section 2.5.

When there is a tax perturbation in city p, denoted by $\Delta T(p)$, it first affects firms locational choices in that city and around it.[19] The optimal locational choice will then feed back into cities around the perturbation and propagate to further ones in $\mathscr{B}$, followed by the adjustment of policies. Within the sorting environment, accordingly, behavioral effects in all cities corresponding to a tax perturbation in a specific city can be characterized by a curve with a kink at the purturbation city. Figure 2.3 illustrates examples of behavioral effects corresponding to two cities: when tax perturbation city is at 25 perecentile in terms of ex ante TFP level in whole distribution (blue solid curve) and when tax perturbation city is at 75 percentile (red dashed curve). This is an example when ex ante TFP follows uniform distribution. Take when $p = 1.25$ as an example, when city $p = 1.25$ increase one unit of tax, cities around 1.25 see larger size increase than that in cities far away, i.e., it decays from the tax perturbation city to two sides- cites with TFP above 1.25 and below 1.25. The intuition behind this is that when tax changes in a specific city, nearby cities (in terms of ex ante TFP) changes more than cities far away. For example, when city of Pittsburgh changes tax, it affects cities such as Boston more than cities such as Yuma. This result comes from positive sorting and it also holds in discrete location case.

[19] We illustrate the group perturbations with smaller and larger measures in the Appendix, see Figure 2.18 and 2.19.

Similarly, Figure 2.4 illustrates examples of behavioral effects corresponding to three tax perturbation cities with p at the 10, 50, and 90 percentiles respectively, when ex ante TFP follows Pareto distribution.

At the same time, different cities have heterogeneous distributional effects on all other cities. For example, when the City of Pittsburgh changes its tax, it has a different behavioral effect from all other cities compared to the case when New York City changes its tax. Thus we see behavioral effect curves are different among different tax perturbation cities. In both uniform and Pareto cases, we could analytically show that the response surfaces are different when conducting tax perturbations in different cities.

Figure 2.3: An Illustration of Behavioral Effects Corresponding to Two Example Tax Perturbations ($b \sim Uniform$)

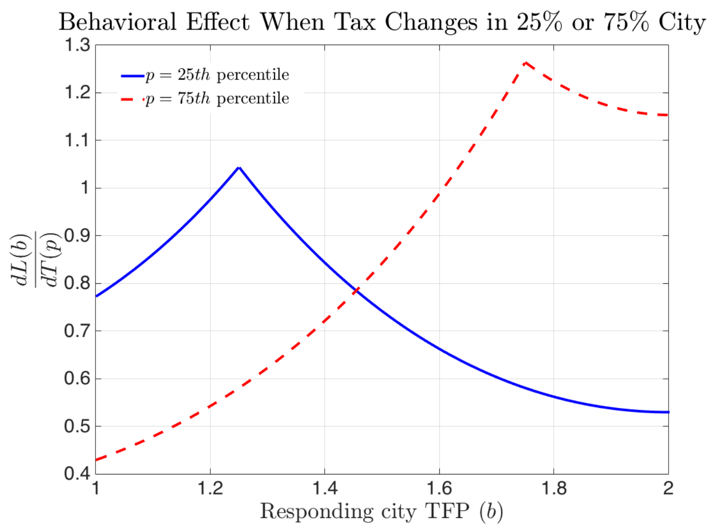
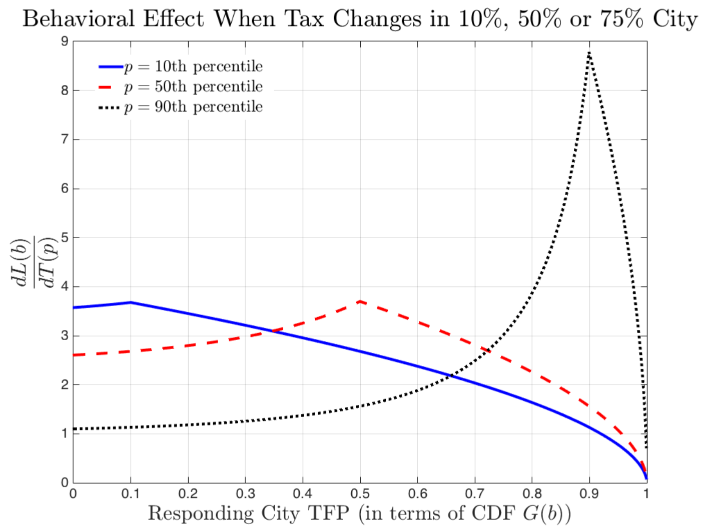


Figure 2.4: An Illustration of Behavioral Effects Corresponding to Three Example Tax Perturbations ($b \sim Pareto$)



2.5 Cooperative and Uncooperative Equilibrium

In this section, we compare some properties of cooperative and uncooperative equilibrium outcomes and provide intuitions to understand the government's behavior. We start with the following lemma for distributional effects.

Lemma 3. Shape of productivity distribution is determined by relative heterogeneity between firm and location and extent of congestion cost in all regimes. ($s^*(\sigma, G_a, G_b)$). Firms' location decisions at the top remain the same regardless of tax competition and coordination regimes.

Proof. See appendix. □

Lemma 3 indicates that the tax competition hurts firms more at the bottom. This is an intuitive result in the sense that most productive firms are not really driven by tax competition since they are capable, and would have the motivation when conditions permit, to enter most productive area anyways. Our model suggests that what matter to these productive ones are relative hetergeneity between firm and city TFP, as well as how extent of congestion cost dampens this relative term. In all three regimes, these three primitives will directly affect concentration pattern of firms at the top. This will is also related to the following proposition about inequality.

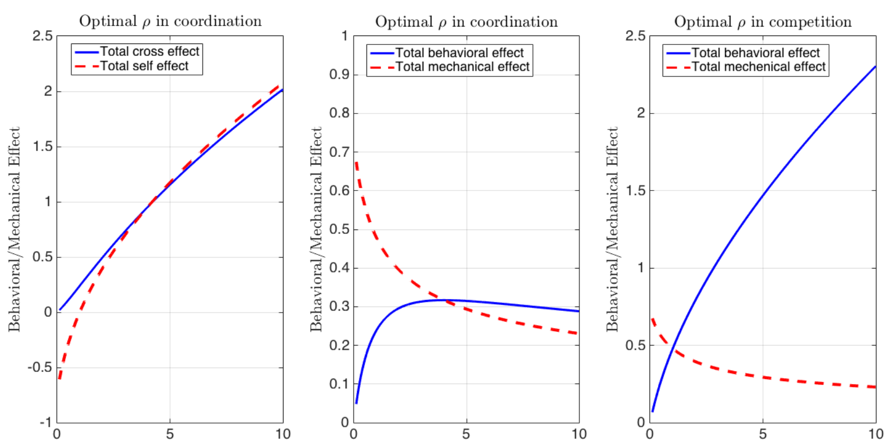
Proposition 4. *(Equilibrium Tax Comparison)* In tax coordination, a central government will increase taxes across all locations, compared to uncooperative regime, i.e., $T(b)_{coord} > T(b)_{comp}$, $\forall b \in \mathscr{B}$.

Proof. See appendix. □

The above difference comes from the fact that the central government takes into account all cross behavioral effects, whereas local governments do not. In particular, consider comparing policy condition in two regimes. To maximize the total revenue, a social planner internalizes a full set of elasticities. With a unit increase in local tax in p, fixing size $L(p)$, there will be a $L(p)$ revenue increase mechanically. However, the central government also realizes a $\frac{T(p)}{w'(L(p))}$ decrease in revenue from local behavioral effect– firms exiting this area, and an increase in revenue from cross behavioral effect– firms entering other areas, thus varying (BE_C^U, BE_C^L).

Now we turn to the difference of ρ in two policy regimes in Figure 2.5. Consider a tax increase in tax perturbation city p. The left subplot shows when total revenue benefit from extra inflow in other locations ("Total cross effect"$=BE_C^L + BE_C^U$) dominates that in tax perturbation city ("Total self effect"$= -BE_S + ME$). The middle subplots shows when total behavioral effect-including the loss from tax perturbation city to other cities plus the gain from inflows in other cities ("Total behavioral effect"$= BE_C^L + BE_C^U - BE_S$) dominates that of total mechanical effect in perturbation city ("Total mechanical effect"$= ME$). The right subplot shows that, in competition, local governments equalize self-behavioral effect ("Total behavioral effect"$-BE_S$) with self-mechanical effect ("Total mechanical effect"$= ME$). In competition, local governments do not take into account of how firms leaving the perturbation city increases revenue in other cities on average.

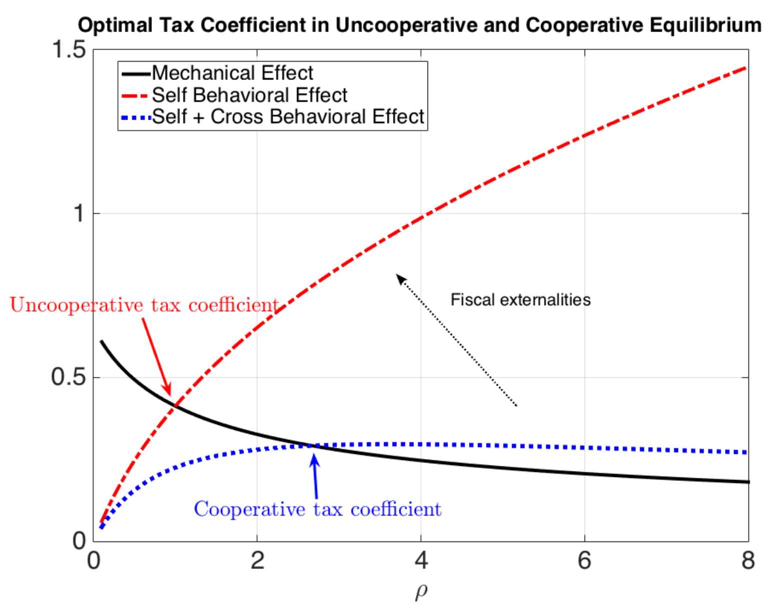
Figure 2.5: Equilibrium Tax Coefficient (ρ) in Two Regimes



When ρ increases, self mechanical effect increases together with the general equilibrium effects on work, thus revenue in a menu of locations increases. This effect is taken into consideration by the planner, reflected by how the optimal ρ shifts to the right in the left and middle plots. ρ^* under coordination is the one in which all general equilibrium effects are embededed, indicating marginal beneficiary in other cities from tax perturbation. In other words, the planner is able to utilize the cross-effects when maximizing total revenue by levying high taxes in a menu of locations.

Why can taxes necessarily be set higher in the coordination? When a planner takes into account all cross-elasticities, benefits from behavioral effects are extended up until a point when all general equilibrium effects entered in optimality condition are exhausted. With same self-BE and ME, what determines the tax differential between two regimes is how total cross-BE pins down the tax coefficient ρ. This is shown in middle and right subplot in Figure 2.5. If we combine the middle and the right plot, we can show how behavioral effects are dampend in the existence of fiscal externalities. In Figure 2.6, we plot the self-BE curve and total BE curve and compare the equilibrium ρ in one plot. The cooperative tax coefficient ρ is determined by the intersection of total BE and ME, highlighted by the blue arrow. This plot highlights the *extent of fiscal externalities* by the difference between the red dashed curve and blue dotted curve.

Figure 2.6: Equilibrium Tax Coefficient in Two Regimes and the Extent of Fiscal Externalities



General equilibrium effects enter all terms in the coordinated condition 2.12. The intuition is that the central government knows when increasing tax in an arbitrary location in the open interval, revenues from other locations will increase (two cross-behavioral effects from the lower and upper parts). To maximize the total revenue, a planner will take advantage of its power to increase taxes in all locations. Fixing city sizes, this first decreases the productivity of firms that could afford to any given city. On the other hand, market-clearing induces a decrease in the number of firms in all cities, including the most productive ones. However, since only the most productive ones will stay in the market, the higher tax levied on them will balance off the lower cost (because of the adjusting scales). This further prevents less productive firms to enter. In other words, the central government is able to manage city sizes and the city distribution to be relatively small in general and with less dispersion. The only equilibrium to sustain this is to leave the most productive firms to stay in the economy. Since cities are smaller, the city distribution shifts up at the bottom.[20] Therefore, we observe that the productivity increases in all cities, and thus a high-productivity economy on average.[21]

A central government's tradeoff between cross-behavioral effects and self-behavioral effects is characterized in the elasticity function. Figure 2.7 shows that elasticity surface shifts up compared to that in competition (with associated contour plots shown in Figure 2.8.). With higher cross-elasticity, with any tax perturbation, a central government can gain more when firms are more elastic to move. In other words, the way that a government can earn high profits across locations is to leave firms who can afford such tax in the economy and take advantage of the high behavioral effects.[22]

---

[20]This shift is driven by a change not in the shape, but the scale paramter of the Pareto distribution.

[21]This is reflected by a upward shifts in productivity distribution in Figure 2.9

[22]Note that sorting effect is a key to guarantee such coordination strategy. This is because when increasing $T(p)$ in any arbitrary location, the only way to have firms not exiting the market, given that other places also have high tax, is that all firms in the market are most productive ones. In the coordinated regime, the planner levys high taxes across locations and can effectively uses the entry margin to control who are excluded.

Figure 2.7: Comparision of Elasticity Surface in Two Regimes

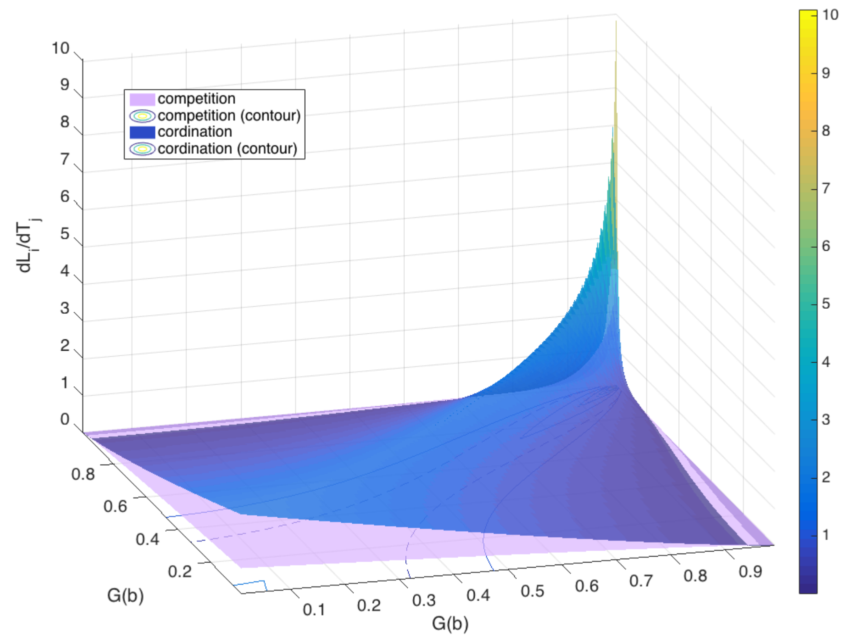
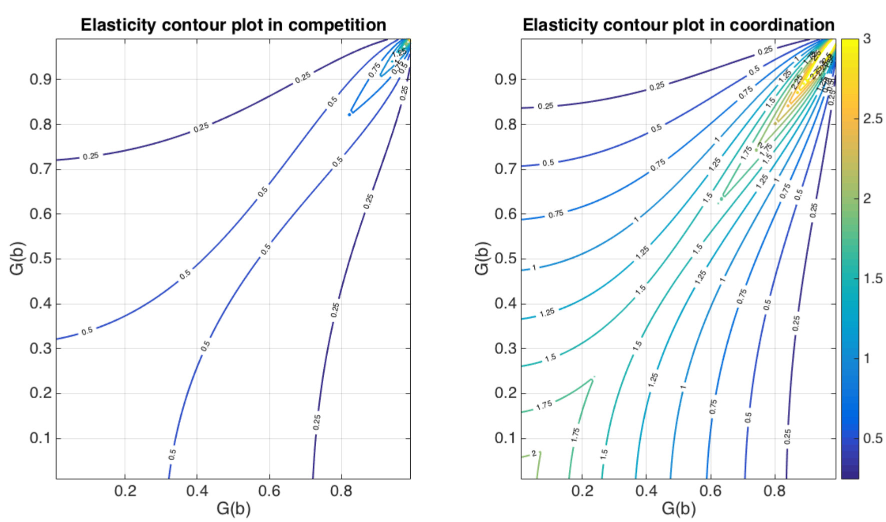


Figure 2.8: Comparison of Contour Plots



We now turn to policy effects on equilibrium variables, summarized in the following lemma. Note that due to firms' production function, city ex post TFP is equivalent to $a(b)b \equiv A(b)$.

Lemma 4. Under coordination, equilibrium city size distribution is first order stochastic dominated by that in competition. With lower city taxes in competition ($T^*(b)^{comp} < T^*(b)^{coord}, \forall b \in \mathscr{B}$), city sizes in competition are larger than that in coordination; city-level TFP in competition is lower than that in coordination. Size dispersion in terms of range is larger and productivity dispersion is smaller in competition.

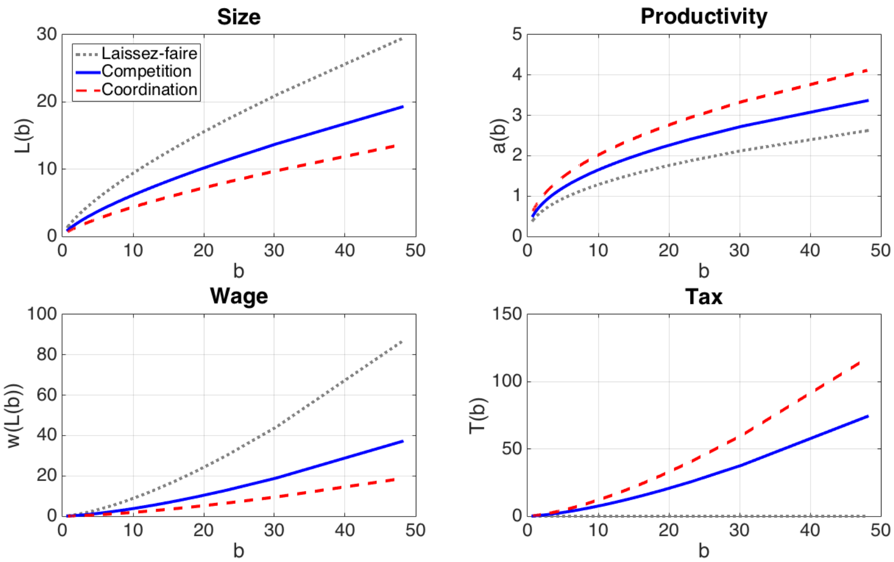
$$A^*(b)^{comp} < A^*(b)^{coord}, \ \forall b \in \mathscr{B}$$
$$L^*(b)^{comp} > L^*(b)^{coord}, \ \forall b \in \mathscr{B}$$

Proof. See appendix. □

The intuition of the above lemma is the following. In competition, all local governments set lower tax to lure firms. This is consistent to standard tax competition results. This is effectively decreasing local fixed costs for all firms in all cities. Consequently, fixing a city b, we see less productive firms can afford to produce in this city. At the top, since positive sorting still holds, with such race-to-the-bottom triggered by tax competition, all firms are now "upgrading" ex ante higher TFP cities. At the bottom, the entry margin is driven down. In other words, tax competition allows least productive firms who could not afford to enter the market at the beginning to enter. The above process reflects how taxes affect firms' optimal locational choices. This will then feed back into the size distribution. Sizes across locaitons will increase until the decreased tax cost is balanced off by increasing city size, thus increasing congestion and inducing fiscal externalities.

With the above mechanism, wage and city size distributions differ between regimes due to general equilibrium effects. The reason why a central government can set a high tax schedule across locations in coordination is that it can control city size distribution and entry margin. In Figure 2.9, we illustrate how size, productivity, wage and tax as a function of ex ante local TFP changes when switching regimes. Functional form assumptions induce equilibrium tax in coordination also proportional to wage, but with a higher coefficient.

Figure 2.9: Equilibrium Allocations in Three Regimes



Therefore, Proposition 3 and Proposition 4 indicate that, in competition, city sizes are relatively bigger, whereas local TFPs are lower (i.e., firms are less productive) across all cities. Because the Pareto shifts in scale parameter, this means that the aggregate TFP decreases. In tax competition, with lower taxes across locations, governments are effectively creating a divergence across locations vis-à-vis productivity. With lower taxes, the least productive firms could enter the economy and cities become inefficiently larger than that in coordination. In a constrained policy regime where at least the jurisdictional power is at work to some extent, and when less productive locations attract firms, this could be the case that cities are controlled to be smaller and the least productive firms are exluded outside of the market.[23]

2.6 Comparative Statics

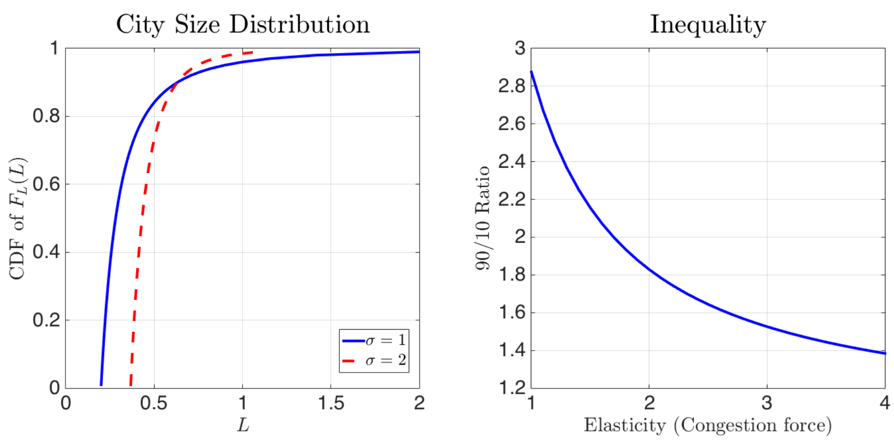
2.6.1 Inequality under Tax Competition

We start by exploring how a shock in the extent of congestion drives changes in size inequality. Under a competitive regime, the size inequality increases when

[23]Examples of group perturbations are shown in the Appendix Figure 2.18 and 2.19. Also note that the most productive firms always enter the market and stay at high TFP cities. What matters for them is how large city they end up being with a specific slope of city size distribution and how the size differential is relative to the tax differential.

the wage elasticity [24] of labor supply decreases. This is because when the wage elasticity is low, firms are able to afford to locate themselves in ex ante higher TFP cities. At the same time, the entry margin is driven down. As in tax effect, a decreasing congestion exponent is balanced by the increasing size in distribution. In the left plot in in Figure 2.10, we present a case where σ increases from 1 to 2, the city size distribution (cdf) has a thinner tail, i.e., due to an economy-wise increase in wage cost, small cities at the bottom are bigger, while big cities at the top are smaller. This induces a direct shrink in inequality. Meanwhile, the right plot in in Figure 2.10 shows how inequality, measured by a 90/10 ratio in city size distribution, decreases with a continuous increase in wage elasticity (σ).

Figure 2.10: Comparative Statics to City Size Distribution and Inequality with respect to Wage Elasticity (σ)

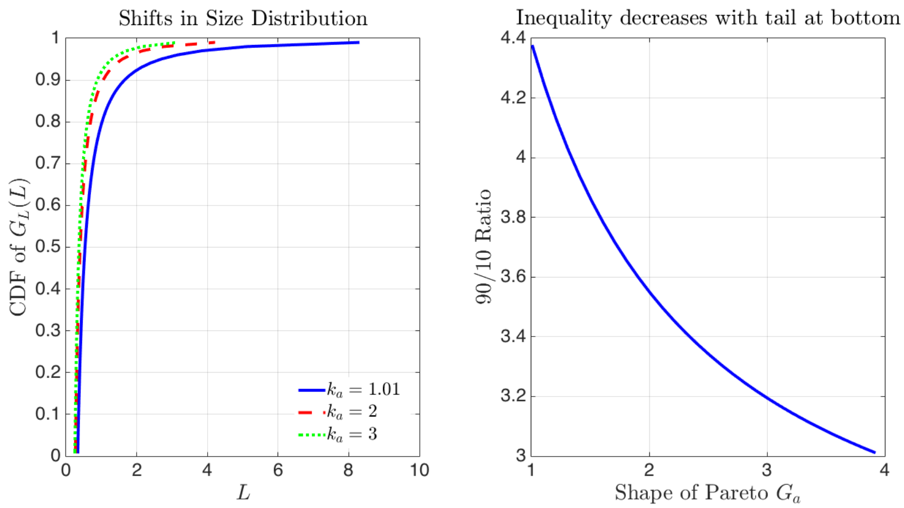


Then, we conduct experiments on shapes of firms productivity distribution. Under a competitive regime, the size inequality decreases when the firms' productivity distribution have thinner tails. This is because, with a thinner tail, there are fewer superstar firms at the top. Therefore, there are less firms to effectively form mega cities in equilibrium, i.e., the city size inequality decreases. As shown in Figure 2.11, in the left subplot, when k_a increases, i.e., the density of firms' productivity distirbution skews to the left, city size distribution moves toward the left. This means large cities cannot be particularly large compared to the case when more productive firms exist. The leftward shift, together with the shrink in support, result in a decrease in the city-size inequality, as shown in the right subplot (90/10 ratio). This comparative statics is helpful to understand what may drive size and income inequality when there is a structural change in the firms' productivity distribution. Recent literature has discussed

[24] More generally, it can be regarded as the extent of congestion elasticity. Existing studies discuss possible sources as in Duranton and Turner (2018) [70] and Hsieh and Moretti (2019) [71].

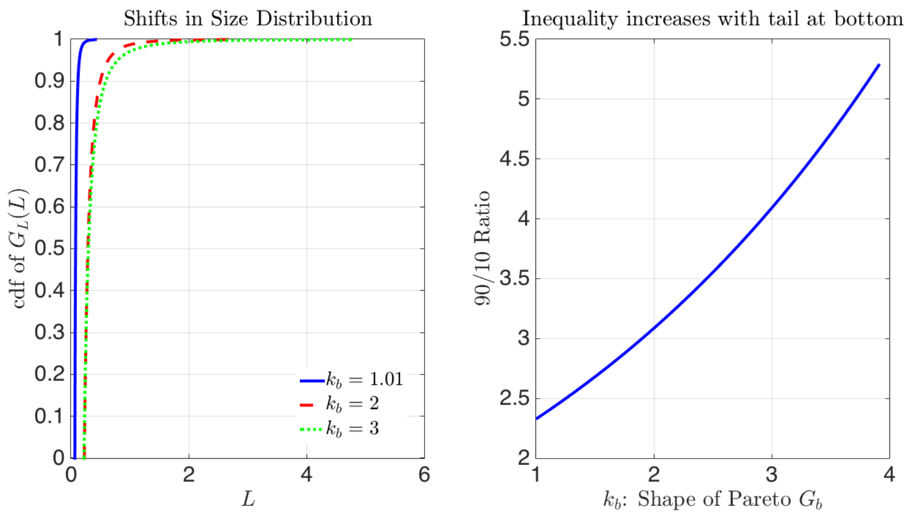
such changes. One example is the rise of superstar firms, as studied by Autor et al. (2020) [72].

Figure 2.11: Comparative statics with respect to shape parameter of firm productivity distribution (k_a)



At last, we also consider varying the shape of city ex ante TFP distribution. Under the competitive regime, the size inequality increases when the city ex ante TFP distribution has thinner tails. With a thinner TFP tail at the top, there are fewer mega cities in the economy. Conditional on the same firm productivity distribution, the lower number of these cities determines the fact that more productive firms crowd into the limited set of high TFP cities. As a result, inequality is boosted. In the left subplot in Figure 2.12, we observe that when k_b increases, i.e., the density of the city TFP distribution has a thinner tail at the bottom, city size distribution is stretched horizontally and shifts to the right. This means there are very large cities formed compared to the case where there are many ex ante high TFP cities. Therefore, as we see in the right subplot, with a thinner tail in TFP distribution, i.e., larger k_b, the inequality increases. Estimating local TFP distribution is challenging in general. This comparative statics reveals that how exogenous shocks in TFP distribution could drive inequality. One handy property of the equilibrium under the Pareto case is that we can theoretically identify the aggregate TFP loss and show the independence between its level and the ex ante city TFP distribution, as summarized in Lemma 6.

Figure 2.12: Comparative statics with respect to shape parameter of city ex ante TFP distribution (k_b)



2.6.2 The Cost of Tax Competition

We now study the driving factors of the inefficiency created by uncooperative game, and compare the uncoorperative policy regime with the coorperative one.

Lemma 5. Competition equilibrium is closer to that in coordination when the extent of congestion cost (σ) is higher and productivity distribution on firm side features thinner tail (large $|\alpha_s|$).

Proof. See appendix. □

When the congestion elasticity is high, regardless of competition and coordination regimes, only the most productive firms can afford to enter. It is as if the power of policies is dominated by the entry margin. On the other hand, when there are less *ex ante* heterogeneity in terms of firms productivity in the economy, the discrepancy between tax competition and coordination diminishes. This is beacause coordination results in an exclusion to relatively high productive firms staying in the market. When firms productivity has a fatter tail at the bottom, the competition cannot induce much difference by lowering tax across the economy with only firms with advantages left. In other words, in both cases (high σ or α_s), effects of taxes on the sorting mechanism are dampened since governments confront a greater challenge when distorting productivity distribution. The comparison of the elasticity between competition and coordination is shown in Figure 2.13, and the contour plots in the two regimes are shown in Figure 2.14.[25]

[25]The double-side Pareto assumption induces tractablility in the nonlinear cost model. In the

Figure 2.13: Comparision of Elasticity Surface between Competition and Coordination

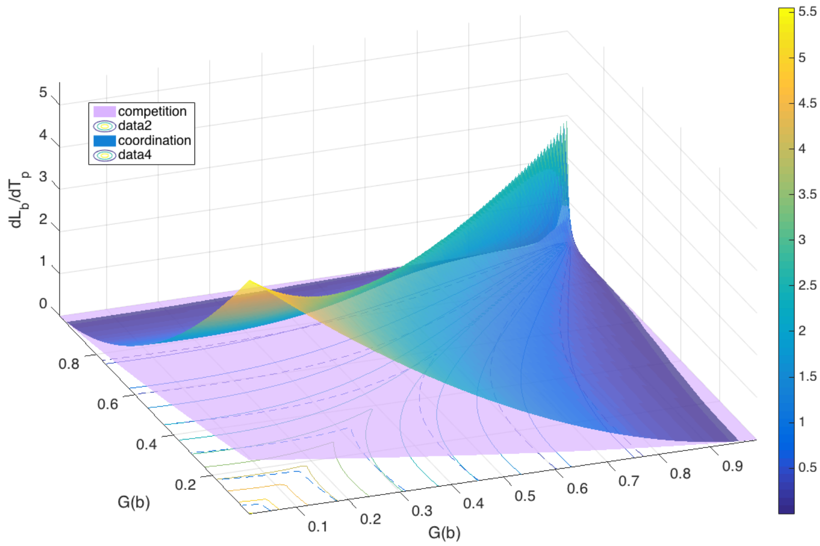
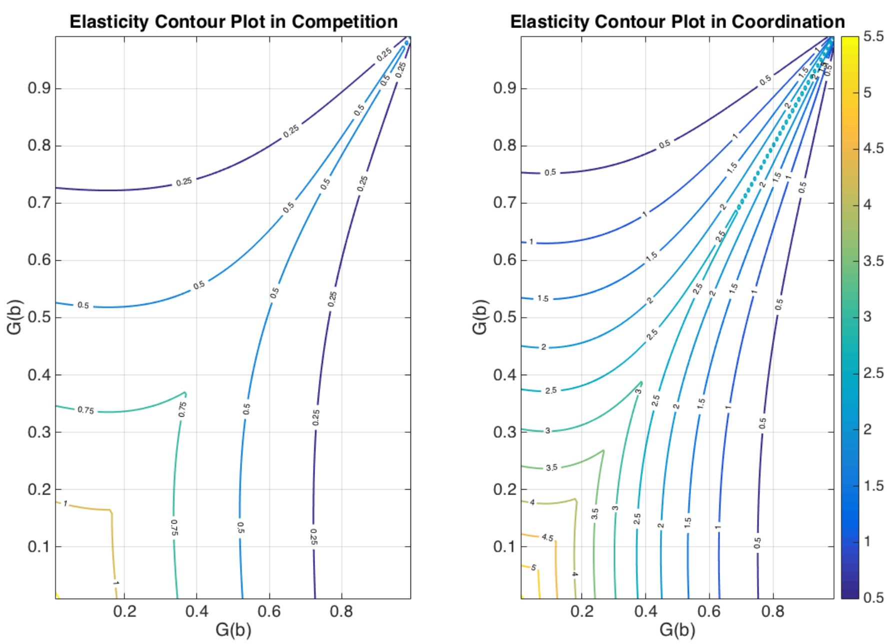


Figure 2.14: Comparison of Contour Plots



Appendix, we also provide analysis with elasticity illustrated when we assume linear cost with uniform TFP and firm productivity distribution. The decay pattern of the elasticity surface still holds.

Lastly, we analyze the aggregate cost of tax competition in terms of the aggregate TFP loss. We first characterize it with the following lemma.

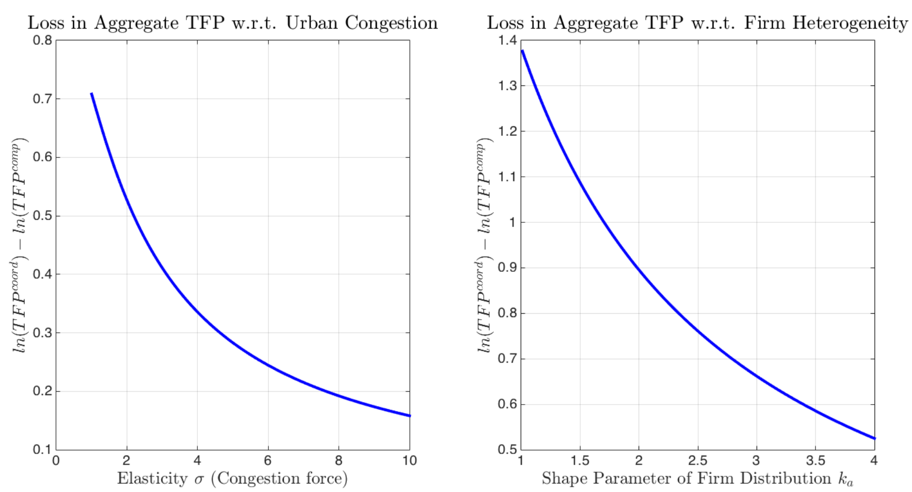
Lemma 6. In equilibrium, the aggregate TFP in a policy regime is $TFP = \int_{\mathscr{B}} a(b)bdG_b(b)$. The loss in the aggregate TFP from tax competition is:

$$\text{TFP Loss} = \frac{1}{1-(\alpha_s+1)\sigma} ln\left(\frac{\rho\sigma+1}{\sigma+1}\right)$$

Proof. See Appendix 6. □

The mechanism behind this identification is the dampening effect of the tax competition on local and aggregate productivity ($a(b)$) with a coefficient shifter in the *ex post* Pareto distribution so that firms at the top are not affected effectively.[26] Figure 2.15 shows the comparative statics of the loss of tax competition with respect to congestion elasticity and firm heterogeneity.

Figure 2.15: Comparative statics of the loss of tax competition



The extent of concentration of the productivity distribution is determined by the relative heterogeneity between the firm and the location as well as by the extent of the congestion cost. The fact that firms' spatial decisions at the top are not affected by regimes is what policymakers can leverage when desiging fiscal policies to diminish the aggregate costs of the existence of decentralized competition.

The above analyses shed light on ways for policymakers to design fiscal policy regimes. The first is about choosing a proper fiscal policy regime. Our results

[26] We do not need the information on the specific ex ante distribution of city TFP to pin down the aggregate loss. This is an advantage of our model specifications.

suggest that, to lower the aggregate costs, tax competition induces a higher cost when the congestion elasticity is lower (less responsive to wage differentials), or the shape parameter is smaller. If there are institutional constraints to switch to a fully centralized fiscal regime, a centralized fiscal policy regime is more beneficial to the aggregate economy when the economy is closer to the above circumstances. The second takeaway is about the distributional effects of a decentralized fiscal regime. Our results show that the existence of tax competition reduces aggregate TFP and city-level TFP, where big cities are affected more than small cities vis-à-vis the level of size and productivity. In a decentralized regime, the size inequality is higher when the wage elasticity is lower, when there are more superstar firms, or when there are fewer mega cities.

2.7 Conclusion

This paper studies the macroeconomic effects of the existence of tax competition. We develop a new theory to study how uncooperative local governments use local tax incentives to lure heterogeneous firms. In equilibrium, productive firms enter ex ante productive and ex post larger cities, which is consistent with the urban premium. In our model, inequality can be driven by exogenous shocks, such as an urban congestion force, or a shock in the ex ante distribution on both firms' productivity and city TFPs.

This paper contributes to the current literature on both spatial production and tax competition by analyzing how different policy regimes interact with endogenous locational choice in firm sorting. We analytically characterize cross-behavioral effects in the general equilibrium with a heterogeneous-agent and continuum-space model, which has not been fully analyzed before. This cross-effect is a key endogenous term in understanding the difference between coordinated and uncoordinated policy, and how conventional fiscal externalities can be accelerated by general equilibrium effects.

Through the lens of the model, we show how firms' locational choices are distorted in an uncoordinated policy regime compared to a coordinated one. Equilibrium taxes under competition is lower than that in coordination, which is consistent with a race to the bottom, widely documented empirically in literature. With the theoretical framework, we find that the existence of tax competition dampens aggregate TFP and city-level TFPs. With double-side Pareto distributions, big cities are affected more vis-à-vis the level of size and productivity. Regarding the aggregate costs of the tax competition, we find that the competitive equilibrium converges to a planner solution when urban congestion cost is higher or the firm productivity distribution features a thinner tail. When the economywise congestion cost elasticity is higher, the cost of tax competition is lower. Thus, the ability of a planner to manage the entry margin vanishes compared to that in competition. Moreover, we find that productive firms respond to policy perturbations more sensitively.

The policy takeaways from this paper are twofold. First, policymakers should carefully implement a fiscal regime and consider the aggregate cost of a decentralized fiscal regime determined by country circumstances. This paper provides some key factors to consider in a tractable way. Second, conditional on a decentralized fiscal regime, for example, under structural constraints to moving to a coordinated regime, there are distributional effects of an interjurisdictional tax competition. Therefore, policy regimes could accelerate the stagnation in productivity growth within most firms and/or dampen a minority of high-performers, or superstar firms.

Our study can also be used as a benchmark for both theoretical and empirical analysis. With the growing literature on spatial decisions and firm sorting (e.g., Eeckout and Guner, 2017 [73]), our mechanism will help researchers and policymakers to consider the implications of fiscal regimes with either endogenous or exogenous policies. Distortion in firm sorting through tax competition indicates that previous estimations could be subject to caveats. In the future, it would be interesting to generalize our current theory and think about quantitative applications.

2.8 Appendix

2.8.1 Proof of Propostition 1 and Lemma 1

Proof. We start from discrete conditions with tax perturbation (2.5)-(2.8). Equivalently, we have the following equations featuring how an arbitrary cutoff productivity responds to a tax perturbation in an arbitrary city $\frac{\partial a_p}{\partial T_r}$. This behavioral response is a function of size of all cities with TFP below the tax perturbation city. Summation comes from general equilibrium effect, representing how this response is accumulated from all cities with higher TFP than the optimal choice of this cutoff firm.

$$-\Lambda_n g\left(a_p\right)\frac{\partial a_p}{\partial T_r} = \sum_{i=0}^{n(1-p)} \frac{\partial L_{p+i\Delta}}{\partial T_r}$$

Therefore, we summarize both optimality and market clearing conditions in the following system of linear equations which determines the Jacobian matrix

$\left(\frac{\partial L_p}{\partial T_r}\right)_{p,r\in\{\frac{1}{n},\cdots,1\}}$ in discrete location model.

$$-\sum_{i=0}^{n(1-p)}\frac{\partial L_{p+i\Delta}}{\partial T_r}\frac{\phi_a\left(a_p,b_p\right)-\phi_a\left(a_p,b_{p-\Delta}\right)}{\Lambda_n g\left(a_p\right)}-w'\left(L_p\right)\frac{\partial L_p}{\partial T_r}+w'\left(L_{p-\Delta}\right)\frac{\partial L_{p-\Delta}}{\partial T_r}=\mathbf{1}\left[r=p\right]-\mathbf{1}\left[r=p+\Delta\right] \tag{2.13}$$

$$-\frac{\phi_a\left(a_\Delta,b_\Delta\right)}{\Lambda_n g\left(a_\Delta\right)}\sum_{i=1}^{n}\frac{\partial L_{i\Delta}}{\partial T_r}-w'\left(L_\Delta\right)\frac{\partial L_\Delta}{\partial T_r}=\mathbf{1}\left[\Delta=r\right] \tag{2.14}$$

Denote self-elasticity and cross-elasticity in n city case $\left(\frac{\partial L(p)}{\partial T(p)}\right)^{(n)}$ and $\left(\frac{\partial L(p)}{\partial T(p')}\right)^{(n)}$. We conjecture the following on self- and cross-elasticity in the limit:

$$\lim_{n\to\infty}\left(\frac{\partial L(p)}{\partial T(p)}\right)^{(n)}=-\frac{1}{w'(L(p))},\ \forall p\in\mathscr{I}$$
$$\lim_{n\to\infty}\left(\frac{\partial L(p)}{\partial T(p')}\right)^{(n)}=0,\ \forall p\neq p'\in\mathscr{I}$$
$$\lim_{n\to\infty}\left(n\frac{\partial L(p)}{\partial T(p')}\right)^{(n)}=\frac{dL(p)}{dT(p')},\ \forall p\neq p'\in\mathscr{I}$$

To see this, if we divide equation (2.5) by Δ and consider $\Delta\to 0$, at $r=\Delta$, the third term of left hand side of (2.8) will be infinity, while the only way to have equation hold is to have self-derivative $\frac{\partial L}{\partial T}$ to be a constant as well. Thus, to have self-derivative cancel the blowing-up Dirac function at $r=\Delta$, this constant must be $\frac{1}{-w'(L(p))}$. The idea is the same for equation (2.6) at $r=p$ or $r=p+\Delta$.

Then from equation (2.14), we take limit when $\Delta\equiv\frac{1}{n}\to 0$ as $n\to\infty$, $\forall r\neq\Delta$:

$$\lim_{\Delta\to 0}\frac{(2.14)}{\Delta}=-\frac{\phi_a(a_\Delta,b_\Delta)}{\lambda g(a_\Delta)}\left(\int_0^1\frac{dL_{\hat{p}}}{dT_r}d\hat{p}-\frac{1}{w'(L_r)}\right)-w'(L_0)\frac{\partial L_\Delta}{\partial T_r}$$

Similarly, from equation (2.13), consider an arbitrary $p\neq r$, we have:

$$\lim_{\Delta\to 0}\frac{(2.13)}{\Delta^2}=\lim_{\Delta\to 0}\left(-\frac{\phi_a(a_p,b_p)-\phi(a_p,b_{p-\Delta})}{\lambda n\Delta^2 g(a_p)}\sum_{i=0}^{n(1-p)}\frac{\partial L_{p+i\Delta}}{\partial T_r}-\frac{w'(L_p)\frac{\partial L_p}{\partial T_r}-w'(L_{p-\Delta})\frac{\partial L_p}{\partial T_r}}{\Delta^2}\right)$$
$$(\forall p>r)=-\frac{\phi_{ab}(a_p,b_p)b'_p}{\lambda g(a_p)}\int_p^1\frac{dL_{\hat{p}}}{dT_r}d\hat{p}-w''(L_p)L'_p\frac{dL_p}{dT_r}-w'(L_p)\frac{d}{dp}\frac{dL_p}{dT_r}$$
$$(\forall p<r)=-\left[-\frac{1}{w'\left(L_r\right)}+\int_p^1\frac{dL_{\hat{p}}}{dT_r}d\hat{p}\right]\frac{\phi_{ab}\left(a_p,b_p\right)b'_p}{\lambda g\left(a_p\right)}-w''\left(L_p\right)L'_p\frac{dL_p}{dT_r}-w'\left(L_p\right)\frac{d}{dp}\frac{dL_p}{dT_r}$$

Thus we can rewrite elasticity conditions as follows. Notice $p, r \in [0,1]$ are both in the inverse cdf scale.

$$-\int_p^1 \frac{dL_{\hat{p}}}{dT_r} d\hat{p} \frac{\phi_{ab}\left(a_p, b_p\right) b_p'}{\lambda g\left(a_p\right)} - w''\left(L_p\right) L_p' \frac{dL_p}{dT_r} - w'\left(L_p\right) \frac{d}{dp}\frac{dL_p}{dT_r} = 0, \forall p > r \tag{2.15}$$

$$-\left[-\frac{1}{w'(L_r)} + \int_p^1 \frac{dL_{\hat{p}}}{dT_r} d\hat{p}\right] \frac{\phi_{ab}\left(a_p, b_p\right) b_p'}{\lambda g\left(a_p\right)} - w''\left(L_p\right) L_p' \frac{dL_p}{dT_r} - w'\left(L_p\right) \frac{d}{dp}\frac{dL_p}{dT_r} = 0, \forall p < r \tag{2.16}$$

$$-\frac{\phi_a(a_0, b_0)}{\lambda g(a_0)} \left[\int_0^1 \frac{dL_{\hat{p}}}{dT_r} d\hat{p} - \frac{1}{w'(L_r)}\right] - w'(L_0)\frac{dL_0}{dT_r} = 0 \tag{2.17}$$

Equation (2.17) gives the first boundary constraint since it essentially captures behavior at entry margin. And the second constraint holds by definition of $F^L(b,p)$. The above system is what we see in Proposition 1.

Then, we denote in continuum location model, $a \sim G_a(a), b \sim G_b(b)$. Both cdf are Pareto distribution $G_a = \mathscr{P}(\underline{a}, \alpha), G_b = \mathscr{P}(\underline{b}, \beta)$, with pdf $g_a(a), g_b(b)$ respectively, where $a \in [\underline{a}, \infty) \equiv \mathscr{A}$, $b \in [\underline{b}, \infty) \equiv \mathscr{B}$. Note $a(b)$ is endogenous although G_a exogenous (through entry margin and functional form of endogenous matching $a(b)$). Note that $b_p : [0,1] \mapsto R^+$ is the inverse cdf of b. Let $p = G_b(b)$, thus we can obtain

$$b_p' = \frac{d}{dp}\left[G_b^{-1}(p)\right] = \left(\frac{dp}{db}\right)^{-1} = \frac{1}{g_b(b)}$$

For convenience to change units, we redefine the following transformed functions: $\hat{a}(b) := a(G_b(b)) = a_p \equiv a(p)$, $\hat{L}(b) := L(G_b(b)) = L_p \equiv L(p)$, and $\hat{T}(b) := T(G_b(b)) = T_p \equiv T(p)$. Consider rewriting equilibrium conditions (2.15), (2.16), (2.17) with transformed measures simplified. These give us the following system.

$$-\int_b^\infty \frac{dL(u)}{dT(b')} dG_b(u) \frac{\phi_{ab}(a(b), b)}{\lambda g(a(b))} - \frac{dw'(L(b))}{db}\frac{dL(b)}{dT(b')} - w'(L(b))\frac{d}{db}\frac{dL(b)}{dT(b')} = 0, \forall \tag{2.18}$$

$$-\left[-\frac{1}{w'(L(b'))} + \int_b^\infty \frac{dL(u)}{dT(b')} dG_b(u)\right] \frac{\phi_{ab}(a(b), b)}{\lambda g(a(b))} - \frac{dw'(L(b))}{db}\frac{dL(b)}{dT(b')} - w'(L(b))\frac{d}{db}\frac{dL(b)}{dT(b')} = 0, \forall \tag{2.19}$$

For the rest of this chapter, we will use the TFP measure unless specified. The above equations (2.18) and (2.19) characterize the system in Lemma 1. □

2.8.2 Proof of Lemma 2

Proof. First, we show how productivity exponent is related to size exponent:

$$a(b) \propto w'(L)L'(b) \propto b^{s_L\sigma-1}$$
$$\Rightarrow s_A = \underbrace{s_L(\sigma-1)}_{\text{wage effect } w'(L)} + \underbrace{(s_L-1)}_{\text{dampened by size effect } L'(b)} = s_L\sigma - 1$$

Key exponent of city size distribution is pinned down by the market clearing condition:

$$\underbrace{s_L}_{\text{GE effect from size}} + \underbrace{\beta_s}_{\text{GE effect from } G_b} = \underbrace{s_A - 1}_{\text{endogenous choice } a'(b)} + \underbrace{s_A\alpha_s}_{\text{dampened by distribution } G_a}$$
$$\Rightarrow \underbrace{(1-(\alpha_s-1)\sigma)}_{\text{Total GE effect}} s_L = -\alpha_s - \beta_s - 2$$
$$\Rightarrow s_L^* = \frac{\alpha_s + \beta_s + 2}{\alpha_s\sigma + \sigma - 1}$$

Derivations for other coefficients and components are obvious from equilibrium conditions. The total general equilibrium effect is shown in all denominators in shape parameters. In addition, with the above setup we have the following key restrictions on parameters:

$$s_L\sigma - 1 > 0, 1 - \sigma(\alpha_s + 1) > 0$$

2.8.3 Proof of Propostition 2

Proof. Consider $L(b) = c_L b^{S_L}, w(L) = \kappa(L(b))^{\sigma}, T(b) = c_T b^{s_T}, a(b) = c_A b^{s_A}$. "$C$" denotes coefficients, "S" denotes exponents. Equilibrium allocation parameters in competition are pinned down by the following location games and policy games.

$$\textbf{P1:} \; -\frac{T(b)}{w'(L(b))} + L(b) = 0$$
$$\textbf{L1:} \; a(b) - w'(L)L'(b) - T'(b) = 0, \forall b \in \mathscr{B}$$
$$\textbf{L2:} \; L(b)g_b(b) - \lambda g_a(a(b))a'(b) = 0, \forall b \in \mathscr{B}$$

With condition $P1$, we obtain $c_T(c_L), s_T(s_L)$:

$$T(b) = \kappa\sigma c_L^{\sigma} b^{s_L\sigma} \equiv c_T b^{s_T} \Rightarrow T'(b) = \kappa\sigma^2 c_L^{\sigma} s_L b^{s_L\sigma-1}$$

With condition $L1$, we obtain $c_A(c_L), s_A(s_L)$:

$$a(b) = \kappa\sigma \underbrace{(1+\sigma)}_{\text{distortion from tax competition}} c_L^{\sigma} s_L b^{s_L\sigma - 1} \equiv c_A b^{s_A}$$

With condition $L2$,

$$\beta_c c_L b^{s_L+\beta_s} = \lambda\alpha_c(\kappa\sigma(1+\sigma)s_L)^{\alpha_s+1} c_L^{\sigma(\alpha_s+1)} s_A b^{s_A\alpha_s+s_A-1}$$

Thus we can solve for parameters of equilibrium allocation as in propostiton.

