# Title
Essays on Human Capital



Author
Wang, Yang

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

**Essays on Human Capital**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Yang Wang

Committee in charge:

Professor Gordon Hanson, Co-Chair
Professor Karthik Muralidharan, Co-Chair
Professor David Lagakos
Professor Craig McIntosh
Professor Yiqing Xu

2020

The dissertation of Yang Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California San Diego

2020

DEDICATION

To my beloved parents and Jing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisors, Karthik Muralidharan, Gordon Hanson, David Lagakos, Craig McIntosh and Yiqing Xu, for their patient advice and constant encouragement. I am grateful to have the opportunity to learn from them, both academically and personally. Without their effort, I could not make the progress today. I also thank my friends for their help and support in the past five years. I benefited a lot from their suggestions, and I appreciated it very much.

I thank comments and suggestions from all seminar participants at UCSD. I thank Social Sciences Computing Facility (SSCF) at UCSD for computing resources, Michael Hatch for technical support, and Yue Chu for excellent research assistance.

VITA

| | |
|---|---|
| 2012 | B. A. in Economics, Tsinghua University |
| 2014 | M. A. in Economics, Tsinghua University |
| 2014-2020 | Teaching Assistant, University of California San Diego |
| 2020 | Ph. D. in Economics, University of California San Diego |

ABSTRACT OF THE DISSERTATION

**Essays on Human Capital**

by

Yang Wang

Doctor of Philosophy in Economics

University of California San Diego, 2020

Professor Gordon Hanson, Co-Chair
Professor Karthik Muralidharan, Co-Chair

This dissertation studies the role of human capital in urbanization and labor market in China.

Chapter 1 models and quantifies the importance of within-firm skill complementarity in explaining cross-city productivity gaps in China. I argue that skill complementarity is an important driver of skill concentration which augments these productivity gaps with agglomeration economies. I develop a spatial general equilibrium model that captures an economy inhabited by heterogeneous individuals who form production teams through assortative matching and sort across cities in these teams. I structurally estimate the model using firm-level census data.

Through counterfactual analysis, I find that within-firm skill complementarity accounts for 18% of cross-city productivity gaps in China. I further examine the general equilibrium effects of place-based policies: subsidizing skilled individuals to reside in second-tier cities. The simulated equilibrium shows local gains from such policies at the expense of other cities, suggesting an equity-efficiency trade-off in a spatial economy.

Chapter 2 estimates the income gains from migrating for jobs after graduation using survey data on college graduates. I apply propensity score matching and compare students who have similar propensity to move. I find 12-15% gains in starting salary from this geographic mobility. The effect does not vary significantly across family background and education. Further analysis on mechanisms suggests that the migration premium is mainly attributed to local agglomeration factors at the destination.

Chapter 3 turns to one type of human capital and studies the impact of retaking in English test on the labor market. I draw evidence from a national English test and exploit a manipulated regression discontinuity at the passing cutoff for certificates. While there is a positive relationship between scores and wages, I find a 10% jump in starting salary after graduation for those who barely pass the test and bunching just above the score threshold. Among students at risk of failing, retakers are positively selected in terms of abilities unrelated to English skills. Analysis from other job outcomes suggests that the wage gap at cutoff is associated with access to larger firms and state-owned firms.

# Chapter 1

# Skill Complementarity in Teams: Matching, Sorting and Agglomeration in China

## 1.1 Introduction

Urbanization has played a major role in powering economic growth among developing countries over the last few decades. The share of urban population surged from 29% in 1980 to over 50% among developing countries (United Nations, 2018). At the same time, skilled workers and productive firms are increasingly concentrating in large cities (Dingel et al., 2019). This internal integration has led to an uneven distribution of economic activities and is central to overall wage inequality in many countries.[1] While a large body of literature has been devoted to explain the productivity advantages in large cities, there is limited evidence on the relative

[1] See Rosenthal and Strange (2004) and Combes and Gobillon (2015) for reviews on empirical evidence of agglomeration. For inequality, see for example Moretti (2013), Baum-Snow and Pavan (2013) and Song et al. (2018) for US; Combes et al. (2008) for France; de la Roca (2017) for Spain; Rice et al. (2006) for UK and Chauvin et al. (2017) for China, Brazil and India.

importance of spatial sorting of skills and its underlying mechanisms.[2]

In this paper, I focus on one potential driver of skill concentration and productivity differences across cities: skill complementarity at the workplace. This idea of skill complementarity is closely related to knowledge spillovers (Marshall, 1890). A classic example is the O-Ring theory (Kremer, 1993), in which skilled workers team up to produce high-value outputs.[3] Generalizing the case to a spatial system of cities, it is then important to consider how coworker interdependencies shape the spatial allocation of skills. More importantly, if there are also city-level agglomeration externalities such as technology spillovers between more productive teams (firms), then skill complementarity could further augment the productivity gaps through locational sorting of heterogeneous teams. This paper addresses this mechanism by asking: what share of the observed cross-city productivity gaps is due to firm spatial sorting that rises from skill complementarity at the workplace?