Boundary value conditions in location games can pin down entry margin. If all enter, we have

$$\pi(\underline{a}, \underline{b}) = \underline{a}\underline{b} - \kappa L(\underline{b})^{\sigma} - T(\underline{b}) = 0 \tag{2.20}$$

$$\int_{\underline{b}}^{\infty} L(b) dG_b(b) = c_L^{*}\beta_c \frac{1}{s_L^{*}+1}\left[b^{s_L^{*}+1}\right]_{\underline{b}}^{\infty} = \lambda \tag{2.21}$$

If not all firms enter, $a^{*}(\underline{b}) \neq \underline{a}$, we have

$$\pi(a^{*}(\underline{b}), \underline{b}) = a^{*}(\underline{b})\underline{b} - \kappa L(\underline{b})^{\sigma} - T(\underline{b}) = 0$$

$$\int_{\underline{b}}^{\infty} L(b) dG_b(b) = c_L^{*}\beta_c \frac{1}{s_L^{*}+1}\left[b^{s_L^{*}+1}\right]_{\underline{b}}^{\infty} = \lambda\left(1 - G(a^{*}(\underline{b}))\right)$$

Denote:

$$f^{U}(b,p) \equiv \frac{dL(b)}{dT(p)}, F^{U}(b,p) \equiv -\int_{b}^{\infty} f^{U}(u,p) g_b(u) du$$

Given $g_a(\cdot), g_b(\cdot), w(L)$, equilibrium elasticity in competition is $\{f^{L*}(b,p), f^{U*}(b,p)\}$ determined by the following integro-differential equations (p is tax perturbation city) for $p > b$ and $p < b$ respectively.

$$-\int_{b}^{\infty} f^{U}(u,p) dG_b(u) \frac{1}{g_a(a(b))} - w''(L(b))L'(b) f^{U}(b,p) - w'(L(b)) f_b^{U}(b,p) = 0$$

$$-\left[-\frac{1}{w'(L(p))} + \int_{b}^{\infty} f^{L}(u,p) dG_b(u)\right]\frac{1}{g_a(a(b))} - w''(L(b))L'(b) f^{L}(b,p) - w'(L(b)) f_b^{L}(b,p) = 0$$

Assume $\phi_{ab} = 1, w(L) = \kappa L^{\sigma}, w'(L) = \kappa\sigma L^{\sigma-1}, w''(L) = \kappa\sigma(\sigma-1)L^{\sigma-2}$. The above equations become:

$$\textbf{E1:}\ -\int_b^{\infty} f^U(u,p)\beta_c u^{\beta_s} du \frac{1}{\alpha_c (c_A b^{s_A})^{\alpha_s}} - \kappa\sigma(\sigma-1)L(b)^{\sigma-2}L'(b) f^U(b,p) - \kappa\sigma L(b$$

$$\textbf{E2:}\ -\left[-\frac{1}{\kappa\sigma L(p)^{\sigma-1}} + \int_b^{\infty} f^L(u,p)\beta_c u^{\beta_s} du\right] \frac{1}{\alpha_c (c_A b^{s_A})^{\alpha_s}} - \kappa\sigma(\sigma-1)L(b)^{\sigma-2}L'(b) f^L(b,p) - \kappa\sigma L(b$$

Taking into account power forms in equilibrium allocations, move two terms without integration to right hand side, and multiply by $g(a(b))$, $E1$ give:

$$LHS = F^U(b,p) = -\int_b^{\infty} f^U(x,p)\beta_c x^{\beta_s} dx \tag{2.22}$$

$$RHS = \lambda \underbrace{\alpha_c (c_A b^{s_A})^{\alpha_s}}_{g_a(b)} \left(\underbrace{\kappa\sigma(\sigma-1)(c_L b^{s_L})^{\sigma-2}}_{w''(L)} \left(\underbrace{c_L s_L b^{s_L - 1}}_{L'(b)} \right) f^U(b,p) + \underbrace{\kappa (c_L b^{s_L})^{\sigma-1}}_{w'(L)} f_b^U(b,p) \right) \tag{2.23}$$

From $F^U(b,p)$, we know $f^U(b,p) = \frac{F_b^U(b,p)}{\beta_c b_s^{\beta}}$. Denote the term with f^U to be RHS_1 and the term with f_b^U to be RHS_2 in (2.23) , we have:

$$RHS_1 = \left(\kappa\sigma(\sigma-1)c_L^{\sigma-2} c_L s_L \beta_c^{-1} b^{s_L(\sigma-2)+s_L-1-\beta_s} F_b^U(b,p)\right)$$
$$RHS_2 = \lambda \underbrace{\alpha_c (c_A b^{s_A})^{\alpha_s}}_{g_a(b)} \left(-\kappa\sigma c_L^{\sigma-1}\beta_c^{-1}\beta_s b^{s_L(\sigma-1)-\beta_s-1+s_A\alpha_s} F_b^U(b,p) + \kappa\sigma c_L^{\sigma-1}\beta_c^{-1} b^{s_L(\sigma-1)+\beta_s+s_A\alpha_s} F_{bb}^U(b,p)\right)$$

Denote the exponent of b before $F_b^U(b,p)$ as δ_1^U, we can conduct the following decomposition of economic forces:

$$\delta_1^U \equiv \underbrace{s_L(\sigma-1)}_{\text{marginal cost exponent}} - \underbrace{\beta_s}_{\text{dampened by } g_b} + \underbrace{s_A\alpha_s}_{\text{dampened by } g_a(a(b))} -1 \tag{2.24}$$

Similarly, denote the exponent of b before $F_{bb}^U(b,p)$ as

$$\delta_2^U \equiv s_L(\sigma-1) - \beta_s + s_A\alpha_s \tag{2.25}$$

Notice that $\delta_1^U = \delta_2^U - 1$, which will be useful to see why exponents match in LHS and RHS. Now we can combine RHS_1, RHS_2 to be RHS, so we simplify

(2.23) to be

$$LHS = F^U(b,p) = \underbrace{\lambda \alpha_c c_A^{\alpha_s} \kappa \sigma c_L^{\sigma-1} \beta_c (s_L(\sigma-1)-\beta_s)}_{\equiv \theta_1^U} b^{\delta_1^U} F_b^U(b,p) + \underbrace{\lambda \alpha_c c_A^{\alpha_s} \kappa \sigma c_L^{\sigma-1}}_{\equiv \theta_2^U} \beta_c b^{\delta_2^U} F_{bb}^U(b,p) = RHS$$

Thus we summarize (2.23) with a second-order PDE with derivative of only b dimensions.

$$F^U(b,p) = \theta_1^U b^{\delta_1^U} F_b^U(b,p) + \theta_2^U b^{\delta_2^U} F_{bb}^U(b,p) \tag{2.26}$$

where $\theta_1 = (s_L(\sigma-1)-\beta_s)$, $\theta_2 = \frac{\alpha_c \kappa \lambda \sigma (x)^{\sigma-1}(-(\kappa\sigma(\sigma+1)(-\alpha_s-\beta_s-2)(x)^{\sigma}))}{\beta_c(\alpha_s\sigma+\sigma-1)}$, and

$$x \equiv \left(\frac{\alpha_c \lambda \left(-\frac{\sigma(-\alpha_s-\beta_s-2)}{\alpha_s\sigma+\sigma-1} - 1\right)\left(-\frac{\kappa\sigma(\sigma+1)(-\alpha_s-\beta_s-2)}{\alpha_s\sigma+\sigma-1}\right)^{\alpha_s+1}}{\beta_c} \right)^{\frac{1}{1-(\alpha_s+1)\sigma}}$$

Moreover, our functional assumptions result in $\delta_1 = 1, \delta_2 = 2$,i.e., a Cauchy-Euler type if we take p as some constant. To see this, with previous information, we have:

$$\begin{aligned} \delta_2^U &= s_A\alpha_s + s_L^*(\sigma-1) - \beta_s = (s_L^*\sigma - 1)\alpha_s + s_L^*\sigma - s_L^* - \beta_s \\ &= s_L^*(\sigma\alpha_s + \sigma - 1) - \alpha_s - \beta_s = \beta_s + \alpha_s + 2 - \alpha_s - \beta_s = 2 \end{aligned}$$

Now we show how p enters F^U and F^L. Assume seperable form for F^U and F^L in b and p. We guess there exists some $\varphi^U(p), \varphi^L(p), r_1^U, r_2^U, r_1^L, r_2^L$ with the following form which can solve the system.[27]

$$F^U(b,p) \equiv -\int_b^\infty f^U(u,p)dG_b(u) = \varphi_1^U(p)b^{r_1^U} + \varphi_2^U(p)b^{r_2^U},\ \forall p > b \in \mathscr{B}$$
$$F^L(b,p) \equiv -\int_b^\infty f^L(u,p)dG_b(u) = \varphi_1^L(p)b^{r_1^L} + \varphi_2^L(p)b^{r_2^L} - \frac{1}{w'(L(p))},\ \forall p < b \in \mathscr{B}$$

Consider rewriting relevant differential equations as

$$\varphi_1^U(p)b^{r_1^U} + \varphi_2^U(p)b^{r_2^U} = \theta_1^U b\left(\varphi_1^U(p)r_1^U b^{r_1^U-1} + \varphi_2^U(p)r_2^U b^{r_2^U-1}\right) + \theta_2^U b^2\left(\varphi_1^U(p)r_1^U(r_1^U-1)b^{r_1^U-2} + \varphi_2^U(p\right.$$
$$= \left(\theta_1^U r_1^U + \theta_2^U r_1^U(r_1^U-1)\right)b^{r_1^U}\varphi_1^U(p) + \left(\theta_1^U r_2^U + \theta_2^U r_2^U(r_2^U-1)\right)b^{r_2^U}\varphi_2^U(p)$$

Thus, exponents in foundamental solutions are $r_{1,2} = \frac{\theta_2-\theta_1 \pm \sqrt{(\theta_1-\theta_2)^2+4\theta_2}}{2\theta_2}$. For notation consistency later, we denote two roots as $r_1 > 0 > r_2$.

Now we turn to $f^L(b,p)$ and start from $E2$ (Equation 2.19). Denote the above 2nd order ODE (2.26) $\mathscr{L}(b) = 0$. If $y_1(b)$ and $y_2(b)$ are linearly independent solutions of the equation $\mathscr{L}(b) = 0$ on an interval $\mathscr{B} \subset \mathbf{R}$, where variable coefficients are continuous functions on $\mathscr{B}$, then there are unique constants c_1, c_2 such that every solution $y(b)$ of the differential equation $\mathscr{L}(b) = 0$ on can be written as a linear combination. By conjecture of F^U, F^L, we know

$$f^j = \sum_{i=1}^{2} \frac{\varphi_i^j(p)r_i}{\beta_c} b^{r_i-\beta_s-1},\ j \in \{U, L\}$$

[27] To double-check whether the form we obtained is indeed the solution, first take partial derivatives with respect to b gives:

$$F^U(b,p) = -\int_b^\infty f^U(x,p)\beta_c x^{\beta_s}dx \equiv \beta_c\varphi(p)\frac{b^{\epsilon^U}}{\epsilon^U}$$

where $\epsilon^U \equiv 1 + r^U + \beta_s$. Together with $F_b^U(b,p)$ and $F_{bb}^U(b,p)$, consider rewriting (2.26) as

$$\frac{\beta_c\varphi(p)}{\epsilon^U}b^{\epsilon^U} = \theta_1^U\beta_c\varphi(p)b^{\delta_1^U+\epsilon^U-1} + \theta_2^U\beta_c\varphi(p)(\epsilon^U-1)b^{\delta_2^U+\epsilon^U-2} \tag{2.27}$$

where we clearly see the above would hold (if hold) regardless of $\varphi(p)$ and ϵ^U. Now to verify our guess, we need to check whether the power match in b, so we want to see exponent on the left of b equals that in the right terms. Recall form of δ_1^U and δ_2^U in (2.24) and (2.25), we need to check the equation:

exponent of b on LHS of (2.27) $= \epsilon^U = s_L(\sigma-1) - \beta_s + \epsilon^U - 2 + s_A\alpha_s =$ exponent on RHS

With the equilibrium allocation parameters (we use s_A^*, s_L^* here) solved at the beginning, we confirm the above is true. Since the exponents match, the power form is indeed a solution to the differential equations.

By BV4, $\varphi_1^L(p) = \left(\frac{\frac{B}{A}\frac{r_2}{\beta_c}\underline{b}^{r_2-\beta_s-1} - \underline{b}^{r_2}}{-\frac{B}{A}\frac{r_1}{\beta_c}\underline{b}^{r_1-\beta_s-1} - \underline{b}^{r_1}} \right) \varphi_2^L(p) \equiv \zeta \varphi_2^L(p)$.

By Continuity 1,

$$\begin{aligned} F^U(b,b) &= \varphi_2^U(p) b^{r_2^U}|_{p=b} = \varphi_2(b) b^{r_2^U} \\ &= \left(\zeta b^{r_1^L} + b^{r_2^L}\right) \varphi_2^L(b) - \frac{1}{w'(L(b))} = F^L(b,b) \end{aligned}$$

By Continuity 2,

$$f^U(b,b) = \frac{\varphi_1^L(p) r_1^L}{\beta_c} b^{r_1^L - \beta_s - 1}|_{p=b} + \frac{\varphi_2^L(p) r_2^L}{\beta_c} b^{r_2^L - \beta_s - 1}|_{p=b} = f^L(b,b)$$

Thus we have $\forall p \in \mathscr{B}$,

$$\varphi_1^L(p) = \frac{r_2^L c_L^{1-\sigma} p^{-r_1^L - (\sigma-1)s_L}}{\kappa\sigma(r_2^L - r_1^L)} \tag{2.28}$$

$$\varphi_2^U(p) = \frac{c_L^{1-\sigma}\left(r_2^L p^{r_2^L} + r_1^L \zeta p^{r_1^L}\right) p^{-r_1^L - r_2^L - (\sigma-1)s_L}}{\kappa\sigma\zeta(r_2^L - r_1^L)} \tag{2.29}$$

Substitute into conjectured form, we have:

$$F^U(b,p) = \frac{b^{r_2} c_L^{1-\sigma}(r_2 p^{r_2} + r_1 \zeta p^{r_1}) p^{-r_1 - r_2 - \sigma s_L + s_L}}{\kappa\sigma\zeta(r_2 - r_1)}$$

$$F^L(b,p) = \frac{c_L^{1-\sigma} p^{-r_1 - \sigma(s_L+1) + s_L}\left(r_2 b^{r_2} p^{\sigma} + r_2 \zeta b^{r_1} p^{\sigma} + \zeta(r_1 - r_2) p^{r_1 + (\sigma-1)s_L + 1}\right)}{\kappa\sigma\zeta(r_2 - r_1)}$$

Therefore, elasticity in competition is characterized by the following pieacewise surface.

$$f^U(b,p) = \frac{r_2 c_L^{1-\sigma} b^{-\beta_s + r_2 - 1}(r_2 p^{r_2} + r_1 \zeta p^{r_1}) p^{-r_1 - r_2 - \sigma s_L + s_L}}{\beta_c \kappa\sigma\zeta(r_2 - r_1)}, \ \forall b > p$$

$$f^L(b,p) = \frac{r_2 b^{-\beta_s - 1} c_L^{1-\sigma}(r_2 b^{r_2} + r_1 \zeta b^{r_1}) p^{-r_1 - \sigma s_L + s_L}}{\beta_c \kappa\sigma\zeta(r_2 - r_1)}, \ \forall b < p$$

□

2.8.4 Proof of Proposition 3

Proof. First, consider tax has a variable part which is proportional to wage with an unkown coefficients as the following. We preserve σ before $w(L)$ for convenience to compare with competition regime later.

$$T(b) = c_T b^{s_T} + t = \rho\,(\sigma w(L(b))) + t$$

The method is similar as in competition except for the fact that c_T, c_A, c_L are different, where they are pinned down at the end when we close the model by the policy conditon. Due to similarity, we skip the part for solving PDE. Note that the foundamental solution of PDE is the same as in competition, with all parameters in coordination $c^*(\rho; s^*)$. We use f^U, f^L solved from PDE into condition $P2$ to pin down ρ, t. LHS and RHS of condition $P2$ can be simplied to the following with same structure of equations (2.28)-(2.29):

$$LHS = \frac{c_T c_L^{1-\sigma} r_1 r_2}{\kappa\sigma(r_2 - r_1)}\left[\frac{1}{(s_T + r_1)} - \frac{1}{(s_T + r_2)}\right] p^{s_L} + \frac{t c_L^{1-\sigma}}{\kappa\sigma} p^{-s_L(\sigma-1)} - (C_1 + C_2) p^{-s_L(\sigma-1)-r_1}$$
$$RHS = \left(\frac{c_T}{\kappa\sigma c_L^{\sigma-1}} - c_L\right) p^{s_L} + \frac{t c_L^{1-\sigma}}{\kappa\sigma} p^{-s_L(\sigma-1)}$$

where

$$C_1 = \frac{c_T c_L^{1-\sigma} r_1 r_2 \underline{b}^{s_T+r_1}}{(s_T + r_1)\kappa\sigma(r_2 - r_1)} - \frac{c_T c_L^{1-\sigma} r_1 r_2 \underline{b}^{s_T+r_2}}{(s_T + r_2)\zeta\kappa\sigma(r_2 - r_1)}$$
$$C_2 = \frac{t\underline{b}^{r_1} r_2 c_L^{1-\sigma}}{\kappa\sigma(r_2 - r_1)} - \frac{t\underline{b}^{r_2} r_2 c_L^{1-\sigma}}{\zeta\kappa\sigma(r_2 - r_1)}$$

Thus we have $1 - \frac{1}{\rho^*} = \frac{r_1(\rho^*) r_2(\rho^*)}{(s_T + r_1(\rho^*))(s_T + r_2(\rho^*))}$ in equilibrium for ρ^* where

$$r_1(\rho) = \frac{\theta_2(\rho) - \theta_1(\rho) + \sqrt{(\theta_1(\rho) - \theta_2(\rho))^2 + 4\theta_2(\rho)}}{2\theta_2(\rho)} > 0$$
$$r_2(\rho) = \frac{\theta_2(\rho) - \theta_1(\rho) - \sqrt{(\theta_1(\rho) - \theta_2(\rho))^2 + 4\theta_2(\rho)}}{2\theta_2(\rho)} < 0$$

and

$$\theta_2 = \frac{\lambda\kappa\sigma\alpha_c c_A(\rho)^{\alpha_s} c_L(\rho)^{\sigma-1}}{\beta_c}$$
$$\theta_1 = (s_L(\sigma-1) - \beta_s)\theta_2 = \frac{\lambda\kappa\sigma\alpha_c c_A(\rho)^{\alpha_s} c_L(\rho)^{\sigma-1}(s_L(\sigma-1) - \beta_s)}{\beta_c}$$

Recall θ are from PDE, we have

$$F^U(b,p) = \underbrace{\beta_c^{-1}\lambda\alpha_c c_A^{\alpha_s}\kappa\sigma c_L^{\sigma-1}(s_L(\sigma-1) - \beta_s)}_{\equiv\theta_1^U} b F_b^U(b,p) + \underbrace{\beta_c^{-1}\lambda\alpha_c c_A^{\alpha_s}\kappa\sigma c_L^{\sigma-1}}_{\equiv\theta_2^U} b^2 F_{bb}^U(b,p)$$

In addition, from fundamental solutions, $r_1 r_2 = -\frac{1}{\theta_2}$, $r_1 + r_2 = 1 - (s_L(\sigma-1) - \beta_s)$.

After combining terms and some algebra, we obtain:

$$\rho^* = \frac{\alpha_s\sigma + \sigma - 1}{(\alpha_s + 2)\sigma} \tag{2.30}$$

And thus we also solve for t and obtain:

$$t^* = \frac{\frac{c_T r_1 \underline{b}^{s_T + r_1}}{s_T + r_1} + \frac{c_T r_2 \underline{b}^{s_T + r_2}}{s_T + r_2}}{\underline{b}^{r_1} + \frac{\underline{b}^{r_2}}{\zeta}} \tag{2.31}$$

where $A \equiv \frac{b}{\lambda \alpha_c c_A^{\alpha_s} \underline{b}^{s_A \alpha_s}}$, $B \equiv \kappa\sigma c_L^{\sigma-1} \underline{b}^{s_L(\sigma-1)}$, and $\zeta \equiv -\left(\frac{\frac{B}{A}\frac{r_2}{\beta_c}\underline{b}^{r_2-\beta_s-1} - \underline{b}^{r_2}}{\frac{B}{A}\frac{r_1}{\beta_c}\underline{b}^{r_1-\beta_s-1} - \underline{b}^{r_1}}\right)$. □

2.8.5 Proof of Proposition 4

Proof. By solution from equation (2.30), since $\alpha_s < -2$ and $\sigma > 0$, we know $\rho > 1$.

By solution from equation (2.31), since $s_T + r2 < 0$ (we know from convergence requirement in boundary value conditions), $r_2 < 0 < r_1$. By definition of ζ, we can simplify $\underline{b}^{r_1} + \frac{\underline{b}^{r_2}}{\zeta} = 1$. Thus we know $t^* > 0$.

$$\frac{\partial c_T}{\partial \rho} = \frac{\kappa\sigma(\sigma(\alpha_s - \rho + 1) - 1)\left(\left(\frac{\alpha_c \lambda(\beta_s\sigma + \sigma + 1)\left(\frac{\kappa\sigma(\alpha_s + \beta_s + 2)(\rho\sigma + 1)}{\alpha_s\sigma + \sigma - 1}\right)^{\alpha_s + 1}}{\beta_c(\alpha_s\sigma + \sigma - 1)}\right)^{\frac{1}{1-(\alpha_s+1)\sigma}}\right)^{\sigma}}{(\alpha_s\sigma + \sigma - 1)(\rho\sigma + 1)}$$

$$= (1 + \rho\sigma)^{\frac{\sigma(1+\alpha_s)}{1-\sigma(1+\alpha_s)} - 1}(\sigma(\alpha_s - \rho + 1) - 1) \cdot Constant > 0$$

since $\alpha_s\sigma + \sigma - 1 < 0, \alpha_s + \beta_s + 2 < 0$ and $\sigma, \rho > 0$. □

2.8.6 Proof of Lemma 3

To show firms' behavior at the top do not depend on regime change, we only need to show shape parameters are the same, i.e., exponents of power form are the same.

In competition, by $P1$, we have $T(b) = \sigma w(L(b)) = \kappa\sigma c_L^{\sigma} b^{s_L\sigma}$. Thus $s_T = \sigma s_L^{LF}$. By $L1, L2$, we know $s_A = s_L\sigma - 1$ and $s_L = s_L^{LF}$ (from $c_L \beta_c b^{s_L + \beta_s} = \lambda \alpha_c c_A^{\alpha_s + 1} s_A b^{s_A\alpha_s + s_A + 1}$). Thus competition has the same equilibrium exponents as in laissez-faire.

In coordination, same relationships hold for s_L, s_A, s_T when we guess and verify proportionality:

$$T(b) - t^*(\rho;\Theta) \propto w(L(b)), a(b) \propto w'(L(b))L'(b)$$

At the same time, market clearing condition is independent of regimes, which pins down s_L in all regimes. Thus exponents for b hold in tax coordination as in laissez-faire. Note that although market clearing conditions remain the same form in all regimes, general equilibrium effects enter through coefficients c_L, c_A, c_T. □

2.8.7 Proof of Lemma 4

Proof. To see why city size distribution in coordination is driven down, we start from comparative statics for coefficient of equilibrium size. To show the lemma, it is equivalent to show $\frac{\partial c_L^*}{\partial \rho} < 0, \frac{\partial c_A^*}{\partial \rho} > 0$ which can then indicate $\textbf{Var}(L)_{\text{coord}} \leq \textbf{Var}(L)_{\text{comp}} \leq \textbf{Var}(L)_{\text{laissez-faire}}$. To see this:

$$\frac{\partial c_L^*}{\partial \rho} = \frac{1+\alpha_s}{1-\sigma(\alpha_s+1)}(1+\rho\sigma)^{\frac{\alpha_s+1}{1-\sigma(\alpha_s+1)}-1}\sigma c_L^{LF} < 0$$

To see how scale and shape parameter of size and productivity distribution change, we need to transform power form to standard form. Denote cumulative and probability distribution function of size distirbution $G_L(L), g_L(L)$. Denote l_s, l_{min} the shape and scale parameters of size distribtution. From $G_b(b) = G_L(L(b))$, we know $g_L(L) = \frac{\beta_c}{s_L} c_L^{-\frac{\beta_s+1}{s_L}} L^{\frac{\beta_s+1}{s_L}-1}$. Thus city size distribution follows Pareto with shape and scale parameters:

$$\begin{cases} l_s = -\left(\frac{\beta_s - s_L + 1}{s_L}\right) - 1 \\ l_{min} = \left(\frac{\frac{\beta_c}{s_L} c_L^{-\frac{\beta_s+1}{s_L}}}{l_s}\right)^{\frac{1}{l_s}} = \left(\frac{\beta_c}{s_L l_s}\right)^{\frac{1}{l_s}} c_L \end{cases}$$

where variance of size distribution is $\textbf{Var}(L) = \frac{l_c^2 l_s}{(l_s-1)^2(l_s-2)}$. Thus we know

$$sgn\left(\frac{\partial \textbf{Var}(L)}{\partial \rho}\right) = -sgn\left(\frac{\partial \textbf{Var}(L)}{\partial c_L}\right) = -sgn\left(\frac{\partial l_{min}}{\partial c_L}\right) < 0$$

Since $\alpha_s < -2$ and $\rho, \sigma, c_L^{LF} > 0$. Thus when variable tax increases by multiplied a constant coefficient ρ, scale parameter is driven down. By property of Pareto, city size distribution has lower variance. Therefore income inequality is driven down, i.e.,

$$\textbf{Var}(L)_{coord} \leq \textbf{Var}(L)_{comp} \leq \textbf{Var}(L)_{Laissez-faire}$$

Similarly, productivity distirbution is featured by:

$$g_a(a(b)) = \alpha_c c_A^{\alpha_s} b^{s_A \alpha_s}$$

We again start from comparative statics for coefficient of equilibrium productivity:

$$\frac{\partial c_A^*}{\partial \rho} = \frac{1}{1-\sigma(\alpha_s+1)}(1+\rho\sigma)^{\frac{1}{1-\sigma(\alpha_s+1)}-1}\sigma c_A^{LF} > 0$$

since $1-\sigma(\alpha_s+1) > 0$ (as derived in Laissez-faire) and $\sigma, c_A^{LF} > 0$.

To see how productivity distribution changes on average, consider the following tranformation:

$$g_a(a(b)) = \alpha_s c_A^{\alpha_s} b^{s_A \alpha_s}$$

Thus productivity distribution follows Pareto with shape (a_s) and scale parameters (a_{min}):

$$\begin{cases} a_s = -s_A\alpha_s - 1 \\ a_{min} = \left(\frac{\alpha_s c_A^{\alpha_s}}{-s_A\alpha_s}\right)^{\frac{1}{-s_A\alpha_s-1}} \end{cases}$$

Thus we can derive average productivity:

$$sgn\left(\frac{\partial \mathbf{E}(a)}{\partial \rho}\right) = \frac{\alpha_s}{-s_A\alpha_s-1}\left(\frac{\alpha_c}{-s_A\alpha_s-1}\right)^{\frac{1}{-s_A\alpha_s-1}} c_A^{\frac{\alpha_s}{-s_A\alpha_s}-1} > 0$$

since $\alpha_s < -2, s_A > 0, \alpha_c > 0$. Therefore, comparing among three regimes gives:

$$\mathbf{E}(a)_{\text{coord}} \geq \mathbf{E}(a)_{\text{comp}} \geq \mathbf{E}(a)_{\text{laissez-faire}}$$

In summary, size is larger and endogenous productivity is lower in all cities under tax competition compared to coordination. □

2.8.8 Proof of Lemma 5

Proof. From Proposition 3 and Proposition 4, we know that tax competition can be summarized as a special case where $\rho = 1$ and $t = 0$. Therefore, when uncooperative results are close to that in cooperative game is equivalent to when ρ^* in coordination is close to 1. Thus we simply take limits,

$$\lim_{\alpha_s\to\infty} = 1, \lim_{\sigma\to\infty} \rho^* = \frac{\alpha_s+1}{\alpha_s+2}$$

Note $\alpha_s < -2$ thus when extent of congestion cost is high and productivity distribution has thinner tail ex ante, uncooperative equilibrium is close to that in cooperative game. □

2.8.9 Proof of Lemma 6

Proof. Denote aggregate TFP in tax coordination and tax competition as TFP^{coord} and TFP^{comp} respectively. Since each firm hire unit labor and proudce unit numeraire good, city level ex post TFP can be written as $A(b) \equiv a(b)b$ and thus we can write equilibrium aggregate TFP as:

$$TFP = \int_{\mathscr{B}} a(b) b dG_b(b)$$

Denote endogenous city productivity level $a^{TC}(b)$ in tax competition and $a^C(b)$ in tax coordination. Loss in aggregate TFP in competition from coordination in log scale is:

$$\frac{TFP^C}{TFP^{TC}} = \frac{\int_{\mathscr{B}} a^C(b) b dG_b(b)}{\int_{\mathscr{B}} a^{TC}(b) b dG_b(b)}$$

By Proposition 2 and 3 where all cities are filled (i.e. no ex post empty city), we know

$$\frac{a^C(b)}{a^{TC}(b)} = \frac{c_A^C}{c_A^{TC}}$$

Thus

$$\Delta TFP = ln\left(\frac{\kappa\sigma(\rho\sigma+1)s_L^*\left(\left(\frac{\alpha_c\lambda(\sigma s_L-1)(\kappa\sigma(\rho\sigma+1)s_L)^{\alpha_s+1}}{\beta_c}\right)^{\frac{1}{1-(\alpha_s+1)\sigma}}\right)^{\sigma}}{\kappa\sigma(\sigma+1)s_L^*\left(\left(\frac{\alpha_c\lambda(\sigma s_L^*-1)(\kappa\sigma(\sigma+1)s_L^*)^{\alpha_s+1}}{\beta_c}\right)^{\frac{1}{1-(\alpha_s+1)\sigma}}\right)^{\sigma}}\right)$$
$$= \frac{1}{1-(\alpha_s+1)\sigma} ln\left(\frac{\rho\sigma+1}{\sigma+1}\right)$$

Therefore, when $\rho > 1$, $1-\sigma(\alpha_s+1) > 0$ (as proved in Proposition 2 and 4), we have:

$$\frac{\partial \Delta TFP}{\partial \sigma} = \frac{-(\alpha_s+1)\log\left(\frac{\rho\sigma+1}{\sigma+1}\right)(\sigma+1)(\rho\sigma+1)+(\rho-1)(\alpha_s\sigma+\sigma-1)}{(\alpha_s\sigma+\sigma-1)^2(\sigma+1)(\rho\sigma+1)} < 0$$
$$\frac{\partial \Delta TFP}{\partial k_a} = -\frac{\sigma \log\left(\frac{\rho\sigma+1}{\sigma+1}\right)}{(\alpha_s\sigma+\sigma-1)^2} < 0$$

Note that we use α_s to illustrate distribution in a clearer way agains local TFP (b), while the Pareto shape parameter for productivity distribution is k_a. □

2.8.10 Uniform Distribution with Linear Congestion

In this section, we present ways to solve discrete equilibrium and elasticity surface. Denote competition ("TC") and coordination ("C") two regimes $R \in \{TC, C\}$. $\hat{a}_{k-1}$ is the cutoff proucitivity between city $k-1$ and city k, $k \in \{1, 2, \ldots, n\}$.

Equilibrium conditions in location game consist of a set of indifference conditions and a set of adding-up conditions. Denote L as a mass of firms before normalization, thus $\frac{L_{\frac{i}{n}}}{\Lambda_n}$ is the share of firms with locational choice of i.

$$\begin{cases} \hat{a}_{\frac{1}{n}} b_{\frac{1}{n}} - \kappa L_{\frac{1}{n}} - T_{\frac{1}{n}} = 0 \\ \hat{a}_{\frac{2}{n}} b_{\frac{1}{n}} - \kappa L_{\frac{1}{n}} - T_{\frac{1}{n}} = \hat{a}_{\frac{2}{n}} b_{\frac{2}{n}} - \kappa L_{\frac{2}{n}} - T_{\frac{2}{n}} \\ \vdots \\ \hat{a}_1 b_{\frac{n-1}{n}} - \kappa L_{\frac{n-1}{n}} - T_{\frac{n-1}{n}} = \hat{a}_1 b_1 - \kappa L_1 - T_1 \end{cases} \tag{2.32}$$

$$\begin{cases} L_{\frac{1}{n}} = \Lambda_n [G(\hat{a}_{\frac{2}{n}}) - G(\hat{a}_{\frac{1}{n}})] \\ L_{\frac{2}{n}} = \Lambda_n [G(\hat{a}_{\frac{3}{n}}) - G(\hat{a}_{\frac{2}{n}})] \\ \vdots \\ L_1 = \Lambda_n [1 - G(\hat{a}_1)] \end{cases} \tag{2.33}$$

Rearranging the above system using the telescoping pattern, and denote measure of firms exiting the market as L_0, we can obtain the system:

$$\begin{aligned} \max_{\{L_{\frac{i}{n}}, T_{\frac{i}{n}}, \hat{a}_{\frac{i}{n}}\}_{i=1,\ldots,n}} & \sum_{i=1}^{n} L_{\frac{i}{n}} T_{\frac{i}{n}} (\frac{1}{n}) \\ \text{s.t. } & \hat{a}_{\frac{1}{n}} b_{\frac{1}{n}} + \sum_{i=p}^{n} \hat{a}_{\frac{i}{n}} \left(b_{\frac{i}{n}} - b_{\frac{i}{n}-\Delta} \right) = \kappa L_{\frac{p}{n}} + T_{\frac{p}{n}}, \ \forall p \in \{2, \ldots, n\} \\ & \sum_{i=p}^{n} L_{\frac{i}{n}} = \Lambda_n (1 - G(\hat{a}_{\frac{p}{n}})), \ \forall p \in \{2, \ldots, n\} \end{aligned}$$

Equilibrium can be solved by combining all FOC from Lagrangian into one matrix where $\mathbf{x}' = [\mathbf{L}|\mathbf{a}|\vec{\mu}]$

$$M\mathbf{x} \equiv \begin{bmatrix} -2\kappa & 0 & b_1 & 0 & 1 & 0 \\ 0 & -2\kappa & b_1 & b_2 - b_1 & 1 & 1 \\ 0b_1 & b_1 & 0 & 0 & \frac{\Lambda_n}{\bar{a}} & 0 \\ 0 & b_2 - b_1 & 0 & 0 & 0 & \frac{\Lambda_n}{\bar{a}} \\ 1 & 1 & \frac{\Lambda_n}{\bar{a}} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0\frac{\Lambda_n}{\bar{a}} & 0 & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \Lambda_n \\ \Lambda_n \end{bmatrix} \equiv \mathbf{c}$$

In general, when there are n cities, the constant term is $\mathbf{c}_{3n\times 1} = [\vec{0}|\vec{0}|\Lambda_n]'$.

Thus we can solve all equilibrium variables

$$\mathbf{x} = [\mathbf{L}|\mathbf{a}|\vec{\mu}]' = M^{-1}\mathbf{c}$$

An alternative way to solve the system is to identify the cross-behavioral effect. To see this, with uniform distribution assumption of firms' productivity, we can write out $G(x) = \frac{x}{\bar{a}}$ by plugging in $\hat{a}$'s into market clearing conditions. The mechanism of general equilibrium is clear throught two sets of systems. By conditions 2.32, when t_i, L_i increase, $\hat{a}_i$ increases whereas $\hat{a}_{i+1}$ decreases. By conditions 2.33, we know L_i decreases $\forall i \in \{1, 2, ..., n-1\}$. Change in n will only induce a change in $\hat{a}_n$.

$$\begin{cases} L_{\frac{1}{n}} = \Lambda_n\left(\frac{\kappa L_{\frac{2}{n}} - \kappa L_{\frac{1}{n}} + t_{\frac{2}{n}} - t_{\frac{1}{n}}}{\bar{a}(b_{\frac{2}{n}} - b_{\frac{1}{n}})}\right) - \Lambda_n\left(\frac{\kappa L_{\frac{1}{n}} + t_{\frac{1}{n}}}{\bar{a} b_{\frac{1}{n}}}\right) \\ L_{\frac{2}{n}} = \Lambda_n\left(\frac{\kappa L_{\frac{3}{n}} - \kappa L_{\frac{2}{n}} + t_{\frac{3}{n}} - t_{\frac{2}{n}}}{\bar{a}(b_{\frac{3}{n}} - b_{\frac{2}{n}})}\right) - \Lambda_n\left(\frac{\kappa L_{\frac{2}{n}} - \kappa L_{\frac{1}{n}} + t_{\frac{2}{n}} - t_{\frac{1}{n}}}{\bar{a}(b_{\frac{2}{n}} - b_{\frac{1}{n}})}\right) \\ \vdots \\ L_1 = \Lambda_n - \Lambda_n\left(\frac{\kappa L_1 - \kappa L_{\frac{n-1}{n}} + t_1 - t_{\frac{n-1}{n}}}{\bar{a}(b_1 - b_{\frac{n-1}{n}})}\right) \end{cases} \tag{2.34}$$

Denote I as the standard identity matrix and the following vectors for convenience:

$$\mathbf{L} \equiv \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{bmatrix}, \mathbf{t} \equiv \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}, \mathbf{b} \equiv \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \mathbf{c} \equiv \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \Lambda_n \end{bmatrix}$$

Equilibrium conditions can be summarized by the following:

$$\mathbf{L} = S\mathbf{L} + T\mathbf{t} + \mathbf{c} \tag{2.35}$$

where T can be shown to be a symmetric matrix, which has negative diagonal and positive off-diagonal elements.

$$\begin{bmatrix} -\frac{\Lambda_n}{\bar{a}(b_2-b_1)} - \frac{\Lambda_n}{\bar{a}b_1} & \frac{\Lambda_n}{\bar{a}(b_2-b_1)} & 0 & \cdots & 0 & 0 \\ \frac{\Lambda_n}{\bar{a}(b_2-b_1)} & -\frac{\Lambda_n}{\bar{a}(b_3-b_2)} - \frac{\Lambda_n}{\bar{a}(b_2-b_1)} & \frac{\Lambda_n}{\bar{a}(b_3-b_2)} & \cdots & 0 & 0 \\ 0 & \frac{\Lambda_n}{\bar{a}(b_3-b_2)} & -\frac{\Lambda_n}{\bar{a}(b_4-b_3)} - \frac{\Lambda_n}{\bar{a}(b_3-b_2)} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{\Lambda_n}{\bar{a}(b_{n-1}-b_{n-2})} - \frac{\Lambda_n}{\bar{a}(b_n-b_{n-1})} & \frac{\Lambda_n}{\bar{a}(b_n-b_{n-1})} \\ 0 & 0 & 0 & \cdots & \frac{\Lambda_n}{\bar{a}(b_n-b_{n-1})} & -\frac{\Lambda_n}{\bar{a}(b_n-b_{n-1})} \end{bmatrix}$$

Thus, we know $I - S$, where $S = \kappa T$, is symmetric and has positive diagonal and negative off-diagonal elements:

$$
\begin{bmatrix}
\frac{\kappa\Lambda_n}{\bar a(b_2-b_1)}+\frac{\kappa\Lambda_n}{\bar a b_1}+1 & -\frac{\kappa\Lambda_n}{\bar a(b_2-b_1)} & 0 & \cdots & 0 & 0 \\
-\frac{\kappa\Lambda_n}{\bar a(b_2-b_1)} & \frac{\kappa\Lambda_n}{\bar a(b_3-b_2)}+\frac{\kappa\Lambda_n}{\bar a(b_2-b_1)}+1 & -\frac{\kappa\Lambda_n}{\bar a(b_3-b_2)} & \cdots & 0 & 0 \\
0 & -\frac{\kappa\Lambda_n}{\bar a(b_3-b_2)} & \frac{\kappa\Lambda_n}{\bar a(b_4-b_3)}+\frac{\kappa\Lambda_n}{\bar a(b_3-b_2)}+1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{\kappa\Lambda_n}{\bar a(b_{n-1}-b_{n-2})}+\frac{\kappa\Lambda_n}{\bar a(b_n-b_{n-1})}+1 & -\frac{\kappa\Lambda}{\bar a(b_n-b} \\
0 & 0 & 0 & \cdots & -\frac{\kappa\Lambda_n}{\bar a(b_n-b_{n-1})} & \frac{\kappa\Lambda_n}{\bar a(b_n-b_{n-1}}
\end{bmatrix}
$$

Notice $I-S$ is a row/column strictly diagonally dominant and our assumption on distribution of b guarantees that it is also irreducible. By Levy–Desplanques Theorem, this matrix is non-sigular and invertible. Therefore we can solve for the equilibrium considering the following decomposition forces below:[28]

$$
\mathbf{L}(\mathbf{t}) = \underbrace{\underbrace{(I-S)^{-1}}_{\text{sorting effect on size}}\underbrace{T}_{\text{tax effect on size}}}_{\text{tax distortion on size (through sorting)}}\mathbf{t} + \underbrace{\underbrace{(I-S)^{-1}}_{\text{sorting effect on size}}\underbrace{\mathbf{c}}_{\text{trivial size (without sorting): all in city } n}}_{\text{Laissez-faire size with sorting}}
$$

Consider denoting the first term of RHS above as $E\mathbf{t}+\mathbf{k}$, where E is the elasticity matrix for the distortion effect on size $E \equiv \frac{d\mathbf{L}}{d\mathbf{t}} = (I-S)^{-1}T$. Thus elements in $diag(E)$ are self-elasticity in each city with respect to change in its own tax. Denote each element in E as $\xi_{i,j}$ at the i^{th} row and the j^{th} column, $\forall i,j \in \{1,2,...,n\}$. Furthermore, assume b are uniformly distribuited as in our context, i.e., $b_1 = b_{i+1}-b_i \equiv b = \frac{1}{n}$ a constant. The above two matrices are both almost-Toeplitz type and can be further simplified to the following.

$$
I-S = \frac{\kappa\Lambda_n}{\bar a b}
\begin{bmatrix}
2+\frac{\bar a b}{\kappa\Lambda_n} & -1 & & & & & \\
-1 & 2+\frac{\bar a b}{\kappa\Lambda_n} & -1 & & & & \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
 & & & & -1 & 2+\frac{\bar a b}{\kappa\Lambda_n} & -1 \\
 & & & & & -1 & 1+\frac{\bar a b}{\kappa\Lambda_n}
\end{bmatrix}
$$

$$
T = \frac{\Lambda_n}{\bar a b}
\begin{bmatrix}
-2 & 1 & & & & & \\
1 & -2 & 1 & & & & \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
 & & & & 1 & -2 & 1 \\
 & & & & & 1 & -1
\end{bmatrix}
$$

[28] By Neumann series expansion, we can write the sorting effect part as

$$
(I-S)^{-1} = \sum_{n=1}^{\infty} S^n
$$

where each element in the series can be regarded as one round of general equilibrium adjustment of city size until $\mathbf{L}^{k+1} = \kappa T\mathbf{L}^k + T\mathbf{t} + \mathbf{c}$ converges to a fixed point.

Note that I, as the key general equilibrium term, plays an important role in generating distributional effect. $(-S)^{-1}$ will generate a trivial distribution of elasticity but $(I-S)^{-1}$ realize the sorting so that elasticity will decay when we look at two otherwise very different cities.

Under tax competition, local government i solves the following problem:

$$\max_{\mathbf{t}} \mathbf{t} \circ \mathbf{L} = \max_{\{t_i\}_{i=1}^n} t_i L(t_i; t_{-i}) = t_i \left[E\mathbf{t} + \mathbf{k}\right]_i \tag{2.36}$$

FOC with respect to t_i gives:

$$\sum_j E_{ij} t_j + \mathbf{k}_i + t_i \frac{d\left(\sum_j E_{ij} t_j + \mathbf{k}_i\right)}{d t_i} = E_{ii} t_i + \sum_{j \neq i} E_{ij} t_j + \mathbf{k}_i + E_{ii} t_i = 0 \tag{2.37}$$

$$\Rightarrow t_i^{TC} = (2E_{ii})^{-1}(-\mathbf{k}_i - \sum_{j \neq i} E_{ij} t_j^{TC}) \tag{2.38}$$

We can write it in matrix form to decompose mechanical and behavioral effects:

$$\underbrace{\underbrace{\begin{bmatrix} E_{11} & E_{12} & \dots & E_{1n} \\ E_{21} & E_{22} & \dots & E_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ E_{n1} & E_{n2} & \dots & E_{nn} \end{bmatrix}}_{ME_t = E\text{: mechanical effect vis-à-vis tax}} \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} + \underbrace{\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}}_{ME_c\text{: mechanical effect}}}_{\text{Total mechanical effect}} + \underbrace{\begin{bmatrix} E_{11} & 0 & \dots & 0 \\ 0 & E_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & E_{nn} \end{bmatrix}}_{BE\text{: Behavioral effect in } TC} \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \mathbf{0} \tag{2.39}$$

Similarly, under tax coordination, a central government solves the following problem:

$$\max_{\mathbf{t}} \mathbf{t}^\top \mathbf{L} = \max_{\{t_i\}_{i=1}^n} \sum_i t_i L(t_i; t_{-i}) = \sum_i t_i \left[E\mathbf{t} + \mathbf{k}\right]_i$$

FOC with respect to t_i gives:

$$\sum_j E_{ij} t_j + \mathbf{k}_i + \sum_k t_k \frac{d\left(\sum_j E_{kj} t_j + \mathbf{k}_k\right)}{d t_i} = \mathbf{k}_i + 2E_{ii} t_i + \sum_{j \neq i} E_{ij} t_j + \sum_{j \neq i} E_{ji} t_j = 0$$

$$\Rightarrow t_i^C = (2E_{ii})^{-1}(-\mathbf{k}_i - 2\sum_{j \neq i} E_{ij} t_j^C)$$

System of FOCs can be written as:

$$\underbrace{\underbrace{\begin{bmatrix} E_{11} & E_{12} & \dots & E_{1n} \\ E_{21} & E_{22} & \dots & E_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ E_{n1} & E_{n2} & \dots & E_{nn} \end{bmatrix}}_{ME_t = E\text{: mechanical effect}} \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} + \underbrace{\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}}_{ME_c\text{: mechanical effect}}}_{\text{Total mechanical effect}} + \underbrace{\begin{bmatrix} E_{11} & E_{21} & \dots & E_{n1} \\ E_{12} & E_{22} & \dots & E_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ E_{1n} & E_{2n} & \dots & E_{nn} \end{bmatrix}}_{BE\text{: Behavioral effect in } C} \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \mathbf{0} \tag{2.40}$$

We can summarize FOCs in two regimes $R \in \{TC, C\}$ as:

$$\mathbf{t}^R = (\underbrace{ME_t + BE^R}_{\equiv F^R})^{-1}(-\underbrace{ME_c}_{=\mathbf{k}}) \tag{2.41}$$

Therefore, equilibrium tax under tax competition is:

$$\mathbf{t}^{TC} = (F^{TC})^{-1}(-\mathbf{k}) = (E + diag(E))^{-1}(-\mathbf{k}) \tag{2.42}$$

Equilibrium in coordination is:

$$\mathbf{t}^C = (2(I-S)^{-1}T)^{-1}(I-S)^{-1}\mathbf{c} = \frac{\bar{a}}{2}\mathbf{b} \tag{2.43}$$

Equation (2.43) states that coorperative tax in each location is only a function of its own TFP, while uncooperative tax is a function of TFP across all locations. This is because $BE^C = ME^\top$. Notice F^C symmetric, $(F^C)^{-1}$ also symmetric. Though we cannot write $(F^{TC})^{-1}$ analytically for large n, we know $(F^C)^{-1}_{ij} < (F^{TC})^{-1}_{ij} < 0, \forall i, j$.

$$\underbrace{(F^C)^{-1}}_{n\times n} = \begin{bmatrix} -\frac{\bar{a}b_1}{2\Lambda_n} - \frac{\kappa\Lambda_n}{2\Lambda_n} & -\frac{\bar{a}b_1}{2\Lambda_n} & -\frac{\bar{a}b_1}{2} & \dots & -\frac{\bar{a}b_1}{2\Lambda_n} \\ -\frac{\bar{a}b_1}{2\Lambda_n} & -\frac{\bar{a}b_2}{2\Lambda_n} - \frac{\kappa\Lambda_n}{2\Lambda_n} & -\frac{\bar{a}b_2}{2\Lambda_n} & \dots & -\frac{\bar{a}b_2}{2\Lambda_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\bar{a}b_1}{2\Lambda_n} & -\frac{\bar{a}b_2}{2\Lambda_n} & -\frac{\bar{a}b_3}{2\Lambda_n} & \dots & -\frac{\bar{a}b_n}{2\Lambda_n} - \frac{\kappa\Lambda_n}{2\Lambda_n} \end{bmatrix} \tag{2.44}$$

Note that

$$F^C = F^{TC} + \begin{bmatrix} 0 & E_{21} & \dots & E_{n1} \\ E_{12} & 0 & \dots & E_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ E_{1n} & E_{2n} & \dots & 0 \end{bmatrix} = F^{TC} + \left[F^{TC} - diag(F^{TC})\right]^\top$$

Since $\mathbf{k}_i < 0\ \forall i$, and notice $BE^{TC} = diag(ME_t), BE^C = ME_t^\top, \left[(F^C)^{-1} - (F^{TC})^{-1}\right]_i < 0,\ \forall i$. Thus with invertibility, $\mathbf{t}_i^{TC} < \mathbf{t}_i^C,\ \forall i \in \{1, 2, ..., n\}$.

We can also show that absolute value of self-elasticity is bounded above. Denote elements in $I-S$ as $c_{i,j}$. Define neighborhood effects as $\mu_i := \frac{\sum_{j\neq i}|c_{i,j}|}{|c_{i,i}|}$. Since $I-S$ is a row strictly diagonally dominant (SDD), thus it is non-sigular and invertible (by Levy–Desplanques Theorem). By Nabben (1999) that applys to any SDD matrix, for any $i \in \mathbf{N}$

$$|\xi_{i,i}| \in \left[\frac{1}{|c_{i,i}|(1+\mu_i)}, \frac{1}{|c_{i,i}|(1-\mu_i)}\right]$$

and

$$|\xi_{i,j}| \leq \mu_i |\xi_{j,j}| \leq |\xi_{j,j}|$$

By the formation of $I-S$, we know that the absolute value of the self-elasticity is bounded above by

$$|\xi_{i,i}| \leq \frac{1}{|c_{i,i}|(1-\mu_i)} = \left(|c_{i,i}|\left(1-\frac{\sum_{j\neq i}|c_{i,j}|}{|c_{i,i}|}\right)\right)^{-1} = 1 \tag{2.45}$$

Equilibrium Allocation

In discrete version, equilibrium allocations can be characterized by inverse matrices. And it can be seen as an approximation of continuum location. On the other hand, given the above, we also can solve for continuum equilibrium allocation in uniform distribution directly. In what follows, we show an approach in solving equilibrium by not identifying elasticity functions from one location to another. This can be regarded as a dual approach to what we use above, but will be handy to solve equilibrium variables (size, productiity and wage functions).

An optimal control approach.