After years of rapid urbanization and the expansion of higher education, China sets a live stage to examine this question. Over the past few decades, millions of people have been able to relocate across places in the interest of pursuing economic profit in larger cities. The skill concentration creates a great opportunity to study skill complementarity and its role in regional disparity.[4] Moreover, understanding skill concentration in China has crucial policy implications. Recognizing the brain drain, local governments have put in place a series of individual subsidies to attract skilled people to reside in specific cities since 2017. These place-based policies aim to influence the locational choice of young college graduates, but little is known about their impacts

[2]See Duranton and Puga (2004) for a review on agglomeration sources, which include learning (Moretti, 2004; de la Roca and Puga, 2017; Eckert et al., 2019b; Davis and Dingel, 2019), matching (Card et al., 2013; Dauth et al., 2019), specialization (Tian, 2019), consumer cities (Glaeser et al., 2001), networks (Arzaghi and Henderson, 2008), etc. Another large body of literature focuses on factor misallocation and potential gains from improved allocation of resources in developing countries. See Caselli (2005), Restuccia et al. (2008), Hsieh and Klenow (2009), Vollrath (2014), Restuccia and Rogerson (2017) and more recently Desmet et al. (2018), Lagakos et al. (2018), Bryan and Morten (2019), Hsieh and Moretti (2019), Fan (2019) and Tombe and Zhu (2019).

[3]Skill complementarity may also come from a knowledge-based economy (Garicano and Rossi-Hansberg, 2015). I do not distinguish the micro-foundation of skill complementarity within firms for simplicity.

[4]The share of urban population surged from 20% in 1980 to 60% in 2019. Despite hukou system, over 250 million people are considered to be migrant workers. Education has also been improved. The number of college students graduated in a year increased from 0.17 million in 1978 to 8.2 million in 2019.

on the aggregate economy by altering the allocation of skills.

This paper proposes a new heterogeneous agent model that nests team production into a spatial economy with agglomeration externalities. It is the first paper to structurally quantify the role of within-firm skill complementarity in driving cross-city productivity gaps in China. I argue that within-firm skill complementarity boosts individual productivity as a team and allows teams to locate in larger cities and gain more from agglomeration. I estimate the model sector by sector using restricted firm census data in China that contains unique information on firm-level skill mixes. After recovering key parameters regarding the strength of skill complementarity in each sector, I conduct counterfactual analysis to quantify the importance of skill complementarity mechanisms in driving productivity difference across cities. In the model based counterfactual analysis, I find that skill complementarity accounts for a substantial share of these gaps in China, about 18% of the observed productivity differences across cities. I also evaluate the effect of placed-based subsidies on college graduates, and find local gains but overall efficiency distortion. Overall, the results suggest that the government should be much more careful with policy interventions on skill allocation in a spatial economy.

To guide my investigation, I develop a spatial general equilibrium model that jointly determines the skill composition and location of firms in a monopolistic competitive market. The model captures an economy populated by a continuum of skill-heterogeneous individuals who self-select into occupations, match with others to establish production teams, and choose team locations at the same time.[5] A firm is then defined as a production team. Each firm has one manager who claims profit and many workers who earn type-specific wages. Crucially, there is strong skill complementarity between manager and workers within a team. In other words, a team becomes even more productive when a skilled manager teams up with skilled workers.

[5]In my context, the (assortative) *matching* means teaming between workers and managers. The (team) *sorting* means the locational choices made by teams. The (occupation) *selection* refers to the occupational choices between workers and managers. The term *matching* may refer to different processes in different studies. In early literature, matching may stand for random matches between firms and labor. When complementarity is assumed between city size and individual skills, matching may refer to the relation between cities and agents.

This within-team complementarity generates positive assortative matching between skills of two occupations, and along with other conditions, determines who becomes a manager, with whom to establish a team and where the team locates.