We start from the same model setup with continuum locations. Denote $\mathscr{B} \equiv [\underline{b}, \infty)$. With change of variables and send number of cities to infinity, we can show that optimization problem in continuum case can be analagously written as the following:

$$\max_{T(b),L(b),a(b),a(\underline{b})} \int_{\mathscr{B}} T(b)L(b)dG_b(b) \tag{2.46}$$

$$\text{s.t. } a(\underline{b})\underline{b} + \int_{\underline{b}}^{b} a(\hat{b})d\hat{b} = w(L(b)) + T(b), \ \forall b \in \mathscr{B} \tag{2.47}$$

$$\int_{b}^{\infty} L(\hat{b})dG_b(\hat{b}) = \lambda(1-G_a(a(b))), \ \forall b \in \mathscr{B} \tag{2.48}$$

From (2.48), $L(b)g_b(b) = \lambda g_a(a(b))a'(b) \Rightarrow L(b)g_b(b)db = \lambda d(G_a a(b)) =$

$-\lambda d(1-G_a(a(b)))$. Use this for integration by part (IBP) and we can rewrite $T(b)$ and objective function as

$$T(b) = a(\underline{b})\underline{b} + \int_{\underline{b}}^{b} a(\hat{b})d\hat{b} - w(L(b))$$

$$\Rightarrow \int_{\mathscr{B}} T(b)L(b)dG_b(b) = \int_{\mathscr{B}} \left(a(\underline{b})\underline{b} + \int_{\underline{b}}^{b} a(\hat{b})d\hat{b} - w(L(b)) \right) L(b)dG_b(b)$$
$$= a(\underline{b})\underline{b}\lambda(1-G_a(a(\underline{b}))) + \lambda \int_{\mathscr{B}} a(b)(1-G_a(a(b)))db - \int_{\mathscr{B}} w(L(b))L(b)dG_b($$

Thus we can define

$$\mathscr{W}(b) = a(\underline{b})\underline{b}\lambda(1-G_a(a(\underline{b}))) + \lambda \int_{\underline{b}}^{\infty} a(b)(1-G_a(a(b)))db - \int_{\underline{b}}^{\infty} w(L(b))L(b)dG_b(b) \tag{2.49}$$

Rewrite (2.48) for all $b \in \mathscr{B}$. For any function $\mu(b) : \mathscr{B} \mapsto \mathbf{R}$, we have

$$\int_{\mathscr{B}} \mu(b) \left(\int_{b}^{\infty} L(\hat{b})dG_b(\hat{b}) - \lambda(1-G_a(a(b))) \right) db = 0 \tag{2.50}$$

Suppose costate $\mu(\cdot)$ continuous differentiable. Combine objective (2.49) and (2.50) for $\{L(\cdot), a(\cdot), s\} \subset \mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{R} \equiv \mathbf{X}$ in Banach space.

$$\max_{L(\cdot),a(\cdot),s} \mathscr{F} = \int_{\mathscr{B}} \left(T(b)L(b)g_b(b) + \mu(b) \left(\int_{b}^{\infty} L(\hat{b})dG_b(\hat{b}) - \lambda(1-G_a(a(b))) \right) \right) db$$
$$= \lambda a(\underline{b})\underline{b}(1-G_a(a(\underline{b}))) + \lambda \int_{\mathscr{B}} a(b)(1-G_a(a(b)))db - \int_{\mathscr{B}} w(L(b))L(b)dG_b(b)$$
$$+ \int_{\mathscr{B}} \mu(b) \left(\int_{b}^{\infty} L(\hat{b})dG_b(\hat{b}) - \lambda(1-G_a(a(b))) \right) db$$

Use IBP again:

$$\int_{\mathscr{B}} \mu(b) \left(\int_{b}^{\infty} L(\hat{b})dG_b(\hat{b}) \right) db = \int_{\mathscr{B}} \int_{b}^{\infty} L(b)dG_b(b)d \int_{\underline{b}}^{b} \mu(\hat{b})d\hat{b} \tag{2.51}$$
$$= \left[\int_{b}^{\infty} L(b)dG_b(b) \int_{\underline{b}}^{b} \mu(\hat{b})d\hat{b} \right]_{\underline{b}}^{\infty} + \int_{\mathscr{B}} \left(\int_{\underline{b}}^{b} \mu(\hat{b})d\hat{b} \right) L(b)g_b(b)db \tag{2.52}$$

To see above Gateaux outcomes, first relabel terms:

$$\mathscr{F} = \underbrace{\lambda a(\underline{b})\underline{b}(1-G_a(a(\underline{b})))}_{I} + \underbrace{\lambda \int_{\mathscr{B}} a(b)(1-G_a(a(b)))db}_{II} - \underbrace{\int_{\mathscr{B}} w(L(b))L(b)dG_b(b)}_{III}$$
$$+ \underbrace{\int_{\mathscr{B}} \mu(b)\left(\int_b^{\infty} L(\hat{b})dG_b(\hat{b}) - \lambda(1-G_a(a(b)))\right)db}_{IV}$$

Then we use definition of Gateaux derivatives to solve $a(b)$ (need terms II and IV):

$$\frac{d}{dt}II = \frac{d}{dt}\left(\lambda \int_{\mathscr{B}} (a(b)+tc(b))(1-G_a(a(b)+tc(b)))\,db\right)|_{t=0}$$
$$\Rightarrow 0 = \lambda \int_{\mathscr{B}} c(b)(1-G_a(a(b))) + \lambda \int_{\mathscr{B}} c(b)a(b)(-g_a(a(b)))db + \lambda \int_{\mathscr{B}} c(b)\mu(b)(g_a(a(b)))db$$

for all function $c(b)$. Thus

$$\lambda(1-G_a a(b)) + \lambda a(b)(-g_a(a(b))) + \mu(b)g_a(a(b)) = 0$$

Similarly, we obtain the following conditions from Gateaux derivatives for an arbitrary $\tilde{b} \in \mathscr{B}$, with associated boundary constraint from the free entry condition. This version can be extended to a version with agglomeration externalities in a tractable way.

$$\lambda(1-G_a(a(\tilde{b})) - g_a(a(\tilde{b}))a(\tilde{b})) + \lambda\mu(\tilde{b})g_a(a(\tilde{b})) = 0$$
$$-\left((w'(L(\tilde{b})))L(\tilde{b}) + w(L(\tilde{b}))\right)g_b(\tilde{b}) + g_b(\tilde{b})\int_{\underline{b}}^{b} \mu(\hat{b})d\hat{b} = 0$$

Analytical case with uniform distribution. Assume $w(L(b)) = \kappa L(b)$, $a \sim U[0,\bar{a}]$, $b \sim U[\underline{b},\bar{b}]$. Denote $\Theta \equiv \{\bar{a},\underline{b},\bar{b},\kappa,\lambda\}$, we can solve for equilibrium allocations:

$$a(b) = c_1 e^{\frac{\sqrt{\bar{a}}b}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}} + c_2 e^{-\frac{\sqrt{\bar{a}}b}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}} + \frac{\bar{a}}{2}$$

$$L(b) = \frac{\lambda(\bar{b}-\underline{b})}{\bar{a}}\left(\frac{\sqrt{\bar{a}}c_1 e^{\frac{\sqrt{\bar{a}}b}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}} - \sqrt{\bar{a}}c_2 e^{-\frac{\sqrt{\bar{a}}b}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}}}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}\right)$$

$$T(b) = \frac{e^{-\frac{\sqrt{\bar{a}}\underline{b}}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}}\left(\bar{a}^{3/2} b e^{\frac{\sqrt{\bar{a}}\underline{b}}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}} + 2\sqrt{\bar{a}}\underline{b}\left(c_1 e^{\frac{2\sqrt{\bar{a}}\underline{b}}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}} + c_2\right)\right)}{2\sqrt{\bar{a}}} + c_3$$

$$\mu(b) = 2e^{-\frac{\sqrt{\bar{a}}b}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}}\left(c_1 e^{\frac{2\sqrt{\bar{a}}b}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}} + c_2\right)$$

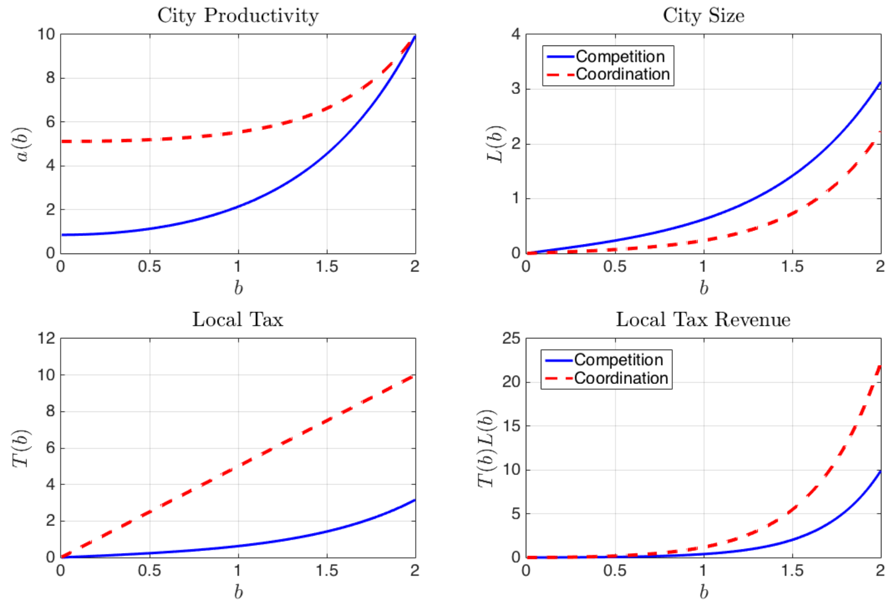
where

$$c_1 \equiv \frac{\bar{a} e^{\frac{\sqrt{\bar{a}}\bar{b}}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}}}{2\left(e^{\frac{2\sqrt{\bar{a}}\bar{b}}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}} - e^{\frac{2\sqrt{\bar{a}}\underline{b}}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}}\right)}$$

$$c_2 \equiv \frac{\bar{a} e^{\frac{\sqrt{\bar{a}}(2\underline{b}+\bar{b})}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}}}{2\left(e^{\frac{2\sqrt{\bar{a}}\underline{b}}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}} - e^{\frac{2\sqrt{\bar{a}}\bar{b}}{\sqrt{(\bar{b}-\underline{b})\sqrt{\kappa}\sqrt{\lambda}}}}\right)}$$

$$c_3 \equiv \frac{2\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}\left(c_2 e^{-\frac{\sqrt{\bar{a}}\underline{b}}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}} - c_1 e^{\frac{\sqrt{\bar{a}}\underline{b}}{\sqrt{\kappa}\sqrt{\lambda}\sqrt{\bar{b}-\underline{b}}}}\right)}{2\sqrt{\bar{a}}}$$

Figure 2.16: Equilibrium Allocation in Uniform Distribution and Linear Congestion Cost



Equilibrium Elasticity

In continuum model with linear congestion cost and uniform TFP and productivity distribution, we solve for elasticity from:

$$-\frac{\bar{a}}{\lambda}\int_p^1 \frac{dL_{\hat{p}}}{dT_r}d\hat{p} - \kappa\frac{d}{dp}\frac{dL_p}{dT_r} = 0, \forall p > r$$

$$-\frac{\bar{a}}{\lambda}\left[-\frac{1}{\kappa} + \int_p^1 \frac{dL_{\hat{p}}}{dT_r}d\hat{p}\right] - \kappa\frac{d}{dp}\frac{dL_p}{dT_r} = 0, \forall p < r$$

$$-\frac{\bar{a}b_0}{\lambda}\left[\int_0^1 \frac{dL_{\hat{p}}}{dT_r}d\hat{p} - \frac{1}{\kappa}\right] - \kappa\frac{dL_0}{dT_r} = 0, \forall r > 0$$

$$F^U(1,r) = 0$$

$$F^U(0,0) + \frac{1}{\kappa} = 0$$

$$F^U(r,r) = F^L(r,r)$$

$$F_p^U(r,r) = F_p^L(r,r)$$

When $p > r$:

$$F(p,r) = c_1(r)e^{\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}} + c_2(r)e^{-\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}}$$
$$\frac{dL_p}{dT_r} = \frac{\sqrt{\bar{a}}c_1(r)e^{\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}}}{\sqrt{\kappa\lambda}} - \frac{\sqrt{\bar{a}}c_2(r)e^{-\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}}}{\sqrt{\kappa\lambda}}$$

When $p < r$:

$$F(p,r) = c_3(r)e^{\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}} + c_4(r)e^{-\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}} - \frac{1}{\kappa}$$
$$\frac{dL_p}{dT_r} = \frac{\sqrt{\bar{a}}c_3(r)e^{\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}}}{\sqrt{\kappa\lambda}} - \frac{\sqrt{\bar{a}}c_4(r)e^{-\frac{\sqrt{\bar{a}}p}{\sqrt{\kappa\lambda}}}}{\sqrt{\kappa\lambda}}$$

Where constant functions $c(r)$ are the following

$$c_1(r) = \frac{e^{-\frac{\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}\left(\sqrt{\bar{a}}b_0\left(e^{\frac{2\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}+1\right)+\sqrt{\kappa\lambda}\left(e^{\frac{2\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}-1\right)\right)}{2\kappa\left(\sqrt{\bar{a}}b_0\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}-1\right)+\sqrt{\kappa\lambda}\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}+1\right)\right)}$$
$$c_2(r) = -\frac{e^{-\frac{\sqrt{\bar{a}}(r-2)}{\sqrt{\kappa\lambda}}}\left(\sqrt{\bar{a}}b_0\left(e^{\frac{2\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}+1\right)+\sqrt{\kappa\lambda}\left(e^{\frac{2\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}-1\right)\right)}{2\kappa\left(\sqrt{\bar{a}}b_0\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}-1\right)+\sqrt{\kappa\lambda}\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}+1\right)\right)}$$
$$c_3(r) = \frac{\left(\sqrt{\bar{a}}b_0+\sqrt{\kappa\lambda}\right)e^{-\frac{\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}+e^{\frac{2\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}\right)}{2\kappa\left(\sqrt{\bar{a}}b_0\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}-1\right)+\sqrt{\kappa\lambda}\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}+1\right)\right)}$$
$$c_4(r) = \frac{\left(\sqrt{\kappa\lambda}-\sqrt{\bar{a}}b_0\right)e^{-\frac{\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}+e^{\frac{2\sqrt{\bar{a}}r}{\sqrt{\kappa\lambda}}}\right)}{2\kappa\left(\sqrt{\bar{a}}b_0\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}-1\right)+\sqrt{\kappa\lambda}\left(e^{\frac{2\sqrt{\bar{a}}}{\sqrt{\kappa\lambda}}}+1\right)\right)}$$

The following is an example for uniform distribution of TFP and productivity. X-axis and y-axis are in terms of CDF, thus both are in $[0,1]$, denoted by p, r. Z-axis is the change in size in city r (or p) when there is a unit tax perturbation in city p (or r).

Figure 2.17: Elasticity Surface in Uniform Distribution and Linear Congestion Cost Model

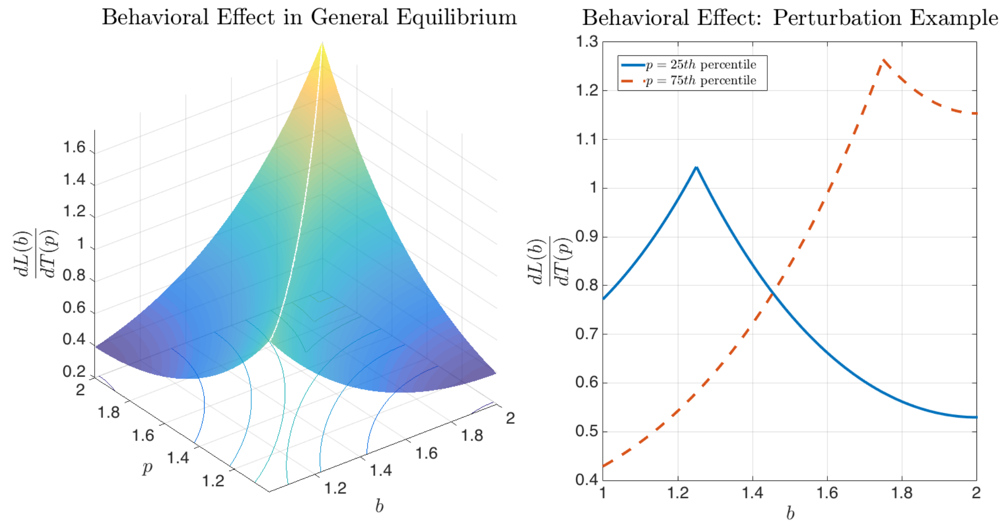


Figure 2.18: Equilibrium with a wider group perturbation

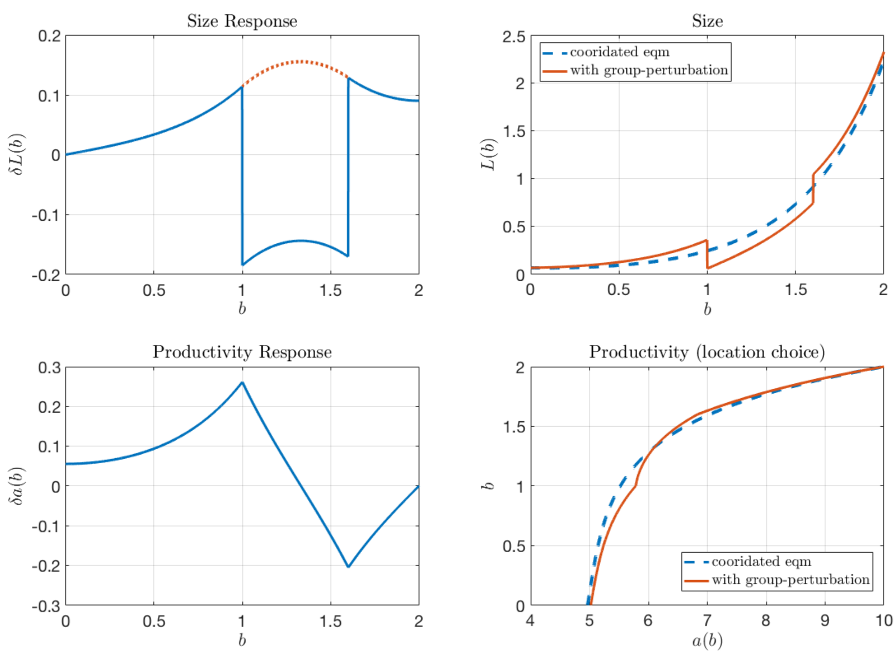
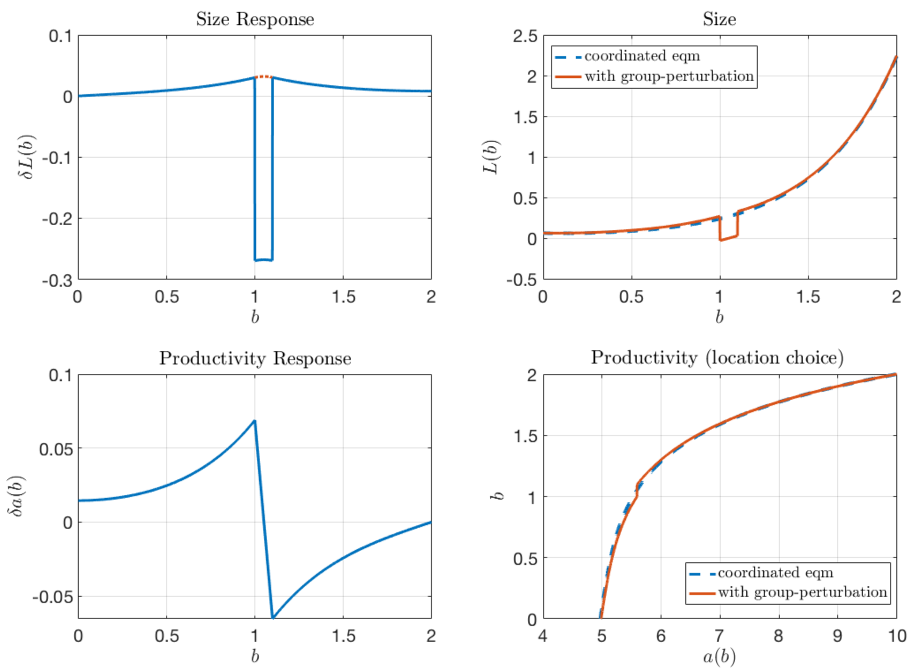


Figure 2.19: Equilibrium with a narrower group perturbation



2.8.11 Empirical Plots and Tables

Facts on Tax Incentive Packages and Related Public Expenditures

Figure 2.20: Mega deals between state/local governments with specific firms in the U.S.

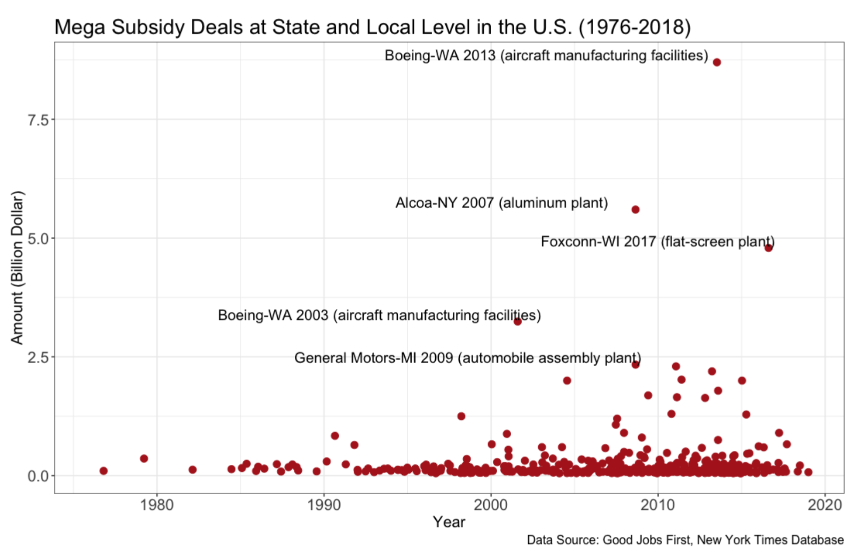


Figure 2.21: Mega deals between state/local governments with specific firms in the U.S. (log scale)

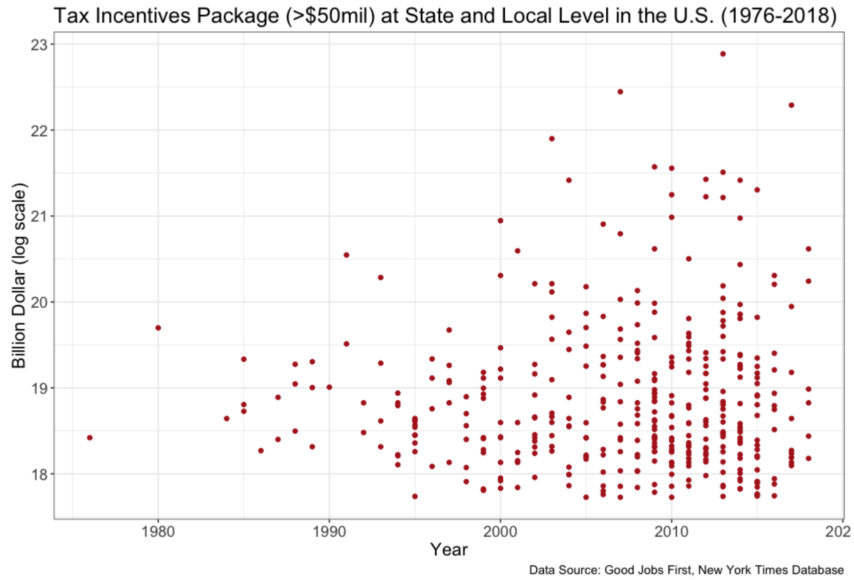
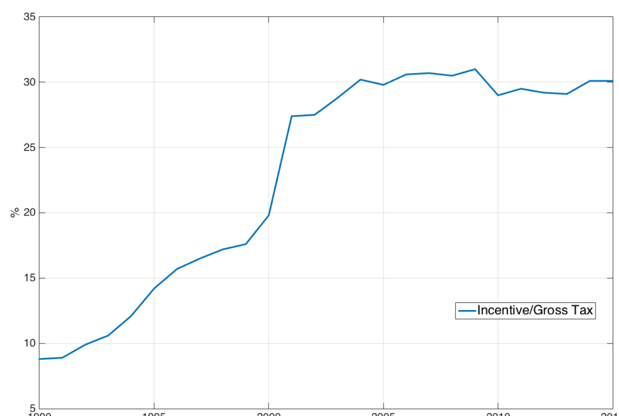
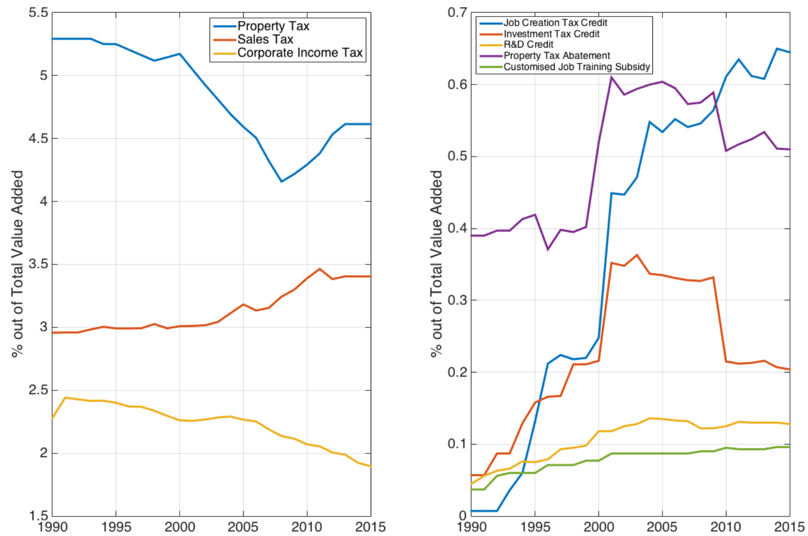


Figure 2.22: Incentive as percentage out of business taxes at state level



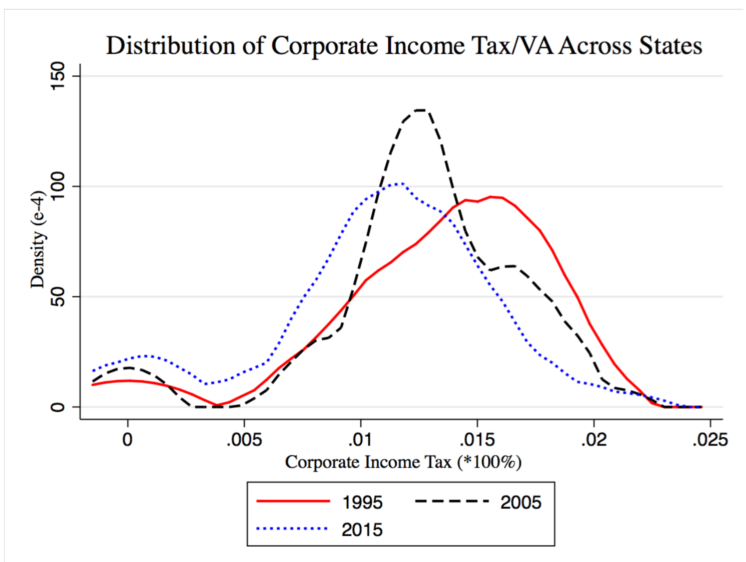
Source: PDIT. Incentives include property tax abatements, customized job training subsidies, investment tax credits, job creation tax credits, and R&D tax credits.

Figure 2.23: Structure of tax incentives



Source: PDIT

Figure 2.24: Distribution of tax shifts to the left



Source: PDIT

International trade and policy incentives

Table 2.1: Correlation between Export and Fiscal Incentives in 2014

Variable	Total Share	Sector Share	Local Share
	(1)	(2)	(3)
Policy Measure			
- ITR	0.000814	0.00508	0.00102
- Total Tax	-0.0405*	-0.634	-0.245
- Total Incentives	-0.00214	-0.375	-0.492
State Fixed Effect	✓	✓	✓
Sector Fixed Effect	✓	✓	✓
Obs.	594	594	594
R^2	0.355	0.628	0.384

Notes: *** p<0.01, ** p<0.05, *p<0.1. “ITR” stands for incentive tax ratio (share of total incentive out of total tax revenue at state level). R^2 in this table is on average across difference policy measures for display simpllicity (same for the plot below on import) Source: PDIT, Census USA Trade.

Table 2.2: Correlation between Import and Fiscal Incentives in 2014

Variable	Total Share	Sector Share	Local Share
	(1)	(2)	(3)
Policy Measure			
- ITR	0.000987	-0.0433*	0.0429
- Total Tax	0.00139	-0.192	-0.372
- Total Incentives	0.0171	-0.477	-0.675
State Fixed Effect	✓	✓	✓
Sector Fixed Effect	✓	✓	✓
Obs.	594	594	594
R^2	0.535	0.648	0.443

Notes: *** p<0.01, ** p<0.05, *p<0.1. “ITR” stands for incentive tax ratio (share of total incentive out of total tax revenue at state level). Source: PDIT, Census USA Trade.

Chapter 3

External Economies of Scale and Place-Based Industrial Policies

Abstract

This chapter studies place-based industrial policies in the presence of production networks and local-sectoral external economies of scale. I document new empirical evidence on the effects of a place-based industrial policy in China. The policy drove employment growth in targeted industries and targeted locations. However, the policy also led to the expansion of untargeted industries in targeted locations, so that the industrial mix in targeted locations did not change significantly. To rationalize these facts, I develop a quantitative general equilibrium model with spatial firm sorting. The model also incorporates policies, production networks, trade, and local sectoral agglomeration externalities. Using numerical simulations, I elucidate model predictions that are in line with empirical evidence on policy effects. In the quantitative analysis, I take the model to Chinese data designing an approach that combines sorting and the exact-hat algebra. I find that targeted industries have higher externalities on average. And results show that the current policy increases total welfare. This gain arises from a combination of externalities and inter-sectoral/-regional linkages that result in net higher competitiveness in both targeted and non-targeted sectors.

3.1 Introduction

Many government policies seek to boost local growth by targeting specific industries and specific regions, i.e., place-based industrial policies (PBIPs).[1] Are these

---

[1]PBIPs abound in developed and developing economies. Examples include, in the United States, *"American Leadership in Advanced Manufacturing,"* implemented at the state and local levels to promote competitive industrial sectors, the *"California Enterprise Zones Program"*, and the *"Texas Enterprise Zones Program"*; in the United Kingdom, the *"Local Industrial Strategies"*. In developing countries, examples include the *"Second Entrepreneurship Initiative in High-Tech Industrial Development Zones"* in China and the *"Finance Act of 1993"* (a program promoting manufacturing in districts designated as backward) in India (Hasan et al., 2017 [74]). These

PBIPs effective? What are the aggregate consequences of PBIPs when industries and regions are interconnected with each other? To design such policies, policymakers would want to understand the likely winners and losers. One common rationale behind these policies is to increase local productivity by taking advantage of local external economies of scale. However, local productivity is affected not only by the total scale but also by the industrial structure. When policies are industry- and location-specific, and when there are both industrial linkages and regional agglomeration, the aggregate effects of policies are ambiguous. Moreover, an environment with PBIPs can be a laboratory in which to study intra- and intersectoral and -regional spillovers. And this provides us a specific perspective to understand industrial policies. In this chapter, I study the aggregate consequences of PBIPs and the potential welfare gains because of the presence of external economies of scale together with industrial and regional linkages.

My context is a specific PBIP in China, the *"Second Entrepreneurship Initiative in High-Tech Industrial Development Zones"* program (hereafter 2002 PBIP). This program is a set of policies mainly about subsidies and preferential tax rates, together with other supports (such as infrastructure and business environment improvements) implemented in high-tech sectors in designated special economic zones. The goal is to promote both local and aggregate economic growth. I first document the fact that 2002 PBIP drives exports in targeted sectors significantly more than other sectors. On top of it, the increase is more significant in policy-targeted zones compared to outside of zones, within policy-targeted cities. On the other hand, the spillover effects along industrial linkages are more significant outside of policy-targeted zones within policy-targeted cities. Then, I provide empirical evidence on changes in the spatial distribution of industry employment in response to these policies using a propensity-score-weighted difference-in-difference estimation and an event-study design. This set of analyses reveals that the 2002 PBIP drives employment in targeted industries to expand in targeted locations. However, the policy also draws untargeted industries to the targeted locations; as a result, the industrial mix there does not change significantly. Specifically, the 2002 PBIP increases employment growth in targeted industries and targeted locations by 16.8 percentage points, but there is a tension between agglomeration and specialization. Targeted industries' shares of employment in targeted locations increase by 0.41 percentage points of the national total in response to the policy. In contrast, targeted industries' shares of local employment do not respond significantly.

---

policies normally offer reduced tax rates, rebates, and/or subsidies. Some are industrial and implemented at the local level, whereas some are industrial but place-neutral. Examples of recent industrial policies are *"Made in China 2025"* (Ju et al., 2021 [75]), *"Manufacturing USA,"* and *"Energiewende,"* in Germany. More general industrial policies include *"Industry 4.0"* in Germany, the *"Franco-German Manifesto for a European Industrial Policy Fit for the 21st Century"* to focus on the industries of the future: renewable energy, artificial intelligence, and robotics in Europe, the *Technology and Innovation Policy in Asian Miracles* (Cherif and Hasanov, 2019 [76]), and the *"Heavy Chemical and Industry (HCI) Drive"* in South Korea (Lane, 2021 [77] and Kim et al. (2021) [78]).

To rationalize the above empirical evidence on policy effects and to guide future policy understandings, I build a quantitative general equilibrium trade model with multiple locations and industries. The model features spatial firm sorting, agglomeration externalities, input-output (I/O) linkages between industries, and trade costs across locations. Firms choose locations and production, considering local agglomeration benefits, costs of congestion, and proximity to suppliers of intermediate goods. Within an industry, more productive firms are disproportionately more productive in larger cities. I also use numerical simulations to illustrate how key statistics change with policies. And comparative statics exercises are taken vis-à-vis the extent of agglomeration externalities, and the extent of I/O linkages. Then, I propose an approach to quantify agglomeration externalities permitting the exact-hat algebra. The idea is to take advantage of equilibrium outcomes determined by model features at the location, industry, and location-cross-industry levels in terms of I/O linkages, policies, externalities, and local sectoral variations. My results show that targeted industries have relatively higher sectoral externalities, which provides a perspective to understand the rationale behind current policies.

Finally, I measure the aggregate welfare effect of current policies through the lens of the model. With a quantified model, I compare the prepolicy and postpolicy scenarios, and I find that the policy generates a 1.71% welfare increase at the aggregate level. This welfare increase comes from the fact that the average local sectoral productivity is boosted, and this quantitatively outweighs the increase in local sectoral costs. The co-existence of firm-sorting and agglomeration accelerates firms' profitability in general while keeping changes in costs relatively small so that economic potential is realized. Sectoral competitiveness increases in both policy-targeted sectors, their close suppliers and heavy industries.

This study sheds light on ways to design and implement (place-based) industrial policies. First, the existence of local sectoral external economies of scale *together* with policies, inter-sectoral linkages, and inter-regional trade provides a rationale for policies. A policy favoring a specific sector in a region affects other sectors in the same region through the local agglomeration externalities and price indices. In addition, it affects all sectors in other regions through firms' choices between locations at the margin, and firms' production choices upon their location choices, conditional on the aggregate price distribution determined in equilibrium. Whether these policies can enhance total welfare depends on how externalities interact with spatial sorting and I/O linkages. Second, policymakers should be careful in trying to support a specific industry-location pair. Further studies will explore welfare analysis based on more appropriate data sets to better extend our understandings on industrial policies.

This chapter makes contributions to literature in the following aspects. On the empirical side, I find new evidence suggesting spillover effects of PBIPs. Other studies estimate the effects of policies, but they focus on either place-based policies (industry-neutral) or industry-based policies (place-neutral). So, we lack studies that evaluate effects of PBIPs. One strand of research explores the effects of place-based policies. For example, Hyun and Ravi (2018) [79] uses the spatial variation in the timing of zonal operations to illustrate the spatial distribution of the impact of special economic zones in India. Lu et al. (2019) [80] find that special economic zones in China have a positive effect on capital investment, employment, output, and productivity. Capital-intensive industries benefit more than labor-intensive ones. However, production networks are not included in these studies. Greenstone et al. (2010) [63] use Million Dollar Plants to estimate the magnitude of local agglomeration. They do not account for industrial and locational heterogeneities together. An exception is a recent study by Kim et al. (2021) [78] that empirically test the effect of the Heavy and Chemical Industry (HCI) Drive in South Korea with both the industrial and the place-based variation. A relevant quantitative study on the HCI Drive is by Choi and Levchenko (2021) [81]. In addition, the potential inefficiencies in my framework come from the agglomeration externalities *together with* industrial linkages and firm sorting. This differs from the policy-neutral results in Klein and Moretti (2014) [82] that relied strong assumption of place-neutral agglomeration.

On the theoretical side, my proposed model combines endogenous spatial firm sorting, agglomeration externalities, I/O linkages between industries, trade costs across locations, and PBIPs. The model inherits agglomeration elements as in Behrens et al. (2014) [6] and Gaubert (2018) [7]. Moreover, I add a multilocation trade framework and I/O linkages as in Caliendo and Parro (2015) [83]. Relevant studies such as Allen and Arkolakis (2014) [84], Costinot and Rodríguez-Clare (2014) [85], Bartelme et al. (2019) [86], Liu (2019) [87], Liu and Ma (2021) [88], Carvalho et al. (2021) [89] and Bakker (2021) [90] incorporate different subsets of the elements in this chapter, but not all together. For example, Liu (2019) [87] studies industrial linkages in environments without agglomeration externalities. Carvalho et al. (2021) [89] study Japanese earthquake shock to identify agglomeration effects, but they do not include variation at the local level. Tian (2021) [91] explores city productivity advantages and the division of labor based on occupational complexity with Brazilian establishments. But this study focuses more on labor specialization instead of sectoral specialization, and thus do not further speak to the scenario when I/O linkages intertwine with agglomeration externalities. Bakker (2021) [90] finds that export intensity is higher in more densely populated areas. Firms in larger cities are more productive and less low-skill-intensive because of comparative advantages, but there are no I/O linkages. Regarding the international trade theories, my model inherits the gravity-type results, but it further extends current parametric forms found in the seminal work by Costinot and Rodríguez-Clare (2014) [85] by embedding firm sorting and agglomeration externalities. Caliendo et al. (2018) [92] consider the aggregate and disaggregate effects of regional and sectoral shocks to productivity and infrastructure. The shock this chapter considers is similar vis-à-vis location-sector combination, but it is different from Caliendo et al. (2018) [92] in the sense that I allow externalities, policies, and locational choices of heterogeneous firms.

The rest of this chapter proceeds as follows. Section 3.2 introduces institutional background and details the empirical facts on local and aggregate effects of the policy. Section 3.3 develops a quantitative spatial general equilibrium model. Section 3.4 shows numerical simulations with supplementary evidence. Section 3.5 quantifies the model and measures the aggregate and distributional effects. Section 3.6 concludes.

3.2 Empirical Evidence on Policy Effects

3.2.1 Institutional Background

This chapter is the first to attempt to evaluate the place-based industrial policy (PBIP) known as the *"Second Entrepreneurship Initiative in High-Tech Industrial Development Zones" (2002)* in China. This policy targets at special economic zones that are designated to focus on the development of high-tech industries (HTIs). Since 2002, local governments in these national high-tech zones (HTZs) have issued local policies to boost the growth of HTIs in such ways as providing subsidies or tax rebates and building infrastructure to improve local business environments especially for high-tech sectors.[2] This initiative was documented in a nation-wide departmental regulatory document titled *"Decision of the Ministry of Science and Technology on Further Supporting the National Development Zones for New and High-Technology Industries"* promulgated by the Ministry of Science and Technology under the supervision of the State Council in January 2002.[3] Following this national guideline, the Ministry of Science and Technology issued *"Some Opinions of the Ministry of Science and Technology Concerning the Reform and Innovation of the Administration System of the National Development Zones for New and High Technology Industries"* in March 2002.[4] Guided by these two national policy documents, local governments in National New- and

[2]I thank useful institutional background information from the Development Research Center of the State Council, P.R.C.

[3]This departmental regulatory document was promulgated by the Ministry of Science and Technology on January 31, 2002 in the *No.32 [2002] of the Ministry of Science and Technology*, with an English version is available. (Archived online with the *"China Law Info"* (CLI) citation code CLI.4.48812(EN) at en.pkulaw.cn) The official explanation for this initiative wrote as *"For the future 5 to 10 years, the national new and high-tech zones shall come into the secondary period of starting the undertaking, during which the innovation of science and technology, and that of the system shall be taken as the momentum, and the main task shall be the fostering of the new- and high-tech industries."*

[4]This policy document was promulgated by the Ministry of Science and Technology on March 5, 2002 in the *No.61 [2002] of the Ministry of Science and Technology*. The English version is available with the CLI citation code CLI.4.48813(EN) at en.pkulaw.cn.

High-Tech Zones responded and initiated local policies to boost the development of high-tech industries.[5]

In what follows, I explain why this 2002 policy is both a place-based policy and an industrial policy. As a *place-based* policy, this PBIP targets the *"National Development Zones for New and High Technology Industries"* (hereafter, HTZs). Between 1994 and 2007, there were 53 HTZs in 53 major prefectures or municipalities in China (hereafter, HTCs).[6] These cities are mostly the capital and major cities in each province. Figure 3.1 shows the geographical distribution of high-tech cities (HTCs) and a sample map for HTZs in one of HTCs. Specifically, Figure 3.1-Left shows the geographical distribution of cities with HTZs in each of them in a map of China. Figure 3.1-Right is an example comparing HTZs (yellow) and non-HTZs (green) on a map of Beijing,[7] which is one of the earliest HTCs. The HTZ in Beijing is also known as Zhongguancun Science Park. It includes several sub-science parks that are all managed by the Administrative Committee of the Beijing HTZ. Each of the 53 HTZs has a similar administrative structure under the leadership of a Management/Administrative Committee of their own. Each committee is a government agency with responsiblity to macro-manage the high-tech zone, to connect municipal government and private sectors, and to implement related local policies.[8]

[5]I have not found systematic records for local implementation across all high-tech cities. But I found some evidence in official and relevant reports on which local governments responded to these national guidelines, see Appendix 3.7.

[6]These zones are in municipalities including Beijing, Tianjin, Shanghai, and Chongqing; the prefectures include Shijiazhuang, Baoding, Taiyuan, Baotou, Shenyang, Dalian, Anshan, Changchun, Jilin, Harbin, Daqing, Nanjing, Changzhou, Wuxi, Suzhou, Hangzhou, Hefei, Fuzhou, Xiamen, Nanchang, Jinan, Qingdao, Zibo, Weifang, Weihai, Zhengzhou, Luoyang, Wuhan, Xiangfan, Changsha, Zhuzhou, Guangzhou, Shenzhen, Zhuhai, Huizhou, Zhongshan, Foshan, Nanning, Guilin, Haikou, Chengdu, Mianyang, Guiyang, Kunming, Xi'an, Baoji, Xianyang, Lanzhou, and Urumqi. There was only one addition to the set of high-tech cities during this period. Note that these zones are at national level. After 2007, many new provincial and local high-tech cities/zones were designated.

[7]The Beijing HTZ map comes from Zhongguancun Administrative Agency website (http://zgcgw.beijing.gov.cn/zgc/zwgk/sfqgk/sfqjs/yqdy/index.html). This is an updated graph for 2019. The analogous graph of 2002 has not been found in public archives.

[8]The Management Committee of a high-tech zone has comprehensive functions in finance, economic development, territory management, planning, investment, talent service, etc. Some are more market-oriented (such as Zhongguancun in Beijing), whereas some others have greater authorization from their local governments (such as the HTZ in Xiangfan).

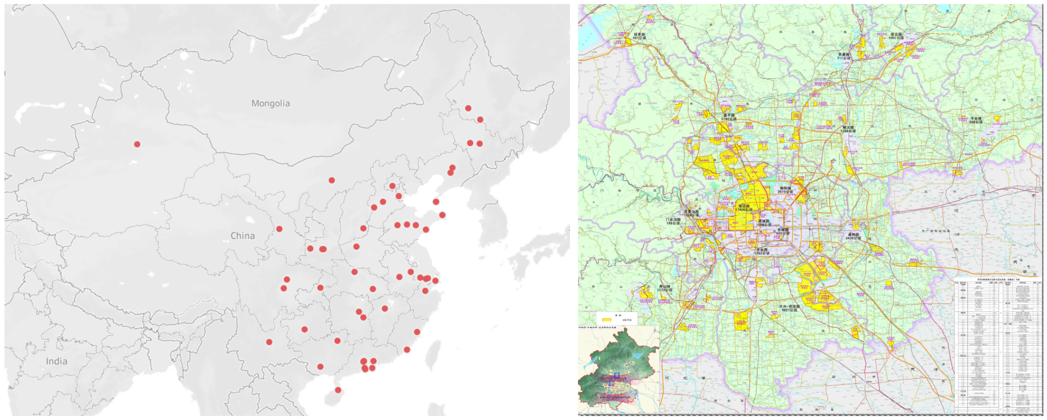


Figure 3.1: Left: Geographic distribution of high-tech cities (HTCs) with high-tech zones (HTZs) in China. Right: Geographic distribution of HTZ (yellow) v.s. non-HTZ (green) in Beijing

As an *industrial* policy, the 2002 policy emphasizes the goal of promoting high-tech industries. These high-tech industries are mostly in the manufacturing sector and in high-value-added industries. Specifically, there are eight fields designated as policy targets: (1) electronic information technology; (2) medical science and biomedical engineering; (3) aerospace and aeronautical technology; (4) new materials technology; (5) high-tech service; (6) new energy science and energy conservation technology; (7) ecology and environmental science; and (8) other new process or new technology applicable in the traditional industries.[9] This is the official list based on Chinese Industry Classification (hereafter, CIC) with 62 industries at CIC (2011 version, GB/T 4754-2011) 4-digit level.[10]

[9]These fields were not emphasized by the Chinese government for the first time in 2002. Instead, if tracing the historical policy documents, we can see that the Chinese government had prepared itself in implementing this policy and moved forward in upgrading on the global value chains. In fact, high-tech development zones are an important part of China's Torch Program strategized by the State Council in 1980s. These fields can be found in *"Notice of the Ministry of Science and Technology on Issuing the Conditions and Measures for Accreditation of High and New Technology Enterprises in National High Technology and New Technology Industry Development Zones (2000 Revision)"* This is an adjusted version issued by the State Council in mid-2000. The official Chinese version can be found at http://www.gov.cn/gongbao/content/2001/content_60688.htm, with English version search code CLI.4.205008(EN) available at en.pkulaw.cn. The original version *"Conditions and Measures on the Designation of High and New Technology Enterprises in National High and New Technology Industry Development Zones"* was approved by the State Council on March 6, 1991. And it was promulgated by the State Science Commission in March 1991. The English version search code CLI.2.5033(EN) at en.pkulaw.cn.

[10]There are 42 sectors described as the "third layer" in the official document of HTI (Manufacturing) Classifcation (2013). And there are 62 specific CIC-4 codes specified in the above document based on the GB/T 4754-2011 system, see Appendix 3.7 Table 3.4. The design of the high-tech-sector classification in China refers to the OECD ISIC Rev.3 Technology Intensity Definition (2011), which classifies manufacturing industries into categories based on R&D intensities. In ISIC Rev.3, *High-technology Industries* consist of (1) aircraft an spacecraft, (2) pharma-

In the empirical analysis, I assigned the designated industries based on this official document with industry codes manually crosswalked with the 2002 version, which include 60 industries at the CIC (2002 version, GB/T 4754-2002) with both 3-digit and 4-digit levels.

So far, there is no public official documents systematically recording the data on this policy implemented at the local level. In the Appendix 3.7, I list some examples from summary reports by local governments. As noted in the 2002 implementation reports by local governments, one goal of this PBIP is to *"extend the industry chain"* within the city and to *"promote technological advancement in related industries"*. These official statements suggest that local governments intend to take advantage of industrial linkages and encourage economic expansion and technological changes in industries that are connected to targeted ones.

In the rest of this chapter, for notation simplicity, I use the following abbreviations to refer to targeted locations or targeted industries in Section 3.2. HTI refers to high-tech industries. These are 4-digit CIC codes that are categorized as high-tech (targeted) industries by the National Bureau of Statistics. HTP refers to high-tech products. These are 8-digit Harmonized System codes that are categorized as high-tech (targeted) industries by the General Administration of Customs of China. HTC refers to high-tech cities. These are 4-digit prefecture-level geography codes (under the Chinese National Bureau of Statistics coding system) that contain an HTZ within the prefecture border. HTZ refers to high-tech zones. These are 5-digit zone/district-level geographical codes (under the Customs coding system) that have been officially approved as National Development Zones for New and High Technology Industries by the Ministry of Science and Technology of China.

3.2.2 Facts

This section presents main facts about the spatial distribution of economic activities corresponding to the implementation of the 2002 policy. Main takeaways are as follows. First, at the granular level, 2002 PBIP drives exports in targeted sectors significantly more than exports in non-targeted sectors. And this disproportionate increase is more significant in targeted zones within targeted cities. Second, at the aggregate level, the PBIP drives employment in targeted industries to expand in targeted locations. However, the PBIP also drives up employment in untargeted industries and targeted locations, so that the industrial mix in targeted locations does not change significantly.

First, I utilize the location information of export production in the detailed customs data from China to analyze the total export performance within HTCs. This information is more granular than prefecture labels, as, so far, the most relevant and systematic one allowing us to seperate the associated 2002 PBIP

ceuticals, (3) office, accounting and computing machinery, (4) radio, TV and communciations equipment, and (5) medical, precision and optical instruments.

targets. Within a prefecture, it specifies the production areas and names special economic zones explicitly.[11] This allows me to seperate zones within HTCs, which effectively helps control for idiosyncratic shocks at above-zone level, including common shocks that standard literature concerns, such as the international trade exposure.[12]

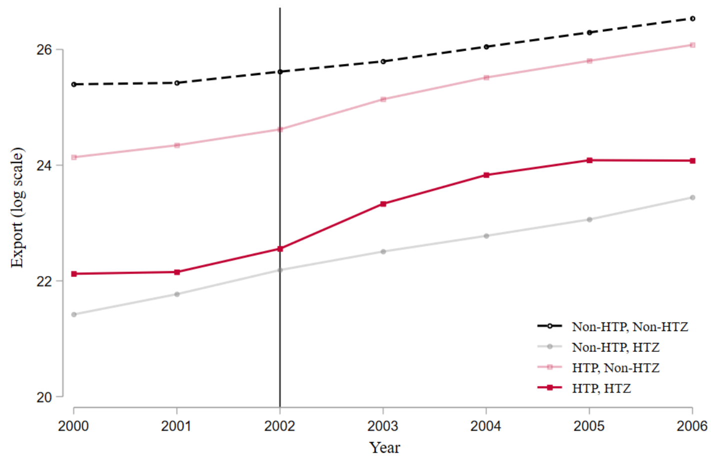


Figure 3.2: Export in HTC by policy groups

For trade pattern, I use a difference-in-difference design to estimate the effects of the 2002 PBIP within HTCs as follows:

$$ln(Export)_{ijt} = \alpha Target_{ij} \times Post_t + FE + \epsilon_{ijt}$$

The dependent variable is the total exporting value (in log scale) at the zone-product-year level. $Target_{ij}$ indicates whether observables belongs to policy-targeted group $\mathbf{1}(HTZ_i \times HTI_j)$, and $i \in HTC$. FE denotes a full set of fixed effects $(\delta_i, \delta_j, \delta_t, \eta_{it}, \eta_{ij}, \eta_{jt})$. The coefficient of interest is α. Regression results

[11] Within a few prefectures, the exact names changed, while the Customs flag categories kept the original name. A particular example is Beijing, where the zone name associated with the current “Zhongguancun” in the Customs data is “Beijing New Tech Industry Development Zone”, which was designated as “Zhongguancun” in the official document from the State Council (national level) to Ministry of Science and Technology and Government of Beijing in 1999. This is partially due to a differentiation between the major raw policy-zone name with some sub-zones (such as “Changping Science Park”). In 2006, some of these sub-zones were re-organized to “Zhongguancun” while keeping the total zone area the same, together with a new wave of newly-approved local special economic zones in other prefectures. The main empirical analysis in this chapter uses the conservative measurement. In Appendix 3.7.4 Figure 3.14, I supplement alternative measurement adding sub-zones in Beijing. Main empirical facts remain the same. Furthermore, current empirical analysis used the conservative measure conditional on HTZs given the major source of the discrepancy.

[12] As a supplementary check, I use the World Integrated Trade Solution (WITS) data set for changes in tariffs. It turns out that HTIs do not receive a disproportionate decrease in tariffs on the international trade market, compared to these in other industries.

are shown in Table 3.1. These results suggest that the total exporting value has been increased by around 20.11 percentage points on average responding to the implementation of the policy, if a product-zone observation belongs to targeted zones and targeted industries, compared to those that are not in policy-targted categories. This number is statistically significant and is robust when using different levels of fixed effects, where column (2) in Table 3.1 is at more granular level.[13]

Table 3.1: Regression results with granular trade data

Variables	Exports	Exports
	(1)	(2)
$Target_{ij} \times Post_t$	0.2152**	0.2011***
	(2.4000)	(2.6686)
One-way FEs	✓	✓
Two-way FEs	✓	✓
Obs.	1,354,809	1,211,321
R^2	0.1089	0.7679

(t statistics in parentheses, with clustered s.e.)
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Notes: In column (1), geographical fixed effects are at the prefecture level, and the sectoral fixed effects are at the HS-4 level, so as the interacting FEs. In column (2), geographical fixed effects are at the zone level, and the sectoral fixed effects are at the HS-8 level, so as the interacting FEs.

I then use event studies with the following specification:

$$ln(Export)_{ijt} = \sum_{t=2000}^{2006} \beta_{t-2002} Target_{ij} \times Year_t + FE + \epsilon_{ijt}$$

Figure 3.3 shows the event study coefficient β_t from 2000 ("−2" means 2 years before policy) to 2006 ("+4" means 4 years after policy).[14] As standard in literature, I normalize the effect in the year before policy implementation (-1) to be 0. Standard errors are two-way clustered at 95% level.[15] The results indicate that the average effect starts to be statistically significant after the policy implementation. Conditional on targeted prefectures, the PBIP drives up the export of high-tech products (HTPs) in high-tech zones (HTZs) significantly. This effect

[13]There are no systematic granular (zone-level) controls available for this analysis.