On top of this team structure, there are agglomeration externalities at the city-level. These between-team externalities include many benefits of large cities such as technology spillover (Ellison et al., 2010). I remain agnostic to the sources of these externalities and assume that high-productivity teams benefit more from larger city sizes.[6] At the same time, larger cities also have higher costs due to the congestion of population. The locational choice of teams is made by managers. Since team productivity depends on both manager skills and worker skills, heterogeneity in skill composition drives team to sort across space and generate cities of different sizes on a continuum of ex-ante identical sites. In addition, I embed a non-tradable sector that produces necessities to anchor the price level in each city. I also consider model extensions that include trade costs, discrete cities, idiosyncratic preferences, and imperfect sorting due to migration barriers.

In equilibrium, the model entails three key features in each free-trade sector. First, the set of skills available in each occupation is endogenously determined and characterizes by a unique skill cutoff as in Lucas (1978). Individuals with skills below the cutoff become workers, and those with skills above become managers. Second, there is a positive assortative matching between the skill type of managers and the skill type of workers. Each firm has one type of manager and one type of workers. Lastly, skilled managers and skilled workers co-locate in larger cities. Cities are collections of organized teams stemming from mutual best responses. In light of the team structure, larger cities tend to host more productive firms and have higher within-firm inequality. Cross-city skill premium and inequality can be similarly derived and positively associated with city size. I show that through the properties of assortative matching, the framework is useful to discuss the impact of overall skill distribution shocks on inequality and internal migration.

[6]See Ellison et al. (2010) for other reasons of industry coagglomeration including proximity to customers and suppliers, labor pooling and natural advantages.

Given the population, the equilibrium is unique in terms of the city size distribution and the mass (number) of cities but not in sector composition or which site becomes occupied.

The model highlights the way within-firm skill complementarity shapes the spatial allocation of human capital and the size distribution of cities at the same time. The setup unifies individual and firm sorting as in Behrens et al. (2014) but extends efficiency units formulation of labor skills to heterogeneous workers who match with heterogeneous managers in the labor market. Moreover, I adopt tools from the assignment model to govern the skill matching (Costinot and Vogel, 2010; Sampson, 2014). Instead of exogenous domains on each side of the matching, I endogenize the skill set via a simple Roy model. This framework allows me to sink the analysis from city-level complementarity (Eeckhout et al., 2014) to firm-level while replicating stylized cross-city facts on price levels, income, and inequality. The fact that the city size distribution feeds back to the occupational cutoff and hence the matching patterns illuminates the trade-off between "the first in a village" and "second in Rome". It characterizes a much more realistic case that differs from many previous studies that sequentially isolate occupational selection and spatial sorting. Skill concentration in large cities is thus the results of anticipated teamwork rather than the coincidence of independent sorting. I take the model to restricted firm census data in China covering the universe of all legal entities in the manufacturing and service sectors during 2008. The dataset is ideal because it provides the educational composition of employees at the firm-level. For each firm in service sectors, all employees are categorized into one of five education levels. The span of education allows me to approximate skill mixes in the model. In the main estimation, I use two tails of education distribution in each firm to capture team skill composition. I document supporting evidence of positive assortative matching across two parallel dimensions: between cities and within cities. I find that for most sectors, the average education at both the right and left tails of the education distribution are increasing in city size (co-location). At the same time, firm-level regressions indicate that the positive correlation between left and right tails of firm skill mixes holds not only across but also within cities (co-work). This result is robust after controlling for firm characteristics, using finer sectors, counties, and alternative proxies for skill mixes.

I structurally estimate the model sector by sector through the simulated method of moments (McFadden, 1989). I parameterize an extended version of the model which incorporates idiosyncratic shocks and traditional agglomeration of density in a discrete set of cities. I calibrate eight parameters for each sector. These parameters represent the strength of agglomeration forces, the degree of complementarity, the shape of matching function, and the variances of occupational skill distribution and idiosyncratic shocks. The key premise of the parameters' identification is that the strength of skill complementarity determines the co-location pattern of skills across cities. Intuitively, given agglomeration forces, a small difference in manager skill will lead to a large difference in his locational choice when skill complementarity is strong and the matching function is steep. Similarly, the sensitivity of city size to the skill of workers should inform identical values of parameters. Since team productivity can be expressed in skills from either occupation, parameters within teams are jointly identified.

This paper is the first one to structurally quantify the importance of firm sorting and within-firm skill complementarity in productivity gaps across cities in China. Using the calibrated model, I conduct counterfactual analysis to measure the share of sorting and within-firm skill complementarity in the productivity gaps. I estimate the changes in the elasticity of firm productivity to city size when systematic sorting of firms is hypothetically shut off. The estimated contribution of firm sorting, around 33%, is smaller than the 50% estimated in Gaubert (2018) among French firms. I then deepen the analysis by quantifying the importance of skill complementarity at the workplace. I turn on manager sorting but keep skill complementarity off so that there is no assortative matching. The difference between the above two scenarios informs me of the contribution of skill complementarity in spatial sorting. I find that within-firm complementarity accounts for about half of the sorting mechanisms, or equivalently 18% of the overall firm productivity gaps observed across cities. The results suggest that within-firm skill complementarity is an important driver of regional disparity.

Finally, I make the first attempt to analyze the general equilibrium effects of trending place-based policies in China: subsidizing young talents for residing in second-tier cities. To boost the local economy, city governments in China started a race to lure educated individuals through cash bonuses, housing subsidies, tax deductions, etc. As a reference point, I evaluate the effect of a 25% income subsidy to new college graduates for residing in 32 middle-sized cities in China. The simulation shows positive local gains at the expense of other cities. Compared to the case without subsidies, the number of incoming firms increases by 1.3% and the corresponding new employment increases by 2.5%. Besides immediate local effects, the simulated equilibrium shows a non-linear welfare loss as subsidies increase. More specifically, while there is nearly zero impact with a 10% subsidy, a 25% subsidy leads to a 0.77% decrease in the aggregate output among the entire new employment. The distributional effect suggests that the local gains from such policies should be carefully reviewed against the expenditure and efficiency loss from displacement of firms across the country. With falling migration barriers, the central government may face a stronger trade-off between overall efficiency and regional equity.

**Related literature.** This paper relates to several strands of literature. First, it contributes to the ongoing work studying agglomeration with sorting and selection of heterogeneous agents (Baldwin and Okubo, 2006; Combes et al., 2012; Eeckhout et al., 2014; Behrens et al., 2014; Davis and Dingel, 2017, 2019; Diamond, 2016; Gaubert, 2018).[7] Using French firm data, Gaubert (2018) develops a model of firm sorting based on complementarity between firm productivity and city size. It estimates that sorting accounts for about 50% of the firm productivity gap across commuting zones. However, it remains silent on the sources of such agglomeration. In contrast, Davis and Dingel (2019) models cities as the places of costly idea exchanges and fully attributes individual sorting to city-wide knowledge spillover. My paper stands between their work and

[7] See Behrens and Robert-Nicoud (2015) for a review on this topic. Conversely, some studies attempt to control for individual fixed effects and estimate the wage premium caused by higher density, e.g., Gould (2007), Glaeser and Maré (2015) and Combes et al. (2008).

dives into one particular mechanism while remaining open to other mechanisms. In terms of occupational selection, Behrens and Robert-Nicoud (2014) builds a model in which agents sort into cities based on their innate talent, followed by ex-post random serendipity shocks that decide who establish firms as entrepreneurs in a closed economy. In contrast, I study a more realistic case where occupation selection and sorting are simultaneously determined in a tradable market, emphasizing the general equilibrium aspects of sorting and selection with heterogeneous agents.[8]

The team structure also relates to the urban literature on how agents organize production in cities. It parallels to work studying the internal organization of firms and its implications on agglomeration such as hierarchical layers (Garicano and Rossi-Hansberg, 2015; Santamaría, 2019; Spanos, 2019), fragmentation (Rossi-Hansberg et al., 2009), labor division (Tian, 2019), etc. I explicitly formulate team production with occupation selection by comparative advantage – following Lucas (1978), Jovanovic (1994) and more recently Eeckhout and Jovanovic (2012) – and at the same time, adopt the assignment model to match heterogeneous managers with heterogeneous workers.[9] The insight of teamwork builds on a large body of literature on human capital externalities at various levels.[10] In this regard, this paper is a generalization of the O-Ring theory (Kremer, 1993) to heterogeneous team composition in a spatial system.

The paper complements with studies on the spatial aspect of wage inequality (Baum-Snow and Pavan, 2013; Song et al., 2018; Baum-Snow et al., 2018; Eckert et al., 2019a) and more broadly the geographic distribution of economic activities (Allen and Arkolakis, 2014; Ahlfeldt et al., 2015). It provides a framework for emerging empirical work on firm-worker assortative matching (Andersson et al., 2007; Mion and Naticchioni, 2009, Card et al., 2013, Dauth et al., 2019). The model depicts assortative matching as an intrinsic driver of inequality and offers

[8]There are two types of selection: individual occupation selection and firm exit selection. My paper focuses on the former one. See Baldwin and Okubo (2006) and Combes et al. (2012) for the latter type of selection.

[9]The assignment model follows the line of theoretical work by Becker (1973), Sattinger (1975), and recently Chade et al. (2017), Eeckhout and Kircher (2018). It is also widely used in trade literature, e.g., Antràs et al. (2006); Costinot (2009), Sampson (2014), Costinot and Vogel (2015), Grossman et al. (2016).