[14]Compared to ASIE, the smaller time range in both IPW-DID and event studies (shown later) is because of the data availability. The China Customs Trade Data starts the trade transactions record from 2000, while ASIE starts from 1998.

[15]I conducted robustness checks by varying clustered s.e. and the event study plots do not change significantly for both set of regressions.

is significant in HTZ and HTP within HTC, compared to HTP outside of HTZ in HTC, and compared to non-HTP within HTZ in HTC.

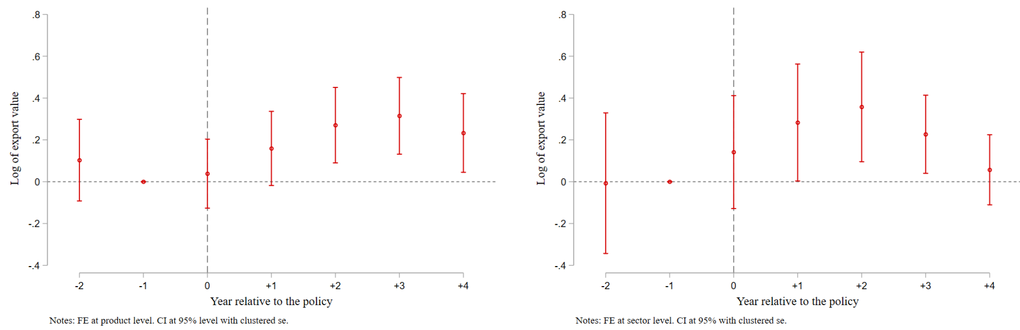


Figure 3.3: Plots of event study coefficients. Left: with granular FEs. Right: with broader level FEs. (See notes in plots)

Apart from the HTC-HTZ decomposition, I also seperate sectors that are “closely-related” to policy-targeted sectors based on the Input-Output (I/O) Table. Note that the raw I/O Table does not capture the specialty of policy-targeted industries in the network intuitively. Some recent studies have shown that targeted industries picked in industrial policies in trade wars are special in terms of substitutability (Ju et al., 2021 [75]). This requires a measure, potentially corresponding to the policy, to capture how “closely-related” industries respond to PBIP differently from others in terms of their relation to policy targets. There are many measures in standard I/O literature.[16] The measure illustrated for simplicity is the relative demand coefficient.[17] I compute key set of coefficients from Leontief inverse b_{ij}, the element of $B \equiv (I-A)^{-1} - I$, where $a_{ij} = \frac{x_{ij}}{X_j}$ (the element of A) indicates how industry j is consuming industry i out of total input. It essentially measures how demand from other industries increases when there’s a unit of a final good consumed from an arbitrary industry j: $F_j = \frac{\frac{1}{n}\sum_{i=1}^{n} b_{ij}}{\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} b_{ij}}$. Result suggests that targeted industries are at far downstream. This means high-tech industries are more inclined to be consumed as final goods instead of intermediate inputs in the supply chain. Appendix 3.7.8 Figure 3.27 shows the relative demand coefficients across industries. Red bars indicate high-tech industries and they rank relatively high in the supply chain. Green bars are that in

[16] Alternatively, one could utilize the accounting identity of usage of goods in an industry for an *upstreamness* (U) measure, proposed by Antras et al. (2012) [93]: $U = (I-\Delta)^{-1}\mathbf{1}$, where Δ is the matrix where entry (i,j) represents the share of sector i’s total output that is purchased by industry j. This statistic essentially summarizes average distance between an industry and final-good consumers. Details are in Appendix 3.7.8. I also refer to industry research for high-tech industries on supply chain analyses. These are similar regarding the ranking of industries. Sectors closely related to targeted high-tech industries turn out to be mostly heavy industries that are direct upstreams of the targeted ones.

[17] This has been widely used in actual input-output statistics in country reports and handbooks for national accounting.

manufacturing sector but not high-tech industries. Blue bars are other sectors. Because of the different coding system between CIC and HS, I refer to both the measure computed above and equity research on high-tech industries to include the heavy industries excluding wastes, utility and resource mining sectors.[18]

In Figure 3.4, I plot the total export (log scale) within and outside of HTZs within HTCs by sector groups. Within HTZs and HTCs, the total export in HTPs (the dark red line) increases more than other sectors. Within non-HTZs and HTCs, the total export in both HTPs and sectors closely related to HTPs (the dark red and light red lines) increase more significantly than the "Far" sectoral group (the gray line).[19] This is within expectation since HTZ is a small subset of HTC in terms of geographic areas. When looking at granular zone-level statistics, it makes sense that only policy-relevant industries respond. Notice that the growth of "close" industries starts to catch up HTI from 2003 and they exhibit similar trends in growth afterwards.[20]

[18]These close-related sectors are petroleum processing, coking and nuclear fuel manufactur, chemical feedstock and chemical manufacturing industry, chemical fiber manufacturing industry, rubber production industry, plastic industry, non-metallic minerals product industry, ferrous metal smelting and extrusion, non-ferrous smelting and extrusion, metalwork industry, general-purpose equipment manufacturing industry. If the granular industry codes are already in the HTI catergory, they are excluded from the midsteam sectors. The NBS categories between heavy and light are based on production and its usage.

[19]Formally, I use the following regression specification within HTC and outside of HTZ.

$$ln(Export)_{ijt} = \beta Close_{ij} \times Post_t + FEs + \epsilon_{ijt}$$

where $Close_{ij} = 1$ if observations are in HTC and the closely related industry group of HTI. FEs consist of the complete set of one-way fixed effects (with j at both sector and product level respectively). Results are shown in Appendix Table 3.5.The estimate is around 14.47 percentage points increase in relevant sectoral group compared to the control group.

[20]As a supplementary check, I also plot the average change from the base year (2000) so that we can get rid of the level difference. Figure 3.24 shows the average growth for non-HTP within HTC (inside HTZ v.s. outside HTZ.) Average growth is current-year trade volume divided by base-year trade volume. This deal with the magnitude difference of absolute value in each group and focus on the growth trend difference. Figure 3.24-left graph suggests that within HTZ in HTC, growth rate of both "Close" and "Far" group experienced similar trends. Figure 3.24-right graph suggests that outside of HTZ (in HTC), growth rates in "Close" group exceeds that in "Far" group significantly from 2003.

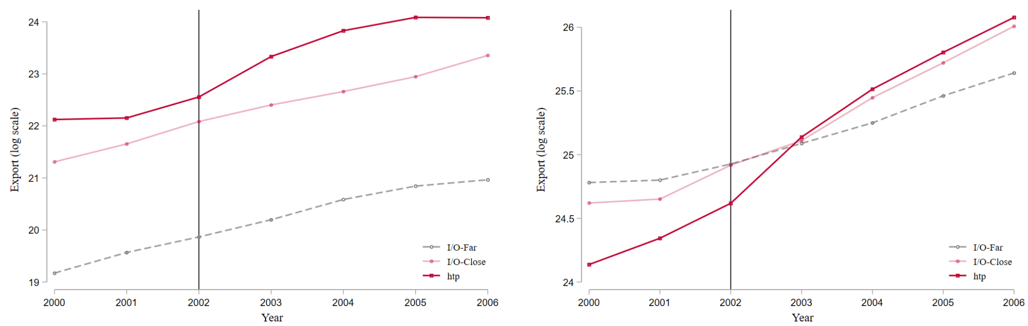


Figure 3.4: Left: Export in HTZ, in HTC. Right: Export outside of HTZ, in HTC

As a placebo check, I also show the average growth outside of high-tech cities in Figure 3.25. It suggests that "close" group still exceeds that in "far" group. This can be an evidence on the spillover effects of the PBIP that happens not only in targeted locations, but also outside of them.

Regarding the distribution of employment within and outside of targeted regions/sectors, I start with plotting the empirical cumulative distribution of targeted industries (HTIs) across prefectures in Figure 3.5.[21] Each dot represents the total employment of each prefecture in HTIs in 1998 and 2007. This plot shows a concentration of employment at the top, reflected by a right shift of the empirical cumulative distribution from the red dotted curve (1998) to the green curve (2007). The median employment of the total distribution does not change significantly. Moreover, at the bottom, the reduction of employment is reflected by a left shift below the median.

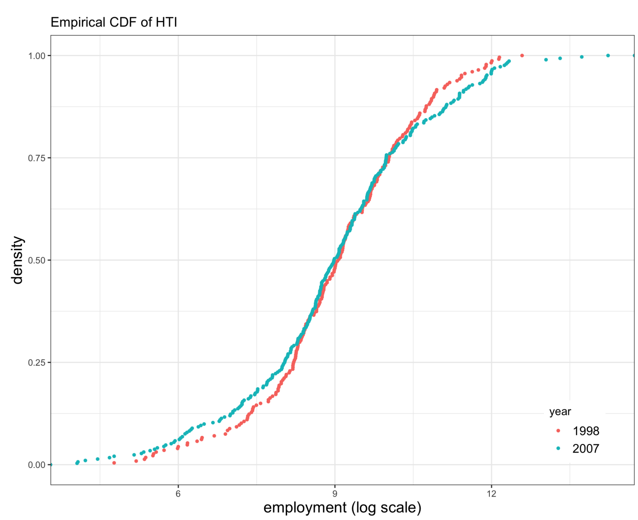


Figure 3.5: Empirical distribution of high-tech industries

[21]This distribution comes from the author's aggregation from the firm-level dataset (Annual Survery of Industrial Enterprises).

The plot of the distributional changes of non-targeted sectors is shown in Appendix 3.7.3 Figure 3.15. The non-HTIs experienced a decrease in most locations, and there was a less significant increase in employment at the top, compared to that of HTIs. This is reflected by a marginal twist at the very top driven by the top cities.

Furthermore, I use the following additional descriptive facts to illustrate this concentration pattern in HTIs compared to that in non-HTIs. First, I use the Herfindahl–Hirschman Index (HHI) at a more aggregated level to show a similar concentration pattern of HTI-concentrated industries.[22] Second, I show the scale-rank plots for the truncated local sectoral employment for HTIs (see Appendix 3.7.3 Figure 3.17) and non-HTIs (Figure 3.18) to compare the distributional change at the top. These two plots essentially show the tail distribution of Figure 3.15. Figure 3.17 shows the significant concentration of the HTIs is driven by the top prefectures, followed by the subsequent ones ranked by size. Besides distribution by industrial targets, I also decompose the data by locational targets. In Figure 3.19 (left), I decompose the employment within HTIs by whether they are located in targeted cities. That is, I show two groups of employment, based on whether it occurs in HTIs. The red line shows total employment in HTCs (denoted by "P" in the plot) and the black solid line shows employment in non-HTCs ("NP"). Both levels increase over time, but employment in HTIs in HTCs increases faster than it does outside of policy-targeted cities. In Figure 3.19 (right), I decompose the employment within HTCs by whether they are categorized as targeted industries. The red line is the same as the left plot, targeted location-industry pairs. The black solid line is the employment level in non-HTIs ("NH" on the y-axis), conditional on whether they are in HTCs. This plot shows increases in both groups that have more aligned paces after policy implementation while not parallel before policy.

Therefore, I conduct a difference-in-difference estimation with a propensity-score approach embedded at the aggregate level. The designated HTCs are mostly the capitals in the provinces, together with a set of relatively developed prefectures. To help address endogeneity, I use the pre-treatment period to generate the propensity scores for being in the treatment group from observables.

$$\mathbf{P}\left(\mathbf{1}(treatment)_{ij}|\mathbf{X}\right) = f\left(\beta_0 + \mathbf{X}_{ij}\beta + \mathbf{X}_i\eta\right)$$

where $\mathbf{1}(treatment)_{ij}$ indicates whether ij is in HTP and HTZ before 2002.[23]

[22]Industry-level concentration measured by HHI in this chapter is specified below,

$$HHI_{jt} = \sum_i \left(\frac{\sum_e x_{eijt}}{\sum_i \sum_e x_{eijt}} \right)^2$$

where x_{eijt} is employment in firm e, location i, industry j, year t. See results in Appendix 3.7.3 Figure 3.16.

[23]Robustness checks are done in using different combinations of pre-policy year to estimate propensity scores.

$f(x)$ is the logistic function. $\mathbf{X}_{ij}$ are observables aggregated from firm level in NBS dataset, including city-industry-time level leverage, number of export, number of SOEs. $\mathbf{X}_i$ are aggregates from prefecture-level Statistical Yearbook, where I include local total population, employment in 2nd (industrial) sector, employment in 3rd (service) sector, local total public expenditures, local public expenditures in science and technology, and some other variables so that we obtain a propensity score for each city-industry aggregates in terms its probability of being in treatment group.[24] Then, I construct inverse propensity weights (IPW) as follows

$$\mathscr{L}_{ijt} = \begin{cases} \frac{1}{\hat{P}(X_{ijt})}, & \text{if } ijt \in \text{Treatment group} \\ \frac{1}{1-\hat{P}(X_{ijt})}, & \text{if } ijt \in \text{Control group} \end{cases}$$

Figure 3.6 shows propensity scores for treatment and control groups before and after adjustments. Differences in raw and weighted covariates are summarized in Section 3.7.5 Figure 3.23. I also plot some other variables before and after inverse propensity weighting.[25]

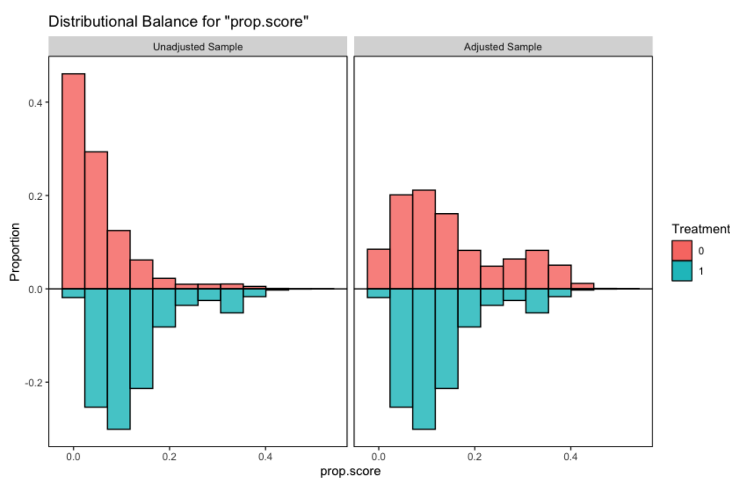


Figure 3.6: Comparison of Propensity

I then estimate the equation with IPW:

$$y_{ijt} = \alpha Target_{ij} \times Post_t + FE + \mathbf{X}_{ijt}\gamma + \epsilon_{ijt}$$

Dependent variables are the level of employment in HTIs in HTCs, the concentration ratio (shares of employment in HTCs conditional on industries), and the

[24]Supported by the export-led policy intention, the propensity assignment precludes observations without export values.

[25]To give a more extensive picture of the data, in Appendix 3.7.5, I plot other propensity-weighted aggregates. See Figure 3.20 for sales, Figure 3.21 for public expenditure in science and technology, and Figure 3.22 for export intensity.

specialization ratio (shares of employment in HTIs conditional on cities). These two sets of ratios are specified in detail together with the regression results. $Target_{ij} \equiv \mathbf{1}\left(HTC_i \times HTI_j\right)$, i.e., 1 if an observation (a location-sector pair) belongs to the PBIP targets. $Post_t \equiv \mathbf{1}(t > 2002)$ indicating post-policy periods. FE consists of one-way (location, sector, time) fixed effects and a full set of two-way interacting fixed effects. $\mathbf{X}_{ijt}$ indicates a set of controls. Specifically, ijt-level export intensity is computed with $\frac{\sum_e \mathbf{1}(x_{eijt} \in (\{Export\}, ijt))}{\sum_e \mathbf{1}(x_{eijt} \in ijt)}$. ijt-level state-owned enterprise (SOE) intensity is computed with $\frac{\sum_e \mathbf{1}(x_{eijt} \in (\{SOE\}, ijt))}{\sum_e \mathbf{1}(x_{eijt} \in ijt)}$. And I also include mean leverage ratio at ijt-level aggregated from firms' information. Outcome variables are either levels or shares of employment. Regression results are shown in Table 3.2.

Table 3.2: IPW-DID Estimation

Variables	**Share** (j) (1)	**Share** (j) (2)	**Share** (i) (3)	**Share** (i) (4)	**Level (log)** (5)	**Level (log)** (6)
$Target_{ij} \times Post_t$	0.0042**	0.0041**	0.0001	-0.0000	0.1636**	0.1677**
	(2.2139)	(2.1574)	(0.0699)	(-0.0057)	(2.2129)	(2.3255)
Year FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Interacting FEs	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓
Obs.	36,292	36,260	36,292	36,260	36,292	36,260
R^2	0.8896	0.8903	0.9195	0.9196	0.8806	0.8879

(t statistics in parentheses, with clustered s.e.)
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

This table reports three types of outcome variables. First four columns are relative shares, and the last two columns are levels (log scale). Specifically, in Column (1) and (2), $Share(j) \equiv s_{ijt|j} \equiv \frac{\sum_e x_{eijt}}{\sum_i \sum_e x_{eijt}}$ are the shares of employment in city i, industry j and year t, conditional on industry j. This measures spatial concentration of an industry. I use concentration ratio, λ^{ind}, to refer to these shares in later sections. In Column (3) and (4), $Share(i) \equiv s_{ijt|i} = \frac{\sum_e x_{eijt}}{\sum_j \sum_e x_{eijt}}$ are the shares of employment in city i, industry j and year t, conditional on city i. This measures industrial specialization within a location. I use specialization ratio, λ^{loc}, to refer to these shares hereafter. In Column (5) and (6), $Level(log) \equiv y_{ijt} = \sum_e x_{eijt}$ are total employment in city i, industry j and year t in log scale.

Since our purpose is to evaluate the effects of the policy, the key difference-in-difference coefficient (α) is what we are interested in. It measures the effects of the PBIP on the treated group (targeted industries and targeted locations) compared to that in the non-treated group, considering the probability that each industry-location-year observation receives treatment. Results indicate that the average effect of being in the targeted group is around 0.41 percentage points increase in the local share of the nation total for the industry, and an increase of around 16.8 percentage points in the growth of employment level. However, the share of HTIs out of the local total employment does not respond significantly.

Then, I conduct event studies with following specification:

$$Share_{ijt} = \sum_{t=1998}^{2007} \beta_{t-2002} Policy_{ij} \times Year_t + FEs + \mathbf{X}_{ijt}\gamma + \epsilon_{ijt}$$

where $Share_{ijt}$ indicates conditional shares (either sectoral or regional, specified in the following figure). FEs consists of the complete set of one-way and two-way fixed effects $(\delta_i, \delta_j, \delta_t, \eta_{it}, \eta_{ij}, \eta_{jt})$. $\mathbf{X}_{ijt}$ are controls (export share, SOE share, and average leverage). Results are shown in the following plots.

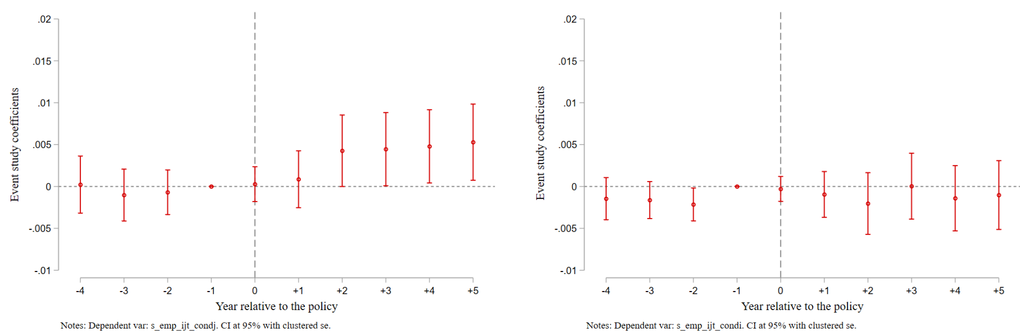


Figure 3.7: Left: Conditional on industry (i.e., concentration ratio λ^{ind}). Right: Conditional on location (i.e., specialization ratio λ^{loc})

Figure 3.7-Left and Figure 3.7-Right show the share conditional on the industry ($s_{ijt|j}$) and the share conditional on location ($s_{ijt|i}$) respectively. The coefficient of the year before the policy was implemented has been normalized to zero. Standard errors are clustered at the city-industry level. It shows a relatively significant increase in concentration ratio in the post-policy period, compared to that of specialization ratio on average.

In summary, the above empirical evidence show that the 2002 PBIP in China drives employment and productivity growth in targeted industries and targeted locations. However, it also drives non-targeted industries to targeted locations so that the industrial mix in targeted locations does not change significantly. Granular export analysis supplements the main empirics from a zone-level perspective, suggesting that the PBIP generates regional spillover effects along the production network. Moreover, these effects mainly occur in outside of the zone within prefectures, suggesting that the spillover effects could occur up to a specific spatial granularity.

3.3 Model

In this section, I develop a quantitative general equilibrium model that incorporates spatial sorting of heterogeneous firms, input-output linkages across sectors, trade between locations, and agglomeration externalities.

3.3.1 Environment

Locations are indexed by $i \in \mathscr{I} = \{1, ..., I\}$. I use L_{ij} to indicate local sectoral employment. Locations are ex-ante identical. The number of locations is assumed to be discrete for tractability.[26]

Sectors are indexed by $j \in \mathscr{J} = \{1, ..., J\}$. Within each sector j, there is a continuum of heterogeneous firms. Each firm in this sector draws a productivity z_j from $F\left(z_j\right)$. Upon entering location i, a firm with productivity z_j takes the set of local sectoral size $\{L_{ij}\}$ as given and choose local labor inputs $\ell\left(z_j\right)$ and composite inputs $x_{i,j'j}\left(z_j\right)$. Production function is Cobb-Douglas

$$y\left(z_j; L_{ij}\right) = A\left(z_j, L_{ij}\right)\ell\left(z_j\right)^{\alpha_j}\prod_{j' \in \mathscr{J}} x_{i,j'j}\left(z_j\right)^{\left(1-\alpha_j\right)\gamma_{j'j}}$$

$A\left(z_j, L_{ij}\right)$ is the firm-specific productivity with agglomeration externalities embedded. Assume the cost of labor is constant across locations. The cost of composites P_{ij} is at location-industry level to be determined in general equilibrium. Given the production structure, unit costs take the following form

$$c_{ij} \equiv w_i^{\alpha_j}\prod_{j'} P_{ij'}^{(1-\alpha_j)\gamma_{j'j}}$$

Intermediate products are in monoplistic competition. They take the demand from all other locations (referred as "global" demand) as given and set price for varieties. Thus the profit maximization problem upon entry is

$$\max_{p_i\left(z_j\right)} \chi_{ij}\left(p_i\left(z_j\right) - \frac{\kappa_{c,j} c_{ij}}{A\left(z_j, L_{ij}\right)}\right)\underbrace{\sum_{i'} q_{i'}^{i}(z_j)}_{\text{global demand } i \to i'}$$

where $\kappa_{c,j} \equiv \frac{\alpha_j^{-\alpha_j}}{\prod_{j'}(\gamma_{j'j}\left(1-\alpha_j\right))^{\gamma_{j'j}(1-\alpha_j)}}$ is a sector-wise constant. χ_{ij} is the place-based industrial policy on a firm's profits. The intermediate inputs above are assembled

[26] In Appendix 3.7.9, I extend this setup to a continuum environment.

with a CES aggregator as follows:

$$Q_{ij} = \left(\sum_{i'} \int_{\mathscr{Z}_{i'j}} q_i^{i'}\left(z_j\right)^{1-\frac{1}{\sigma_j}} dF\left(z_j\right) \right)^{\frac{\sigma_j}{\sigma_j - 1}}$$

$q_i^{i'}\left(z_j\right)$ is the intermediate goods used to produce Q_{ij} shipped from i' to i. From the final goods producers' perspective, these will be the composite inputs $x_{i,j'j}\left(z_j\right)$ in the Cobb-Douglas production function. These composite goods are sold at price $\tau_{i'i} p_{i'}\left(z_j\right)$.[27] The set of local sectoral price indices P_{ij} is determined in equilibrium. $\mathscr{Z}_{ij}$ denotes the set of producers in sector j choosing to locate at city i in equilibrium. We will revisit this when characterizing firms' locational choices.[28] With monoplistic competition, we can obtain the demand for intermediate z_j at location i from i'

$$q_i^{i'}\left(z_j\right) = \left(\frac{p_{i'}(z_j)\tau_{i'i}}{P_{ij}} \right)^{-\sigma_j} Q_{ij}$$

where sectoral-local price index is $P_{ij} = \left(\sum_{i'} \int_{\mathscr{Z}_{i'j}} \left(p_{i'}(z_j)\tau_{i'i}\right)^{1-\sigma_j} dF\left(z_j\right) \right)^{\frac{1}{1-\sigma_j}}$. Input choices are

$$\ell_i\left(z_j; L_{ij}\right) = \kappa_{\ell,j} \left(\frac{A\left(z_j, L_{ij}\right)}{c_{ij}} \right)^{\sigma_j - 1} \frac{G_{ij}^D}{w_i}$$

$$x_{i,j'j}\left(z_j; L_{ij}\right) = \kappa_{x,j'j} \left(\frac{A\left(z_j, L_{ij}\right)}{c_{ij}} \right)^{\sigma_j - 1} \frac{G_{ij}^D}{P_{ij'}}$$

where $\kappa_{\ell,j}$ and $\kappa_{x,j'j}$ are constants.[29] Thus, pre-policy profits for firms choosing location L_{ij} are

$$\pi_i\left(z_j; L_{ij}\right) = \kappa_{\pi,j} \left(\frac{A\left(z_j, L_{ij}\right)}{c_{ij}} \right)^{\sigma_j - 1} G_{ij}^D$$

[27] One can think of a competitive market of the assembling intermediate goods into composite goods together with the shipping process, which is captured by iceberg trade costs $\{\tau_{i'i}\}$.

[28] In Appendix 3.7.11, I discuss the model extension of entry costs.

[29] Specifically,

$$\kappa_{\ell,j} \equiv \left(\frac{\sigma_j}{\sigma_j - 1} \right)^{-\sigma_j} \alpha_j^{1-\alpha_j(1-\sigma_j)} \prod_{j' \in \mathscr{J}} \left((1-\alpha_j)\gamma_{j'j} \right)^{-(1-\sigma_j)(1-\alpha_j)\gamma_{j'j}}$$

$$\kappa_{x,j'j} \equiv \left(\frac{\sigma_j}{\sigma_j - 1} \right)^{-\sigma_j} \alpha_j^{\alpha_j(\sigma_j - 1)} (1-\alpha_j)\gamma_{j'j} \prod_{j'' \in \mathscr{J}} \left((1-\alpha_j)\gamma_{j''j} \right)^{(1-\alpha_j)\gamma_{j''j}(\sigma_j - 1)}$$

where $A\left(z_j, L_{ij}\right)$ is the composite term of agglomeration externalities. $c_{ij} \equiv w_i^{\alpha_j} \prod_{j'} P_{ij'}^{(1-\alpha_j)\gamma_{j'j}}$ summarizes inter-sectoral effects, $G_{ij}^D \equiv \sum_{i'} \tau_{ii'}^{-\sigma_j} P_{i'j}^{\sigma_j} Q_{i'j}$ stands for a local-secotral global demand shifter, which summarizes inter-regional effects. $\kappa_{\pi,j}$ is a sectoral constant[30]. Locational choices are characterized by firm sorting: $L_{ij}^*\left(z_j\right) = \arg\max_{L_{ij}} \pi\left(z_j, L_{ij}\right)$. In laissez-faire, firms choose locations to produce where the marginal benefit from agglomeration offsets the marginal production costs conditional on the industry[31]. Welfare loss comes from a firm's distorted choice of location, leading to sub-optimal locations in decentralized equilibrium because of the existence of agglomeration. Thus, policies can potentially enhance total welfare in the status quo.[32]

Consumers are identical and freely mobile. Assume perfect elastic labor supply. They earn wage and transfers/pay lump sum tax, and maximize utility by choosing a Cobb-Doublus bundle of final goods.

$$\max_{\{C_{ij}\}} U_i = \prod_j C_{ij}^{\eta_j}$$

where $\sum_j \eta_j = 1$, subject to the budget constraint. Note that C_{ij} is the consumption of composite goods in location i and sector j. These composites share the same CES aggregator over differentiated variaties.

Policies on firms are in form of wedges denoted by χ_{ij}. Governments levy tax on (or give transfers to) consumers to finance the subsidies to (or levy taxes on) firms.[33]

$$\sum_j \chi_{ij} \int_{\mathscr{Z}_{ij}} \pi_i(z_j) dF(z_j) = t_i \sum_j L_{ij}$$

3.3.2 Equilibrium

Given fundamentals and policy wedges of an economy, a general equilibrium consists of a vector of local sectoral prices, quantities, scales, and firms' location choices such that: (1) firms maximize profits at the location they choose, i.e., $i^*(z_j) = \arg\max_{i\in\mathscr{I}}\{\pi_i(z_j;\Theta), 0\}$. Once entering a location, they choose prices

[30]Specifically, $\kappa_{\pi,j} \equiv \frac{(\sigma_j-1)^{\sigma_j-1}}{\sigma_j^{\sigma_j}} \alpha_j^{\alpha_j(\sigma_j-1)} \prod_{j'\in\mathscr{J}} \left(\left(1-\alpha_j\right)\gamma_{j'j}\right)^{(1-\alpha_j)\gamma_{j'j}(\sigma_j-1)}$

[31]A mathematical derivation can show this intuition more explicitly when extending the model to continuum locations. See Appendix 3.7.9.

[32]In my framework, because of endogenous spatial sorting, the firm-level equilibrium differs from quantitative trade models with random draws from the Fréchet distribution of observing the lowest price as in Eaton and Kortum (2002) [94].

[33]If χ_{ij} is the rate of taxes on firm's profits, and t_i is local transfers on consumers. The model is general in terms of which side to levy tax on. More wedge specifications can be found in Appendix 3.7.10.

$p_i(z_j)$, as well as labor and intermediate inputs $\{\ell_i(z_j), x_{i,j'j}(z_j)\}$. (2) Consumers choose $\{C_{ij}\}$ to maximize utility. (3) Government budgets are balanced. (4) Trade flows are balanced. (5) Goods markets clear. (6) Labor markets clear. (7) Free mobility condition is satisfied.

With optimal locational choice function implicitly determined above, labor, composite input choices, and revenue can be written as follows.

$$\ell_i(z_j) = \frac{\kappa_{\ell,j}}{\chi_{ij}\kappa_{\pi,j}} \frac{\pi_i\left(z_j, L_{ij}^*(z_j)\right)}{w_i}$$
$$x_{i,j'j}(z_j) = \frac{\kappa_{x,j'j}}{\chi_{ij}\kappa_{\pi,j}} \frac{\pi_i\left(z_j, L_{ij}^*(z_j)\right)}{P_{ij'}}$$
$$r_i(z_j) = \frac{\kappa_{p,j}^{1-\sigma_j}}{\chi_{ij}\kappa_{\pi,j}} \pi_i\left(z_j, L_{ij}^*(z_j)\right)$$

The above indicates that productive firms sort into larger locations. This is consistent with what we observe in data. I show the sorting pattern in the Appendix 3.7.7 Figure 3.26, which depicts the relationship between estimated firm productivity and local sectoral scale. This also provides rationale in Section 3.5.

Moreover, sectoral bilateral trade flows in equilibrium can be expressed as follows

$$X(i,i')|j = \underbrace{\left(\underbrace{(\mathscr{A}_{ij}(z_j^*))^{\sigma_j-1}}_{\text{productivity index}} \underbrace{c_{ij}^{1-\sigma_j}}_{\text{unit cost}}\right)}_{\text{home market effect}} \underbrace{\tau_{ii'}^{1-\sigma_j}}_{\text{trade effect}} \underbrace{\underbrace{P_{i'j}^{\sigma_j}}_{\text{price charged in sales}} \quad Q_{i'j}}_{\text{global market effect}}$$

where $\mathscr{A}_{ij}(z_j^*)$ is at location-sector-level instead of firm-level, and z_j^* is the sectoral productivity threshold in equilibrium. In the rest of this chapter, I do not carry (z_j^*) for simplicity unless specified. And it turns out to be handy to use $\mathscr{A}_{ij}$ to denote the composite productivity term with both externalities and sorting embedded, incorporating agglomeration and inter-regional effects. Specifically,

$$\mathscr{A}_{ij} \equiv \left(\int_{\mathscr{Z}_{ij}^*} A\left(z_{ij}^*, L_{ij}\right)^{\sigma_j-1} dF_j(z)\right)^{\frac{1}{\sigma_j-1}}$$

Local sectoral outflows of goods $X(i,j)$ exhibit the general gravity form as in international trade literature, but augmented with agglomeration externalities, spatial sorting mechanism, and production network terms. The following equation represents the total trade outflow from location i in industry j in this model.

$$X(i,j) = \left[\sum_{r\in\mathcal{I}} \left(\frac{\tau_{ri}c_{rj}}{\mathcal{A}_{rj}P_{ij}}\right)^{1-\sigma_j}\right]\Psi_{ij}\mathscr{W}$$

where c_{ij} denotes local sectoral unit cost. Note that this cost incorporates inter-sectoral effects. $\mathscr{W}$ is the general equilibrium effects through $[\eta\tilde{w}L]$, which embeds size distribution and disposible income. Ψ_{ij} (with dimension $1\times IJ$) is a row of $(\mathbf{I}-\Phi)^{-1}$ where

$$\Phi \equiv \begin{bmatrix} \kappa_{x,11}CA_{11}\tau_{11}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,11}CA_{11}\tau_{1I}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,1J}CA_{1J}\tau_{1I}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \\ \vdots & & \vdots & & \vdots \\ \kappa_{x,11}CA_{I1}\tau_{I1}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,11}CA_{I1}\tau_{II}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,1J}CA_{IJ}\tau_{II}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \\ \vdots & & \vdots & & \vdots \\ \kappa_{x,J1}CA_{11}\tau_{11}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,J1}CA_{11}\tau_{1I}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,JJ}CA_{1J}\tau_{1I}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \\ \vdots & & \vdots & & \vdots \\ \kappa_{x,J1}CA_{I1}\tau_{I1}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,J1}CA_{I1}\tau_{II}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,JJ}CA_{IJ}\tau_{II}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \end{bmatrix}$$

It is as if a generalized "Leontief inverse", where Φ is a composite matrix summarizing inter-sectoral/regional effects.[34] The advantage of writing the Φ as the above form is that it can be easily decomposed into

$$\Phi \equiv (\mathscr{L}(\kappa, CA) \circ \mathscr{G}(\tau, P))_{(I\times J)\times(I\times J)}$$

where $\mathscr{L}$ denotes the local pricing power effect with $\kappa_{x,jj'}$ reflecting local I/O structure, i.e., how j is technologically required by industry j'. CA_{ij} measures competitiveness, i.e., firm-level productivity devided by city-sector level cost, dampended by sectoral elasticity. $\mathscr{G}$ denotes the global market buyer effect with $\tau_{ii'}$ denotes trade from i to i' and $P_{i'}$ in the last subscript denotes local price index in the destination markets.[35] Note that the benchmark model presented above does not consider factor-based PBIP. In Appendix 3.7.10, I differ the policy wedges on different factors of production. And I show that the above general equilibrium *absorption matrix* (Φ) preserves a similar form that adds an outer-layer policy filter.

3.4 Numerical Simulation

In this section, I use simple economy examples to illustrate equilibrium local sectoral scales with comparative statics. Then I provide relevant supplementary

[34] In Appendix 3.7.13, I also explain how it relates to Acemoglu et al. (2012) [95] and derive analagous expressions for concentration ratio and specialization ratio.

[35] See Appendix 3.7.13 for derivation details. An illustration of this matrix in a simple economy is also shown.

empirical evidence with the Chinese data.

First, I simulate the model in a 2 × 2 economy to illustrate how policies can drive industrial clustering. Figure 3.8 shows a graph to highlight inter-regional and inter-sectoral linkages in the simple economy. Red arrows highlight within-location and inter-sectoral flows (through input-output linkages) of composites (Q_{ij}) at cost P_{ij}. Blue arrows highlight across-location and intra-sectoral (through trade) flows of varieties ($y_i(z_j)$) at cost $p_{i'}(z_j)\tau_{i'i}$. Solid lines are production processes, while dashed lines are consumption.[36] In Appendix 3.7.13, I seperately illustrate inter-sectoral flows and inter-regional flows, and show derivation for key conditions with agglomeration and sorting mechanisms embedded.

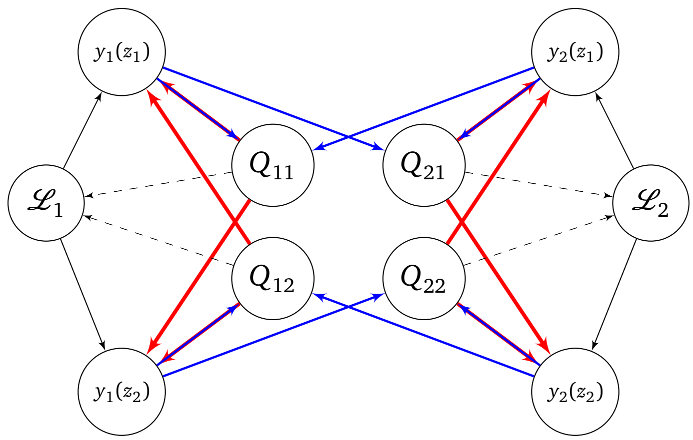


Figure 3.8: *An illustrative example with $I = 2$ and $J = 2$.*

Consider local subsidies given to a specific location-industry pair ($i = 2, j = 2$), with the rate of linear firm subsidies ranging from 1% to 20%. Figure 3.9 shows the result. The x-axis shows the change of policy. And y-axis shows different equilibrium outcomes. L_{ij} indicates the employment in location i and sector j. The upper-left panel shows the employment in location 1, in which the dashed line indicates sector 1 and the solid line indicates sector 2. The upper-right panel shows employment in location 2, in which the dashed line indicates sector 1 and the solid line indicates sector 2 (PBIP target). The lower-left panel shows the concentration ratio of the PBIP target ($\lambda_{22}^{ind} \equiv \frac{L_{22}}{\sum_i L_{i2}}$). The lower-right panel shows the specialization ratio of the PBIP target ($\lambda_{22}^{loc} \equiv \frac{L_{22}}{\sum_j L_{2j}}$).[37] These numerical exercises are in line with the empirical patterns we show in empirical evidence. The policy is targeted at a specific industry-location pair[38], and

[36] With a bit abuse of notation, $\mathscr{L}_i \equiv \sum_j L_{ij}$ in the illustrative graph.

[37] More details on numerical experiment are in Appendix 3.7.14 where I show equilibrium conditions and related results.

[38] In this example, $i = 2, j = 2$. The choice of the location-sector pair here is without loss of generality.

the policy also drives the expansion in non-targeted industries. As a result, we see that targeted industries are concentrated in targeted locations (from around 61.8% to 66.2%). However, the industry mix is not dominated disproportionately more by targeted industries. Instead, this number has dropped, although by a relatively small amount (from around 53.6% to 52.5%).

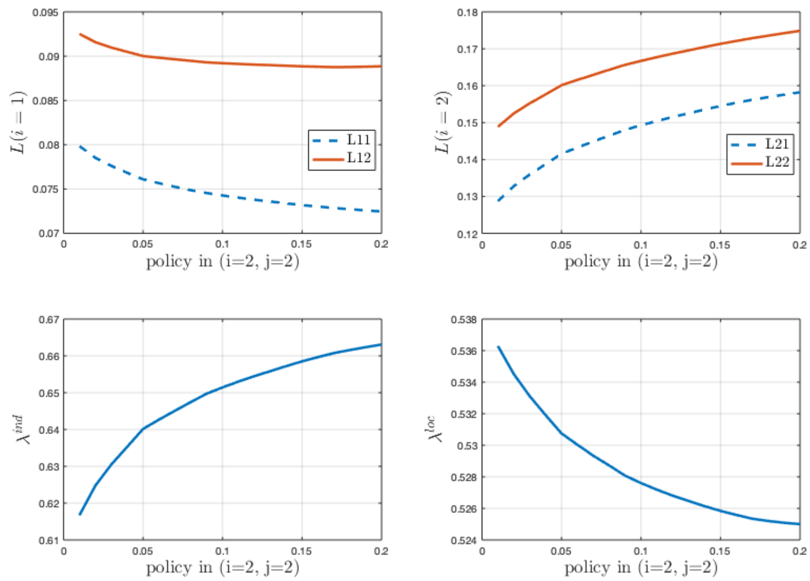


Figure 3.9: Numerical experiments on key statistics varying PBIP

Furthermore, I use the simulation to explore how concentration ratio and specialization ratio respond to the change in policy intensity ($\Delta\chi_{22}$ as above) when varying agglomeration externalities and production network technologies. First, I consider an increase in the agglomeration externality parameter, and plot two ratios against policy changes as in Figure 3.10. The left plot shows the change in concenration ratio (λ^{ind}) with respect to an increase in PBIP. And the right plot shows the change in specialization ratio (λ^{loc}) with respect to an increase in PBIP. The positive slope in the left plot suggests a positive elasticity of the concentration ratio with respect to policies, which is dampened when increasing the extent of externalities. And the right plot suggests that the specialization ratio decreases in policies, although both the elasticity, and its change between comparative statics, are relatively small, compared to these of the concentration ratio.[39]

[39]Equilibrium scales in location-sector pairs are shown in Appendix 3.7.14 Figure 3.30.

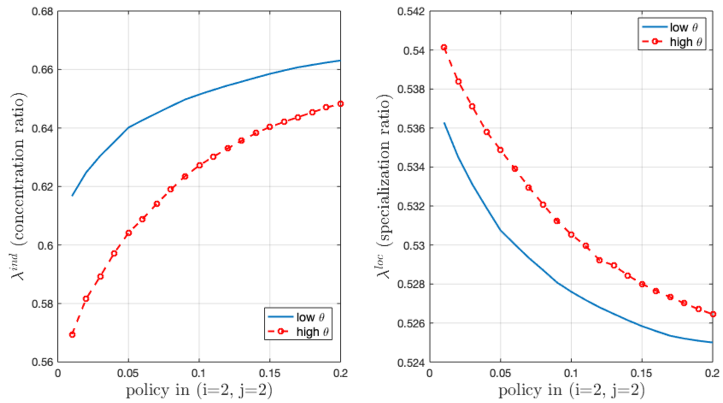


Figure 3.10: Numerical experiments varying PBIP and agglomeration externalities

Then, I consider changes in the production network and plot two ratios against policy changes as in Figure 3.11. The left plot shows the concentration ratio increases when production networks are less effective, i.e., I/O matrix more diagonally dominant (denoted by "I/O stronger diag" in the plot). However, the right plot shows that when this happens, there is a decrease in the specialization ratio, reflected by λ^{loc} becoming flatter (the red dashed curve). This is because the extent of the expansion in other industries (that are otherwise "relevant" to the targeted industry) decreases. In other words, when I/O linkages play an important role, the expansion of targeted industry in targeted location is accompanied by at least equally strong expansion of other related industries. However, if industries other than targeted ones are less connected with the targeted industry, the industrial mix can be dominated by the change in targeted industry, or do not show strong spillover effects along the production network.[40]

[40]Details about this figure are shown in Appendix 3.7.14 Figure 3.31.

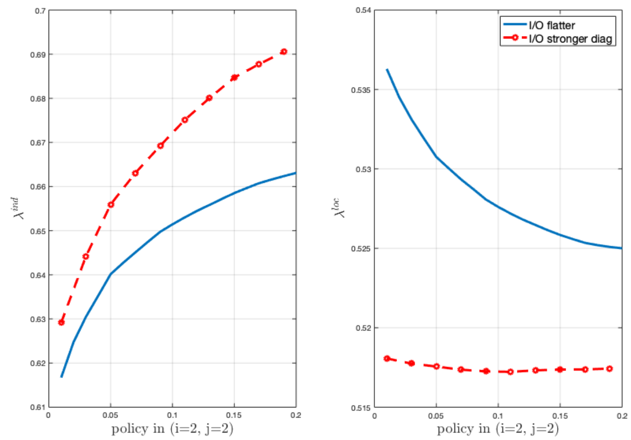


Figure 3.11: Numerical experiments varying PBIP and production networks

At last, I present supplementary evidence on aggregate employment effects. These give a more complete picture of the empirical facts, and potentially enlight further studies. I have documented that high-tech industries experience more significant employment increase than non-high-tech industries when responding to the place-based policy. Furthermore, the place-based effects are more significant for industries that are less connected with high-tech industries, conditional on non-HTI. This suggests the relatively more important role of agglomeration. Specifically, I apply the propensity approach in Section 3.2.2 together with the following difference-in-difference estimation conditional on sectors.

$$y_{ijt} = \alpha HTC_i \times Post_t + FE + \mathbf{X}_{ijt}\gamma + \epsilon_{ijt},\ j \in G$$

where G indicates aggregation groups (HTI, non-HTI and their further decomposed groups). Regression results are shown in Appendix 3.7.6 Table 3.6 and Table 3.7. Table 3.6 column (1) and (2) are regression results for the average effects of being in HTC after policy implementation on total employment in HTI in each city-industry pair. Column (3) and (4) are effects for non-HTI with the same set of controls. Table 3.6 shows results for similar econometric specifications but further decomposes the non-HTI in sectors with their relation with HTI based on I/O relevance (as in Section 3.2.2). Table 3.7 Column (1) and (2) are for that in non-HTI but closely connected with HTI (referred as "Close"). Column (3) and (4) are for that in non-HTI but far from the targeted ones (referred as "Far").[41] Note that the aggregate evidence relies on the city as the geographical unit, instead of more granular level such as zones. The city is the relevant level of granularity supported by the trade analysis shown in 3.2.2 when seperating within- versus out-of-zone patterns conditional on high-tech cities. And

[41]In addition, I analyze the effects by SOE groups. It turns out that non-SOEs respond more than SOEs. Results are shown in Appendix 3.7.6 Table 3.8.

this gives a rationale to quantify externalities by taking the model to data at aggregate spatial levels.

3.5 Quantitative Analysis

In this section, I take the model to Chinese data designing an approach to quantify externalities, and analyze aggregate effects through the lens of the model.

The model allows us to measure key primitives Θ^* by changes in the sorted distribution of local sectoral equilibrium outcomes in response to policy changes, i.e.,

$$\Theta^* = \mathscr{F}^{-1}\left(\hat{L}, \hat{\chi}, \hat{P}, \hat{Q}\right)$$

This differs from previous research by using the spatial sorting moments derived from firm sorting and taking into account the agglomeration and production networks. It is new on top of related studies, for example, Bartleme et al. (2021) [86] without the choice of location, Caliendo and Parro (2015) [83] without agglomeration and the choice of location, and Gaubert (2018) [7] without production networks. Some primitives are first calibrated as in standard literature. I/O matrix Γ comes from the Chinese I/O Table at the national level.[42] Based on the Chinese industry coding system, the I/O Table includes 135 sectors, with 80 manufacturing sectors relevant to this chapter, instead of a complete set of sectors at the CIC 4-digit level. For the purpose of this quantification, I conduct a crosswalk of industry codes between CIC and the Chinese Input-Output Table with an aggregation. I use Figure 3.38 in Appendix 3.7.16 to show the intermediate input usage from an industry (along the vertical axis) from any other industries (along the horizontal axis) in the original I/O Table. The labor share (0.23) is calibrated from the wage bills in the Chinese NBS for manufacturing sectors. Inter-regional trade costs $\mathscr{T}$ are taken from the mean of Tombe and Zhu (2019) [96].[43]

Then, to quantify agglomeration externalities, I utilize the distribution of agglomeration-embedded TFP in terms of distribution of policies, inter-sectoral linkages, and inter-regional trade to back out externalities. The simplified equilibrium conditions are the following:

$$\hat{A}^+_{ij} = \mathscr{E}\left(\underbrace{\hat{\chi}^+_{ij}}_{\text{Policy}}, \underbrace{\hat{c}^+_{ij}}_{\text{I/O}}, \underbrace{\hat{G}^+_{ij}}_{\text{Trade}}; \Theta\right), \ \forall i \in \mathscr{I} \setminus \{I\}, j \in \mathscr{J}$$

[42]In this section, by industry, I mean sectors categorized in the I/O Table. The I/O Tables at national level across time can be found on the NBS website, which are different from regional tables.

[43]See Appendix 3.7.18 Figure 3.42 for values of these costs at the regional level.

where $\hat{A}^{+}_{ij} \equiv \frac{\hat{A}(\tilde{z}_{i,j}, L_{i,j})}{\hat{A}(\tilde{z}_{i+1,j}, L_{i+1,j})}$, $\hat{\chi}^{+}_{ij} \equiv \frac{\hat{\chi}_{i,j}}{\hat{\chi}_{i+1,j}}$, $\hat{c}^{+}_{ij} \equiv \frac{\hat{c}_{i,j}}{\hat{c}_{i+1,j}}$, and $\hat{G}^{+}_{ij} \equiv \frac{\hat{G}_{i,j}}{\hat{G}_{i+1,j}}$. c_{ij} is the local sectoral unit cost taking into account I/O linkages as specified in the model section. $G_{ij} \equiv \sum_{i'} \tau_{ii'}^{-\sigma_j} P_{i'j}^{\sigma_j} Q_{i'j}$ represents global demand taking into account inter-regional trade frictions. The i in the above set of moments denotes the sorted location index, i.e., conditional on each industry, variables are sorted by local sectoral scales (L_{ij}). This sorted distribution features the spatial sorting equilibrium when firms make locational choices. And I take these to a nonlinear least square estimation. There are a couple of notes regarding bringing the model to data. First, these i's are at the regional level because of data availability for inter-regional trade costs, and effective values in aggregation. In this quantification, there are eight broad domestic regions in total.[44] Second, as for agglomeration externalities, I adopt the functional form $A_{ij}(z, L) = z_j L_{ij}^{\theta_j}$ as in Bartelme et al. (2019) [86]. In general, with a separability assumption between a firm's productivity and a local sectoral agglomeration externality, the above set of moments can be simplified to be reshaped scales as a function of the composite terms.[45]

The mean of agglomeration externalities obtained with my model is 0.115. Figure 3.34 shows the results by sector, where HTIs turn out to have relatively larger externalities, led by the pharmaceutical industry as the top one (0.494).[46] I also use Figure 3.35 to compare the model-based EES from my approach with that in some well-cited recent studies vis-à-vis comparable manufacturing sectors. Specifically, I compare the distribution with Lashkaripour and Lugovskyy (2017) [97] (the curve noted as "$LL(2017)$" in the graph), Gaubert (2018) [7], Bartleme et al. (2019) [86] (the curve noted as $BCDR$, 2019 in the graph), and Bakker (2021) [90]. Both Gaubert (2018) [7] and Bakker (2021) [90] (curves in red) have different parametric forms of agglomeration externalities with my setup.[47] Lashkaripour and Lugovskyy (2017) [97] and Bartleme et al. (2019) [86] are more comparable in terms of parametric forms.[48]

[44] Specifically, these regions are Northeast, Beijing/Tianjin, North Coast, Central Coast, South Coast, Central, Northwest, and Southwest.

[45] An alternative approach is to pin down policy changes $\hat{\chi}_{ij}$ instead of directly using additional moments. The aggregated observables in the time series appear in Appendix 3.7.16, Figure 3.32. The way to obtain policies is not directly in policy levels, but rather in changes. Theoretically, we could use model structure to back out the distribution of changes of policies ($\hat{\chi}^{+}_{ij} \equiv \frac{\hat{\chi}_{i,j}}{\hat{\chi}_{i+1,j}}, \forall i \in \{1, ..., I-1\}$) conditional on externalities backed out through inter-regional trade flows. However, it works worse because of the necessary higher-level aggregation. Details are shown in the appendix.

[46] The second highest industry within HTIs is the electronic equipment industry (0.301).

[47] They attempt to separate log-supermodularity externalities. The comparison plot uses log-linear EES.

[48] The sector standing out in the estimation of Lashkaripour and Lugovskyy (2017) [97] is the broad Petroleum Industry. The data used in my chapter does not include the raw exploration of petroleum while includes the petroleum processing.

The main reason for not using conventional exact-hat algebra in my paper is that it does not provide sufficient quantification convenience in this specific environment.[49] Compared to the conventional exact-hat algebra, my approach has several benefits specific to the question and data counterparts. First, focusing on the distributional change allows me to quantify agglomeration externalities without relying on the estimated TFP and inter-regional trade. Bilateral trade flow data are normally hard to get at the micro level (such as zone, city, or county). Second, relative terms summarize changes in distribution. This can help address the extent to which the policies are targeted to HTI-HTC. If measuring in levels, I see nonzero policies in most locations and industries. However, when using changes, I focus on the location-sector pairs that are receiving disproportionately more policies. With a level-removed change in policies, it is more convincing to use the observed distributional changes across locations within a sector associated with a policy change to obtain the magnitudes of agglomeration externalities. Accordingly, from an empirical concern, the data counterpart of $\hat{\chi}_{ij}$ can be a better measurement for policy.[50] However, a major cost of this approach is the necessary aggregation. According to the data, not all locations and industries export at the granular level.[51] So in my context, the aggregation helps preserve relevant variations to some extent.

Then, I measure the aggregate effects of the 2002 PBIP, using the parameters quantified in the above section and the data counterparts. To do this, I first depict the pattern of positive sorting at the sector level. Figure 3.36-upper plot shows the revenue (log scale) of each location-industry pair (each cell) in 2007. Locations are sorted from the smallest to the largest (1 for the smallest, 8 for the

[49]To see this, we can derive the exact-hat algebra-equivalent key conditions as follows:

$$f\left(\hat{L}_{ij};\theta_j,\sigma_j\right)=\hat{\kappa}_{\ell,j}\hat{FE}_i T\hat{\bar{F}}P_{ij}\left(\prod_{j'}\hat{P}_{ij'}^{(1-\alpha_j)(1-\sigma_j)\gamma_{j'j}}\right)\sum_{i'}\phi^G_{ii',j}\hat{\tau}_{ii'}^{-\sigma_j}\hat{P}_{i'j}^{\sigma_j-1}\hat{R}_{i'j}$$

where $\phi^G_{ii',j}\equiv\frac{\tau_{ii'}^{-\sigma_j}P_{i'j}^{\sigma_j-1}R_{i'j}}{\sum_k\tau_{ik}^{-\sigma_j}P_{kj}^{\sigma_j-1}R_{kj}}$. Note that domestic bilateral trade flows $\pi^G_{ii',j}$ are not available at disaggregated level. In addition, we need to use estimated $T\bar{F}P$.

[50]As a supplementary support, I utilized a novel dataset on local-sectoral infrastructure investments to help illustrate policy variations. This is by far the most complete record on Chinese infrastructure investments. Changes in the distribution of infrastructure investment aggregated by policy targets (location-industry pairs) are shown in Appendix 3.7.2 Figure 3.13. Although we have to admit the concerns on measured policy levels, these plots show that there are more-significant changes in infrastructure investment distribution in HTI-HTC, especially compared to those in non-HTIs.

[51]In Appendix 3.7.19 Figure 3.43, I show the distribution of positive exporting value in Automobile and Motor Vehicle Equipment Manufacturing (CIC = 3725) at the prefecture-level. Many prefectures do not have exports in specific sectors. With aggregation at the province level (see Figure 3.44), there are still provinces with no effective data. This decreases effective total observations, although it helps eliminate sparsity when dealing with granular variations.

largest) from up to down.[52] These appear in brighter yellow on the bottom compared to the darker blue on the top conditional on sectors. Figure 3.36-lower plot shows the average revenue increasing in regions across sectors, indicating the positive sorting by industry holds on average. Bringing together prepolicy and postpolicy outcomes in data counterparts and the aggregate welfare measure through the model structure, I find that the current PBIP increases aggregate welfare by 1.71%. The positive welfare gain at the aggregate level provides a rationale to implement the policy by taking advantage of local agglomeration externalities together with intersectoral linkages, inter-regional trade, and firm sorting. This gain at the aggregate level comes from the fact that local sectoral nominal consumption is boosted, accompanied by an limited increase in the composite costs. The intuition is that costs are relatively stable when productive firms are not drastically crowded out. This, in turn, makes it possible that inter-regional trade accelerates the price advantages in policy targets.

At last, I use the model to quantify the distributional changes in the general equilibrium effects. Consider rewriting the change in local sectoral competitiveness (referred as composite advantages) with the following expression

$$\hat{CA}_{ij} \propto \underbrace{\left(\mathscr{A}'_{ij}/\mathscr{A}_{ij}\right)}_{\text{sorting w/ EES}} \underbrace{\left(c'_{ij}/c_{ij}\right)^{-1}}_{\text{inter-sectoral linkages}} \underbrace{\left(G'_{ij}/G_{ij}\right)}_{\text{inter-regional trade}}$$

where $\mathscr{A}_{ij} = \left(\int_{\mathscr{Z}_{ij}} A(z_j, L_{ij})^{\sigma_j - 1} dF_j(z)\right)^{\frac{1}{\sigma_j - 1}}$ and $G_{ij} = \sum_{i'} \tau_{ii'}^{-\sigma_j} P_{i'j}^{\sigma_j - 1} Y_{i'j}$. Figure 3.12-upper plot shows the composite advantage in changes (pre- and postpolicy). The blue bars are CA computed with G as local demand ($\tilde{G}_{ij} = P_{ij}^{\sigma_j - 1} Y_{ij}$), while the red bars are the CA computed with G as global demand (G_{ij}). Figure 3.12-lower plot shows the difference between two cases with local demand and global demand. It turns out that the existence of inter-regional trade in general equilibrium (i.e., when triggering the global demand shifter), the relative composite advantage $\Delta CA \equiv \hat{CA}_{ij}^{global} - \hat{CA}_{ij}^{local}$ reverts a set of sectors from a decrease to an increase in CA.[53] Special chemical products, automobile, electromechanics, other electronic equipment, normal chemical products are sectors benefiting the most. In Figure 3.37, I normalize the distribution of ΔCA to illustrate sectoral beneficiaries. I highlight HTI industries with blue bars.[54] These suggest

[52]Eighty sectors based on the the China Input-Output Table. The employment levels and changes are shown in Appendix 3.7.16 Figure 3.33.

[53]Vegetable oil processing is the only sector standing out for a higher change in CA (18.98) that is excluded, followed by Sugar (2.88) as the second (in terms of the change in CA) excluded food-related manufacturing sector in the non-HTI group.

[54]Figure 3.37 shows sectors with increase in CA that are non-food manufacturing, sorted from the smallest to the largest. Note that this plot normalizes the distribution of changes. Regarding the raw ΔCA, most industries experienced an increase (74 out of 80), except for (1) 20-Textile and clothing, shoes, hat; (2) 17-Hemp textile, silk textile, and fine processing; (3) 6-Aquatic

that both targeted and non-targeted industries' competitiveness are boosted on average, especially other industries that are users or suppliers of HTIs.

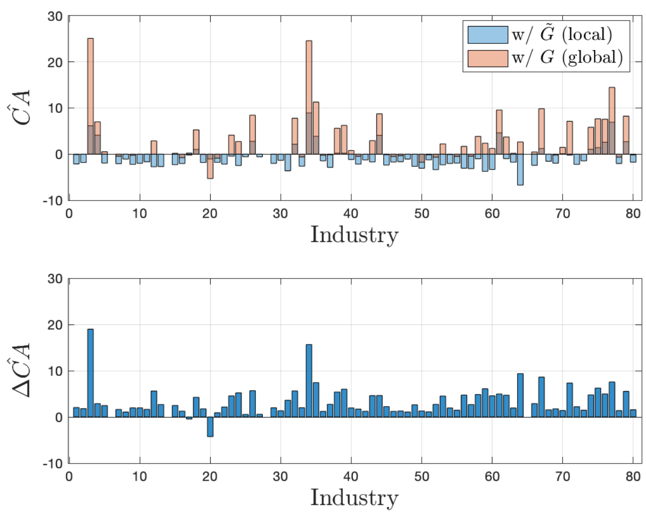


Figure 3.12: Upper: Sectoral Changes in Competitiveness with and without Market Assess. Lower: Differences in Competitiveness ($\hat{CA}$)

3.6 Conclusion

In this chapter, I study the aggregate consequences of place-based industrial policies (PBIPs) from both empirical and model-based perspectives.

Using Chinese firm-level data sets and a policy shock (the *Second Entrepreneurship Initiative in High-Tech Industrial Development Zones*), I find that this policy induced significant cross-industry, within-location spillovers. The 2002 PBIP drives employment growth in targeted industries and targeted locations. But it also drives nontargeted industries to targeted locations so that the industrial mix in targeted locations does not change significantly. In addition, the spillover effects is at a broader area compared to observed granular outcomes. I provide supportive empirical evidence that responding to the 2002 PBIP, total export is most significant in policy-targeted zones within policy-targeted cities, while the spillover effects along the production network are more significant outside of policy-targeted zones within policy-targeted cities.

Then, I build a quantitative general equilibrium model incorporating firm sorting, external economies of scale, and production networks to understand productivity growth under PBIPs. In this multisector and multilocation framework, each firm endogenously sorts into different locations and products based on the idiosyncratic productivity, production networks, externalities, and trade

products processing; (4) 14-Tobacco products; (5) 28-Coking; (6) 65-Shipbuilding and floating equipment.

frictions across locations. Firms trade off between local agglomeration benefits and the cost of congestion driven by proximity to suppliers of intermediate goods. Agglomeration externalities imply that productive firms are disproportionately more productive in larger locations. Agglomeration externalities in this framework are at the location-sector level and are scale-based. The intuition is that both the total scale and the industry structure respond in general equilibrium at the local level, and both channels can absorb unintended policy consequences. To quantitatively analyze the effect of PBIPs, I propose a new approach to quantify agglomeration externalities by using the change in equilibrium distribution and the exact-hat algebra in the literature on international trade. The idea is to use the changes in the distribution of sectoral employment across locations that respond to policy shocks to obtain sectoral agglomeration externalities. Quantitative analysis shows that current PBIPs increase aggregate welfare. This comes from the fact that the gain in competitiveness to policy beneficiaries in targeted locations and targeted sectors outweighs the loss from nontargeted sectors in both targeted and nontargeted locations. In addition, the increase in nominal income outweighs the increase in price indices. The coexistence of firm-sorting and agglomeration accelerates profitability in general while keeping cost increases relatively low, so that more economic potential is realized.

This study has the following policy implications. First, the existence of externalities provides a rationale for place-based policies and/or industry-based policies. Second, policymakers should be careful in supporting specific industry-location pairs. Scale and structure are not independent. Whether policies targeted at specific regions and sectors can enhance total welfare depends on how externalities interact with spatial sorting and input-output technologies. A policy favoring a specific sector in a region may eventually hurt other sectors in the same region through agglomeration together with sorting channels.

3.7 Appendix

Appendix: Supplementary Evidence on Local Responses to the National Guidelines

The following are local policies implemented responding to the *"Second Entrepreneurship"* national guideline. The most detailed set of policies found is in Xi'an. Selected local policies[55] responding to the national guidelines are shown below.

- Interim Policy Implementation of Housing Subsidies for R&D Institutions

[55] Implementation on the Price of Industrial Land, by the Administrative Committee of Xi'an High-Tech Zone.

- Policy incidents: companies/centers that are establised in the Xi'an High-tech zones on and after Janurary 1st, 2003. Specifically, these consist of Fortune 500 companies, R&D institutions of well-known international companies, R&D institutions of domestic publicly listed companies, R&D institutions (designated as national-level, including technology development centers, engineering centers, and technology testing centers).
- Subsidies:
  1. R&D institutions of Fortune 500 companies and top 10 well-known companies in the international ranking receive a one-time subsidy of 500,000 RMB.
  2. R&D institutions of domestic listed companies, national-level R&D institutions, including technology development center, engineering center, and technical testing center receive a one-time subsidy of 300,000 RMB.

• Interim Policies of Intellectual Property Subsidy Funds

- Policy incidents: enterprises with the tax relationship is in Xi'an High-tech Zone that obtained patent rights, trademark rights, computer software registration or product registration from January 1, 2003 to December 31, 2006,
- Patent funding standards:
  1. 3,000 RMB per domestic invention patent
  2. 30,000 RMB per foreign invention patent
  3. 1,000 RMB per utility model patent
  4. 300 RMB per design patent
- Trademark rights funding/subsidy standards
  1. 1,200 RMB per domestic trademark
  2. 5,000 RMB per foreign trademark
- Calculator software registration or product registration subsidy standard
  1. 300 RMB per piece of calculator software copyright
  2. 500 RMB per piece of calculator software product registration

• Policies on Preferential Industrial Land Prices

- Regional scope and requirements for preferential land prices:
  1. The scope of the preferential land price is limited to the newly developed areas in the second venture, excluding the first and second phases of the high-tech zone and Chang'an Park.

2. The preferential land price is only applicable to industrial land. Industrial land refers to product manufacturing, software development production and technology development land for development and technology research and development enterprises.

– Basic conditions for preferential land transfer:

1. High-tech, high investment intensity, high-yield industrial projects
2. An independent legal entity registered in the high-tech zone and tax relationship in the high-tech zone
3. The project shall meet the environmental protection standards required by the environmental planning of the high-tech zone

Appendix: Supplementary Institutional Background

Examples of 2002 policy implementation at local level are as follows.

Shanghai HTZ: (1) Business environment: Free registration and approval (approximately equivalent to 8 million RMB). (2) FDI: 0.28 billion USD foreign investment (incraesed by 24% compaed to 2002), with more than 70% have independent R&D department. (3) Financial support: provide fi founding subsidy and housing subsidy to HTI entreprenuers. (4) Labor: HTZ Administry initiates special residence approval for worker hired by high-tech firms in HTZ (promote mobility of high-skilled labors)

Shenzhen HTZ: (1) Business environment: construct software district in HTZ: 1.8 billion RMB; open branches of all major banks in HTZ. (2) Financial support: 20 million RMB grants as HTZ entrepernuer fund, 10 million RMB awards to innovators. (3) R&D infrustructure: Build Industry Technology Research Institute, National Biology-medicine Incubator and Software Training Foundation in HTZ. (4) Labor: Build post-doc station in HTZ; cooperate with local universities.

Suzhou HTZ: (1) Business environment: build investment service center, business registration, tax filing, technology property rights exchange center, attract law/accounting/real estate firms. (2) R&D infrustructure: Build Industry Technology Research Institute, National Biology-medicine Incubator and Software Training Foundation in HTZ. (3) Financial support: Add 4% out of total public expenditure of HTZ zone as special technology develoment fund to support R&D in HTI firms in HTZ.

Appendix: Data

3.7.1 Firm Production Data

The data I use is the firm-level data in industrial sector (manufacuring, construction and utility). It includes production and accounting data at firm level. The period this chapter uses is from 1998 to 2008. The dataset is a survey data conduced by the National Bureau of Statistics (NBS) and includes all state-owned enterprises (SOE) and non-SOEs that have annual total sales of 5 million RMB or above. According to GAAP, pre-screen observations that violate any of the following criteria: total assets greater than variable assets; total assets greater than fixed assets; valid establishing date; total annual sales greater than or equal to 5 million RMB (if non-SOE); annual sales, total assets, total exporting value, intermediate inputs, number of employee positive; interest payment non-negative; total exporting value less than total annual sales.

As for industry information, for each firm, we know the 4-digit level Chinese Industrial Classification (CIC). This chapter only uses the firms that are in manufacturing sector, which have CIC *category code* "C" and 2-digit *sector codes* from 13 to 42. Table 3.3 shows some example CIC codes and names in HTI or non-HTI category, with different digits to give readers an idea on the detailed level of the industry counted in empirical evidence.

Table 3.3: Example industries in HTI v.s. non-HTI

Category	**CIC-2**	**Industry Name**	**CIC-3**	**Industry Name**	**CIC-4**	**Industry Name**
HTI	37	Transportation equipment	376	Aerospace vehicle	3761	Aircraft manuf. and repair
	40	Electronic and computer	405	Electronic device	4052	Semiconductor discrete device
					4053	Integrated circuit
n/ HTI	14	Food	141	Bakery food	1412	Cake and bread
	22	Paper and paper prod.	222	Papermaking	2223	Processed paper
	31	Non-metallic mineral prod.	311	Cement, lime, and gypsum	3111	Cement

The following shows the list of high-tech industries based on GB/T 4754-2011. CIC-2 is at sector level and can be up to 4-digit level. A crosswalk was conducted in my empirical analysis to be consistent with 2002 version classification system. HTI assignments follow the 3-digit level to cover changes or delete/merge of industry code across time.

Table 3.4: Categories of High-tech Industries (Manufacturing)

Category Name	CIC-2 Code	No. of CIC-4 Code
Pharmaceutical manuf.	27	7
Aviation, spacecraft and equipment manuf.	37,42	5
Electronic and communication equipment manuf.	35,38,39	19
Computer and office equipment manuf.	34,39	6
Medical equipment and instrumentation manuf.	35,40	24
Information chemical manuf.	26	1
Total	-	62

3.7.2 Supplementary Evidence on Policy Variations

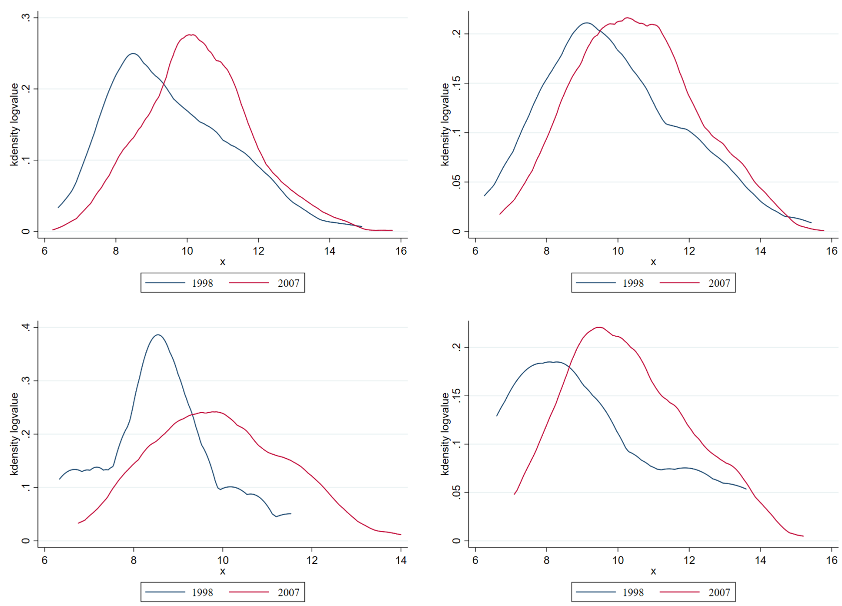


Figure 3.13: Upper Left: nHTI-nHTC. Upper Right: nHTI-HTC. Lower Left: HTI-nHTC. Lower Right: HTI-HTC.

3.7.3 Other Data Sets

The firm innovation data set includes patents filed Chinese firms. This chapter used the matched version with the NBS firm data. For each firm in the dataset, there is the total number of each main type of patents: invention, utility and design.

For industrial linkages, I use the China Input-Output Table covers 135 sectors at nation-level (including agriculture, manufacturing, construction, utility and service), which comes from NBS. Multi-region Input-Output Tables (2002, 2007) cover 17 sectors and 8 regions (among which 11 are manufacturing sectors), which comes from the Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences.

For local aggregates, I use the China Statistical Yearbook that contains annual prefecture-level statistics (share of industrial, service sectors, area, local public expenditure etc.)

Appendix: Supplementary Empirics

3.7.4 Descriptive Facts

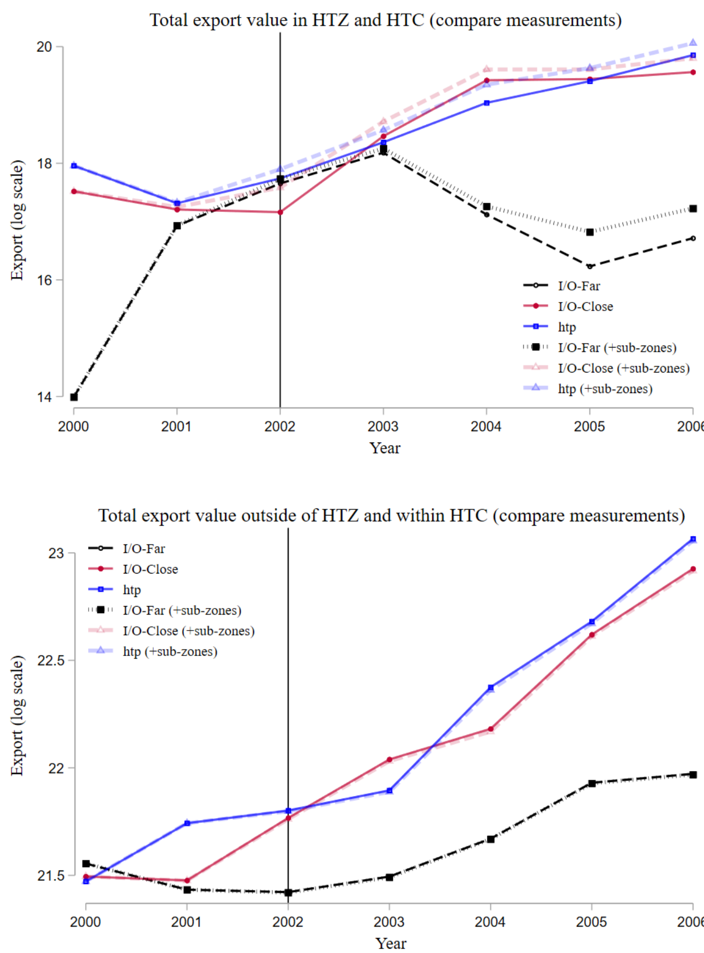


Figure 3.14: Export in Beijing by policy-I/O group, with policy-zone measurements compared

Table 3.5: Estimates for Non-Targeted Industries and Non-Targeted Zones

Variables	**Export Value** (1)	**Export Value** (2)
$Close_{ij} \times Post_t$	0.0971***	0.1447***
	(3.2360)	(14.7221)
FEs	✓	✓
Obs.	1,106,494	1,106,415
R^2	0.0879	0.2811

(t statistics in parentheses, with clustered s.e.)
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Notes: In column (1), location fixed effects are at the prefecture level, and the product fixed effects are at the HS-4 level. In column (2), location fixed effects are at the zone level, and the product fixed effects are at the HS-8 level.

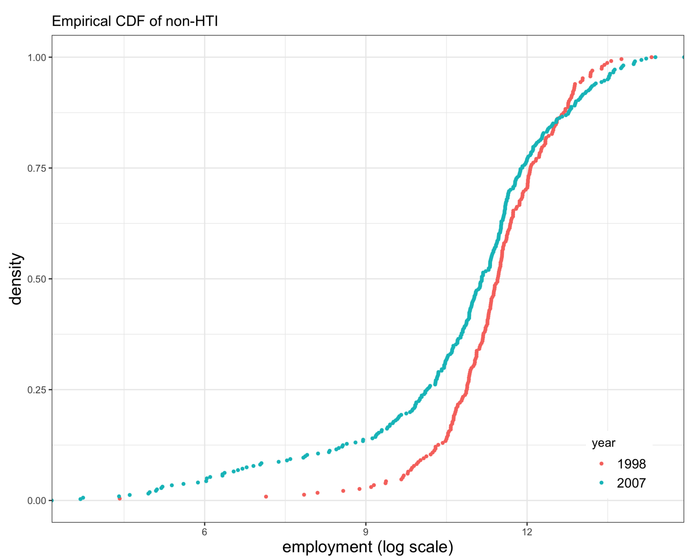


Figure 3.15: Empirical distribution of high-tech industries

Three 2-digit CIC that have most 4-digit HTI categories: C37 (railroad, shipment and aircraft equipment manufacturing), C39 (instruments and meters manufacturing), C40 (computer, communication and other electronic manufacturing).

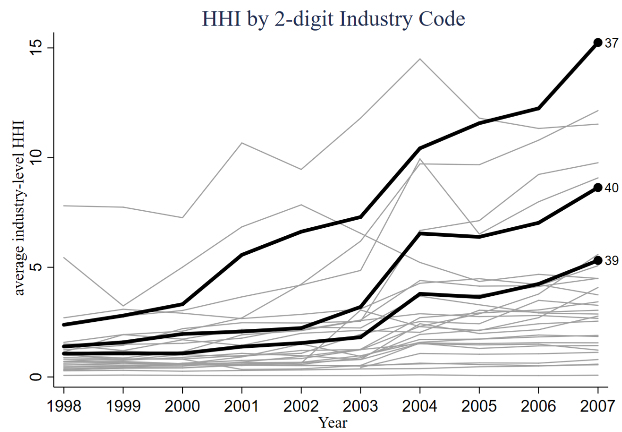


Figure 3.16: HHI at 2-digit CIC level

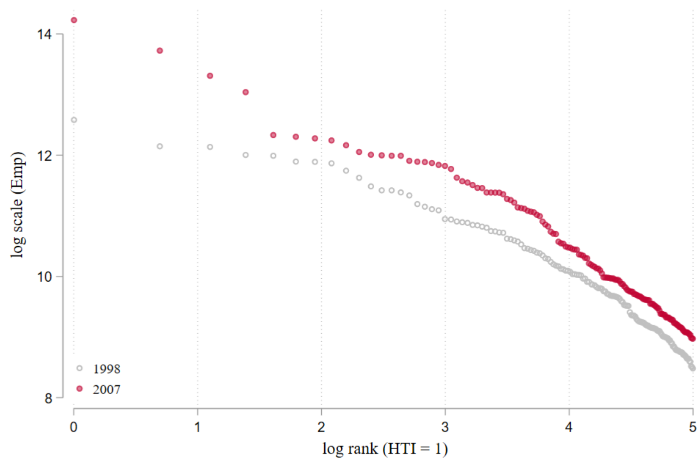


Figure 3.17: Empirical distribution in scale-rank relation for high-tech industry

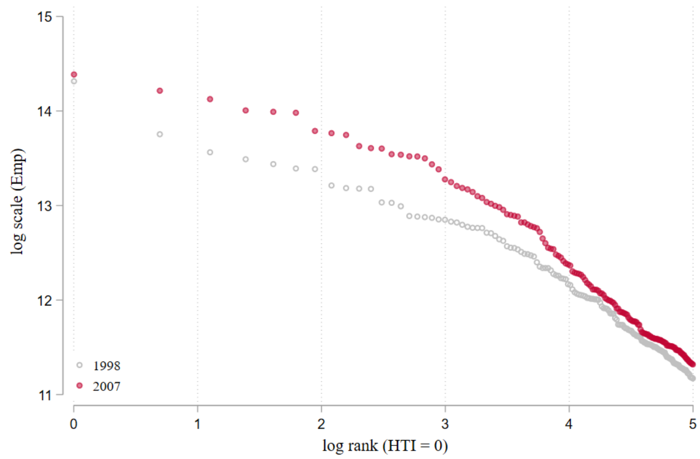


Figure 3.18: Empirical distribution in scale-rank relation for non-high-tech industry

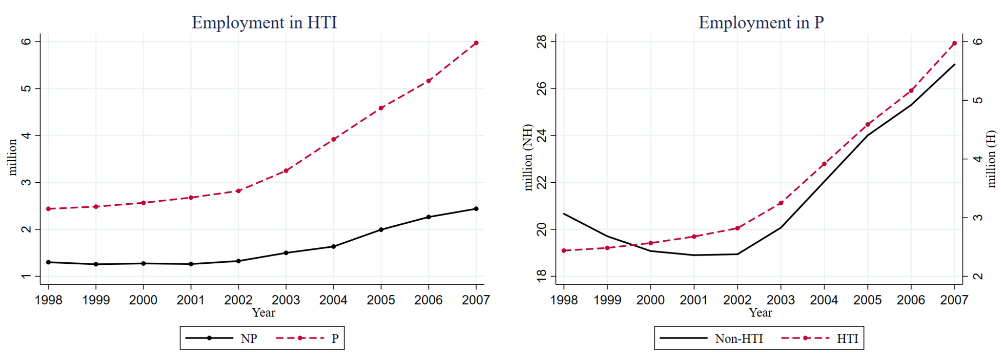


Figure 3.19: Employment by policy groups

3.7.5 Inverse-Propensity Weighting

Figure 3.20, 3.21, and 3.22 show density distributions of the following covariates: sales, public expenditure in science and technology, and export intensity respectively (before and after IPW).

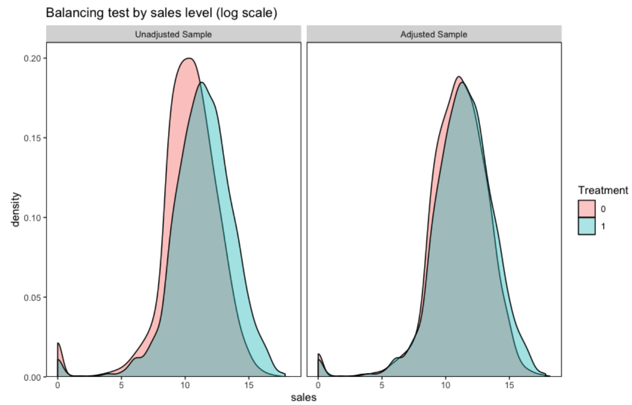


Figure 3.20: Sales level before and after IPW

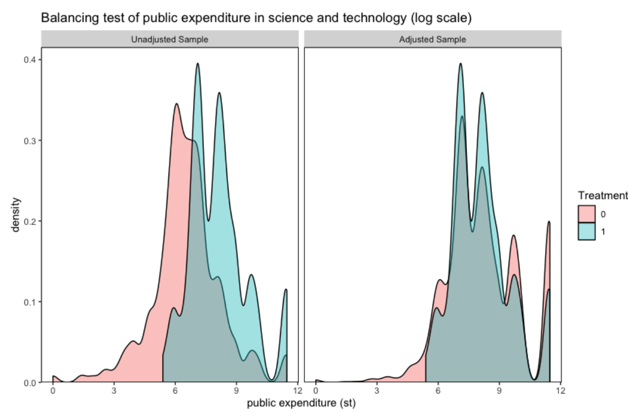


Figure 3.21: Public expenditure in science & technology, before and after IPW

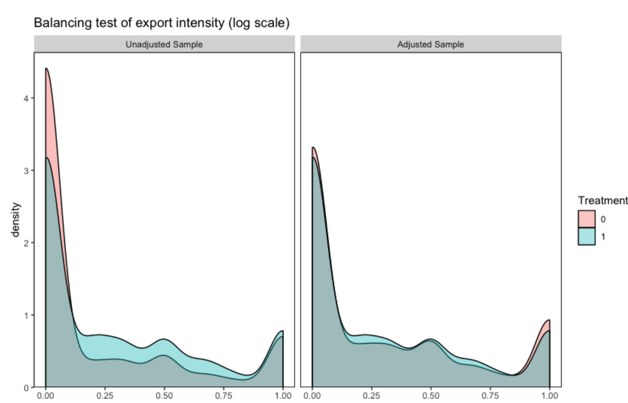


Figure 3.22: Export intensity, before and after IPW

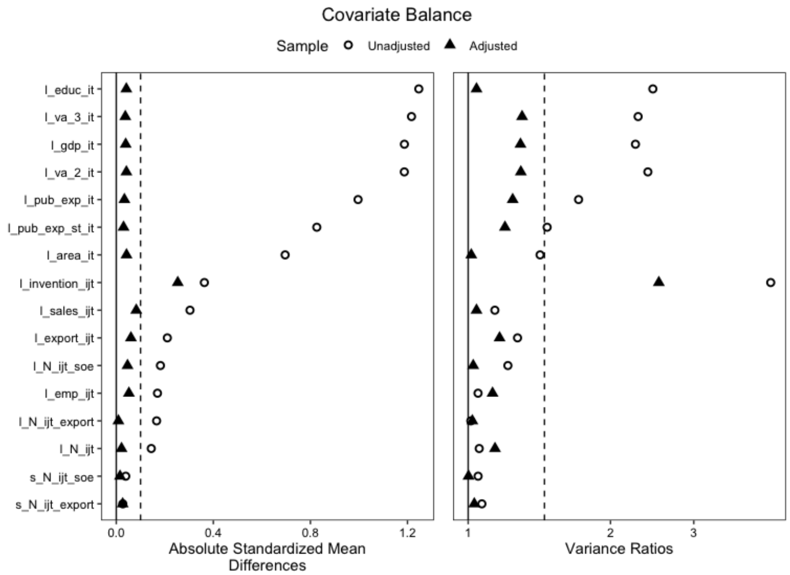


Figure 3.23: Covariate balancing comparison before and after IPW

3.7.6 Supplementary Tables

Table 3.6: Region-Based Estimates by Industry Groups

	HTI	**HTI**	**Non-HTI**	**Non-HTI**
	(1)	(2)	(3)	(4)
$HTC \times Post$	0.5693***	0.5226***	0.5640***	0.5071***
	(5.9085)	(5.7234)	(13.0254)	(12.2580)
Controls	✓	✓	✓	✓
FEs	✓	✓	✓	✓
Obs.	6,830	6,829	29,554	29,551
R^2	0.3450	0.3790	0.3689	0.4040

t statistics in parentheses, with cluseted s.e.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Region-Based Estimates by Industry-I/O Groups

	Close	Close	Far	Far
	(1)	(2)	(3)	(4)
$HTC \times Post$	0.4972***	0.4617***	0.6198***	0.5415***
	(3.4975)	(3.5914)	(5.9475)	(6.0552)
Controls	✓	✓	✓	✓
FEs	✓	✓	✓	✓
Obs.	13,218	13,218	16,335	16,332
R^2	0.3306	0.3569	0.3982	0.4413

t statistics in parentheses, with cluseted s.e.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Region-Based Estimates by Industry-SOE Groups

	H-S	H-NS	NH-S	NH-NS
	(1)	(2)	(3)	(4)
$HTC \times Post$	0.3562***	0.3584***	0.5936***	0.5251***
	(3.3552)	(7.4962)	(5.6469)	(12.0416)
FEs	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Obs.	5,240	22,036	5,868	26,340
R^2	0.4075	0.4653	0.4601	0.4390

t statistics in parentheses, with cluseted s.e. H stands for HTI, S stands for SOE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

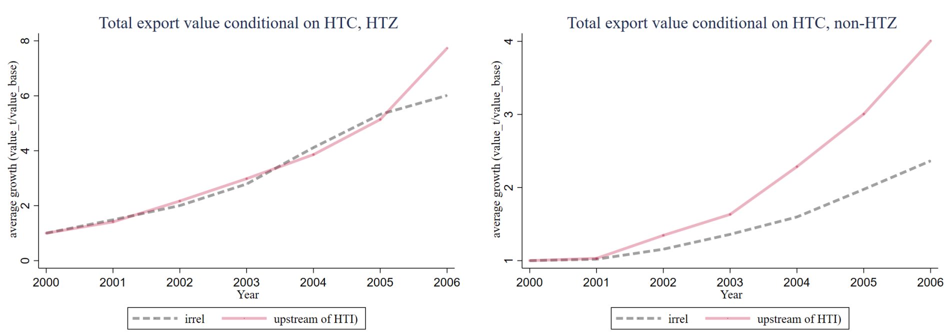


Figure 3.24: Within HTC comparison

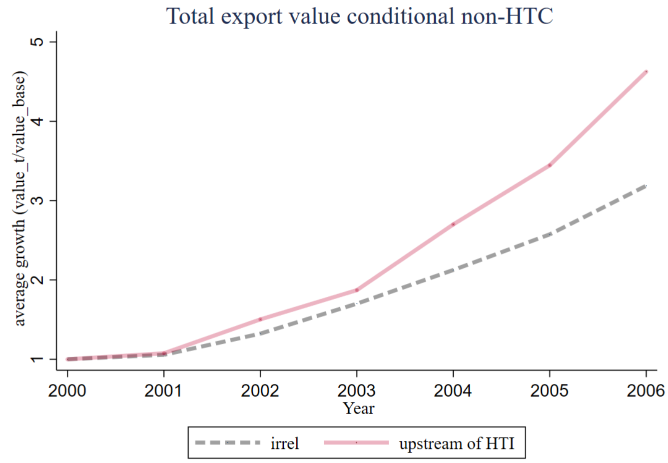


Figure 3.25: Outside of HTC

3.7.7 Sorting Pattern

I use Levinsohn and Petrin (2003) [98] to back out firm-level TFP, then aggregate to region-sector level. The version with TFP backed out with Ackerberg, Caves, and Frazer (2015) [99] approach does not affect the main pattern.

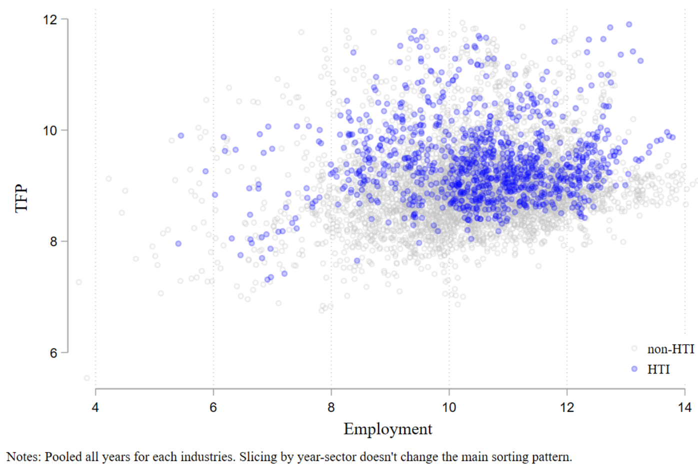


Figure 3.26: Positive sorting between measured TFP and employment

3.7.8 Input-Output Measures

The following plot shows the ranking of sectors by relative demand coefficients.

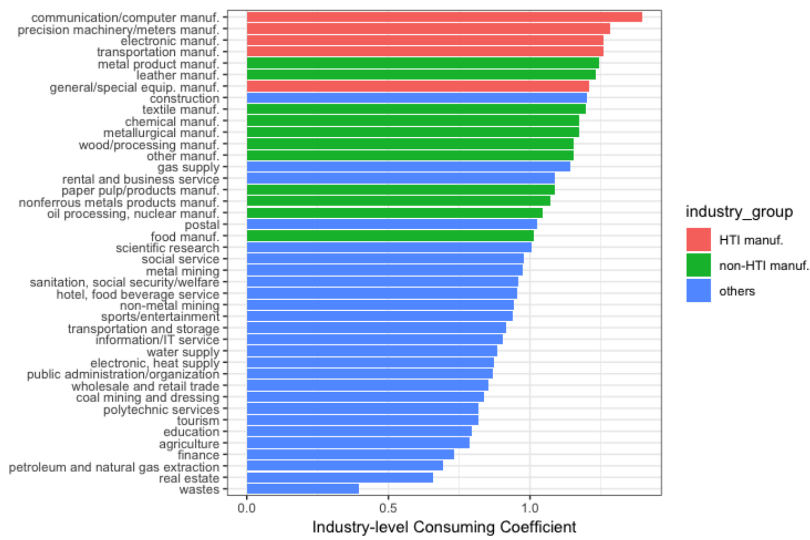


Figure 3.27: Rankings of relative demand coefficients

An alternative way to do it is to apply the upstreamness measurement proposed in Antras et al. (2012) [93].

$$
\begin{aligned}
Y_i &= F_i + Z_i = F_i + \sum_{j=1}^{N} d_{ij} F_j \\
&= F_i + \sum_{j=1}^{N} d_{ij} F_j + \sum_{j=1}^{N}\sum_{k=1}^{N} d_{ik} d_{kj} F_j + \sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} d_{i\ell} d_{\ell k} d_{kj} F_j + \cdots \\
U_i &\equiv 1 \cdot \frac{F_i}{Y_i} + 2 \cdot \frac{\sum_{j=1} N d_{ij} F_j}{Y_i} + 3 \cdot \frac{\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} d_{i\ell} d_{\ell k} d_{kj} F_j}{Y_i} + \cdots \\
U_i &= 1 + \sum_{j=1}^{N} \left(\frac{d_{ij} Y_j}{Y_i} U_j \right)
\end{aligned}
$$

where Y_i denotes the value of gross output in industry i, F_i denotes its use (in terms of value) as a final good. Z_i denotes its use (in terms of value) as an intermediate input to other industries, which can be expressed in a recursive way as an infinite sequence. d_{ij} is the value of sector i's output needed to produce one dollar's worth of industry j's output. The U_i is essentially a weighted average of an industry's output in the value chain, by taking into account of total value used in each step, as well as how each step is close to the end-use stage.

Appendix: Model

3.7.9 Extension: Continuum Locations

In this section I use the continuum version of the current model to highlight the mechanisms of spatial sorting with input-output linkages. Starting with the profit for firm choosing location L_{ij}, which is the same as in the baseline model.

$$\pi_i\left(z_j; L_{ij}\right) = \chi_{ij}\kappa_{\pi,j}\left(\frac{A\left(z_j, L_{ij}\right)}{c_{ij}}\right)^{\sigma_j-1} G^D_{ij}$$

where $\kappa_{\pi,j} \equiv \frac{\left(\sigma_j-1\right)^{\sigma_j-1}}{\sigma_j^{\sigma_j}}\alpha_j^{\alpha_j\left(\sigma_j-1\right)}\prod_{j'\in\mathscr{J}}\left(\left(1-\alpha_j\right)\gamma_{j'j}\right)^{\left(1-\alpha_j\right)\gamma_{j'j}\left(\sigma_j-1\right)}$ is a sectoral constant and local-secotral global demand shifter can also be written as $G^D_{ij} = \int_{i'}\tau_{ii'}^{-\sigma_j}P_{i'j}^{\sigma_j-1}R_{i'j}di'$. Locational choices are charaiceterized by firm sorting:

$$L^*_{ij}\left(z_j\right) = \arg\max_{L_{ij}}\pi\left(z_j, L_{ij}\right)$$

Notice G^D_{ij} is location specific only due to $\tau_{ii'}$ and we integrate price index and quantities over a continuum of locations in "trade effect". Thus, first-order conditions with respect to L_{ij} gives

$$A\left(z_j, L_{ij}\right)^{-1}\frac{\partial A_{ij}}{\partial L_{ij}}c_{ij} = w_i^{\alpha_j}\sum_{j''}\left(\left(1-\alpha_j\right)\gamma_{j''j}P^{-1}_{ij''}\frac{\partial P_{ij''}}{\partial L_{ij}}\prod_{j'}P_{ij'}^{\left(1-\alpha_j\right)\gamma_{j'j}}\right)$$

To write in a more compact form, we have

$$\epsilon_{A,L_{ij}} = \bar{\kappa}_{ij}\sum_{j'}\left(\gamma_{j'j}\epsilon_{P_{ij'},L_{ij}}\right)$$

where $\bar{\kappa}_{ij}$ a local sectoral constant. $\epsilon_{A,L_{ij}}$ is the "self-elasticity" of local sectoral productivity A_{ij} with respect to local sectoral size, and $\epsilon_{P_{ij'},L_{ij}}$ is the "cross-elasticity" of local sectoral price index $P_{ij'}$ with respect to local sectoral size L_{ij}. Namely, $\epsilon_{A,L_{ij}} \equiv \frac{\partial A\left(z_j,L_{ij}\right)}{\partial L_{ij}}\frac{L_{ij}}{A\left(z_j,L_{ij}\right)}$ and $\epsilon_{P_{ij'},L_{ij}} \equiv \frac{\partial P_{ij'}}{\partial L_{ij}}\frac{L_{ij}}{P_{ij'}}$. The first-order condition characterizes the optimal locational choice: given sector j, marginal gain in choosing L_{ij} due to EES equals to the marginal cost reflected by a price elasticity accumulated by all industries interacted with I/O matrix $\left[\gamma_{j'j}\right]$ and factor intensity α_j.

3.7.10 Extension: Factor-Based PBIP

In this section, I show the model with factor-based PBIP. The goal is to show that key moment conditions and the *absorption matrix* (Φ) preserve similar forms as in the baseline model specified in the main text. For simplicity, I assume all sectors share the same level of elasticities of subsititution and external economies of scale. Consider tax on revenue, labor and sector-level material inputs can be potentially different. Normalize labor wedge to be 1, but preserve $t_{y,ij}$ for place-based indusrial revenue tax (or subsidy) and denote $t_{x,ij}$ the place-based indusrial tax (or subsidy) on price of composites. Policy wedge is at sectoral, but do not differ across varieties. In other words, the profit function for each intermediate producer in the market looks like the following:

$$\max_{p_i(\omega_j)} \left(1-t_{y,ij}\right) p_i\left(\omega_j\right) y_i(\omega_j) - w_i \ell_i\left(\omega_j\right) - \sum_{j'} \left(1+t_{x,ij'}\right) P_{ij'} x_{i,j'j}\left(\omega_j\right)$$

Due to the CRS technology, let's first derive unit cost function to easily track the wedge in further steps. Cost minimization with wedge can shown as follows:

$$C\left(y_i(\omega_j)\right) = \min_{\ell(\omega_j),\{x_{i,j'j}(\omega_j)\}_{j'}} w_i \ell_i(\omega_j) + \sum_{j'} \left(1+t_{x,ij'}\right) P_{ij'} x_{i,j'j}\left(\omega_j\right) \text{ s.t. prod. fcn.}$$

Thus we know the unit cost is

$$\begin{aligned} C(y) &= \left(\prod_{j'} \left(\frac{(1+t_{x,ij'})P_{ij'}}{(1-\alpha)\gamma_{j'j}}\right)^{(1-\alpha)\gamma_{j'j}}\right) \left(\frac{w_i}{\alpha}\right)^{\alpha} \frac{y_i(\omega_j)}{A(\omega_j, L_{ij})} \\ &\equiv \chi_{x,ij} c_{ij}^{LF}\left(\{P_{ij'}\}_{j'}, w_i\right) \frac{y_i(\omega_j)}{A(\omega_j, L_{ij})} \end{aligned}$$

where $\chi_{x,ij} \equiv \prod_{j'}(1+t_{x,ij'})^{(1-\alpha)\gamma_{j'j}}$ is the I/O-accumulated wedge. c_{ij}^{LF} is the unit cost in Laissez-faire case as a function of two set of prices for cross-sector composites prices and labor, both at local level.

We can then derive price

$$\begin{aligned} p_i(\omega_j) &= \frac{\sigma}{\sigma-1} \frac{\prod_{j'}(1+t_{x,ij'})^{(1-\alpha)\gamma_{j'j}}}{1-t_{y,ij}} \frac{c_{ij}^{LF}}{A(\omega_j, L_{ij})} \\ &\equiv \chi_{ij} \frac{\sigma}{\sigma-1} \frac{c_{ij}^{LF}}{A(\omega_j, L_{ij})} \equiv \chi_{ij} p_i^{LF}(\omega_j) \end{aligned}$$

where χ_{ij} summarizes effective relative wedges for two inputs and $p_i^{LF}(\omega_j)$ is the price form as that in laissez-faire. We also know factor demand and profits

$$x_{i,j'j}(\omega_j) = \underbrace{(1-\alpha)\gamma_{j'j}}_{\text{factor share w/ CRS}} \underbrace{\frac{\chi_{x,ij}}{1+t_{x,ij'}}}_{\text{IO-induced wedge}} \frac{c_{ij}^{LF} y_i(\omega_j)}{P_{ij'}} \equiv \phi^x_{i,j'j} x^{LF}_{i,j'j}(\omega_j)$$

$$\ell_i(\omega_j) = \alpha\chi_{x,ij}\frac{c_{ij}^{LF} y_i(\omega_j)}{w_i} \equiv \phi^\ell_{ij}\ell_i^{LF}(\omega_j)$$

$$\pi_i(\omega_j) = \left((1-t_{y,ij})p_i(\omega_j) - \chi_{x,ij}\frac{c_{ij}}{A_{ij}}\right)p_i(\omega_j)^{-\sigma}G^D_{ij}$$

$$= \frac{\chi^\sigma_{y,ij}}{\chi^{\sigma-1}_{x,ij}}(\frac{\sigma}{\sigma-1})^{1-\sigma}\left(\frac{A_{ij}}{c_{ij}}\right)^{\sigma-1} G^D_{ij} \equiv \phi^\pi_{ij}\pi_i^{LF}(\omega_j)$$

Thus key equilibrium with tax

$$Q_{ij} = \underbrace{C_{ij}}_{\text{local final usage}} + \underbrace{\sum_{j'}\int_{\mathscr{Z}_{ij'}} \phi^x_{i,j'j} x_{i,jj'}(z_{j'}) dF_{j'}(z)}_{\text{local intermediate usage due to IO matrix}}$$

For example,

$$Q_{11} = C_{11} + \int_0^{z_1^*} \phi^x_{1,11j}\kappa_{x,1}A(z_1,L_{11})^{\sigma-1}\mathscr{P}_{11}^{-1}P_{11}^{-1}G^D_{11}dF_1(z) + \int_0^{z_2^*} \phi^x_{1,12j}\kappa_{x,2}A(z_2,L_{11})^{\sigma-1}\mathscr{P}_{12}^{-1}P_{11}^{-1}G^D_{12}dF_2(z$$

So the demand for composite will be filtered with a tax matrix $\mathscr{T}$.

$$\mathbf{\Phi}_{PBIP} \equiv \left[\begin{array}{cc:cc} T\tilde{L}^S_{1,11}\tilde{G}^D_{11,1} & T\tilde{L}^S_{1,12}\tilde{G}^D_{11,2} & T\tilde{L}^S_{1,11}\tilde{G}^D_{12,1} & T\tilde{L}^S_{1,12}\tilde{G}^D_{12,2} \\ T\tilde{L}^S_{1,21}\tilde{G}^D_{11,1} & T\tilde{L}^S_{1,22}\tilde{G}^D_{11,2} & T\tilde{L}^S_{1,21}\tilde{G}^D_{12,1} & T\tilde{L}^S_{1,22}\tilde{G}^D_{12,2} \\ \hdashline T\tilde{L}^S_{2,11}\tilde{G}^D_{21,1} & T\tilde{L}^S_{2,12}\tilde{G}^D_{21} & T\tilde{L}^S_{2,11}\tilde{G}^D_{22,1} & T\tilde{L}^S_{2,12}\tilde{G}^D_{22,2} \\ T\tilde{L}^S_{2,21}\tilde{G}^D_{21,1} & T\tilde{L}^S_{2,22}\tilde{G}^D_{21,2} & T\tilde{L}^S_{2,21}\tilde{G}^D_{22,1} & T\tilde{L}^S_{2,22}\tilde{G}^D_{22,2} \end{array}\right] \equiv (\mathscr{T}\circ\mathscr{L}\circ\mathscr{G})_{ij\times ij}$$

This form preserves that in the benchmark model presented in the main content, but with a generalization.

We can also characterize sorting cutoff with PBIP:

$$\frac{\chi^\sigma_{y,1j}}{\chi^{\sigma-1}_{x,1j}}\frac{\chi^{\sigma-1}_{x,2j}}{\chi^\sigma_{y,2j}}\left(\frac{L_{1j}}{L_{2j}}\right)^{\theta(\sigma-1)} = \frac{\tau_{21}^{-\sigma}P^\sigma_{1j}Q_{1j} + \tau_{22}^{-\sigma}P^\sigma_{2j}Q_{2j}}{\tau_{11}^{-\sigma}P^\sigma_{1j}Q_{1j} + \tau_{12}^{-\sigma}P^\sigma_{2j}Q_{2j}}\left(\frac{\mathscr{P}_{1j}}{\mathscr{P}_{2j}}\right)$$

Thus, a simplified system with factor-based PBIP (labor wedges normalized) consists of (1) goods markets clear: $[\mathbf{P}\circ\mathbf{Q}]_{ij\times 1} = [I-\mathbf{\Phi}_{\mathbf{PBIP}}]^{-1}_{ij\times ij}[\eta\mathbf{wL}]_{ij\times 1}$, where $P = \mathscr{F}\left(\mathscr{K}\mathscr{W}\tau\tilde{\mathscr{A}}(z_j^*, L_{ij})\right)$. (2) Labor markets clear: $L_{ij}^{1-\theta(\sigma-1)} = \phi^\ell_{ij}\kappa_{\ell,j}\mathscr{P}_{ij}^{-1}\Delta_i(z_j^*)G^D_{ij}$. And (3) sorting $\frac{\phi^\pi_{1j}}{\phi^\pi_{2j}}\frac{L_{1j}}{L_{2j}} = \left[\left(\frac{G^D_{2j}}{G^D_{1j}}\right)\left(\frac{\mathscr{P}_{1j}}{\mathscr{P}_{2j}}\right)\right]^{\frac{1}{\theta(\sigma-1)}}$, $\forall j \in \mathscr{J}$.