[10]See, for example, Rauch (1993), Moretti (2004), Iranzo et al. (2008), Mas and Moretti (2009), Glaeser and Maré (2015), Cornelissen et al. (2017), Neffke (2017), Jarosch et al. (2019), Serafinelli (2019).

plausible explanations for the recent decline of middle-skill jobs in larger cities and falling urban wage premium for low-skill workers documented by Autor (2019).

The paper also adds to the empirics of agglomeration economy in developing countries (Bryan et al., 2019). In the context of China, it is relevant to studies on the spillover effect of human capital on regional economic growth (Au and Henderson, 2006a; Fleisher et al., 2010; Liang and Lu, 2017; Glaeser and Lu, 2018). By quantifying the importance of spatial sorting, it also echoes to the misallocation of factors and potential gains from reducing migration and trade barriers (Donaldson, 2018; Lagakos et al., 2018; Bryan and Morten, 2019), particularly in China (Au and Henderson, 2006b; Hsieh and Klenow, 2009; Bosker et al., 2012; Fan, 2019; Ma and Tang, 2019; Tombe and Zhu, 2019).

The remainder of the paper is organized as follows. Section 2 develops a structural spatial model for the joint determination of skill composition and spatial location of firms. Section 3 introduces the data and documents supporting evidence on assortative matching. Section 4 presents the structural estimation. Section 5 shows the quantification results. Section 6 conducts analysis of place-based subsidizing policies in China. Section 7 concludes.

## 1.2 A Spatial Model of Team Matching of Heterogeneous Individuals

This section presents a spatial general equilibrium model, in which heterogeneous individuals select into occupations and form production teams through assortative matching. Teams, or equivalently, firms sort across cities trading off between agglomeration and congestion forces. The framework shares many features of Behrens et al. (2014), Gaubert (2018), Davis and Dingel (2019) and assortative matching literature reviewed by Costinot and Vogel (2015) and Eeckhout (2018).

### 1.2.1 Environment Setup

The economy consists of a continuum of individuals of mass $\mathbb{L}$, a discrete number of $S$ tradable sectors and one non-tradable ($n$) local sector. Individuals pay a sunk cost and draw a sector $s$ and a single dimensional skill level $\alpha$ from a prior distribution $\mu_s(\cdot)$, which is continuously distributed on a bounded interval $\mathbb{A}_s$. The model abstracts away from endogenous choice of sectors and take each sector population $\mathbb{L}_s$ as given.[11]

There is a continuum of homogeneous sites index by $j$ that may become cities with endogenous population and skill composition. Sites are ex-ante identical in fundamentals and cities emerge under self-organization. I allow for the possibility of a non-integer number of cities of any given size. Therefore, the mass (number) of cities are also endogenously determined in equilibrium. Since there is no amenity differences, the equilibrium do not predict which sites are occupied.[12]

In the economy, a firm is a two-layer team consisted of one manager and many workers. The production location is decided by the manager. Throughout the model, I assume complete information and no friction in searching and matching. For the baseline model, I also assume free mobility and costless trade in tradable sectors. I then extend the model to discrete cities, idiosyncratic locational preferences and migration barriers. I also provide extendable framework to costly trade in the appendix.

### 1.2.2 Preferences

Agents derive utility from consumption of a bundle of freely traded final goods and a fixed per person amount $\bar{n}$ of non-tradable service, which is strict necessities. The indirect utility

[11] Selection into sectors can be either modeled in a Roy model with multiple sectors, e.g., Hsieh et al. (2019), or directly reflected in the exogenous skill distributions of each sector. Since sector selection is not the focus of this paper and does not affect the insights, I take skill distribution and population in each sector as exogenous given.

[12] Amenity difference is important factor in locational choice of individuals. Future work would incorporate such difference.

of a person with income $I$ living in city $j$ is then

$$V_j = I(j) - p_j^n \overline{n}, \tag{1.1}$$

where $p_j^n$ is the endogenously determined price of non-tradable goods and services $n$ at $j$.