3.7.11 Extension: Fixed Costs in Entry and Trade

In this section I discuss a model with fixed costs of entry and exports. The goals are two fold. First, there is extra complication introduced in estimation by introducing these costs, which is beyond of this chapter's estimation scope. Second, I show how this model is generalization of Costinot and Rodríguez-Clare (2014) [85].

Consider two fixed costs with the following setup. In order to get a productivity draw, firms must pay a fixed sector-specific entry cost f_j^e. In order to sell to location i, firms from country must then pay a fixed cost of export f_{ij}^{ex}. Once fixed entry costs and fixed exporting costs have been paid, the constant cost of producing and delivering by z_j from location i in location i' is given by $\frac{c_{ij}\tau_{ii'}}{A(z_j,L_{ij})}$. I follow the Pareto distribution assumption as in Costinot and Rodríguez-Clare (2014) [85]. This is just for tractablility, and one can think of more general distributions. Formally, assume productivity z_j independently drawn across monoplistically competitive firms in each sector upon entry from a Pareto distribution G, with dispersion parameter $\phi_j > \sigma - 1$ and lower bound b (normalized to be 1):

$$G_j(z) = 1 - z^{-\phi_j}, \ \forall z \geq 1$$

My framework exhibits endogenous spatial sorting, thus the productivity draw is not i.i.d. with respect to each location. This is one of the major distinction from Costinot and Rodríguez-Clare (2014) [85]. As shown below, it matters for the key general price equation. From individual firm's revenue and the standard sorting to export implies $\frac{r_i(z_j)}{\sigma_j} = f_{ij}^{ex}$ implies

$$A\left(z_j^*, L_{ij}\right) = \frac{c_{ij}\chi_{ij}^{\sigma_j} f_{ij}^{ex}}{\kappa_{p,j}^{1-\sigma j} \sum_{i'} \tau_{ii'}^{-\sigma_j} P_{i'j}^{\sigma_j - 1} R_{i'j}}$$

With the log-linear assumption on agglomeration externalities, together twith the price index in the main context, we can obtain the sectoral price between locations as

$$\begin{aligned} P_{ii',j} &= \left(M_j \int_{\mathscr{Z}_j^*} \left[\frac{\kappa_{p,j} c_{ij} \tau_{ii'}}{z_j L_{ij}^{\theta}} \right]^{1-\sigma_j} dG_j(z) \right)^{\frac{1}{1-\sigma_j}} \\ &= \left(\left[\frac{M_j \kappa_{p,j} c_{ij} \tau_{ii'}}{L_{ij}^{\theta}} \right]^{1-\sigma_j} \frac{\left[z^{\sigma_j - \phi_j - 1} \right]_{z_j^*}^{z_j^+}}{\phi_j - (\sigma_j - 1)} \right)^{\frac{1}{1-\sigma_j}} \end{aligned}$$

where z_j^+ is the cutoff productivity at the sorted location above i (denoted by i^+). As standard with entry costs derivation in trade literature, free entry costs implies

$$f_j^e = M_j \kappa_{\pi,j} R_j P_j^{\sigma_j - 1} \sum_i \left(\int_{\mathscr{Z}_{ij}} \left(\frac{c_{ij}}{A(z_j, L_{ij})} \right)^{\sigma_j - 1} dF_j(z) \right)$$

With some algebra, we can further obtain

$$P_{ii',j} = \xi_j \underbrace{\frac{c_{ij}\tau_{ii'}}{L_{ij}^{\theta_j}}}_{\text{Intensive margin}} \underbrace{\left(\frac{R_j}{f_j^e}\right)^{\frac{1}{1-\sigma_j}}}_{\text{Extensive Margin (entry)}} \underbrace{\left[\mathscr{E}(G, \tau, c, f^{ex}, P; \Theta)\right]^{\frac{\phi_j - (\sigma_j - 1)}{\sigma_j - 1}}}_{\text{Extensive Margin (selection)}}$$

where the new version of extensive margin (selection) consists of

$$\mathscr{E}(G, \tau, c, f^{ex}, P; \Theta) \equiv \left(\frac{G_{i^+,j}}{f_{i^+,j}^{ex} \chi_{i^+,j}} \right)^{\frac{1}{1-\sigma_j}} \left(\frac{c_{i^+,j}}{L_{i^+,j}^{\theta}} \right) - \left(\frac{G_{ij}}{f_{ij}^{ex} \chi_{ij}} \right)^{\frac{1}{1-\sigma_j}} \left(\frac{c_{ij}}{L_{ij}^{\theta}} \right)$$

and $\xi_j \equiv \kappa_{p,j}^{\sigma_j - \phi_j - 1} \left(\phi_j - (\sigma_j - 1)\right)^{\frac{\sigma_j - \phi_j - 1}{\sigma_j - 1}}$ a sectoral shifter.

There are a couple of notes on this expression as a generalized version of Costinot and Rodríguez-Clare (2014) [85]. First, intensive margin and extensive margin in selection include both cross-sectoral general equilibrium price and local agglomeration externalities. Second, extensive margin (selection) also embeds spatial sorting, reflected by the integration over equilibrium locational choice of firms. Due to data availability, we cannot seperately identify entry costs and elasticities, even though major margins are log-linear. Furthermore, extensive margin of selection also involves ordered statistics and additive terms. These, together with the nonlinear agglomeration externalities, further complicate the estimation.

3.7.12 Extension: Imperfect Sorting

In this section, I extend the model to imperfect sorting. This is to show how spatial sorting holds systematically with shocks, without changing the intuition delivered in the baseline model. The imperfect sorting inherits log-supermodular setup as in Gaubert (2018) [7]. As we will see, log-supermodularity eases the estimation with a parametric decomposition in agglomeration. Assume there exists sector-level sorting s_j, and sector-level agglomeration externalities. Specifically, assume the following productivity function for firm with productivity draw z:

$$log(A_j(z, L_{ij}; s_j, a_j)) = a_j log(L_{ij}) + log(z_i) \left(1 + log\left(\frac{L_{ij}}{L_{0j}} \right) \right)^{s_j} + \epsilon_{i,L_{ij}}$$

if $log(z_i) \geq 0, L_{ij} \geq L_{0j}$, and $log(A_j(z, L_{ij}; s_j, a_j)) = 0$ for $L_{ij} < L_{0j}$. L_{0j} is the minimum effective sectoral size. And we assume $\epsilon_{i,L}$ is i.i.d. across cities and firms from a type-I extreme value distribution with mean zero and variance ζ_j. Thus firms' locational choice is

$$L^*(z_j) = \arg max_{L_{ij}} V_i(z_j) exp\{(\sigma_j - 1)\epsilon_i\}$$

where $V_i(z_j)$ is the determistic part at firm level. Invoking monotone comparative statics, we obtain the positive sorting property of firms along city sizes within each sector. Namely, when two firms within a the sector draw different productivities, for example, $z_{1,j} < z_{2,j}$, we have $F(L_{ij}|z_{2,j})\ FOSD\ (L_{ij}|z_{1,j})$. With idiosyncratic shocks, the probability that a firm locates in a location is the following regardless of the supermodularity assumption.

$$\mathbf{P}(L_{ij}|z_j) = \frac{\left(\chi_{ij}\kappa_{\pi,j}\left(\frac{A\left(z\left(\omega_j\right),L_{ij}\right)}{c_{ij}}\right)^{\sigma_j - 1} G^D_{ij}\right)^{\frac{\zeta_j}{\sigma_j - 1}}}{\sum_i\left(\chi_{ij}\kappa_{\pi,j}\left(\frac{A\left(z\left(\omega_j\right),L_{ij}\right)}{c_{ij}}\right)^{\sigma_j - 1} G^D_{ij}\right)^{\frac{\zeta_j}{\sigma_j - 1}}}$$

3.7.13 Model in a Simple Economy

In what follows, I elaborate the model in a 2×2 economy by decomposing flows of goods, and derive key equations as mentioned in the main text.

From Intermediates to Composites (Cross-Location and Within-Sector)

Figure 3.28 illustrates all flows for these tradable intermediate goods in two layers of production processes. The first subscript denotes location i, followed by industry j. Arrows are pointing from suppliers to buyers.

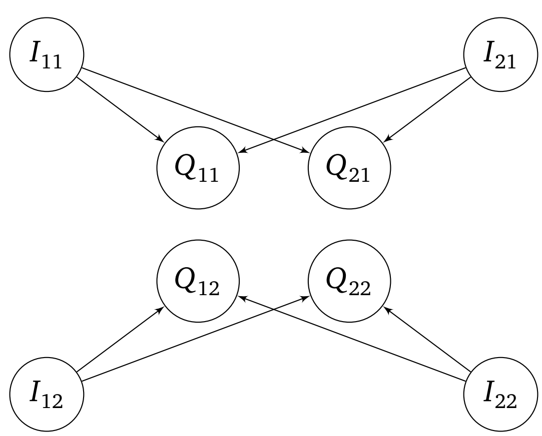


Figure 3.28: *Trade flows (intra- and inter-regional) from intermediate products I_{ij} to composite products Q_{ij}*

In a simple economy, given that assembly of compsite goods uses intermediate inputs from its own sector, we can characterize eight flows. With demand $q_i^{i'}(z_j)$, we can rewrite them in temrs of location-sector level price index, quantity and sector-level cutoffs with effective variety supports. Consider using origin-destination-sector-wise flows seperately, and denote a trade flow from location i to location i' within industry j by $X(i,i')|_j$. For example, conditional on sector $j=1$, flows from $i=1$ are:

$$X(1,1)|_{j=1} = \int_{\mathscr{Z}_{11}} q_1^1(z_1)p_1(z_1)\tau_{11}dF(z_1) = P_{11}^{\sigma}Q_{11}\int_{\mathscr{Z}_{11}} \left(p_1(z_1)\tau_{11}\right)^{1-\sigma} dF(z_1)$$
$$X(1,2)|_{j=1} = \int_{\mathscr{Z}_{11}} q_2^1(z_1)p_1(z_1)\tau_{12}dF(z_1) = P_{21}^{\sigma}Q_{21}\int_{\mathscr{Z}_{11}} \left(p_1(z_1)\tau_{12}\right)^{1-\sigma} dF(z_1)$$

where $p_{i'}\left(z_j\right)\tau_{i'i} = \kappa_{p,j}\frac{w_{i'}^{\alpha}\prod_{j'}P_{i'j'}^{(1-\alpha)\gamma_{j'j}}}{A\left(z_{j|i'},L_{i'j}\right)}\tau_{i'i}$. $\mathscr{Z}_{ij}$ is the effective variety supports in equilibrium as a function of cutoffs.[56]

From Composites to All Uses (Within-location and Cross-sector)

The following graph illustrates the cross-sector flows. Note that these flows occur within location. Bold ones highlight input-output linkages. The first subscript denotes location i, followed by industry j.

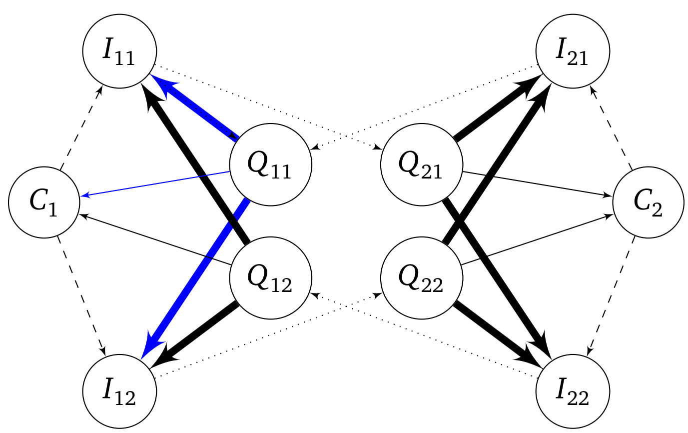


Figure 3.29: *The highlighted bold solid arrows are flows from composite sectors Q_{ij} to intermediate producers I_{ij}. Blue arrows are how Q_{11} flows to all possible buyers, where bold arrows to producers whereas thin arrows are to consumers. Dotted arrows are inter-regional trade in the previous section.*

[56] Conditional on each industry, there are I^2 possible flows. Here I show two examples for $j=1$, which are similar for $j=2$.

Composite goods are sold to local intermediate producers across all sectors, summarized by

$$Q_{ij} = \underbrace{C_{ij}}_{\text{local final usage}} + \underbrace{\sum_{j'} \int_{\Omega_{ij'}} x_{i,jj'}(\omega_{j'}) d\omega_{j'}}_{\text{local intermediate usage due to IO matrix}} \equiv C_{ij} + \sum_{j'} I_{ij'}$$

where $x_{i,j'j}\left(\omega_j\right) = \kappa_{x,j} \frac{A\left(z_i\left(\omega_j\right), L_{ij}\right)^{\sigma-1}}{w_i^{\alpha(\sigma-1)} P_{ij'} \prod_{j''} P_{ij''}^{(\sigma-1)(1-\alpha)\gamma_{j''j}}} G_{ij}^D$ (generic notations in subscipts). Although above reflects local supply chain, $I_{i,j}$ has trade embeded implicitly, as shown in dotted lines across boarders. For an arbitrary Q_{ij}, it equals to the demand from both sectors within location i, including demand from both consumers and intermedaite producers. For clarity in notation, I write more notations in Q_{11} but abbreviate a bit in other expressions.

$$\begin{aligned} Q_{11} = \underbrace{C_{11}}_{\text{direct final usage}} &+ \underbrace{\kappa_{x,1} \mathscr{A}_{11}^{\sigma-1} \mathscr{W}_1 \mathscr{P}_{11}^{-1} P_{11}}_{\text{local supply shifter (LSS) } L^S_{1,11} \text{ with I/O } (Q \to I)} \underbrace{\left(\tau_{11}^{-\sigma} P_{11}^{\sigma} Q_{11} + \tau_{12}^{-\sigma} P_{21}^{\sigma} Q_{21}\right)}_{\text{global demand shifter (GDS) } G^D_{11} \text{ with trade } (I \to Q)} \\ &+ \underbrace{\kappa_{x,2} \mathscr{A}_{12}^{\sigma-1} \mathscr{W}_1 \mathscr{P}_{12}^{-1} P_{11}}_{\text{local supply shifter } L^S_{1,12} \text{ with I/O } (Q \to I)} \underbrace{\left(\tau_{11}^{-\sigma} P_{12}^{\sigma} Q_{12} + \tau_{12}^{-\sigma} P_{22}^{\sigma} Q_{22}\right)}_{\text{global demand shifter } G^D_{12} \text{ with trade } (I \to Q)} \end{aligned}$$

where $\mathscr{W}_i \equiv w_i^{-\alpha(\sigma-1)}$ and $\mathscr{A}_{ij} \equiv \left(\int_{\mathscr{Z}_{ij}} A(z_j, L_{ij})^{\sigma-1} dF_j(z)\right)^{\frac{1}{\sigma-1}}$.

Similarly, we can write decomposition of other composite output supply ($\{Q_{ij}\}$) in terms of final usage, local supply shifter (LSS, or L^S_{ij}), and global demand shifter (GDS, or G^D_{ij}). For example,

$$\begin{aligned} Q_{12} = \underbrace{C_{12}}_{\text{direct final usage}} &+ \underbrace{\kappa_{x,1} \mathscr{A}_{11}^{\sigma-1} \mathscr{W}_1 \mathscr{P}_{11}^{-1} P_{12}}_{\text{LSS } L^S_{1,21} \text{ with I/O}} \underbrace{\left(\tau_{11}^{-\sigma} P_{11}^{\sigma} Q_{11} + \tau_{12}^{-\sigma} P_{21}^{\sigma} Q_{21}\right)}_{\text{GDS } G^D_{11} \text{ with trade}} \\ &+ \underbrace{\kappa_{x,2} \mathscr{A}_{12}^{\sigma-1} \mathscr{W}_1 \mathscr{P}_{12}^{-1} P_{12}}_{\text{LSS } L^S_{1,22} \text{ with I/O}} \underbrace{\left(\tau_{11}^{-\sigma} P_{12}^{\sigma} Q_{12} + \tau_{12}^{-\sigma} P_{22}^{\sigma} Q_{22}\right)}_{\text{GDS } G^D_{12} \text{ with trade}} \end{aligned}$$

Note that there exists interactions between I/O linkages and trade effect through sorting equilibrium. Thus we obtain

$$\mathbf{Q} = (\mathbf{I} - \mathbf{\Phi})^{-1} \mathbf{C}$$

The above shows a recursive form, which can be regarded as a generalized version of input-output analysis with basic Leontif inverse and the upstreamness measure as in Antras et al. (2012) [93].

$$\begin{bmatrix} Q_{11} \\ Q_{12} \\ Q_{21} \\ Q_{22} \end{bmatrix} = \begin{bmatrix} C_{11} \\ C_{12} \\ C_{21} \\ C_{22} \end{bmatrix} + \begin{bmatrix} L^S_{1,11}G^D_{11} + L^S_{1,12}G^D_{12} \\ L^S_{1,21}G^D_{11} + L^S_{1,22}G^D_{12} \\ L^S_{2,11}G^D_{21} + L^S_{2,12}G^D_{22} \\ L^S_{2,21}G^D_{21} + L^S_{2,22}G^D_{22} \end{bmatrix} = \begin{bmatrix} C_{11} \\ C_{12} \\ C_{21} \\ C_{22} \end{bmatrix} + \mathbf{\Phi} \begin{bmatrix} Q_{11} \\ Q_{12} \\ Q_{21} \\ Q_{22} \end{bmatrix}$$

where

$$\begin{aligned}
\mathbf{\Phi} &= \begin{bmatrix} \kappa_{x,11}CA_{11}\tau_{11}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,11}CA_{11}\tau_{1I}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,1J}CA_{1J}\tau_{1I}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \\ \vdots & & & & \\ \kappa_{x,11}CA_{I1}\tau_{I1}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,11}CA_{I1}\tau_{II}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,1J}CA_{IJ}\tau_{II}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \\ \vdots & & \vdots & & \vdots \\ \kappa_{x,J1}CA_{11}\tau_{11}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,J1}CA_{11}\tau_{1I}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,JJ}CA_{1J}\tau_{1I}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \\ \vdots & & & & \\ \kappa_{x,J1}CA_{I1}\tau_{I1}^{-\sigma_1}P_{11}^{\sigma_1-1} & \cdots & \kappa_{x,J1}CA_{I1}\tau_{II}^{-\sigma_1}P_{I1}^{\sigma_1-1} & \cdots & \kappa_{x,JJ}CA_{IJ}\tau_{II}^{-\sigma_J}P_{IJ}^{\sigma_J-1} \end{bmatrix} \\
&\equiv \begin{bmatrix} \tilde{L}^S_{11,11}\tilde{G}^D_{11,11} & \cdots & \tilde{L}^S_{11,11}\tilde{G}^D_{1I,I1} & \cdots & \tilde{L}^S_{1J,1J}\tilde{G}^D_{1I,IJ} \\ \vdots & & \vdots & & \vdots \\ \tilde{L}^S_{11,I1}\tilde{G}^D_{I1,11} & \cdots & \tilde{L}^S_{11,I1}\tilde{G}^D_{II,I1} & \cdots & \tilde{L}^S_{1J,IJ}\tilde{G}^D_{II,IJ} \\ \vdots & & \vdots & & \vdots \\ \tilde{L}^S_{J1,11}\tilde{G}^D_{11,11} & \cdots & \tilde{L}^S_{J1,11}\tilde{G}^D_{1I,I1} & \cdots & \tilde{L}^S_{JJ,1J}\tilde{G}^D_{1I,IJ} \\ \vdots & & \vdots & & \vdots \\ \tilde{L}^S_{J1,I1}\tilde{G}^D_{I1,I1} & \cdots & \tilde{L}^S_{J1,I1}\tilde{G}^D_{II,I1} & \cdots & \tilde{L}^S_{JJ,IJ}\tilde{G}^D_{II,IJ} \end{bmatrix} \\
&\equiv (\mathscr{L} \circ \mathscr{G})_{(I\times J)\times(I\times J)}
\end{aligned}$$

The last expression is a Hadamard product of price-adjusted $\mathscr{L}(z_j^*, \{L_{ij}\}, \{P_{ij}\})$ (with elements $\tilde{L} \equiv \frac{LSS}{P_{ij}}$) and quantity-adjusted form $\mathscr{G}(\{\tau_{ii'}\}, \{P_{ij}\})$ (with elements $\tilde{G} \equiv \frac{GDS}{Q} = \left(\frac{P}{\tau}\right)^{\sigma}$). With the Cobb-Douglas consumption for composites, we know consumers spend constant share of their local income in specific sectors: $C_{ij} = \frac{\eta_j \tilde{w}_i L_i}{P_{ij}}$ where $\tilde{w}$ disposable imcome. Thus we can express composite sales in terms of size, price and the composite term reflecting local supply and global demand:

$$\begin{aligned}
[\mathbf{PQ}]_{ij\times 1} &= (\mathbf{I} - \mathscr{L} \circ \mathscr{G})^{-1}_{ij\times ij}[\eta \mathbf{wL}]_{ij\times 1} \equiv \Psi \eta \mathbf{wL} \\
\Rightarrow \lambda_{ij} &\equiv \frac{P_{ij}Q_{ij}}{GDP} = (\mathbf{I} - \mathscr{L} \circ \mathscr{G})^{-1}_{ij\times ij}\eta_j
\end{aligned}$$

where is ψ_{ij} entry of a generalized "Leontif inverse" Ψ. Propagation channels involve both within-sector/location and across-sector/location on top of sorting. Thus this model differs from trade literature as Allen and Arkolakis (2014) [84], or network literature as Hulten (1978) [100], Acemoglu et al. (2012) [95] and

Liu (2019) [87]. In Acemoglu et al. (2012) [95], influence vetors are defined as $\mathbf{v} = \frac{1}{n}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{1}$. In my case, the analogous influence vector is in the form of a generalized Leontief inverse $\Psi \equiv (\mathbf{I} - (\mathscr{L} \circ \mathscr{G}))^{-1}$, which will be useful in constructing upstreamness/downstreamness. The standard (sales-based) Domar weigths in network literature (such as Carvalho and Tahbaz-Salehi (2018) [101]) is sales share as a fraction of GDP. In my context, define aggregate Domar weight Λ_{agg}, sectoral-based Domar weight Λ_{sec} and location-based Domar weight Λ_{loc} as the follows. This is, first, for expositional and decompositional purposes; and second, to map to empirical moments,

$$\Lambda_{ij}^{\text{agg}} = \frac{\int_{\Omega_{ij}} p_i(\omega_j) y_i(\omega_j) d\omega_j}{\sum_i \sum_j \int_{\Omega_{ij}} p_i(\omega_j) y_i(\omega_j) d\omega_j}$$

$$\Lambda_{ij}^{\text{sec}} = \frac{\int_{\Omega_{ij}} p_i(\omega_j) y_i(\omega_j) d\omega_j}{\sum_i \int_{\Omega_{ij}} p_i(\omega_j) y_i(\omega_j) d\omega_j}$$

$$\Lambda_{ij}^{\text{loc}} = \frac{\int_{\Omega_{ij}} p_i(\omega_j) y_i(\omega_j) d\omega_j}{\sum_j \int_{\Omega_{ij}} p_i(\omega_j) y_i(\omega_j) d\omega_j}$$

Dividing both sides by total income (value added) in the economy gives aggregate GDP-based Domar weights connected with employment-based Domar weights.

$$\Lambda_{ij}^{agg} = \eta_j \lambda_i + \sum_k \mathbf{\Phi}_{2*(i-1)+j,k} \Lambda_{ij}^{agg}, \ \forall i, j$$

Thus we obtain $\Lambda_{agg} = (\mathbf{I} - (\mathscr{L} \circ \mathscr{G}))^{-1}[\eta\lambda]$, where $[\Lambda_{agg}]_{ij} = \frac{P_{ij}Q_{ij}}{\sum_i \sum_j P_{ij}Q_{ij}}, \lambda_i = \frac{L_i}{L}$.[57]

[57] We use capital Λ for final goods production and lowercase λ for labor Domar weights.

3.7.14 A Simple Economy, with Numerical Analysis

In a 2×2 economy, equilibrium consists of $\{L_{ij}, Q_{ij}, P_{ij}, z_j^*, t_i, \bar{u}\}_{\mathscr{I},\mathscr{J}}$ such that

$$w_i L_{ij} = \kappa_{\ell,j} CA_{ij} \sum_{i'} \left(\tau_{ii'}^{-\sigma_j} P_{i'j}^{\sigma_j} Q_{i'j} \right), \ \forall i,j$$

$$P_{ij} Q_{ij} = \eta_j \tilde{w}_i + \sum_{j'} \left[\kappa_{ij'} CA_{ij'} \left(\sum_{i'} \tau_{ii'}^{-\sigma_{j'}} P_{i'j'}^{\sigma_{j'}} Q_{i'j'} \right) \right], \ \forall i,j$$

$$P_{ij}^{1-\sigma_j} = \kappa_{p,j}^{1-\sigma_j} \sum_{i'} \left[\phi_{i'}(z_j^*) \tau_{i'i}^{1-\sigma_j} L_{i'j}^{\theta(\sigma_j-1)} \left(c_{i'j} \right)^{1-\sigma_j} \right], \ \forall i,j$$

$$\left(\chi_j^+ \right) \left(\mathscr{A}_j^+ \right) \frac{\sum_{i'} \tau_{1i'}^{-\sigma_j} P_{i'j}^{\sigma_j} Q_{i'j}}{\sum_{i'} \tau_{2i'}^{-\sigma_j} P_{i'j}^{\sigma_j} Q_{i'j}} = \left(\frac{c_{1j}}{c_{2j}} \right)^{\sigma_j - 1}, \ \forall j$$

$$t_i \sum_j L_{ij} = \sum_j s_{ij} \left[\kappa_{\pi,j} CA_{ij} \sum_{i'} \left(\frac{\tau_{ii'}^{-\sigma_j}}{P_{i'j}^{1-\sigma_j}} (P_{i'j} Q_{i'j}) \right) \right], \ \forall i$$

$$u_i = \bar{u}, \ \forall i$$

where $CA_{ij} \equiv \left(\frac{\phi_i(z_j^*) L_{ij}^{\theta}}{c_{ij}} \right)^{\sigma_j - 1}$, $c_{ij} \equiv w_i^{\alpha_j} \prod_{j'} P_{ij'}^{(1-\alpha_j)\gamma_{j'j}}$. Composite local sectoral level productivities are $\phi_1(z_j^*) \equiv \left(\left(\frac{z_j^*}{\sigma_j} \right)^{\sigma_j} - \left(\frac{\underline{z}_j}{\sigma_j} \right)^{\sigma_j} \right)^{\frac{1}{\sigma_j - 1}}$ and $\phi_2(z_j^*) \equiv \left(\left(\frac{\bar{z}_j}{\sigma_j} \right)^{\sigma_j} - \left(\frac{z_j^*}{\sigma_j} \right)^{\sigma_j} \right)^{\frac{1}{\sigma_j - 1}}$. These expressions are found to be useful in quantitative analysis. $\kappa_{\ell,j}, \kappa_{p,j}, \kappa_{x,j'j}$ are industry-level constants. Specifically, $\kappa_{\ell,j} \equiv \left(\frac{\sigma_j}{\sigma_j - 1} \right)^{-\sigma_j} \alpha^{1-\alpha(1-\sigma_j)} \prod_{j' \in \mathscr{J}} \left((1-\alpha) \gamma_{j'j} \right)^{-(1-\sigma_j)(1-\alpha)\gamma_{j'j}}$, $\kappa_{p,j} \equiv \frac{\sigma_j}{\sigma_j - 1} \frac{\alpha^{-\alpha}}{\prod_{j'} (\gamma_{j'j}(1-\alpha))^{\gamma_{j'j}(1-\alpha)}} \kappa_{x,j'j}$, and

$$\kappa_{x,j'j} \equiv \left(\frac{\sigma_j}{\sigma_j - 1} \right)^{-\sigma_j} \alpha^{\alpha(\sigma_j - 1)} (1-\alpha) \gamma_{j'j} \prod_{j'} \left((1-\alpha) \gamma_{j'j} \right)^{(1-\alpha)\gamma_{j'j}(\sigma_j - 1)}$$

The generalized "Leontif-inverse" in a 2×2 economy is shown below, where I write in a way ready to decompose local and global effects.

$$\mathbf{\Phi} \equiv \left[\begin{array}{cc|cc} \tilde{L}_{1,11}^S \tilde{G}_{11,1}^D & \tilde{L}_{1,12}^S \tilde{G}_{11,2}^D & \tilde{L}_{1,11}^S \tilde{G}_{12,1}^D & \tilde{L}_{1,12}^S \tilde{G}_{12,2}^D \\ \tilde{L}_{1,21}^S \tilde{G}_{11,1}^D & \tilde{L}_{1,22}^S \tilde{G}_{11,2}^D & \tilde{L}_{1,21}^S \tilde{G}_{12,1}^D & \tilde{L}_{1,22}^S \tilde{G}_{12,2}^D \\ \hline \tilde{L}_{2,11}^S \tilde{G}_{21,1}^D & \tilde{L}_{2,12}^S \tilde{G}_{21}^D & \tilde{L}_{2,11}^S \tilde{G}_{22,1}^D & \tilde{L}_{2,12}^S \tilde{G}_{22,2}^D \\ \tilde{L}_{2,21}^S \tilde{G}_{21,1}^D & \tilde{L}_{2,22}^S \tilde{G}_{21,2}^D & \tilde{L}_{2,21}^S \tilde{G}_{22,1}^D & \tilde{L}_{2,22}^S \tilde{G}_{22,2}^D \end{array} \right]$$

We can further map the above PBIP model in a simple economy to empirical moments in the motivating facts. Concentration ratio λ_{ij}^{ind} denotes the share of employment in industry j and location i, conditional on any industry.

$$\frac{L_{1j}}{L_{2j}} = \left(\frac{G_{1j}^{D}}{G_{2j}^{D}} \frac{\Delta_1\left(z_j^*\right)}{\Delta_2\left(z_j^*\right)} \frac{\mathscr{P}_{1j}^{-1}}{\mathscr{P}_{2j}^{-1}} \right)^{\frac{1}{1-\theta(\sigma-1)}} = \frac{\Delta_1\left(z_2^*\right)/\chi_{1j}}{\Delta_2\left(z_2^*\right)/\chi_{2j}}$$

$$\Rightarrow \tilde{\chi}_{2j}^{ind}\lambda_{2j}^{ind} = \left(1 + \left(\frac{G_{1j}^{D}}{G_{12}^{D}} \frac{\Delta_1\left(z_j^*\right)}{\Delta_2\left(z_j^*\right)} \frac{\mathscr{P}_{1j}^{-1}}{\mathscr{P}_{2j}^{-1}} \right)^{\frac{1}{1-\theta(\sigma-1)}} \right)^{-1}$$

The specialization ratio λ_{ij}^{loc} denotes the share of employment in industry j and location i, *conditional on any location*.

$$\frac{L_{i1}}{L_{i2}} = \left(\frac{\kappa_{\ell,1}}{\kappa_{\ell,2}} \frac{G_{i1}^{D}}{G_{12}^{D}} \frac{\Delta_i\left(z_1^*\right)}{\Delta_i\left(z_2^*\right)} \frac{\mathscr{P}_{i1}^{-1}}{\mathscr{P}_{i2}^{-1}} \right)^{\frac{1}{1-\theta(\sigma-1)}} = \frac{\Delta_i\left(z_1^*\right)/\chi_{i1}}{\Delta_i\left(z_2^*\right)/\chi_{i2}}$$

$$\Rightarrow \tilde{\chi}^{loc}\lambda_{i2}^{loc} = \left(1 + \left(\frac{\kappa_{\ell,1}}{\kappa_{\ell,2}} \frac{G_{i1}^{D}}{G_{12}^{D}} \frac{\Delta_i\left(z_1^*\right)}{\Delta_i\left(z_2^*\right)} \frac{\mathscr{P}_{i1}^{-1}}{\mathscr{P}_{i2}^{-1}} \right)^{\frac{1}{1-\theta(\sigma-1)}} \right)^{-1}$$

where $\kappa_{\ell,j} \equiv \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \frac{\alpha^{1-\alpha(1-\sigma)}}{\prod_{j'}\left(\gamma_{j'j}(1-\alpha)\right)^{(1-\alpha)(1-\sigma)\gamma_{j'j}}}$.

The following figure shows the comparative statics when increasing agglomeration externalities and when varying I/O technologies respectively.

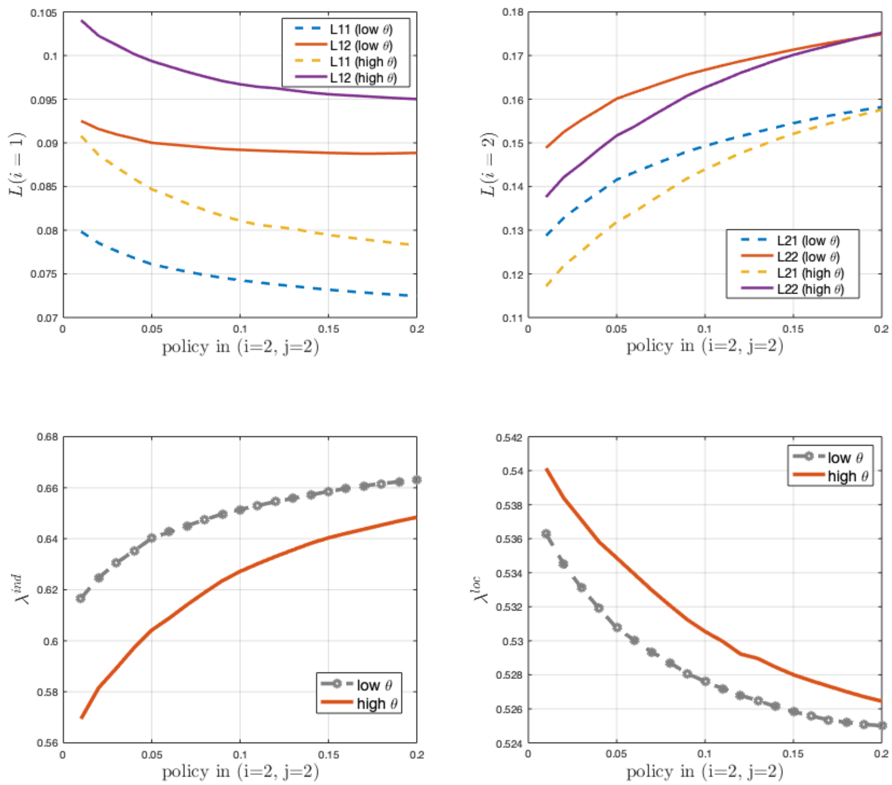


Figure 3.30: Numerical experiment for local sectoral scales and key ratios

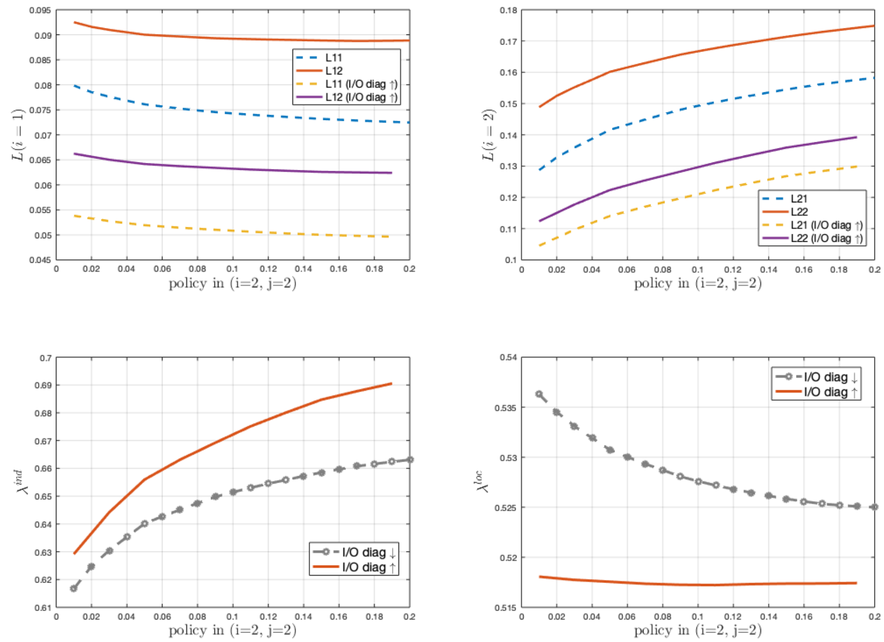


Figure 3.31: Numerical experiment for local sectoral scales and key ratios

WLOG, consider a policy change in industry 2 and location 2 (which is labled to be the location where more productive firms go). Conditional on industry 2 (HTI), share of firms in location 2 (HTC) is

$$\lambda_2^{ind} \equiv \frac{L_{22}}{\sum_i L_{i2}} = \left[1 + \left(\frac{\chi_{22}}{\chi_{12}}\right)^{\frac{1}{\theta(\sigma-1)}} \left(\frac{\tau_{21}^{-\sigma} P_{12}^{\sigma} Q_{12} + \tau_{22}^{-\sigma} P_{22}^{\sigma} Q_{22}}{\tau_{11}^{-\sigma} P_{12}^{\sigma} Q_{12} + \tau_{12}^{-\sigma} P_{22}^{\sigma} Q_{22}}\right)^{\frac{1}{\theta(\sigma-1)}}\right]^{-1}$$

Whereas this ratio in laissez-faire is

$$\lambda_2^{ind} = \left[1 + \frac{\Delta_1\left(z_2^*\right)}{\Delta_2\left(z_2^*\right)}\right]^{-1}$$

When χ_{22} increases, the first-order effect is a decrease in size of more productive city since a smaller size (thus agglomeration benefit) is enough to generate same profit. However, this will then trigger a behavioral change in marginal firms to move to more productive cities (i.e. Δ_{2j} increases, and Δ_{1j} decreases). But the extent to which these two productivity cutoffs change in relative size depends on changes of composite price index, which, together with L_{ij}, pin down all quantities. The change in price index depends on relative dominance of local supply shifter LLS (I/O linkages effect) and global demand shifter GDS (trade effect).

Appendix: Quantitative Analysis

3.7.15 Statistics in Changes

Denote the relative changes of an arbitrary variable as: $\hat{x} = \frac{x'}{x}$. Consider hat algebra on concentration ratio $\lambda_{ij}^{ind} = \frac{L_{ij}}{\sum_i L_{ij}}$. For our empirical facts, we will focus on all potential Domar weights: concentration ratio, specialization ratio or aggregate ratio. The change in concentration ratio in any arbitrary location and industry can be written as

$$\hat{\lambda}_{ij}^{ind} = \frac{\hat{L}_{ij}}{\sum_i \frac{L_{ij}\hat{L}_{ij}}{\sum_k L_{kj}}} = \frac{\hat{L}_{ij}}{\sum_i \lambda_{ij}^{ind}\hat{L}_{ij}}$$

Similarly, the change in specialization ratio can be written as $\hat{\lambda}_{ij}^{loc} = \frac{\hat{L}_{ij}}{\sum_j \lambda_{ij}^{loc}\hat{L}_{ij}}$. Together with market clearing conditions, we have

$$\hat{L}_{ij} = \left(\kappa_{\ell,j}\hat{G}_{ij}^{D}\hat{\mathscr{P}}_{ij}^{-1}\hat{\Delta}_i\left(z_j^*\right)\right)^{\frac{1}{\theta_j\left(\sigma_j-1\right)}}$$

Thus

$$\hat{\lambda}_{ij}^{ind} = \frac{\left(\kappa_{\ell,j}\hat{G}_{ij}^{D}\hat{\mathscr{P}}_{ij}^{-1}\hat{\Delta}_i\left(z_j^*\right)\right)^{\frac{1}{\theta_j\left(\sigma_j-1\right)}}}{\sum_i\left[\lambda_{ij}^{ind}\left(\kappa_{\ell,j}\hat{G}_{ij}^{D}\hat{\mathscr{P}}_{ij}^{-1}\hat{\Delta}_i\left(z_j^*\right)\right)^{\frac{1}{\theta_j\left(\sigma_j-1\right)}}\right]}$$

where the components on right hand side can be obtained as follows:

- Global demand shifter

$$\begin{aligned}\hat{G}_{ij}^{D} &= \frac{\sum_{i'} \tau'^{-\sigma}_{ii'} P'^{\sigma}_{i'j} Q'_{i'j}}{\sum_{i'} \tau^{-\sigma}_{ii'} P^{\sigma}_{i'j} Q_{i'j}} \\ &= \frac{\left(\left(\frac{P'_{i'j}}{\tau'_{ii'}}\right)^{\sigma} Q'_{i'j}\right)_{i'=1} + \cdots + \left(\left(\frac{P'_{i'j}}{\tau'_{ii'}}\right)^{\sigma} Q'_{i'j}\right)_{i'=I}}{\sum_{i'}\left(\frac{P_{i'j}}{\tau_{ii'}}\right)^{\sigma} Q_{i'j}} \\ &= \sum_{i'} \phi_{ij}^{G_{i'}}\left(\frac{\hat{P}_{i'j}}{\hat{\tau}_{ii'}}\right)^{\sigma}\hat{Q}_{i'j}\end{aligned}$$

  where $\phi_{ij}^{G_{i'}} \equiv \frac{\left(\frac{P_{i'j}}{\tau_{ii'}}\right)^{\sigma} Q'_{i'j}}{\sum_{i'}\left(\frac{P_{i'j}}{\tau_{ii'}}\right)^{\sigma} Q_{i'j}}$.

- Price index

$$P_{ij} = \left(\int_{\Omega} \left(p_{i'}(\omega_j) \tau_{i'i} \right)^{1-\sigma} d\omega_j \right)^{\frac{1}{1-\sigma}}$$
$$= \left(\sum_{i'} \left(\tau_{i'i} \kappa_{p,j} w_i^{\alpha} \left(\prod_{j'} P_{i'j}^{(1-\alpha)\gamma_{j'j}} \right) \mathscr{A}_{i'j}^{-1} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

Thus

$$\hat{P}_{ij}^{1-\sigma} = \frac{\sum_{i'} \left(\tau'_{i'i} \kappa'_{p,j} w'^{\alpha}_{i} \left(\prod_{j'} P'^{(1-\alpha)\gamma_{j'j}}_{i'j} \right) \mathscr{A}'^{-1}_{i'j} \right)^{1-\sigma}}{\sum_{i'} \left(\tau_{i'i} \kappa_{p,j} w_i^{\alpha} \left(\prod_{j'} P_{i'j}^{(1-\alpha)\gamma_{j'j}} \right) \mathscr{A}_{i'j}^{-1} \right)^{1-\sigma}}$$
$$= \sum_{i'} \phi_{ij}^{P_{i'}} \left(\hat{\tau}_{i'i} \hat{\kappa}_{p,j} \hat{w}_i^{\alpha} \left(\prod_{j'} \hat{P}_{i'j}^{(1-\alpha)\gamma_{j'j}} \right) \hat{\mathscr{A}}_{i'j}^{-1} \right)^{1-\sigma}$$

where $\phi_{ij}^{P_{i'}} \equiv \frac{\left(\tau_{i'i} \kappa_{p,j} w_i^{\alpha} \left(\prod_{j'} P_{i'j}^{(1-\alpha)\gamma_{j'j}} \right) \mathscr{A}_{i'j}^{-1} \right)^{1-\sigma}}{\sum_{i'} \left(\tau_{i'i} \kappa_{p,j} w_i^{\alpha} \left(\prod_{j'} P_{i'j}^{(1-\alpha)\gamma_{j'j}} \right) \mathscr{A}_{i'j}^{-1} \right)^{1-\sigma}}$. Thus aggregate price index in change is

$$\hat{\mathscr{P}}_{ij} = \prod_{j'} \left(\sum_{i'} \phi_{ij}^{P_{i'}} \left(\hat{\tau}_{i'i} \hat{\kappa}_{p,j} \hat{w}_i^{\alpha} \left(\prod_{j'} \hat{P}_{i'j}^{(1-\alpha)\gamma_{j'j}} \right) \hat{\mathscr{A}}_{i'j}^{-1} \right)^{1-\sigma} \right)^{(1-\alpha)\gamma_{j'j}}$$

- The cost of the input bundle

$$\hat{c}_{ij} = \hat{w}_i^{\alpha_j} \prod_{j'} \left(\sum_{i'} \phi_{ij}^{P_{i'}} \left(\hat{\tau}_{i'i} \hat{\kappa}_{p,j} \hat{w}_i^{\alpha_j} \left(\prod_{j'} \hat{P}_{i'j}^{(1-\alpha_j)\gamma_{j'j}} \right) \hat{\mathscr{A}}_{i'j}^{-1} \right)^{1-\sigma_j} \right)^{(1-\alpha_j)\gamma_{j'j}}$$

- Trade flow

$$\hat{X}_{i'i,j} = \hat{CA}_{ij} \hat{\tau_{ii'}}^{1-\sigma_j} \hat{P}_{i'j}^{\sigma_j - 1} \hat{Y}_{i'j}$$

3.7.16 Supplementary Figures

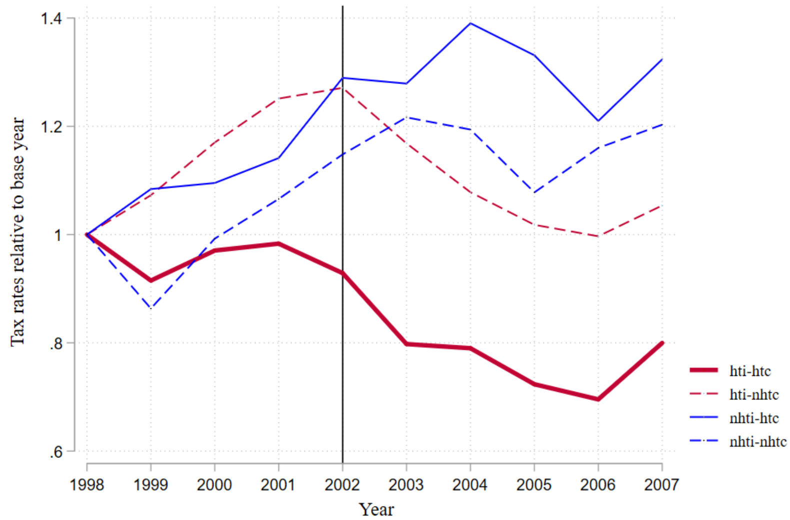


Figure 3.32: Change in tax rates by policy groups with observables aggregated from firm-level data

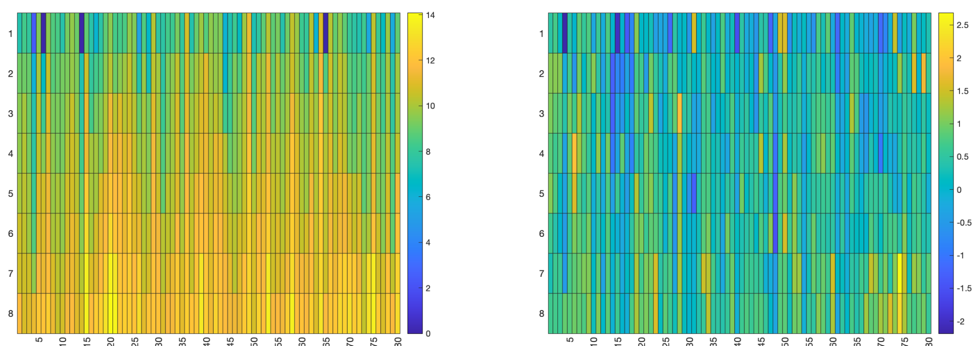


Figure 3.33: local sectoral employment L_{ij}. Left: level in log-scale. Right: change (pre- and post-policy)

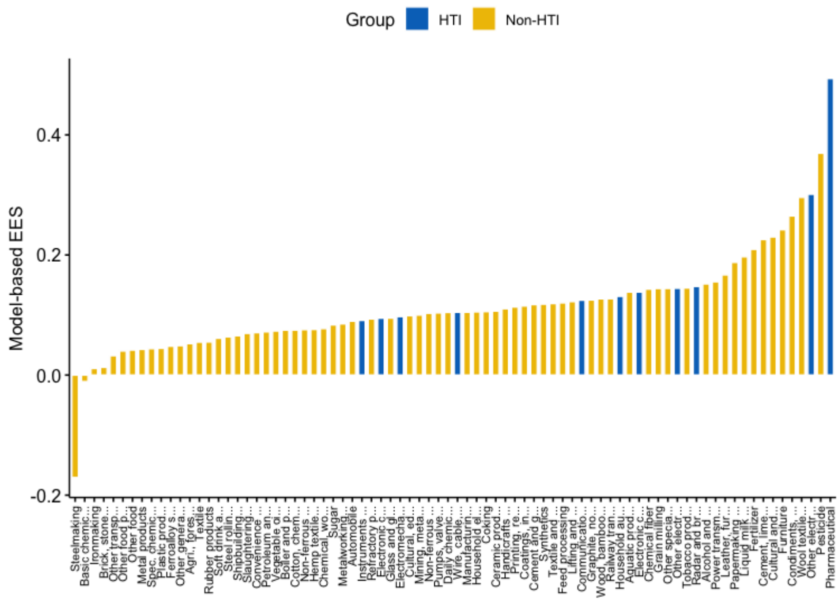


Figure 3.34: Sectoral EES

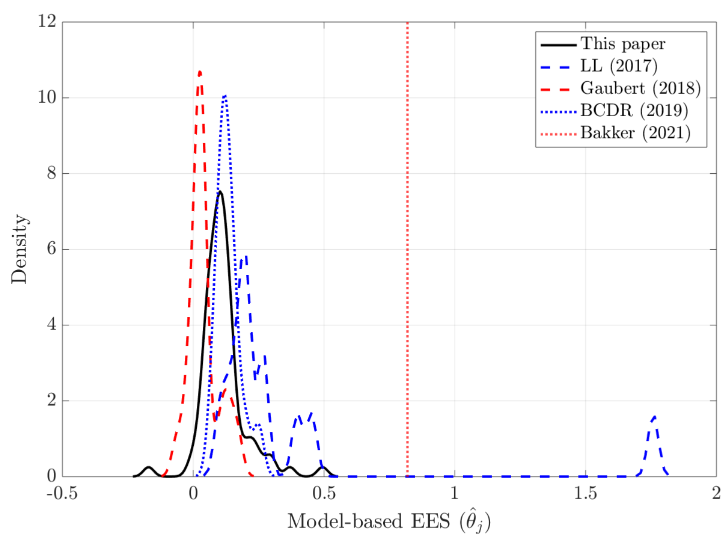


Figure 3.35: Comparison of EES

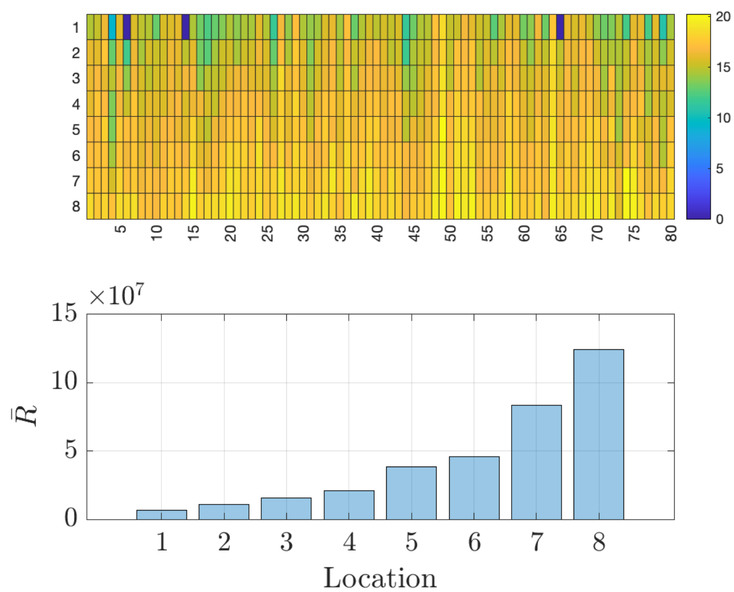


Figure 3.36: Positive sorting pattern (2007)

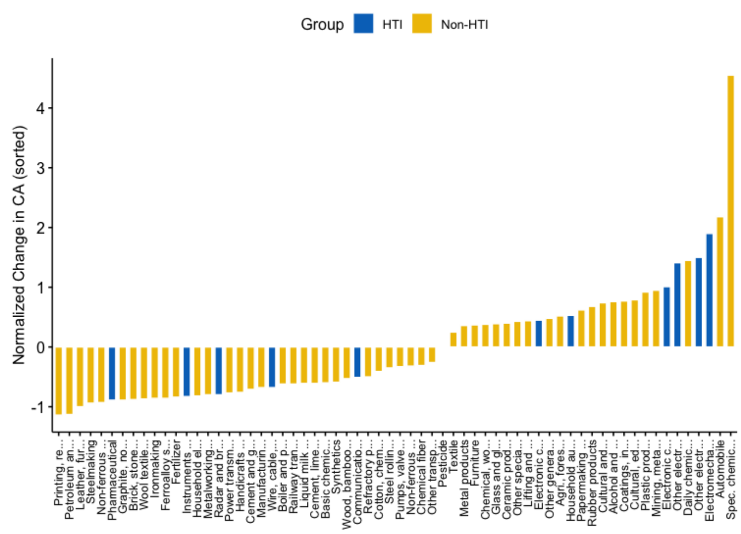


Figure 3.37: Sorted relative $\hat{CA}$ by sector

The following shows the national input-output table in 2002. There are 80 manufacturing sectors in total, which is equivalent to 2/3-digit CIC level.

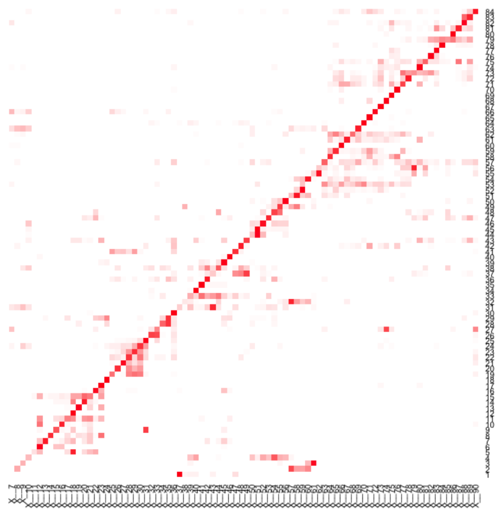


Figure 3.38: Input-Output Matrix in China (2002)

3.7.17 Regional Flow Approach

The alternative approach utilizes trade flows among regions. Specifically, I take the changes in sectoral bilateral trade flows ($\hat{X}(i,i'|j)$) from China's multi-regional input-output table (MRIO) across time (2002 and 2007) to quantify θ_j^*. These MRIO trade flows cross the regions. (The most-detailed level for regions and sectors are contained in the 2002 and 2007 versions of the MRIO.[58]) The variation is within-sector between all origin-destination pairs. This cross-sectional variation incorporates information from both the inter-sectoral and the inter-regional components:

$$X(i,i'|j) = \underbrace{\left(\underbrace{(\mathscr{A}_{ij}(z_j^*))^{\sigma_j-1}}_{\text{agglomeration}} \underbrace{\rho_{ij}^{1-\sigma_j}}_{\text{I/O linkage}} \right)}_{\text{effect from origin}} \underbrace{\tau_{ii'}^{1-\sigma_j}}_{\text{trade effect}} \underbrace{P_{i'j}^{\sigma_j} Q_{i'j}}_{\text{effect from destination}}$$

where $\rho_{ij} \equiv \kappa_{p,j} w_i^{\alpha_j} \prod_{j'} P_{ij'}^{(1-\alpha_j)\gamma_{j'j}}$ and $\kappa_{p,j} \equiv \frac{\sigma_j}{\sigma_j-1} \frac{\alpha_j^{-\alpha_j}}{\prod_{j'}(\gamma_{j'j}(1-\alpha_j))^{\gamma_{j'j}(1-\alpha_j)}}$. The "agglomeration" term is the productivity including agglomeration externalities. Then, I can compute prepolicy and postpolicy changes of trade flows, which is the hat term on the left-hand side of the following equation, and changes of price (P), productivity (TFP), and revenue (R) in each (i,j) pair, which are hat terms on

[58] I thank the Institute of Geographic Science and Natural Resources Research, Chinese Academy of Science for data support.

the right-hand side below.

$$\hat{X}_{i,i',j} = f\left(\underbrace{\{\hat{P}_{i,j'}\}_{j' \in \mathcal{J}}}_{\text{origin, inter-sectoral}}, T\hat{F}P_{ij}, \underbrace{\hat{P}_{i',j}, \hat{R}_{i',j}}_{\text{destination, inter-regional}} \right)$$

In the baseline, there are $(2I-1) \times J$ moments and there are $(I+2) \times J$ parameters to estimate. The following plots show the variation of trade flows at sector level. There are 17 sectors in total in the MRIO 2002 and 2007, among which 11 are manufacturing sectors.

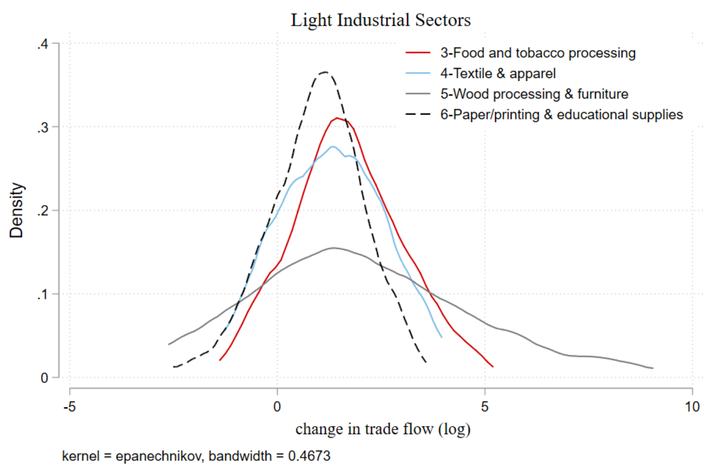


Figure 3.39: Change in trade flow from 2002 to 2007 at sector level: light

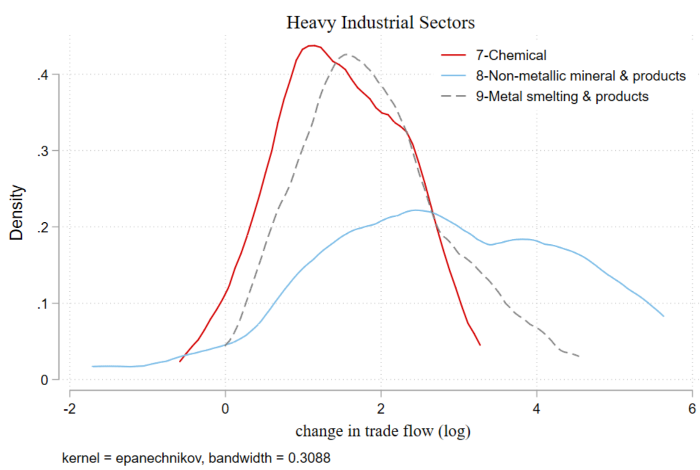


Figure 3.40: Change in trade flow from 2002 to 2007 at sector level: heavy

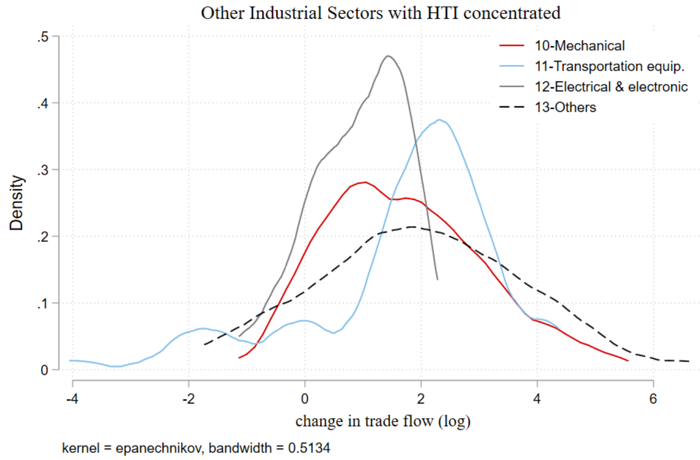


Figure 3.41: Change in trade flow from 2002 to 2007 at sector level: others

All sectors have experienced an increase in flow values on average, but the extent of the increase vary, led by “Non-metallic mineral and products”. Unlike in the baseline quantification, when using a complete set of industries of trade flows with consistent industries and locations, there are $(I^2 + 2I - 1) \times J$ moments and $(I + 2) \times J$ parameters to estimate with J the number of sectors at a more-aggregated level. With 11 aggregated manufacturing sectors in MRIO has a mean EES of 0.82, closer to Bakker (2021) [90] compared to the other approaches. However, it has a much larger dispersion, which is mainly because of the necesary aggregation associated with MRIO.

3.7.18 Trade Costs

Trade costs are taken from Tombe and Zhu (2019) [96], where the inter-regional values are shown by the following plot.

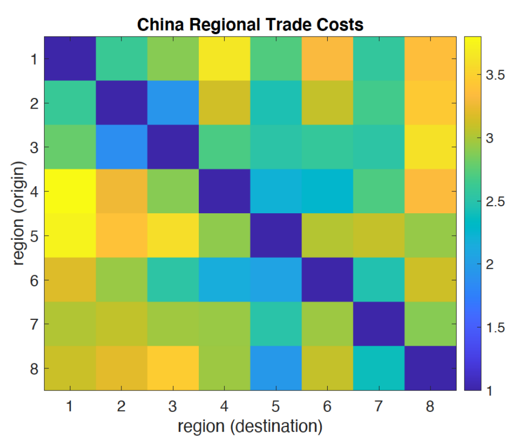


Figure 3.42: Trade costs across regions in China

3.7.19 Limitation

The following shows that change in trade at prefecture and province level for $CIC = 3725$.

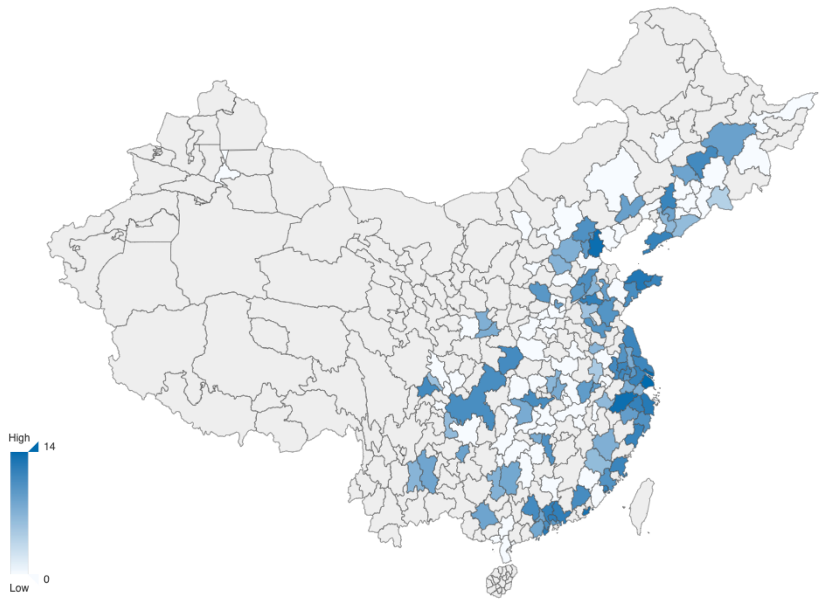


Figure 3.43: Map of China and level of export in $CIC = 3725$ at prefecture level

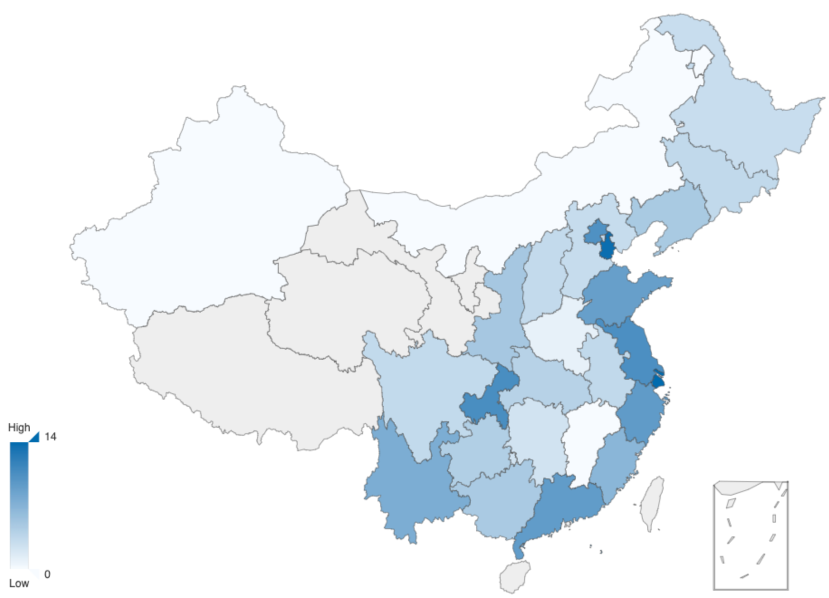


Figure 3.44: Map of China and level of export in $CIC = 3725$ at province level

Chapter 4

High Performance Export Portfolio: Design Growth-Enhancing Export Structure with Machine Learning

(joint with Natasha Che)

Abstract

To study the relationship between export structure and growth performance, we design an export diversification system using a collaborative filtering algorithm based on countries' revealed comparative advantages. The system is used to produce export portfolio covering more than 190 economies spanning over 30 years. We find that economies whose export structure is more aligned with the algorithm-recommended export structure achieve better growth performance, in terms of both higher GDP growth rate and lower growth volatility. These findings demonstrate that export structure matters for obtaining high, and stable growth. Our results provide empirical evidence on export diversification and growth for further quantitative studies.

4.1 Introduction

Over the past decades, many success stories of growth and income convergence—most notably, China and several other East Asian emerging economies—have been export-led. Some of these countries have governments that actively pursue industrial policies that foster strategic export industries; others let the market take the lead. Setting aside the pros and cons of each approach, there is no denying that export diversification and industrial structural change are important for growth (e.g., Aiginger and Rodrik, 2020 [102]). Despite the topic's relevance, however, there is little exporting product analysis in economic literature regarding what types of export product portfolio are possibly suitable for growth, and what type of export portfolio countries could consider diversifying into.[1]

[1] This chapter provides cross-country empirical evidence. However, it shall not be taken as direct country-wise policy advisory. "Recommendation" in this chapter refers to the algorithm-

Classical trade theories suggest that countries should export what they are relatively good at producing, i.e., following comparative advantages. But how exactly does one ascertain comparative advantages? Trade theories predict that many developing countries tend to have comparative advantages in labor-intensive exports and should, to some extent, stay away from capital-intensive industries. But in reality, comparative advantages contain far more dimensions than capital and labor. Some of these dimensions are quantifiable in a more linear way with production technologies, whereas others are not.

The matter becomes even more complicated when we consider the fact that comparative advantages evolve as a country grows. How could the export structure change as a result to achieve better growth performance? General trade theories and empirical studies do not go very far in granular-level insights in explaining the structural change in exports. In a recent study, Che (2020) [103] proposes a novel method to operationalize the concept of comparative advantage and its evolution. It uses collaborative filtering algorithms in machine learning, most commonly applied to product recommendations in e-commerce, to produce export diversification recommendations that reflect a country's *latent comparative advantages* and future potentials in export structure. In Section 4.3, we will go over the details of the methodology, but the basic intuition is that a country is likely to have comparative advantages in those products that are highly related to the products it is already good at exporting (i.e., products with *revealed comparative advantages*), where the "relatedness" between any two products is measured by the similarities among countries that are the main exporters of the two products. Moreover, it turns out that the export structures recommended by the "product-based KNN" algorithm can predict the evolution of actual export structure for several high-growth countries (China, India, Chile and Poland) reasonably well. Here the *export structure* is measured by the number of products recommended by the algorithm, categorized by Standard International Trade Classification (SITC) 4-digit codes, which belong to each of the 10 SITC 1-digit sectors, as a share of the total number of recommended products.

Inheriting the intuition of exporting based on comparative advantages, this chapter further provides cross-country evidence by designing an export diversification system. The rationale for our approach comes from two cross-country observations. First, products that require similar production inputs and know-how tend to show up in an export portfolio together. For example, a country that has successfully exported beef can branch into, with some effort, dairy. A country that has mastered the trade of exporting desktop computer hardware is in a well positioned to produce and export cellphones, than otherwise. Therefore, the products in a country's existing export portfolio contain valuable information regarding what other products the country could get good at producing. Second, countries with similar comparative advantages tend to export similar products. Bangladesh and Vietnam are both successful in exporting garments

based results. Specific policy recommendations need to be in accordance with a country's circumstances and further structural analyses.

because of the countries' shared abundance in low-cost labor. New Zealand and Uruguay both specialize in cattle exports partly because of their abundance of suitable pasture land. In other words, products related to a country's existing exports and export portfolios of similar countries offer useful information about the country's latent comparative advantages, even though the latter cannot always be neatly expressed quantitatively.

Our key index calculation (i.e., the *similarity score*, which will be specified later) is based on *export portfolio*, instead of the *"diversification"* notion as commonly understood. Specifically, the export portfolio is more general in addressing the evolution of trade patterns in a cross-country analysis. It also provides insights on diversification strategies as an application, especially in a single-country analysis. For example, a country can double the number of products it exports, i.e., diversification in numbers, without changing its export structure at all, if the sectoral distribution of its exports stays the same. In contrast, if a country that used to export 100 products all in the food industry now switches to exporting 50 products in the food sector and 50 in the chemical industry, it has not "diversified" in numbers, but its export structure has changed. A country's export portfolio can be potentially improved by diversifying in the number of export products, as well as by adjusting the export structure to fit its evolving comparative advantages. Moreover, the expression *diversification* per se may not be well-defined when comparing countries. For example, it can be controversial whether a country already producing a fair amount of products becomes relatively more diversified when its exporting products expand within a sector, compared to other countries with the same change in the total number of products. Therefore, in our cross-country analyses, we use the *export portfolio* to capture more general features on the evolution of international trade. We do not take a stand regarding whether a specific country should consider a structural change or a diversification. The set of algorithm-based export portfolio provides some possible structures, which facilitates us to analyze how its alignment with actual ones relates to growth performance.

Our primary goal in this chapter is to test the hypothesis that export portfolio[2] based on a collaborative filtering algorithm is positively correlated with a given country's export structure that likely to help it grow better over time. To do this, we first use a product-based k-nearest neighbors algorithm to predict annual export product recommendations in the SITC 4-digit product space, for more than 190 economies spanning over three decades. We then compare the algo-recommended sectoral structure of exports with the actual export structure of each country. At last, we use the full sample countries to test the main hypothesis. See Section 4.3 for details on methodology. The intuition is that, if the export recommendations produced by the algorithm indeed capture some countries' latent comparative advantages, we should observe that countries whose export structure closely aligns with the recommended structure would have bet-

[2]The portfolio refers to the SITC-1 export structure derived from the algorithm-generated product recommendations, where "recommendations" is from an algorithm perspective.

ter growth performance. Here we define “better” as higher growth and lower growth volatility.

A preliminary look at the data appears to support our hypothesis. Figure 4.1 plots the cross-country correlation between average real GDP growth per capita over the 1982–2017 period and average *similarity score* between a country’s actual export structure and recommended export structure produced by our algorithm.[3] Figure 4.2 plots the correlation between the five-year standard deviation of annual growth rate and the similarity score. The charts indicate that countries with an export structure closer to the algo-recommended structure enjoy higher growth, lower growth volatility and higher risk-adjusted growth. The same pattern can be discerned from Figure 4.3, which shows a positive correlation between the similarity score and countries’ risk-adjusted growth, i.e., the five-year average growth divided by standard deviation of growth.

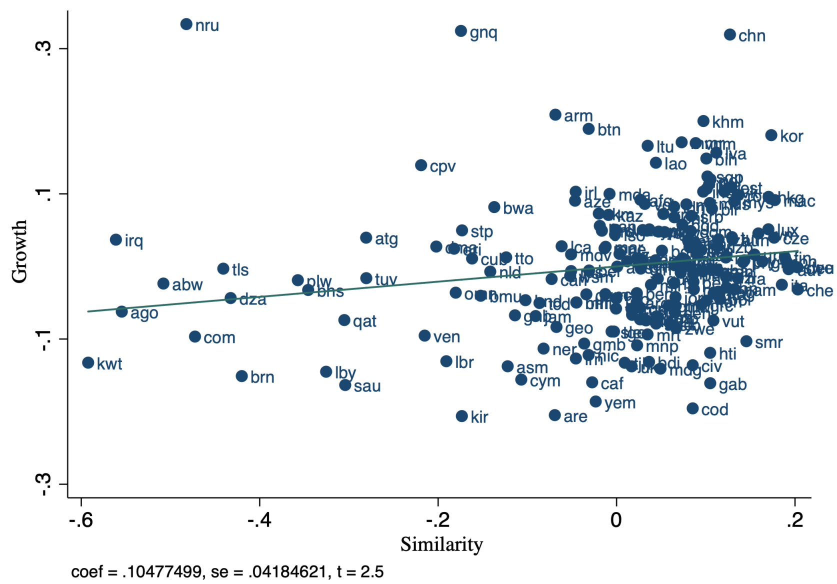


Figure 4.1: Relationship between “Similarity Score” and Growth

[3]The similarity score is calculated as the Pearson correlation between actual and recommended export structures. Thus it has a theoretical range of [-1, 1]. See Section 4.3 for details.

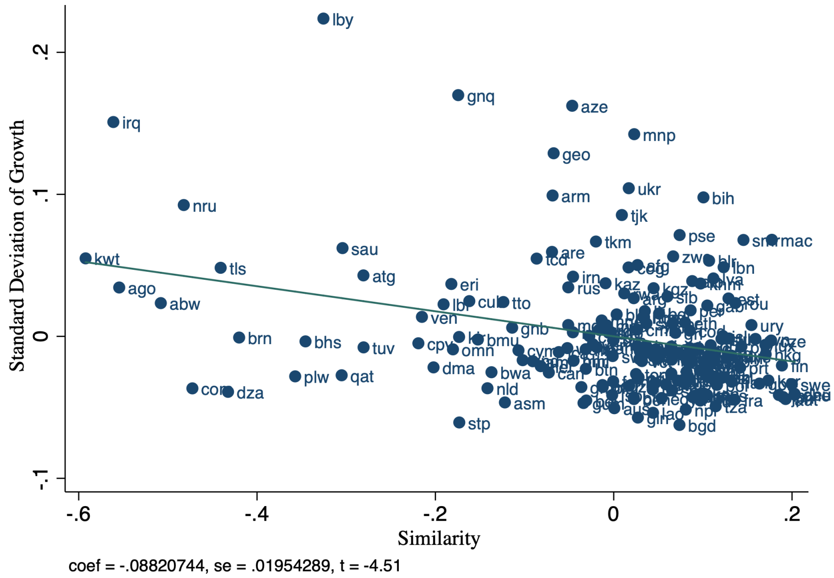


Figure 4.2: Relationship between "Similarity Score" and Growth Volatility

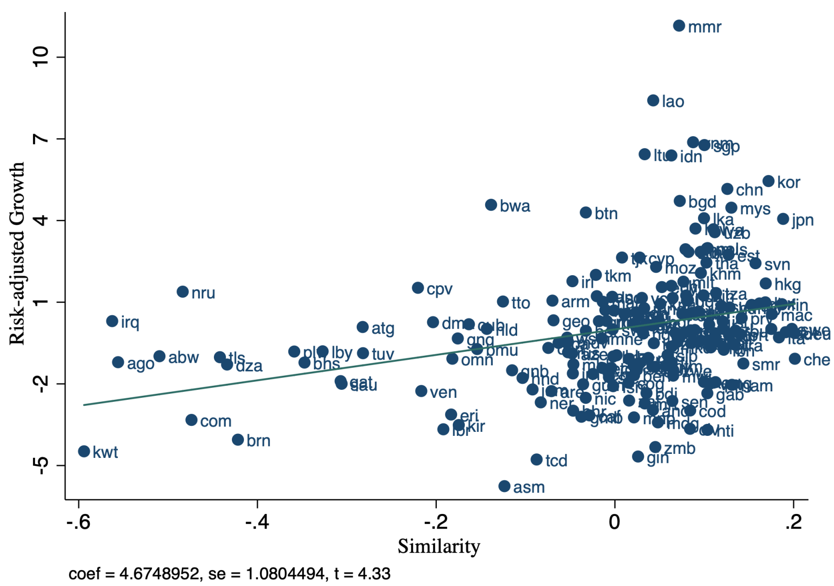


Figure 4.3: Relationship between "Similarity Score" and Volatility-Adjusted Growth

It is interesting to look at some examples of specific countries as well. Figure 4.4 plots the evolution of the similarity score between algorithm-recommended export structure and actual export structure for China, Singapore, South Korea, and Germany.[4] Since the late 1980s, the similarity score for China has increased significantly, from below the world average to the top 3% of the world sample. The magnitude of increase for Singapore is similar. For South Korea, though the similarity score has a decreasing trend in general (despite increasing in the most recent years), it still remains at high levels (in the top 10%) on average. Likewise, Germany has one of the highest similarity scores in the world, which is unsurprising given the country's diversified, and dynamic industrial export base over the past decades.

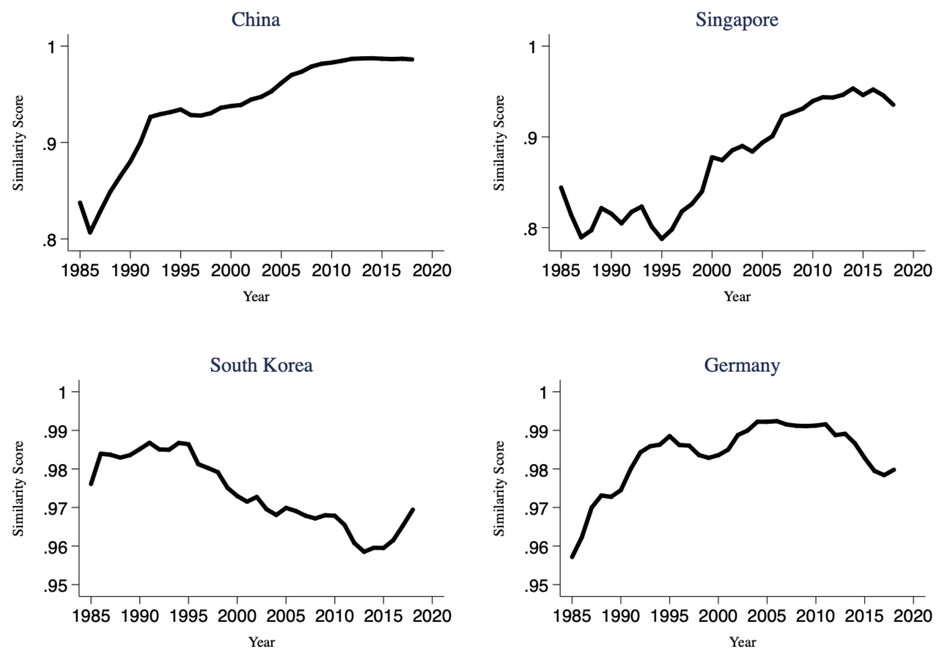


Figure 4.4: Similarity Scores for Selected High-Growth and Developed Countries (5-Year MA)

Figure 4.5 plots the evolution of the similarity score for several developing countries with lower growth—Honduras, Kuwait, Libya, and Venezuela. For Libya and Kuwait, the similarity scores are particularly low. Though Kuwait's score has increased since 2010, it remained below 0.2 until the early 2010s and is still low, below 0.4 in recent years, compared to the world average of 0.82. For Honduras and Venezuela, the similarity scores are higher but still below the world level, and they have dropped significantly since the mid 1990s, likely reflecting a decline in diversification and manufacturing capacity.

[4]The similarity scores are calculated annually and the charts present five-year moving averages of the scores.

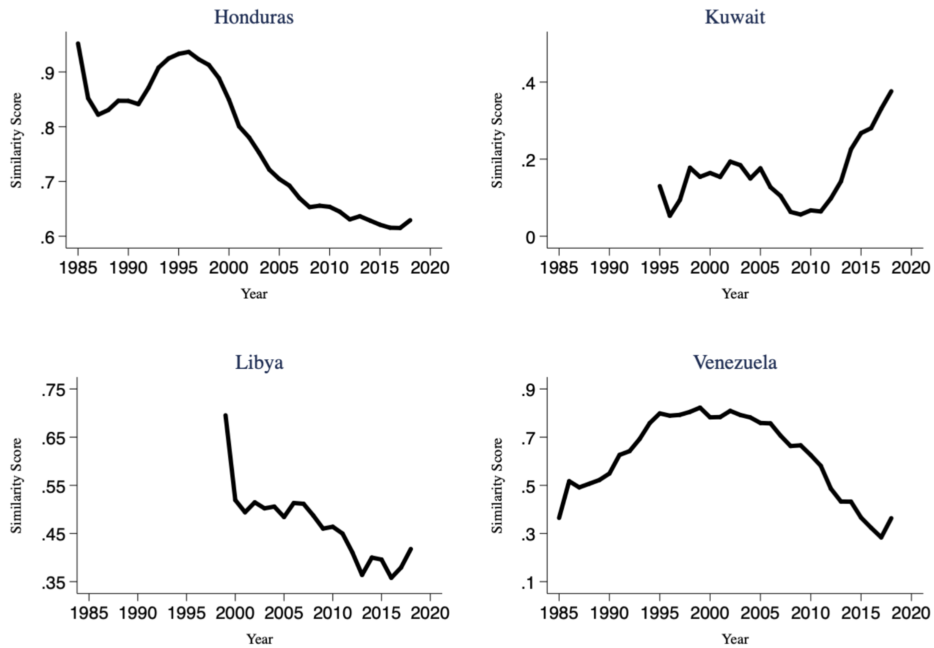


Figure 4.5: Similarity Scores for Selected Low-Growth and Fragile Countries (5-Year MA)

The rest of this chapter is organized as follows. Section 4.2 presents related literature on export structure and diversification. Section 4.3 explains the product-based KNN algorithm and our empirical methodology. Section 4.4 describes the data. Sections 4.5 and 4.6 present the baseline empirical results and robustness checks. Section 4.7 concludes.

4.2 Literature

The literature closest to our paper is the studies on the so-called *product space* and its implication for diversification and growth (e.g., Hausmann & Klinger 2007 [104], Hidalgo & Hausmann 2009 [105]). Like the current paper, this strand of research seeks to understand a country's export structure by looking at the *relatedness* among products.

But there are two key differences. The first is regarding the efficiency in the use of information contained in the trade data. The product-space literature uses a probability formula to represent the relatedness, or *proximity* between two products.[5] While this approach features a clean, easy-to-understand formula and makes subsequent analyses computationally simpler, it is at the cost of not

[5] Specifically, the proximity between product A and product B is defined as the probability that a country exports product A given that it exports product B, or vice versa. For example,

fully exploiting the information contained in the data matrix of country-product exports. In contrast, the product-based KNN algorithm in the present paper makes more efficient use of the data to detect the unique blend of characteristics of countries and products. This leads to potentially better recommendations. To be sure, it is at the cost of requiring more computational resources and forsaking the easily comprehensible linear formula. This is a common drawback of many machine learning algorithms– the nonparametric nature of the approach can make some results seem “magical” and harder to explain with linear logic.

The second, and more important, difference is one of perspectives. The product-space literature makes specific value judgments about the worthiness of different products for diversification purpose. A product’s diversification value is seen as broadly depended on 1) how “complex” it is, meaning, how much sophisticated knowledge is required to product the product, and 2) how closely related the product is to other more complex products. Each product is assigned a complexity level as such. The rationale for doing so is a reasonable one– more complex products have higher value-added, use more human capital, and face less global competition. And products that are “bridges” to the more complex products may be a pathway for a country to move up the global value chain. Some empirical evidence shows that diversifying into these products is supposed to be better for growth (Hausmann, 2007 [104]). However, several issues emerge when this model is used for product-level recommendations. First, there is an underlining tension between this line of thinking and the framework of comparative advantages that the product-space analysis is built on. By assigning each product a score of virtue (e.g., industrial products are “good”, agricultural commodities are “bad”), it leads to a tendency to recommend products that the model deems universally worthy to countries of drastically different fundamentals. In the extreme, though improbable, scenario where all countries internalize the same worthiness ranking of products in developing their export structure, there would be no comparative advantages to speak of. Secondly, to come up with a tractable, universally applicable scoring system for “product complexity”, strong assumptions need to be made that reduce the feature dimensions of reality and throw away valuable country- and product- specific information, which may limit the model’s usefulness in producing realistic export recommendations for individual countries.[6]

In contrast, the approach of the current paper is agnostic about the diversification value of any specific product. Instead, we seek to fully exploit the information contained in the country-product space, and make realistic export recommendations off of a country’s current revealed comparative advantages. One implication is that countries do not necessarily need to chase the “complex”

suppose that 17 countries export wine, 24 export grapes and 11 export both, all with revealed comparative advantage. Then, the proximity between wine and grapes is $11/24 = 0.46$.

[6] For the algo-recommendation scores using more observable data, this is a limit. Strcutural analyses, on the other hand, require necessary simplication for more mechanisms.

exports to achieve better growth performance. As Section 4.5 shows, countries whose export structure closely aligns with the algorithm-recommended structure have higher and more stable growth, even though the algorithm's recommendations do not make any specific judgment regarding product complexity, and are solely based on information from a country's currently revealed comparative advantages.

The chapter is also related to the literature on the relationship between export diversification and countries' economic performance. Existing research asserts that export diversification is a key element in the economic development process, particularly for developing and emerging market economies trying to catch up with their advanced peers. Various studies provide evidence of a positive association between export diversification and economic development (e.g., Imbs and Wacziarg, 2003 [106]; Klinger and Lederman 2004 [107] and 2011 [108]; Cadot et al., 2011 [109]). Numerous country studies also supports the benefits of export diversification. For example, Feenstra and Kee (2008) [110] use data from a large set of countries exporting to the US, to show that a sustained increase in export diversification results in increases in productivity and a notable increase in the GDP of the exporters. IMF (2014) [111] finds that diversification in exports and in domestic production has been conducive to faster economic growth in LICs. Al-Marhubi (2000) provides similar findings within a set of developing economies. Balaguer and Cantavella-Jorda (2004) [112] find that export variety plays a key role in Spain's economic development. And Herzer and Danzinger (2006) [113] report a positive impact of export diversification on economic growth of Chile. Research also points to a positive association between export diversification and macroeconomic stability (e.g., IMF, 2014 [111]).

However, not all types of diversification are created equal, and diversification for its own sake is hardly a recipe for sustainable growth. A fundamental idea of the classical international trade theory is that under free trade, countries will tend to export what they are relatively good at producing, i.e., products they have a comparative advantage in. "Diversifying" into industries that are misaligned with a country's current endowment fundamentals, as the former Soviet-block nations did after World War II through industrial policies that aimed to accelerate industrialization, has negative growth consequences (Lin, 2009) [114]. On the other end of the spectrum, delayed industrialization also leads to negative growth outcomes, as the experience of many resource-rich countries that are entrenched in their over-dependence on commodity exports has shown (e.g., Frankel, 2010 [115]).

A difference in focus between the current paper and the export diversification literature is that the latter sees diversification as mostly in increasing the number of export products, while the current paper emphasizes more on the structure of the export portfolio. Our algorithm does provide a list of recommended products for each country, which offers useful insights for countries looking to increase the number of export items. But our econometric exercise focuses on the growth impact of the appropriate export structure, i.e., sectoral distribution of exports.

4.3 Methodology

The goal of this chapter is to answer the question of whether our algo-based export recommendations are positively correlated and can predict growth-enhancing export structure, in the sense that countries with higher alignment with the algo-based recommendations tend to achieve better growth performance. We go about answering this question in the following steps:

- **STEP 1.** Determine the number of SITC 4-digit products to predict for each country (see Section 4.3.1). These numbers of products are derived from a country's size and development level.
- **STEP 2.** Generate a list of export products[7] using a product-based KNN algorithm, at country-year level in the sample (see Section 4.3.2).
- **STEP 3.** Calculate the *similarity score* between the export structure implied by the list of recommended export products and the actual export structure, for each country-year (see Section 4.3.3).
- **STEP 4.** Estimate the robust correlation between the similarity score and growth, volatility, and risk-adjusted growth (see Section 4.3.4).

An important concept used throughout the chapter is *Revealed Comparative Advantage* (RCA). The RCA score, first introduced by Balassa (1965) [116], is a popular measure to calculate the relative importance of a product in a country's export basket. Formally, the RCA score of country i in product j can be calculated as:

$$RCA_{ij} = \frac{E_{ij}/\sum_j E_{ij}}{\sum_i E_{ij}/\sum_i \sum_j E_{ij}}$$

where E_{ij} is the export value of product j from country i. $\sum_j E_{ij}$ is the total export value from country i. $\sum_i E_{ij}$ is the total export value of product j from all countries around the world. And $\sum_i \sum_j E_{ij}$ is the total world exports.

A *high-RCA product* for country i is defined as a product with its $RCA_{ij} > 1$. Mathematically, it means that the product's share in the country's export portfolio is greater than its share in the total world exports, which can be seen as an indication that the country has a comparative advantage in the product. For example, vehicle exports were about 12 percent of total world exports in 2017, while they constituted 22 percent of total exports from Mexico. Therefore, $RCA_{ij} = 22/12 = 1.8$ for Mexico's vehicle exports in 2017. Since it is

[7]We sometimes refer this with export "recommendation". Note that these are algorithm-recommeded predictions, instead of direct policy advisory.

greater than 1, according to our criteria, Mexico has a revealed comparative advantage in automobiles. To put it another way, automobiles is a high-RCA product for Mexico. The recommendation algorithm that will be introduced in Section 4.3.2 essentially simulates a hypothetical RCA score for each country-product combination, and pick the top products with the highest hypothetical scores as the recommended export portfolio for country i.

4.3.1 Choosing the number of products in algorithm

Examining the export data by SITC 4-digit industry[8] reveals the following empirical regularities. First, more developed economies tend to have a larger number of high RCA products. This reflects the fact that more advanced economies have a wider range of production knowledge and more sophisticated production structures. Figure 4.6 regresses the number of high RCA exports of each country on its real GDP per capita relative to the U.S. level, controlling for the country size, and shows a positive relationship.[9] Second, larger countries tend to have a larger number of high RCA exports. This is unsurprising, as population size correlates highly with the number of firms, the amount of human capital and the amount of other production resources a country may have, enabling the country to viably export a wider range of products. In addition, some industries and products need a minimum scale to be sustainable. Figure 4.7 plots this positive relationship between the number of high RCA exports and country population.

There are obviously other factors that determine how many high RCA exports a country has. But since we are focusing on exploring export structure instead of simply expanding the number of export products, we pin down the number of export products to predict for each country by the country's size and development stage. Specifically, we run the following estimation:

$$K_{rca,it} = \beta_1 GDP_{it} + \beta_2 POP_{it} + \gamma_t + \epsilon_{it}$$

where $K_{rca,it}$ is the number of high RCA exports that country i has in year t. GDP_{it} is GDP per capita and POP_{it} is population size. We add a time fixed effect γ_t in the regression, as the average number of high RCA exports tends to rise overtime around the world, with the increase in product variety brought about by economic growth and technological change. We use $\hat{K}_{rca,it}$, the predicted number of high RCA exports from the regression, as the number of products that the algorithm will recommend at country-year level.

[8] See Section 4.4 for a more detailed description of the underlining data.

[9] This chart reproduces Figure 3 of Che (2020).

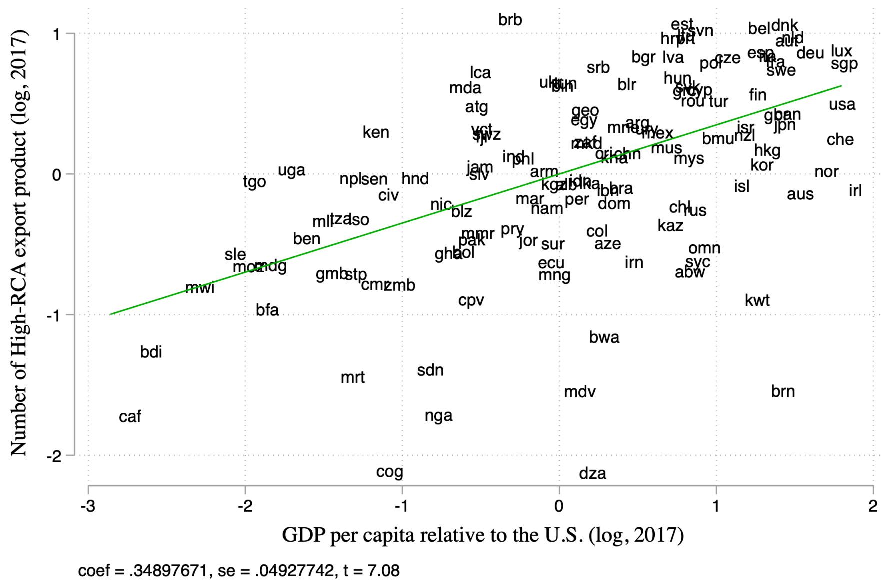


Figure 4.6: Number of High-RCA Exports v.s. Income Level, Partial Regression Plot

4.3.2 The algorithm-based recommendation

Our export product recommendation system[10] employs a product-based K-nearest neighbor (KNN) algorithm that is widely used in the *collaborative filtering recommendation systems* of online commerce. The goal of the exercise is to generate, for each country-year, a list of "top-K recommendations", i.e., K products that a country should export the most of based on empirical regularities.[11]

The algorithm produces the list by computing a *recommendation score* for each product for the underling country, using the training dataset of RCA scores by country and SITC 4-digit product, and recommending the K products with the highest recommendation scores. Here, we set K to be equal to the $\hat{K}_{rca,it}$ estimated in Section 4.3.1 for each country i in year t.

The underlining data used in the recommendation algorithm can be represented as a $m \times n$ matrix R, where m is the number of country-year combinations in the database, and n is the total number of SITC 4-digit products. Each element of R, i.e., r_{ij}, is country i's RCA score in product j. R is a sparse matrix due

[10] "Recommendation" throughout this chapter is from an algorithm perspective for our cross-country empirical analysis in later steps, instead of direct policy advisory.

[11] In our implementation, we mainly include the fundamentals in scale and development stage for the cross-country analysis. Note that these are not (necessarily) optimal levels to implement at country level. Rather, it provides a way to systematically compare portfolio across countries based on one explanatory metrics, supported by empirical regularities. There can be alternative ways to obtain number of products a country could (or even should) export, which are more suitable and relevant when considering specific country cases.

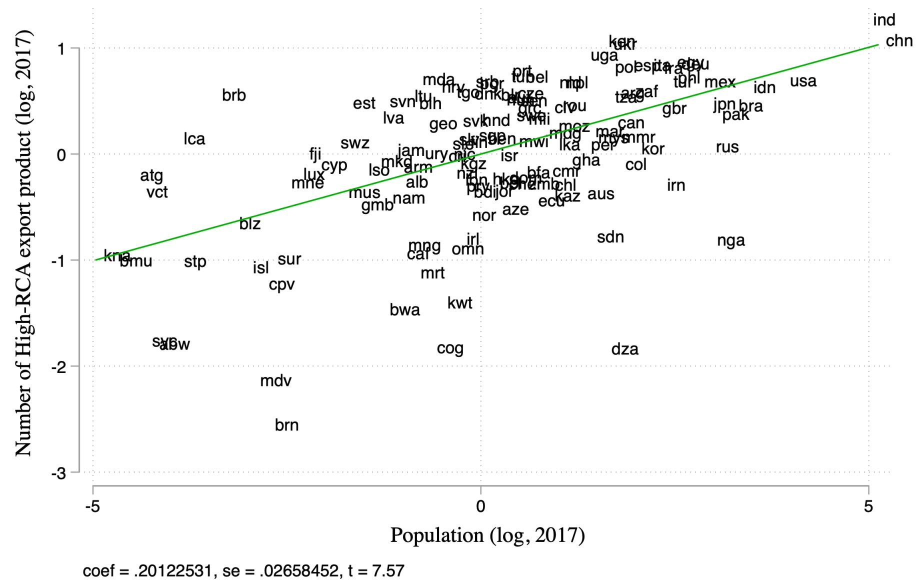


Figure 4.7: Number of High-RCA Exports v.s. Population, Partial Regression Plot

to the fact that each country only exports a subset of the products in the SITC universe.[12] In the case that country i does not export any product j, $r_{ij} = 0$. If in running the algorithm, multiple years of export data are used as the training set, then each country-year is a row in R, i.e., $m = c \times y$, where c is the number of countries in the dataset, and y is the number of years included. In practice, we set $y = 1$. In other words, when we generate scores at product-country level i in 2017, only the cross-country export data for 2017 is included in the training set.[13]

KNN is one of the most frequently used methods in solving classification and pattern recognition problems, and is a popular approach in constructing recommender systems. The basic idea of KNN is learning by analogy– classifying the test sample by comparing it to the set of training samples most similar to it. Different KNN implementations vary in terms of their choices of how the similarity between input vectors is calculated. In the present paper, the cosine similarity score is used as the similarity measure.

The intuition behind the product-based KNN implementation is simple– first look at what products a country already has a revealed comparative advantage in, and then recommend other products that are "related" to those products. To

[12]The dimension of the space depends on the granularity of products.

[13]We experimented with including multiple years of data in the training set, but found no significant improvement in the results, while the model took longer to compute as the size of m increases.

explain the approach in more details, first write the RCA score matrix R as:

$$R = \left[\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n\right]$$

where $\mathbf{p}_j$, an arbitrary element in R, is a vector of length m that represents the RCA scores of product j for all the m countries[14] in the sample:

$$\mathbf{p}_j = \begin{bmatrix} r_{1j} \\ r_{2j} \\ \cdot \\ \cdot \\ r_{mj} \end{bmatrix}$$

The cosine similarity between products j and j' is equal to $\frac{\mathbf{p}_j \cdot \mathbf{p}_{j'}}{\|\mathbf{p}_j\| \|\mathbf{p}_{j'}\|}$, which ranges from -1, when the two vectors are the exact opposite, to 1, when the two are exactly the same. The intuition behind this is that by comparing the two sets of countries that export i and j, and how important the products are in the countries' export baskets, information can be inferred regarding how closely related the two products are.

The implementation of the product-based KNN recommender for country i in year t involves the following steps:

1. Represent each product in the SITC 4-digit product space as a vector of RCA scores, $\mathbf{p}_j$.

2. Select the set of products that country i has a revealed comparative advantage, i.e., $r_{ij} > 1$, which will be referred as the high-RCA product set of country i.

3. For each $j \in [1, n]$, calculate the predicted value of r_{ij} as a weighted average RCA score of the high-RCA product set, weighted by the cosine similarity between product j and the products in the country's high-RCA set.

4. The algorithm-generated results for country i are the $\hat{K}_{rca,it}$ products with the highest predicted r_{ij} values (i.e., scores of the training result, or "recommendation scores" in accordance with the convention in the collaborative filtering literature), where $\hat{K}_{rca,it}$ comes from the estimation in Section 4.3.1.

We repeat the above steps for each country-year pair to obtain the algorithm-generated scores in terms of SITC 4-digit products in our full sample of countries and time range with effective first-stage prediction.

[14] Note that in our implementation, m is effectively the cross-sectional country numbers. In machine learning terminology, each product in the sample has m features.

4.3.3 Calculating similarity scores

For the next step, we compute the similarity between the actual export portfolio of a country and the recommended export portfolio.

We define the portfolio structure of country i's actual exports in time t as the number of high RCA exports (at SITC 4-digit level) that belong to each SITC 1-digit sector,[15] as a share of total number of high RCA exports. In other words, let $N_{l,it}^{actual}$ be the number of high RCA exports in sector l, and

$$s_{l,it}^{actual} \equiv \frac{N_{l,it}^{actual}}{\sum_{l'} N_{l',it}^{actual}}$$

is the share of the number of high RCA exports that belong to sector l in the total number of high RCA exports. Country i's export structure, S_{it}^{actual}, is thus defined as a $L \times 1$ vector: $\left[s_{l,it}^{actual}\right]_{L\times 1}$, where $L = 10$ with the SITC sectoral level aggregation.

Similarly, we define the recommended export structure $S_{it}^{rec} \equiv \left[s_{l,it}^{rec}\right]_{L\times 1}$, as the vector for the number of recommended products that belong to each SITC 1-digit sector as a share of the total number of recommended export products.

The similarity score between the actual and the recommended export portfolios for country i at time t is then calculated as the similarity between the two vectors of actual and recommended structures:

$$Sim_{it} \equiv \frac{\left(S_{i,t-\Delta t}^{rec} - \bar{S}_{i,t-\Delta t}^{rec}\right)\cdot\left(S_{it}^{actual} - \bar{S}_{i,t}^{actual}\right)}{||S_{i,t-\Delta t}^{rec} - \bar{S}_{i,t-\Delta t}^{rec}||||S_{it}^{actual} - \bar{S}_{i,t}^{actual}||}$$

where Δt is a time lag we used when calculating the similarity scores to account for the fact that it takes time for an export structure to evolve.[16] This means that the similarity scores depend on our choices of time lags. In our baseline estimation, we set $\Delta t = 5$ years. Alternative assumptions for the time lags are also adopted in robustness checks presented in Section 4.6.

We calculate the annual Sim_{it} for all country-year pairs in the sample, and then incorporate the scores into the growth, volatility and risk-adjusted regressions that will be specified in the following section.

[15] See appendix Table 4.17 for the full list of SITC 1-digit sectors.

[16] We include this time lag because Che (2020) [103] found that recommendations given by the product-based KNN algorithm are to some extent forward-looking, in that they match the export portfolios of several high-growth countries in their future years better than in the current year.

4.3.4 Growth and volatility estimations

Our main econometric exercise aims to investigate the correlation[17] of the alignment of recommended and actual export structure on growth and volatility of growth. The hypothesis is that countries with an export structure highly aligned with their latent comparative advantages– manifested as a high similarity score between their actual and algorithm-recommended export portfolios, as defined in Section 4.3.3– should see higher and more stable growth over time.

To examine the robust correlation of export structure and growth, we specify the following estimation model:

$$g_{it} = \beta_0 + \beta_1 y_{i,t-\Delta t} + \beta_2 Sim_{it} + \gamma \mathbf{X}_{it} + \epsilon_{it} \tag{4.1}$$

where g_{it} the average annual growth in GDP per capita for country i from $t-\Delta t$ to t. $y_{i,t-\Delta t}$ is the lagged real GDP per capita in log form. Sim_{it} is the similarity score calculated as in Section 4.3.3. $\mathbf{X}_{it}$ is a set of controls, including investment-to-GDP ratio, human capital growth, TFP growth, and world GDP growth in some specifications to account for common external shocks. We also include country and time fixed effects in $\mathbf{X}_{it}$ for some of the regression specifications (see Section 4.5). The similarity scores, as well as most control variables, are annual averages over the Δt time window. In our baseline estimation, we set $\Delta t = 5$ years. The regressions are run with non-overlapping Δt as the time unit. Our main parameter of interest is β_2.

Similarly, we can look at the impact of export structure on the volatility of growth with the following model:

$$vol_{it} = \beta_0 + \beta_1 vol_{i,t-\Delta t} + \beta_2 Sim_{it} + \gamma \mathbf{X}_{it} + \epsilon_{it} \tag{4.2}$$

where vol_{it} is the standard deviation of annual growth of real GDP per capita during the Δt time period. And $vol_{i,t-\Delta t}$ is the lagged dependent variable. Controls ($\mathbf{X}_{it}$) are broadly the same as in the growth regression, except we replace world growth with the growth volatility of world GDP, to control for the level of external volatility.

Alternatively, we can combine the information on the left-hand side of Equations (4.1) and (4.2), and estimate the impact of export structure on countries' "risk-adjusted growth". Here we define country i's risk-adjusted growth, g^{ra}_{it}, as the deviation of country i's growth from the world average growth rate, $g_{it} - \overline{g}_t$, divided by its standard deviation σ_{it}, over the Δt time period. We then estimate the following equation,

$$g^{ra}_{it} = \beta_0 + \beta_1 g^{ra}_{i,t-\Delta t} + \beta_2 Sim_{it} + \gamma \mathbf{X}_{it} + \epsilon_{it}$$

[17]We are aware of the limit using the growth panel addressing causality. In the rest of the chapter, "impact" is interpreted in correlation.

Controls ($\mathbf{X}_{it}$) are the same as in the growth regression, except we replace world growth with the average risk-adjusted growth across countries in $\mathbf{X}_{it}$.