Individuals spend the rest of income on a final bundle $y$ of tradable goods, which is used as the numéraire. The final bundle $y$ is a Cobb-Douglas combination of aggregated tradable goods over $s = 1, ..., S$ sectors

$$y = \prod_{s=1}^{S} y_s^{\eta_s} \quad \text{with} \quad \sum_{1}^{S} \eta_s = 1.$$

Within each sector, individuals have a Dixit-Stiglitz preference over a bundle of tradable varieties

$$y_s = \left[ \int x_s(\omega)^{\rho_s} \, \mathrm{d}\omega \right]^{1/\rho_s}, 0 < \rho_s < 1.$$

Since intermediate goods are freely traded across cities, the final bundle $y$ has identical price index

$$P = \left[ \prod_{s=1}^{S} \left( \frac{P_s}{\eta_s} \right)^{-\eta_s} \right]^{-1} \equiv 1 \quad \text{and} \quad P_s = \left[ \int p_s(\omega)^{1-\sigma_s} \, \mathrm{d}\omega \right]^{\frac{1}{1-\sigma_s}},$$

where $\sigma_s \equiv \frac{1}{1-\rho_s} > 1$ is the sectorial elasticity of substitution. The demand for variety $\omega$ in sector $s$ is

$$x_s(\omega) = \frac{R_s}{P_s} \left[ \frac{p_s(\omega)}{P_s} \right]^{-\sigma_s},$$

where $R_s$ is the aggregate spending on goods from sector $s$.

### 1.2.3 Production

An individual endowed with skill in sector $s$ may produce tradable goods in $s$ or opt out for non-tradable production. Non-tradables are produced individually with uniform productivity regardless of initial sector and skill. Tradable goods, on the other hand, are produced by teams leveraging their skills. Each team is a firm that requires one manager (entrepreneur) as the founder and owner, and many workers who earn type-specific wages. Therefore, the mass of managers equals the mass of firms and varieties in tradable sectors.

In addition to labor, the only fixed cost of production in each $S+1$ sector is one unit of land. The land is purchased by managers in tradable sectors (firm owners) and individuals in non-tradable sector (self-employed). I denote by $c(L)$ the cost of land in a city of size $L$ and assume it is in the form of

$$c(L_j) = \gamma L_j^{\phi}, \tag{1.2}$$

with constant parameters $\gamma$ and $\phi$ across sectors. This is a common form of rent and commuting disutility from a mono-centric inner city structure.[13] The land revenue is fully spent back on the tradables in the economy by atomic landowners out of the sorting system.

**Non-Tradables Sector.** The market of non-tradable goods is perfectly competitive. One unit of labor input leads to one (normalized) unit of non-tradable output.[14] The income of an individual in non-tradable business in city $j$ is thus $p_j^n$ regardless of skill level. With perfect mobility, utilities are equalized across all occupied cities $j$:

$$p_j^n - \gamma L_j^{\phi} - \bar{n} p_j^n = \bar{V}_n,$$

where $\bar{V}_n$ is a constant utility level to be determined in equilibrium. Then the price can be written as

$$p_j^n = \frac{\bar{V}_n + \gamma L_j^{\phi}}{1-\bar{n}}. \tag{1.3}$$

[13] See the appendix of Behrens et al. (2014) or Davis and Dingel (2019) for the micro foundation of this form, which is derived from a standard model of the internal structure of a monocentric city in which commuting costs increase with population size as governed by the technological parameters $\gamma$ and $\phi$.

[14] The uniform productivity can be extended to the case with density agglomeration so the utility becomes: $p_j^n L_j^e - \gamma L_j^{\phi} - \bar{n} p_j^n$, where $L_j^e$ is the externality from higher density. It however does not change the key insight of the model.

This expression is consistent with the fact that larger cities have higher price of local non-tradable goods and services.

**Tradable Sectors.** Each team in a given sector engages in monopolistic competition and produces exact one variety of goods in their own sector. There are two occupations in tradable sectors: managers and workers. The firm productivity depends on manager's skill type, workers' skill type, sector-specific agglomeration externality and city size. Denote the skill of a representative manager by $\alpha$ and the skill of workers by $z$. The team production function is given by

$$x_s(\alpha, z, L) = \int_{z \in \mathbb{A}_s} f_s(\alpha, z) g(c_s, L) l_s(z) \, \mathrm{d}z,$$

where $x_s(\alpha, z, L)$ is the total output of team. $f_s(\cdot, \cdot)$ is a sector-specific team productivity that depends on the skill type of manager $\alpha$ and workers' skill type $z$ chosen from the skill support $\mathbb{A}_s$. $l_s(z)$ is the quantity of type $z$ labor hired. Outputs by different types of workers are perfect substitutes, whereas labor of different types are non-substitutable.

Productivity is also affected by a function $g$ that governs agglomeration externality. It contains a sector-specific agglomeration strength indexed by $c_s$ and city total population $L$.[15] I further assume $f$ and $g$ are multiplicatively separable, which plays a key role to facilitate the analysis below. By doing so, I assume the agglomeration externality works through total outputs rather than individuals. I acknowledge different sources of agglomeration at the city level such as innovation spillover, natural advantages and better business environment. I remain agnostic to them throughout the paper and incorporate them into sector-specific agglomeration $g_s(L)$.