For each equation, the estimation is done using simple OLS, country and time fixed-effect estimator, and a system GMM estimator following Arellano and Bond (1991) [117]. The system GMM estimator is employed to address the endogeneity issue introduced by having the lagged dependent variable on the right hand side, which likely affects the consistency of OLS and fixed-effect estimators. In the system GMM estimation, the lagged dependent variable and country-level controls are treated as endogenous and instrumented as such. Time fixed effect and world-level controls are treated as exogenous. Section 4.5 presents results from all three estimators for each regression.

4.4 Data

The country-product level raw export data and the actual export RCA scores come from the Atlas of Economic Complexity Dataverse (2020 version), which are in turn sourced from UN Comtrade Database. The macroeconomic variables come from the World Bank and Penn World Table[18]. Summary statistics for the main variables used in the regressions are shown in Table 4.1. The data used for growth regressions are based on annual frequency covering 1980-2018. Specific estimations depend on different time-lag experiments as detailed in Section 4.5 and Section 4.6.

Table 4.1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Similarity score	0.817	0.179	-0.203	0.995	1203
GDP per capita	8.390	1.505	5.129	11.663	1414
Investment rate	0.219	0.109	-0.479	0.942	1398
TFP growth	0.002	0.028	-0.184	0.222	868
Human capital growth	0.01	0.007	-0.025	0.043	1107
GDP per capita growth	0.017	0.037	-0.247	0.367	1324
Growth volatility	0.015	0.014	0.001	0.142	1177
Risk-adjusted growth	0.705	4.474	-75.438	50.674	1175

The *similarity score* is calculated at the country-year level, following the steps described in Section 4.3.3. In the appendix, we show summary statistics for RCA scores and recommendation scores used to calculate the similarity score (see Table 4.18 and Table 4.19), as well as the box plots for the distributions of RCA scores and recommendation scores by SITC 1-digit sector (see Figure 4.10). Figure 4.8 presents the similarity score distribution around the world in 2018.

[18]Version 10.0, see Feenstra et al. (2015) for metadata details.

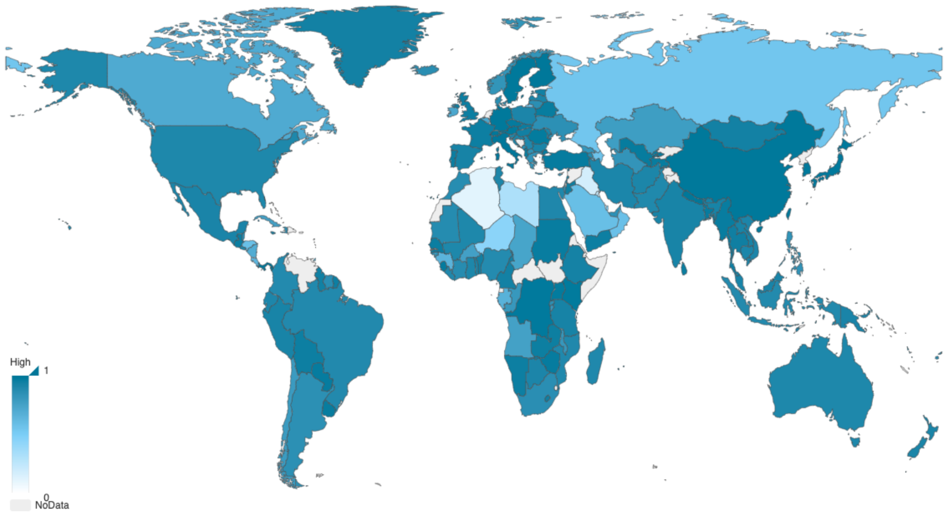


Figure 4.8: The Global Distribution of Similarity Scores

4.5 Estimation Results

4.5.1 Correlation with growth

Table 4.2 presents our baseline results for the growth regression. Columns (1) and (2) are results from the OLS estimation with robust standard errors, without and with controls respectively. Column (3) presents results from the fixed-effect estimation with clustered standard errors. Columns (4) and (5) are results from the system GMM estimation, with year dummies and world GDP growth included in the controls, respectively. Standard errors are given in parentheses. As mentioned earlier, we choose a time window of 5 years in the baseline estimations, to account for the fact that export structure is a slow moving variable whose impact may take some time to show. This gives us over 700 observations in the GMM estimation.

As expected, the control variables– investment rate, TFP and human capital growth, world GDP growth– are positive and mostly significant. The lagged GDP per capita variable is negative and significant in the OLS and fixed effect estimations, consistent with the prediction of economic convergence theory that poorer countries should grow faster than richer countries. But the variable has a negative and insignificant coefficient in the system GMM specifications.

Consistent with our hypothesis, the similarity score variable is positive and significant in all regression specifications, i.e., higher alignment between actual export structure and recommended export structure is good for growth. The magnitude of the coefficient does not vary too much across different specifications. According to the system GMM estimation (Column 5), a 0.1 increase in the similarity score (Sim_{it}) is associated with a 0.22 percentage point increase in the annual growth rate of real GDP per capita. The coefficient is statistically significant at 1% level. This is equivalent to a move from the median to the 90th percentile of the similarity score distribution.

The choice of the time window potentially affects the magnitude and significance of the result.[19] We will explore alternative specifications for the time window in the robustness section (Section 4.6).

4.5.2 Correlation with growth volatility

Table 4.3 gives the results for the volatility regression, where the dependent variable is the standard deviation of annual growth during each Δt period. Similar to Table 4.2, Columns (1)-(3) are results from OLS and fixed effect estimations. And Columns (4)-(5) are system GMM results, with year dummies and volatility of world growth in the controls respectively.

The lagged volatility variable is positive and significant across all regression specifications, even in the fixed effect and system GMM specifications, where persistent and country specific volatility differences are supposed to be accounted for. This suggests significant stickiness in growth volatility. The control variables of investment rate, TFP growth, and human capital growth do not appear to have any material influence on the growth volatility, ceteris paribus. However, the dependent variable is shown to be highly correlated with the external environment, as indicated by the positive and significant coefficient for the variable of world growth volatility.

Turning to the similarity score, the results show a negative and significant coefficient for the variable under two OLS and two system GMM estimation specifications, i.e., higher alignment between actual export structure and recommended export structure helps growth to be more stable. The coefficient is negative but not significant in the fixed effect estimation.

In terms of the magnitude of impact, the system GMM estimator in Column 5 suggests that a 0.1 increase in the similarity score is associated with a 0.0015 decrease in the standard deviation of growth rate in a 5-year window. This is statistically significant at 1% level. In the robustness section, we will explore whether changing Δt alters the sign and significance of the baseline results.

4.5.3 Correlation with risk-adjusted growth

Instead of looking at growth rate and growth volatility separately, we can also summarize a country's growth performance in one variable, which we call the risk-adjusted growth. This is calculated as the average annual growth divided by the standard deviation of growth during Δt period. Table 4.4 presents results

[19] Note that the default setting, with $\Delta t = 5$ years, is by our choice to start with considering the trade evolution and the panel data structure.

for the risk-adjusted growth regression. The columns are structured similarly as in Tables 4.2 and 4.3.

The lagged dependent variable shows up with a positive coefficient in all regression specifications, though it is not significant except in the OLS specification without controls and the fixed effect specification. Among the control variables included, investment rate and TFP growth are shown to be associated with higher risk-adjusted growth, while the human capital variable is insignificant and its sign varying across different estimations. In addition, the dependent variable is strongly correlated with the risk adjusted growth at the world level.

Consistent with our hypothesis, the similarity score variable has a positive coefficient across all specifications. The coefficient is significant in the basic OLS and the system GMM estimations. The system GMM result (Column 5) suggests that a 0.1 increase in the similarity score is associated with 20.95 percentage points increase per standard deviation in the annual growth rate. To put it in an alternative way, an increase of 0.8 in the similarity score moves a country from the world medium in the risk-adjusted growth spectrum to the 75th percentile level. The result is statistically significant at 5% level.

4.6 Robustness

4.6.1 Changing time interval

In the baseline regressions, we set $\Delta t = 5$ years. In this section, we examine how our results may change with different assumptions for Δt. Tables 4.5 to 4.7 show results with $\Delta t = 3, 5$, and 7 years, under the system GMM specification with all controls. Tables 4.8 to 4.13 present the results with these alternative time interval assumptions for all estimation specifications.

As Table 4.5 shows, the signs remain the same and the magnitude and significance level for our variable of interest (Sim_{it}) differ slightly from the baseline case ($\Delta t = 5$). When $\Delta t = 3$, a 0.1 increase in the similarity score is associated with 0.19 percentage point increase in growth, while this coefficient is 0.13 when $\Delta t = 7$. The results for the cases of $\Delta t = 3$ and 7 under OLS and fixed effect specifications can be found in Tables 4.8 and 4.9.

Table 4.6 summarizes results for the volatility regression for different time windows, under the system GMM specification. It shows that a 0.1 increase in the similarity score is associated with 0.0017 and 0.0013 decrease in the standard deviation of growth when $\Delta t = 3$ and 7 respectively. These are mostly consistent with the baseline result. For the estimates of OLS and fixed effect specifications, see Tables 4.10 and 4.11.

The adjusted-growth regression results with alternative time windows using the system GMM estimator are summarized in Table 4.7. And Tables 4.12 and 4.13 present the results for other regression specifications. As Tables 4.7 shows, when $\Delta t = 3$, a 0.1 increase in the similarity score is associated with 21.84

percentage point increase in the risk-adjusted annual growth rate, at 1% significance level. However, this value decreases to 13.79 when $\Delta t = 7$, and it is not statistically significant.[20]

A summary of estimates with time intervals varying between 3 to 7 can be found in Figure 4.9. In this plot, we show confidence intervals for the key coefficient in growth regressions, volatility regressions, and risk-adjusted growth regressions against consecutive time intervals.

4.6.2 Winsorization

We also run a set of regressions where we winsorize the dependent variables by removing the top and bottom 5% from the sample observations. Overall, signs and significance of the similarity score variable do not change markedly from the baseline results.

Table 4.14 presents results of the growth regression with the real GDP growth rate winsorized. The coefficient for the similarity score remains positive and significant across all specifications. In fact, the magnitude of the coefficient is slightly lower in four out of the five specifications compared to the baseline. According to the system GMM estimates (Column 5), a 0.1 increase in the similarity score (Sim) is associated with a 0.18 percentage point increase in real GDP per capita growth.

Table 4.15 show results of the winsorized volatility regression. Overall the estimates for the similarity score variable are somewhat weaker than in the baseline in terms of magnitudes, but still point to a negative impact of the similarity score on growth volatility. The coefficient for Sim_{it} is negative in all five specifications, but not significant in four out of five specifications.

Table 4.16 are results for the winsorized adjusted-growth regression. Similar to the baseline results, the coefficient of similarity score is positive and significant in the system GMM with the world growth control, though the magnitudes of the coefficient is smaller than in the baseline. According to the system GMM estimates (Column 5), a 0.1 increase in the similarity score is associated with 9.8 percentage points increase per standard deviation in the annual growth rate. This is much smaller than 20.95 percentage points increase in the baseline result. The statistical significance level changes from 1% to 10%. But overall, the winsorized results still confirm a positive correlation between the similarity score and countries' comprehensive growth performance.

[20]As a further robustness check, we tried $\Delta t = 8$ and this value is 19.89 percentage point at 5% significance level. Specifically, the estimated coefficient for similarity score in the risk-adjusted growth regression under the system GMM specification is 1.989 with standard error 0.959.

4.7 Conclusion

In this chapter, we try to combine machine learning and economics approaches to shed light on the importance of export structure evolution in the convergence of growth and income. We first leveraged machine learning methods to characterize the complex patterns in countries' latent comparative advantages and create algorithm-based export recommendations accordingly. Specifically, we used a product-based KNN algorithm to provide annual export product recommendations at the SITC 4-digit level for over 190 economies, from 1980 to 2018. We then use a standard empirical growth model to evaluate whether a country's growth performance is better when the actual portfolio is more aligned with the algorithm-based portfolio.

Our results confirm the merits of such algorithm-based export portfolio. We find that economies with a higher similarity score between algorithm-based recommended and actual export portfolios achieve better growth performance. In our baseline estimation, an increase in the similarity score is associated with an increase in the annual growth of real GDP per capita, a decrease in the standard deviation of growth rate over five-year time windows, and an increase in risk-adjusted growth. These results are overall robust with respect to changing time intervals and winsorization.

To understand detailed mechanisms driving comparative advantages, one needs to utlize structural models for further aggregate analyses and decompositions. The algorithm-produced export recommendations are no substitutes for detailed and multidimensional analyses of the viability of any industry in a country. In addition, it goes without saying that knowing which industries a country may have comparative advantages in does not automatically translate into specific policy recommendations. Neither are we advocating for direct policy interventions in shaping a country's export structure. How a country can best support the growth of its tradable sectors to leverage the country's comparative advantages requires model-based analyses and is likely a case-by-case discussion, depending on many country-specific factors. We have considered this in the methodology design, and we acknowledge that specific institutional challenges vary significantly across the world. Therefore, there still remain gaps between the algo-based recommendations and policy recommendations in terms of diversification advisory and policy tools under different country circumstances.

Nonetheless, our paper provides cross-country empirical evidence on directions of diversification with positive correlation with growth performance. Our study also raises interesting questions on potential structural changes and economic reforms for different countries. When considering greater accession to the global market, small open economies with weak fundamentals confront tradeoffs about what products to diversify into. Large emerging markets could face challenges associated with either short-term or long-term bottlenecks. Further quantitative studies can be conducted to identify potential policy gaps and reforms on the path to achieving better growth.

4.8 Appendix

Table 4.2: Growth Regression (baseline)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged GDP Per Capita	-0.002**	-0.008***	-0.016***	-0.001	-0.001
	(0.001)	(0.003)	(0.004)	(0.001)	(0.001)
Similarity Score	0.023***	0.036***	0.035**	0.023***	0.022***
	(0.008)	(0.011)	(0.014)	(0.006)	(0.006)
Inv. Rate		0.016***	0.014***	0.021***	0.021***
		(0.002)	(0.002)	(0.004)	(0.004)
TFP Growth		0.716***	0.689***	0.745***	0.742***
		(0.063)	(0.057)	(0.064)	(0.061)
Human Capital Growth		0.426***	0.420***	0.600***	0.571***
		(0.127)	(0.133)	(0.149)	(0.153)
World Growth		0.381*			0.263
		(0.195)			(0.193)
No. of Obs.	1168	732	732	732	732
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.01	0.02
Hansen-J (p-value)				0.79	0.80

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.3: Volatility Regression (baseline)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Volatility	0.443***	0.272***	0.268***	0.468***	0.468***
	(0.071)	(0.075)	(0.059)	(0.056)	(0.055)
Similarity Score	-0.010***	-0.023**	-0.022	-0.014**	-0.015***
	(0.004)	(0.011)	(0.014)	(0.006)	(0.006)
Inv. Rate		-0.001	-0.001	-0.003**	-0.002
		(0.001)	(0.001)	(0.001)	(0.001)
TFP Growth		-0.011	-0.013	-0.012	-0.011
		(0.035)	(0.026)	(0.023)	(0.027)
Human Capital Growth		0.136*	0.138	0.112	0.082
		(0.080)	(0.090)	(0.083)	(0.078)
World Volatility		0.675***			0.464**
		(0.190)			(0.209)
No. of Obs.	966	614	613	614	614
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.51	0.53
Hansen-J (p-value)				0.58	0.38

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: Risk-adjusted Growth Regression (baseline)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Adj. Growth	0.311^{***}	0.073	0.086^{*}	0.066	0.050
	(0.111)	(0.045)	(0.048)	(0.052)	(0.046)
Similarity Score	3.406^{***}	3.270	2.604	1.985^{**}	2.095^{**}
	(0.627)	(2.589)	(2.568)	(0.863)	(0.853)
Inv. Rate		1.989^{***}	2.247^{***}	1.775^{***}	1.721^{***}
		(0.487)	(0.527)	(0.482)	(0.541)
TFP Growth		33.794^{***}	38.395^{***}	36.424^{***}	29.998^{***}
		(8.458)	(11.167)	(10.365)	(8.329)
Human Capital Growth		1.194	-11.629	-25.097	-14.007
		(29.750)	(28.030)	(34.874)	(34.170)
World Adj. Growth		0.508^{**}			0.578^{***}
		(0.218)			(0.215)
No. of Obs.	964	613	612	613	613
AR1 (p-value)				0.01	0.02
AR2 (p-value)				0.49	0.17
Hansen-J (p-value)				0.49	0.43

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.5: Robustness: Growth

	(1) +3	(2) +5	(3) +7
Lagged GDP Per Capita	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.002)
Similarity Score	0.019***	0.022***	0.013**
	(0.007)	(0.006)	(0.007)
Inv. Rate	0.021***	0.021***	0.018***
	(0.003)	(0.004)	(0.004)
TFP Growth	0.797***	0.742***	0.758***
	(0.067)	(0.061)	(0.076)
Human Capital Growth	0.650***	0.571***	0.530***
	(0.158)	(0.153)	(0.183)
World Growth	0.391***	0.263	0.613***
	(0.085)	(0.193)	(0.210)
Constant	0.034**	0.030**	0.030
	(0.015)	(0.015)	(0.019)
No. of Obs.	1258	732	523
AR1 (p-value)	0.00	0.00	0.01
AR2 (p-value)	0.89	0.02	0.10
Hansen-J (p-value)	1.00	0.80	0.01

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.6: Robustness: Volatility

	(1) +3	(2) +5	(3) +7
Lagged Volatility	0.417***	0.468***	0.243***
	(0.058)	(0.055)	(0.073)
Similarity Score	-0.017***	-0.015***	-0.013***
	(0.005)	(0.006)	(0.004)
Inv. Rate	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)
TFP Growth	-0.116***	-0.011	0.088**
	(0.032)	(0.027)	(0.035)
Human Capital Growth	0.075	0.082	-0.058
	(0.102)	(0.078)	(0.102)
World Volatility	0.526***	0.464**	0.532**
	(0.132)	(0.209)	(0.234)
Constant	0.010*	0.008*	0.009*
	(0.006)	(0.005)	(0.005)
No. of Obs.	1231	614	409
AR1 (p-value)	0.00	0.00	0.00
AR2 (p-value)	0.05	0.53	0.31
Hansen-J (p-value)	1.00	0.38	0.23

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: Robustness: Risk-adjusted Growth

	(1)	(2)	(3)
	+3	+5	+7
Lagged Adj. Growth	0.265***	0.050	0.253***
	(0.055)	(0.046)	(0.059)
Similarity Score	2.184***	2.095**	1.379
	(0.679)	(0.853)	(0.893)
Inv. Rate	1.483***	1.721***	0.442
	(0.320)	(0.541)	(0.400)
TFP Growth	36.657***	29.998***	58.120***
	(7.757)	(8.329)	(11.848)
Human Capital Growth	49.083**	-14.007	73.046*
	(21.635)	(34.170)	(40.394)
World Adj. Growth	0.528***	0.578***	0.715***
	(0.156)	(0.215)	(0.164)
Constant	0.358	1.486	-1.217
	(0.703)	(1.020)	(0.900)
No. of Obs.	1230	613	408
AR1 (p-value)	0.00	0.02	0.00
AR2 (p-value)	0.14	0.17	0.01
Hansen-J (p-value)	1.00	0.43	0.10

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.8: Growth Regression (fwd+3)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged GDP Per Capita	-0.001*	-0.008***	-0.015***	-0.002	-0.001
	(0.001)	(0.002)	(0.005)	(0.001)	(0.001)
Similarity Score	0.025***	0.022**	0.023**	0.020***	0.019***
	(0.007)	(0.010)	(0.010)	(0.007)	(0.007)
Inv. Rate		0.019***	0.018***	0.022***	0.021***
		(0.002)	(0.003)	(0.004)	(0.003)
TFP Growth		0.764***	0.740***	0.779***	0.797***
		(0.050)	(0.069)	(0.070)	(0.067)
Human Capital Growth		0.492***	0.486***	0.566***	0.650***
		(0.121)	(0.167)	(0.196)	(0.158)
World Growth		0.422***			0.391***
		(0.101)			(0.085)
No. of Obs.	2030	1258	1258	1258	1258
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.95	0.89
Hansen-J (p-value)				1.00	1.00

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.9: Growth Regression (fwd+7)

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS-X	FE(t),X	GMM-(t)	GMM-X(t)
Lagged GDP Per Capita	-0.002**	-0.011***	-0.024***	-0.001	-0.001
	(0.001)	(0.003)	(0.006)	(0.002)	(0.002)
Similarity Score	0.015**	0.019	0.021**	0.014**	0.013**
	(0.007)	(0.015)	(0.010)	(0.007)	(0.007)
Inv. Rate		0.014***	0.012***	0.018***	0.018***
		(0.002)	(0.002)	(0.005)	(0.004)
TFP Growth		0.686***	0.622***	0.770***	0.758***
		(0.076)	(0.093)	(0.093)	(0.076)
Human Capital Growth		0.481***	0.496***	0.580**	0.530***
		(0.176)	(0.187)	(0.239)	(0.183)
World Growth		1.054***			0.613***
		(0.301)			(0.210)
No. of Obs.	838	523	522	523	523
AR1 (p-value)				0.01	0.01
AR2 (p-value)				0.13	0.10
Hansen-J (p-value)				0.01	0.01

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.10: Volatility Regression (fwd+3)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Volatility	0.444***	0.304***	0.329***	0.450***	0.417***
	(0.044)	(0.061)	(0.055)	(0.054)	(0.058)
Similarity Score	-0.014***	-0.011	-0.011	-0.015**	-0.017***
	(0.003)	(0.009)	(0.011)	(0.006)	(0.005)
Inv. Rate		-0.001	-0.002	-0.003**	-0.002
		(0.001)	(0.001)	(0.002)	(0.001)
TFP Growth		-0.113***	-0.117***	-0.117***	-0.116***
		(0.031)	(0.029)	(0.033)	(0.032)
Human Capital Growth		0.025	0.059	0.084	0.075
		(0.089)	(0.108)	(0.121)	(0.102)
World Volatility		0.661***			0.526***
		(0.118)			(0.132)
No. of Obs.	1948	1231	1231	1231	1231
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.04	0.05
Hansen-J (p-value)				1.00	1.00

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.11: Volatility Regression (fwd+7)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Volatility	0.453***	-0.013	-0.017	0.235***	0.243***
	(0.098)	(0.096)	(0.058)	(0.069)	(0.073)
Similarity Score	-0.009**	-0.016	-0.017	-0.013***	-0.013***
	(0.004)	(0.010)	(0.011)	(0.005)	(0.004)
Inv. Rate		-0.002	-0.002	-0.003*	-0.002
		(0.001)	(0.001)	(0.001)	(0.001)
TFP Growth		0.065	0.068*	0.102***	0.088**
		(0.039)	(0.038)	(0.033)	(0.035)
Human Capital Growth		0.167	0.129	-0.121	-0.058
		(0.123)	(0.151)	(0.098)	(0.102)
World Volatility		0.908***			0.532**
		(0.247)			(0.234)
No. of Obs.	645	409	402	409	409
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.25	0.31
Hansen-J (p-value)				0.30	0.23

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.12: Risk-adjusted Growth Regression (fwd+3)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Adj. Growth	0.407***	0.156***	0.164***	0.280***	0.265***
	(0.050)	(0.043)	(0.038)	(0.056)	(0.055)
Similarity Score	2.641***	2.259*	2.870**	2.403***	2.184***
	(0.418)	(1.167)	(1.311)	(0.824)	(0.679)
Inv. Rate		2.084***	2.257***	1.520***	1.483***
		(0.322)	(0.321)	(0.318)	(0.320)
TFP Growth		35.252***	37.052***	40.304***	36.657***
		(4.555)	(7.304)	(7.740)	(7.757)
Human Capital Growth		57.144***	38.695**	31.331	49.083**
		(19.311)	(18.272)	(24.025)	(21.635)
World Adj. Growth		0.410***			0.528***
		(0.137)			(0.156)
No. of Obs.	1946	1230	1230	1230	1230
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.16	0.14
Hansen-J (p-value)				1.00	1.00

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.13: Risk-adjusted Growth Regression (fwd+7)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Adj. Growth	0.234**	0.137**	0.153***	0.289***	0.253***
	(0.107)	(0.054)	(0.045)	(0.060)	(0.059)
Similarity Score	3.283***	1.676	0.965	1.189	1.379
	(0.680)	(2.548)	(2.452)	(0.993)	(0.893)
Inv. Rate		1.251***	1.321***	0.457	0.442
		(0.446)	(0.448)	(0.393)	(0.400)
TFP Growth		44.795***	49.179***	65.872***	58.120***
		(11.965)	(15.320)	(13.691)	(11.848)
Human Capital Growth		43.964	25.323	58.270	73.046*
		(53.825)	(55.088)	(42.408)	(40.394)
World Adj. Growth		0.732***			0.715***
		(0.191)			(0.164)
No. of Obs.	643	408	401	408	408
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.03	0.01
Hansen-J (p-value)				0.13	0.10

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.14: Growth Regression (winsorized)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged GDP Per Capita	-0.000	-0.008***	-0.018***	-0.002**	-0.002**
	(0.000)	(0.003)	(0.006)	(0.001)	(0.001)
Similarity Score	0.023***	0.019*	0.018*	0.018***	0.018***
	(0.006)	(0.010)	(0.009)	(0.007)	(0.007)
Inv. Rate		0.017***	0.015***	0.018***	0.018***
		(0.002)	(0.002)	(0.003)	(0.003)
TFP Growth		0.676***	0.651***	0.695***	0.693***
		(0.062)	(0.074)	(0.075)	(0.071)
Human Capital Growth		0.325***	0.306**	0.391***	0.367**
		(0.117)	(0.120)	(0.138)	(0.147)
World Growth		0.339**			0.313*
		(0.168)			(0.165)
Constant	0.001	0.076***	0.177***	0.048***	0.041***
	(0.007)	(0.025)	(0.049)	(0.013)	(0.013)
No. of Obs.	931	631	627	631	631
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.04	0.11
Hansen-J (p-value)				0.85	0.81

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.15: Volatility Regression (winsorized)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Volatility	0.329***	0.141**	0.129***	0.311***	0.300***
	(0.044)	(0.069)	(0.043)	(0.038)	(0.038)
Similarity Score	-0.004*	-0.005	-0.005	-0.008	-0.007
	(0.002)	(0.006)	(0.006)	(0.005)	(0.005)
Inv. Rate		-0.002	-0.002	-0.002**	-0.002
		(0.001)	(0.001)	(0.001)	(0.001)
TFP Growth		-0.011	-0.014	0.001	0.002
		(0.029)	(0.022)	(0.023)	(0.024)
Human Capital Growth		0.029	0.026	-0.023	-0.034
		(0.056)	(0.054)	(0.057)	(0.058)
World Volatility		0.560***			0.449***
		(0.147)			(0.160)
Constant	0.011***	0.014*	0.011*	0.010*	0.004
	(0.002)	(0.007)	(0.006)	(0.005)	(0.005)
No. of Obs.	816	530	524	530	530
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.32	0.29
Hansen-J (p-value)				0.24	0.26

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.16: Risk-adjusted Growth Regression (winsorized)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Adj. Growth	0.435***	0.225***	0.274***	0.388***	0.313***
	(0.038)	(0.051)	(0.046)	(0.058)	(0.058)
Similarity Score	1.878***	0.478	-0.181	0.699	0.980*
	(0.403)	(1.661)	(1.654)	(0.503)	(0.507)
Inv. Rate		0.850***	1.083***	0.467*	0.245
		(0.268)	(0.288)	(0.273)	(0.295)
TFP Growth		28.479***	31.456***	37.499***	32.498***
		(6.663)	(8.919)	(8.468)	(6.919)
Human Capital Growth		22.874	13.397	20.382	31.930
		(18.292)	(16.395)	(25.115)	(25.516)
World Adj. Growth		0.374***			0.438***
		(0.110)			(0.113)
Constant	-1.136***	0.502	2.163	-0.287	-0.680
	(0.331)	(0.666)	(1.588)	(0.731)	(0.710)
No. of Obs.	804	522	517	522	522
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.81	0.22
Hansen-J (p-value)				0.23	0.26

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.17: Classification according to SITC 1 - Section

SITC Code	Sector Name
0	Food and live animals
1	Beverages and tobacco
2	Crude materials, inedible, except fuels
3	Mineral fuels, lubricants and related materials
4	Animal and vegetable oils, fats and waxes
5	Chemicals and related products, n.e.s.
6	Manufactured goods classified chiefly by material
7	Machinery and transport equipment
8	Miscellaneous manufactured articles
9	Commodities and transactions not classified elsewhere in the SITC

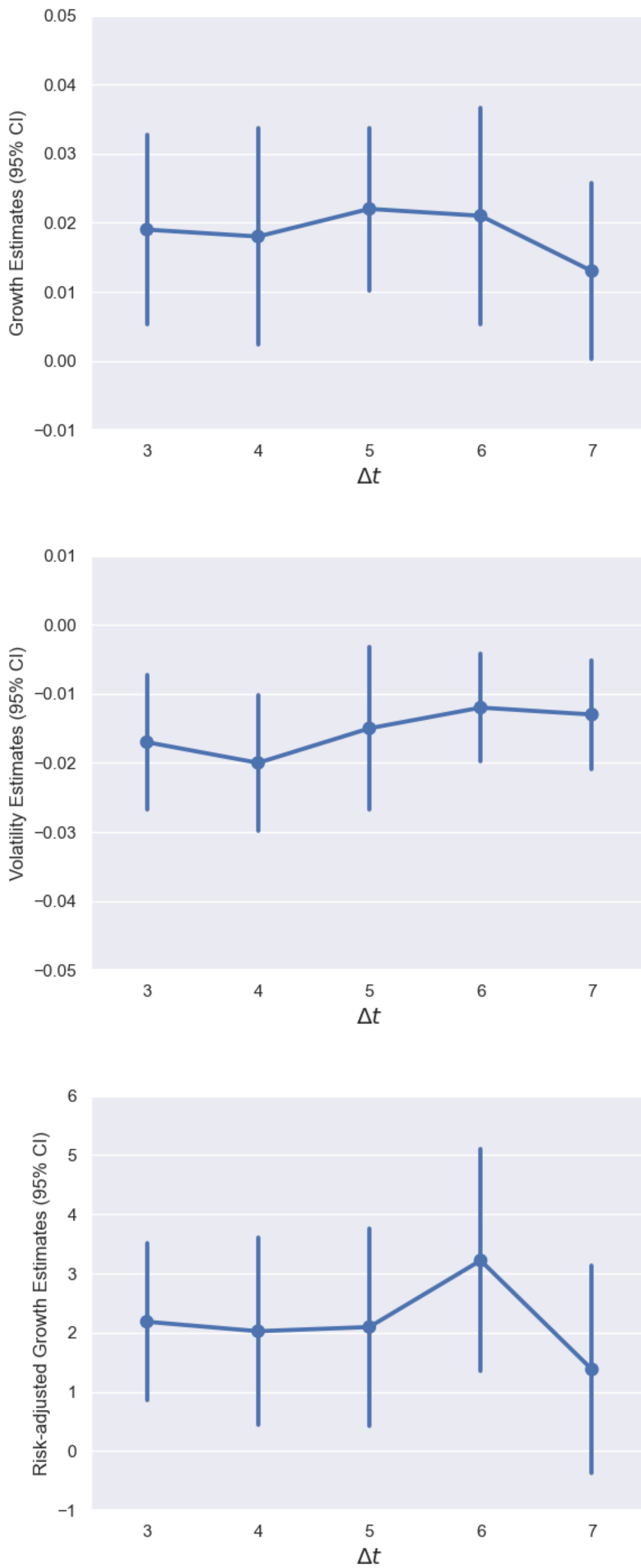


Figure 4.9: Robustness: Confidence Intervals in Main Regressions

Table 4.18: Summary of Actual RCA by Year

year	N	Mean	Q1	Median	Q3
1985	61160	0.136	0.024	0.158	0.844
1990	69130	0.133	0.024	0.156	0.788
1995	84213	0.144	0.027	0.168	0.798
2000	96925	0.138	0.027	0.163	0.780
2005	100931	0.128	0.024	0.155	0.744
2010	104069	0.120	0.023	0.152	0.735
2015	101789	0.117	0.022	0.143	0.708

Table 4.19: Summary of Recommendation Scores by Year

year	N	Mean	Q1	Median	Q3
1985	133722	0.269	0.140	0.328	0.648
1990	134504	0.308	0.161	0.370	0.686
1995	152281	0.350	0.202	0.403	0.712
2000	159444	0.365	0.217	0.438	0.754
2005	160218	0.348	0.198	0.419	0.749
2010	160784	0.342	0.195	0.416	0.747
2015	160576	0.337	0.194	0.393	0.702

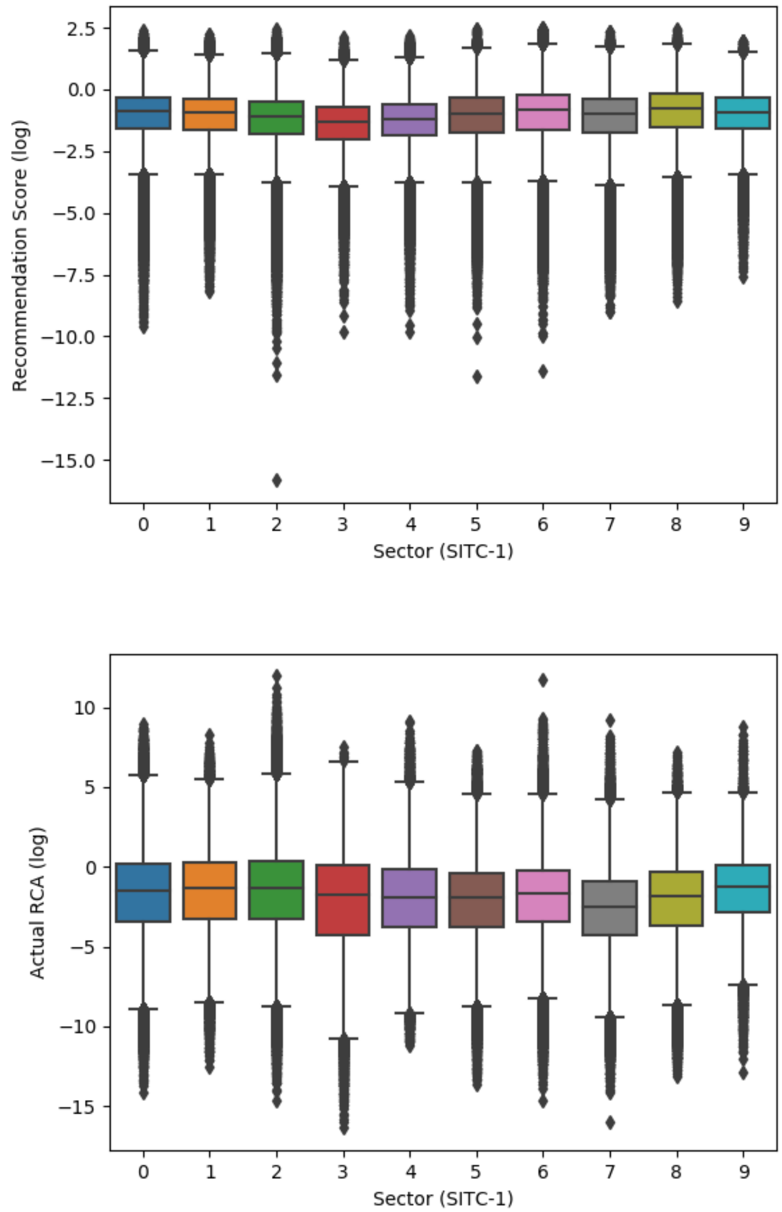


Figure 4.10: Distribution of Actual RCA and Algorithm-Based Recommendation Scores by SITC Sector

Bibliography

1. Baum-Snow N, Pavan R. Understanding the City Size Wage Gap. The Review of Economic Studies. 2011;79(1):88–127.
2. Decker RA, Haltiwanger J, Jarmin RS, Miranda J. Declining Business Dynamism: What We Know and the Way Forward. American Economic Review. 2016;106(5):203–07.
3. Edwards C. Entrepreneurs and Regulations: Removing State and Local Barriers to New Businesses. Cato Institute, Washington, DC; 2021. 916.
4. Chade H, Eeckhout J, Smith L. Sorting through Search and Matching Models in Economics. Journal of Economic Literature. 2017;55(2):493–544.
5. Becker GS. A Theory of Marriage: Part I. Journal of Political Economy. 1973;81(4):813–846.
6. Behrens K, Duranton G, Robert-Nicoud F. Productive Cities: Sorting, Selection, and Agglomeration. Journal of Political Economy. 2014;122(3):507 – 553.
7. Gaubert C. Firm Sorting and Agglomeration. American Economic Review. 2018;108(11):3117–53.
8. Baldwin RE, Okubo T. Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. Journal of Economic Geography. 2005;6(3):323–346.
9. Nocke V. A Gap for Me: Entrepreneurs and Entry. Journal of the European Economic Association. 2006;4(5):929–956.
10. Okubo T, Picard P, Thisse J. The spatial selection of heterogeneous firms. Journal of International Economics. 2010;82(2):230–237.
11. Gabaix X. Zipf's Law for Cities: An Explanation. The Quarterly Journal of Economics. 1999;114(3):739–767.
12. Eeckhout J, Guner N. Optimal Spatial Taxation: Are Big Cities too Small? CEMFI; 2017. wp2017.1705.
13. Ioannides Y, Skouras S. US city size distribution: Robustly Pareto, but only in the tail. Journal of Urban Economics. 2013;73(1):18–29.
14. Rosen KT, Resnick M. The size distribution of cities: An examination of the Pareto law and primacy. Journal of Urban Economics. 1980;8(2):165–186.
15. Soo KT. Zipf's Law for cities: a cross-country investigation. Regional Science and Urban Economics. 2005;35(3):239–263.

16. Rossi-Hansberg E, Wright MLJ. Urban Structure and Growth. Review of Economic Studies. 2007;74(2):597–624.

17. Geerolf F. A Static and Microfounded Theory of Zipf's Law for Firms and of the Top Labor Income Distribution. Society for Economic Dynamics; 2015.

18. Haltiwanger J, Jarmin RS, Kulick R, Miranda J. High Growth Young Firms: Contribution to Job, Output, and Productivity Growth. 2016; p. 11–62.

19. Pugsley B, Sahin A. Grown-up business cycles. Federal Reserve Bank of New York; 2014. 707.

20. Rubinton H. The Geography of Business Dynamism and Skill Biased Technical Change. Federal Reserve Bank of St. Louis; 2020.

21. Garicano L, Lelarge C, Van Reenen J. Firm Size Distortions and the Productivity Distribution: Evidence from France. American Economic Review. 2016;106(11):3439–79.

22. Barwick PJ, Kalouptsidi M, Zahur NB. China's Industrial Policy: an Empirical Evaluation. National Bureau of Economic Research; 2019. 26075.

23. Gu S, Jia C. Firm Dynamics and SOE Transformation During China's Economic Reform. Federal Reserve Bank of Cleveland; 2021. 21-24.

24. Asturias J, Hur S, Kehoe TJ, Ruhl KJ. Firm Entry and Exit and Aggregate Growth. National Bureau of Economic Research; 2017. 23202.

25. Alfred M. Principles of Economics. Palgrave Macmillan: London; 1890.

26. Ellison G, Glaeser EL. The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? American Economic Review. 1999;89(2):311–316.

27. Duranton G, Puga D. Micro-foundations of urban agglomeration economies. In: Henderson JV, Thisse JF, editors. Handbook of Regional and Urban Economics. vol. 4 of Handbook of Regional and Urban Economics. Elsevier; 2004. p. 2063–2117.

28. Gennaioli N, Porta RL, de Silanes FL, Shleifer A. Human Capital and Regional Development. The Quarterly Journal of Economics. 2013;128(1):105–164.

29. Bartik F. Das Gesetz der Bevolkerungskonzentration; 1913.

30. Embrechts KCMT Paul. Modelling Extremal Events for Insurance and Finance. Springer-Verlag Berlin Heidelberg; 1997.

31. Clauset A, Shalizi CR, Newman MEJ. Power-Law Distributions in Empirical Data. SIAM Review. 2009;51(4):661–703.

32. Luttmer EGJ. Selection, Growth, and the Size Distribution of Firms. The Quarterly Journal of Economics. 2007;122(3):1103–1144.

33. Acemoglu D, Angrist J. How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws. In: NBER Macroeconomics Annual 2000, Volume 15. NBER Chapters. National Bureau of Economic Research, Inc; 2001. p. 9–74.

34. Moretti E. Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. American Economic Review. 2004;94(3):656–690.

35. Axtell RL. Zipf Distribution of U.S. Firm Sizes. Science. 2001;293(5536):1818–1820.

36. Rossi-Hansberg E, Sarte PD, Trachter N. Diverging Trends in National and Local Concentration. NBER Macroeconomics Annual. 2021;35:115–150.

37. Moretti E. The New Geography of Jobs. Boston: Houghton Mifflin Harcourt; 2012.

38. Hernandez-Murillo R. Interjurisdictional Competition and Location Decisions of Firms. Federal Reserve Bank of St. Louis; 2014. 2014-36.

39. Saiz A. The Geographic Determinants of Housing Supply*. The Quarterly Journal of Economics. 2010;125(3):1253–1296.

40. Autor DH, Dorn D, Hanson GH. The Geography of Trade and Technology Shocks in the United States. American Economic Review. 2013;103(3):220–25.

41. Autor DH, Dorn D, Hanson GH. The China Syndrome: Local Labor Market Effects of Import Competition in the United States. American Economic Review. 2013;103(6):2121–68.

42. Acemoglu D, Restrepo P. Robots and Jobs: Evidence from US Labor Markets. National Bureau of Economic Research; 2017. 23285.

43. Chen Z, Liu Z, Suárez Serrato JC, Xu DY. Notching R&D Investment with Corporate Income Tax Cuts in China. American Economic Review. 2021;111(7):2065–2100.

44. Keen M, Konrad KA. The Theory of International Tax Competition and Coordination. Tax Law: Tax Law and Policy eJournal. 2014;.

45. Combes P, Duranton G, Gobillon L, Puga D, Roux S. The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. Econometrica. 2012;80(6):2543–2594.

46. Ossa R. A Quantitative Analysis of Subsidy Competition in the U.S. National Bureau of Economic Research; 2015. 20975.

47. Devereux MP, Loretz S. What Do We Know About Corporate Tax Competition? National Tax Journal. 2013;66(3):745–774.

48. Hines JR. Evaluating Tax Harmonization. National Bureau of Economic Research; 2022.

49. OECD. Statement on a Two-Pillar Solution to Address the Tax Challenges Arising from the Digitalisation of the Economy; 2021. Available from: https://www.oecd.org/tax/beps/statement-on-a-two-pillar-solution-to-address-the-tax-challenges-arising-from-the-digitalisation-of-the-economy-october-2021.pdf.

50. Hebous S, Keen M. Pareto-Improving Minimum Corporate Taxation. International Monetary Fund; 2021.

51. Wilson J. Trade, Capital Mobility, and Tax Competition. Journal of Political Economy. 1987;95(4):835–56.

52. Ottaviano GIP, van Ypersele T. Market size and tax competition. Journal of International Economics. 2005;67(1):25–46.

53. Kanbur R, Keen M. Jeux Sans Frontieres: Tax Competition and Tax Coordination when Countries Differ in Size. Economics Department, Queen's University; 1991. 819.

54. Nielsen S. A Simple Model of Commodity Taxation and Cross-Border Shopping. Economic Policy Research Unit (EPRU), University of Copenhagen. Department of Economics; 2001. 98-18.

55. Haufler A, Stähler F. Tax Competition in a Simple Model with Heterogeneous Firms: How Larger Markets Reduce Profit Taxes. CESifo; 2009. 2867.

56. Wilson J, Wildasin D. Capital tax competition: bane or boon. Journal of Public Economics. 2004;88(6):1065–1091.

57. Fajgelbaum PD, Gaubert C. Optimal Spatial Policies, Geography, and Sorting. The Quarterly Journal of Economics. 2020;135(2):959–1036.

58. Suárez Serrato JC, Zidar O. Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms. American Economic Review. 2016;106(9):2582–2624.

59. Fajgelbaum PD, Morales E, Serrato JCS, Zidar O. State Taxes and Spatial Misallocation. Review of Economic Studies. 2019;86(1):333–376.

60. Chirinko RS, Wilson DJ. Tax Competition Among U.S. States: Racing to the Bottom or Riding on a Seesaw? CESifo; 2011. 3535.

61. Giroud X, Rauh JD. State Taxation and the Reallocation of Business Activity: Evidence from Establishment-Level Data. Journal of Political Economy. 2019;127(3):1262–1316.

62. Moretti E, Wilson DJ. The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. American Economic Review. 2017;107(7):1858–1903.

63. Greenstone M, Hornbeck R, Moretti E. Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. Journal of Political Economy. 2010;118(3):536–598.

64. Nallareddy S, Serrato JCS, Rouen E. Do Corporate Tax Cuts Increase Income Inequality? In: Tax Policy and the Economy, Volume 36. NBER Chapters. National Bureau of Economic Research, Inc; 2021.

65. Bartik TJ. A New Panel Database on Business Incentives for Economic Development Offered by State and Local Governments in the United States. Working Paper Prepared for the Pew Charitable Trusts. 2017;.

66. Helpman E. A Simple Theory of International Trade with Multinational Corporations. Journal of Political Economy. 1984;92(3):451–471.

67. Wang Z. Headquarters gravity: How multinationals shape international trade. Journal of International Economics. 2021;131(C).

68. Behrens K, Robert-Nicoud F. Agglomeration Theory with Heterogeneous Agents. In: Duranton G, Henderson JV, Strange WC, editors. Handbook of Regional and Urban Economics. vol. 5 of Handbook of Regional and Urban Economics. Elsevier; 2015. p. 171–245.

69. Keen MJ, Kotsogiannis C. Tax competition in federations and the welfare consequences of decentralization. Journal of Urban Economics. 2004;56(3):397–407.

70. Duranton G, Turner MA. Urban form and driving: Evidence from US cities. Journal of Urban Economics. 2018;108:170–191.

71. Hsieh CT, Moretti E. Housing Constraints and Spatial Misallocation. American Economic Journal: Macroeconomics. 2019;11(2):1–39.

72. Autor D, Dorn D, Katz LF, Patterson C, Van Reenen J. The Fall of the Labor Share and the Rise of Superstar Firms. The Quarterly Journal of Economics. 2020;135(2):645–709.

73. Eeckhout J, Guner N. Optimal Spatial Taxation: Are Big Cities too Small? CEMFI; 2017.

74. Hasan R, Jiang Y, Rafols RM. Place-based preferential tax policy and industrial development: Evidence from India's program on industrially backward districts. Journal of Development Economics. 2021;150(C).

75. Ju J, Ma H, Wang Z, Zhu X. Trade Wars and Industrial Policy along the Global Value Chains. Working paper; 2021.

76. Hasanov F, Cherif R. The Return of the Policy That Shall Not Be Named: Principles of Industrial Policy. International Monetary Fund; 2019. 2019/074.

77. Lane N. Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea. Working paper; 2021.

78. Kim M, Lee M, Shin Y. The Plant-Level View of an Industrial Policy: The Korean Heavy Industry Drive of 1973. National Bureau of Economic Research; 2021. 29252.

79. Hyun Y, Ravi S. Place-based Development: Evidence from Special Economic Zones in India; 2018. dp-306.

80. Lu Y, Wang J, Zhu L. Place-Based Policies, Creation, and Agglomeration Economies: Evidence from China's Economic Zone Program. American Economic Journal: Economic Policy. 2019;11(3):325–60.

81. Choi J, Levchenko AA. The Long-Term Effects of Industrial Policy. National Bureau of Economic Research; 2021. 29263.

82. Kline P, Moretti E. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. The Quarterly Journal of Economics. 2013;129(1):275–331.

83. Caliendo L, Parro F. Estimates of the Trade and Welfare Effects of NAFTA. Review of Economic Studies. 2015;82(1):1–44.

84. Allen T, Arkolakis C. Trade and the Topography of the Spatial Economy. The Quarterly Journal of Economics. 2014;129(3):1085–1140.

85. Costinot A, Rodríguez-Clare A. Trade Theory with Numbers: Quantifying the Consequences of Globalization. vol. 4. Elsevier; 2014. p. 197–261.

86. Bartelme DG, Costinot A, Donaldson D, Rodríguez-Clare A. The Textbook Case for Industrial Policy: Theory Meets Data. National Bureau of Economic Research; 2019. 26193.

87. Liu E. Industrial Policies in Production Networks. The Quarterly Journal of Economics. 2019;134(4):1883–1948.

88. Liu E, Ma S. Innovation Networks and Innovation Policy. National Bureau of Economic Research; 2021. 29607.

89. Carvalho VM, Nirei M, Saito YU, Tahbaz-Salehi A. Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. The Quarterly Journal of Economics. 2020;136(2):1255–1321.

90. Bakker JD. Trade and Agglomeration: Theory and Evidence from France. University College London; 2021.

91. Tian L. Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil. C.E.P.R. Discussion Papers; 2021.

92. Caliendo L, Parro F, Rossi-Hansberg E, Sarte PD. The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy. The Review of Economic Studies. 2017;85(4):2042–2096.

93. Antras P, Chor D, Fally T, Hillberry R. Measuring the Upstreamness of Production and Trade Flows. American Economic Review. 2012;102(3):412–16.

94. Eaton J, Kortum S. Technology, Geography, and Trade. Econometrica. 2002;70(5):1741–1779.

95. Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. The Network Origins of Aggregate Fluctuations. Econometrica. 2012;80(5):1977–2016.

96. Tombe T, Zhu X. Trade, Migration, and Productivity: A Quantitative Analysis of China. American Economic Review. 2019;109(5):1843–72.

97. Lashkaripour A, Lugovskyy V. Profits, Scale Economies, and the Gains from Trade and Industrial Policy. Indiana University; 2017.

98. Levinsohn J, Petrin A. Estimating Production Functions Using Inputs to Control for Unobservables. Review of Economic Studies. 2003;70(2):317–341.

99. Ackerberg DA, Caves K, Frazer G. Identification Properties of Recent Production Function Estimators. Econometrica. 2015;83(6):2411–2451.

100. Hulten CR. Growth Accounting with Intermediate Inputs. The Review of Economic Studies. 1978;45(3):511–518.

101. Carvalho VM, Tahbaz-Salehi A. Production Networks: A Primer. Faculty of Economics, University of Cambridge; 2018. 1856.

102. Aiginger K, Rodrik D. Rebirth of Industrial Policy and an Agenda for the Twenty-First Century. Journal of Industry, Competition and Trade. 2020; p. 1–19.

103. Che N. Intelligent Export Diversification: An Export Recommendation System with Machine Learning. IMF Working Papers. 2020;2020(175).

104. Hausmann R, Klinger B. The structure of the product space and the evolution of comparative advantage. CID Working Paper Series. 2007;.

105. Hidalgo CA, Hausmann R. The building blocks of economic complexity. Proceedings of the national academy of sciences. 2009;106(26):10570–10575.

106. Imbs J, Wacziarg R. Stages of diversification. American Economic Review. 2003;93(1):63–86.

107. Klinger B, Lederman D. Discovery and development: An empirical exploration of "new" products. The World Bank; 2004.

108. Klinger B, Lederman D. Export discoveries, diversification and barriers to entry. Economic Systems. 2011;35(1):64–83.

109. Cadot O, Carrère C, Strauss-Kahn V. Export diversification: what's behind the hump? Review of Economics and Statistics. 2011;93(2):590–605.

110. Feenstra R, Kee HL. Export variety and country productivity: Estimating the monopolistic competition model with endogenous productivity. Journal of international Economics. 2008;74(2):500–518.

111. IMF. Sustaining Long-Run Growth and Macroeconomic Stability in Low-Income Countries-The Role of Structural Transformation and Diversification. IMF Policy Paper. 2014;.

112. Structural change in exports and economic growth: cointegration and causality analysis for Spain (1961–2000);.

113. Herzer D, Nowak-Lehnmann D F. What does export diversification do for growth? An econometric analysis. Applied economics. 2006;38(15):1825–1838.

114. Lin JY, Li F. Development strategy, viability, and economic distortions in developing countries. The World Bank; 2009.

115. Frankel JA. The natural resource curse: a survey. National Bureau of Economic Research; 2010.

116. Balassa B. Trade Liberalisation and "Revealed" Comparative Advantage. The Machester School. 1965;33(2):99–123.

117. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Review of Economic Studies. 1991;58(2):277–297.