The key assumption of the model is that the skill of manager and the skill of workers exhibit strong complementarity.

**Assumption 1.** ***(Strong Skill Complementarity)*** *The team productivity $f_s(\cdot, \cdot)$ and sector agglom-*

[15]In this paper, the agglomeration force depends on the total population of the city. It can be extended to be sector-dependent or education-dependent. It would not affect the insight of model but makes the strength and form of agglomeration complicated.

*eration $g(\cdot,\cdot)$, are twice continuously differentiable, strictly increasing and log-supermodular in all arguments:*

$$\frac{\partial^2 \log f_s(\alpha,z)}{\partial\alpha\partial z} > 0; \quad \frac{\partial^2 \log g(c_s,L)}{\partial c_s \partial L} \geq 0.$$

The first inequality states strictly log-supermodularity between manager skill and worker skill, whereas the second one allows sectors of higher productivity $c_s$ to weakly benefit more from agglomeration externality. This assumption is motivated by both theoretical development in the organization of production (Kremer, 1993; Garicano and Rossi-Hansberg, 2006) and empirical evidence of peer effect (Moretti, 2004; Mas and Moretti, 2009; Cornelissen et al., 2017; Jarosch et al., 2019). This strong complementarity between two occupations is the foundation of the model that generates assortative matching.

**Manager's Problem.** Given a sector $s$, a manager of skill $\alpha$ chooses the type and quantity of workers, and the production location indexed by city size $L$ to maximize utility. One manager is limited to one physical location with no offshoring or fragmentation.

Managers could potentially hire multiple worker types, but it is not optimal to do so because of the log supermodularity and perfect substitution of outputs produced by different workers. Under assumption 1, I show that given city size and sector, a manager hires only one particular type of labor for any upward-sloping wage schedule. The procedure is standard as in Costinot and Vogel (2010), Sampson (2014) and Eeckhout and Kircher (2018).

Let $Q_s = P_s^{\sigma_s} y_s$, so variety demand $x_s(\omega) = Q_s p_s(\omega)^{-\sigma_s}$. Let $w_s(z)$ be the wage schedule of type $z$ workers in sector $s$, then operating profit for a manager of skill $\alpha$ conditioning on city of size $L$ is

$$\pi_s(\omega) = Q_s^{1/\sigma_s}\left(\int_{z\in\mathbb{A}_s} f_s(\alpha,z)g_s(L)l_s(z)\,\mathrm{d}z\right)^{\frac{\sigma_s-1}{\sigma_s}} - \int_{z\in\mathbb{A}_s} w_s(z)l_s(z)\,\mathrm{d}z.$$

First order condition on $l(z)$ implies

$$\frac{\sigma_s - 1}{\sigma_s} Q_s^{1/\sigma_s} x_s^{-1/\sigma_s} f_s(\alpha, z) g_s(L) - w_s(z) \leq 0,$$

with equality holds when $l_s(z) > 0$. Each worker's marginal revenue product inherits the log supermodularity. So higher $\alpha$ managers are willing to pay more for higher worker type $z$. Therefore, only one type of $z$ is hired by each firm.

### 1.2.4 Matching in Teams

For expositional purposes, in this subsection I study the matching problem taking the skill set in each occupation as given. Furthermore, I assume both sets of managers and workers are convex and bounded on an interval. Given the fact that exactly one type of workers is hired, the managers' problem, conditional on city size, is to choose labor (type and quantity) and set prices to maximize operating profit, taking wage schedule as given.

Consider a manager of skill $\alpha$ in sector $s$ in a city of size $L$. Market structure implies that the optimal price is a constant markup over marginal cost for the team, $p_s(i) = \frac{w_s(z)}{\rho_s f_s(\alpha,z) g_s(L)}$. The operating profit of hiring $z$ type workers conditional on city size $L$ is

$$\pi_{is}(\alpha, z|L) = p_s(i) f_s(\alpha, z) g_s(L) l_s(z) - w_s(z) l_s(z) = \Lambda_s Q_s \left[ \frac{f_s(\alpha, z) g_s(L)}{w_s(z)} \right]^{\sigma_s - 1}, \tag{1.4}$$

where $\Lambda_s = \frac{(\sigma_s - 1)^{\sigma_s - 1}}{\sigma_s^{\sigma_s}}$ is a sector specific constant, and $Q_s$ defined above is the sector specific market size taken as given by individuals. This is a standard result in a monopolistic competitive market except productivity now depends on types of both managers and workers, and the wages are type-specific. Equation (4) also implies that the city agglomeration is loaded on managers.

Given the wage schedule, a manager chooses worker type $z$ to maximize above profit.

Taking first order condition with respect to $z$ yields the following constraint:

$$\frac{w_s'(z)}{w_s(z)} = \frac{f_2(\alpha, z)}{f(\alpha, z)}. \tag{1.5}$$

From now on, I suppress the subscript $s$ in $f_s(\alpha, z)$ since all results hold within sectors. The implicit solution of $z$ delivers a matching assignment $M_s(\alpha)$, mapping from manager type $\alpha$ to workers' type. Notice that $L$ does not enter equation (5) due to multiplicativity of $f$ and $g$. In other words, managers' choice of worker type is not directly affected by city size and the matching function is global rather than city specific. This view differs from benefit of labor pooling in large cities where the quality or chance of matching gets improved, but produces equivalent observed outcomes. However, it does not mean that matching and city sizes are independent choices as the size distribution of cities also affects the skill set in each occupation. Before introducing occupation selection, I characterize the patterns of matching in equilibrium.

**Lemma 1.** ***(Positive Assortative Matching)*** *In any sector, equilibrium matching function $M_s(\alpha)$ is a bijection, strictly increasing in $\alpha$. For any $\alpha > \alpha'$, $M_s(\alpha) > M_s(\alpha')$.*

*Proof.* See the appendix. □

Lemma 1 is a direct result of Assumption 1 using classic theorem in monotone comparative statics (Topkis, 1978). With strong complementarity, more skilled managers hire more skilled workers. The optimization of $z$ depends on underlying distribution $\mu_s(\alpha)$ but not $Q_s$ or $\sigma_s$. Due to invertibility of $M_s(\cdot)$, I denote by $M_s^{-1}(\alpha)$ the type of matched manager for $\alpha$-workers. Both $M_s$ and $M_s^{-1}$ are strictly increasing in $\alpha$. In addition, it is easy to show that wage function is continuous and differentiable. In equilibrium, wage function is closely related to matching function by the first order condition:

$$\frac{w_s'(\alpha)}{w_s(\alpha)} = \frac{f_2(M_s^{-1}(\alpha), \alpha)}{f(M_s^{-1}(\alpha), \alpha)}. \tag{1.6}$$

Any changes in wage function will be accompanied by changes in matching. Intuitively, the quality of matching determines the marginal productivity therefore increases the return to skill. This further implies that wage inequality are closely related to matching quality. By log-supermodularity, the elasticity of wage function $w_s(\alpha)$, i.e., return to skill in percentage terms, is increasing in the matched skill of managers

$$\frac{\partial \frac{\alpha w_s'(\alpha)}{w_s(\alpha)}}{\partial M_s^{-1}(\alpha)} = \alpha \frac{\partial^2 \ln f(M_s^{-1}(\alpha), \alpha)}{\partial \alpha \partial M_s^{-1}(\alpha)} > 0.$$

I could now write net profit $\tilde{\pi}_s$ (operating profit net of land cost) and labor demand $l_s$ as a function of manager type $\alpha$ using matching function.

$$\tilde{\pi}_s(\alpha, M_s(\alpha), L) = \Lambda_s Q_s \left[\frac{f(\alpha, M_s(\alpha)) g_s(L)}{w_s(M_s(\alpha))}\right]^{\sigma_s - 1} - c(L), \tag{1.7}$$

$$l_s(\alpha, M_s(\alpha), L) = \frac{(\sigma_s - 1)\Lambda_s Q_s}{w_s(M_s(\alpha))} \left[\frac{f(\alpha, M_s(\alpha)) g_s(L)}{w_s(M_s(\alpha))}\right]^{\sigma_s - 1} \propto \frac{\pi(\alpha, M_s(\alpha), L)}{w(M_\alpha)}. \tag{1.8}$$

Using equation (2) and (3) to substitute necessities, manager's utility can be rewritten as

$$V_s(\alpha, M_s(\alpha), L|\text{manager}) = \tilde{\pi}_s - p_j^n \bar{n} = \Lambda_s Q_s \left[\frac{f(\alpha, M_s(\alpha)) g_s(L)}{w_s(M_s(\alpha))}\right]^{\sigma_s - 1} - \bar{\gamma} L^{\phi} - \frac{\overline{V}_n \bar{n}}{1 - \bar{n}}, \tag{1.9}$$

where $\bar{\gamma} = \frac{\gamma}{1-\bar{n}}$ is a constant and $\overline{V}_n$ is the equalized utility in non-tradable sector taken as given. Notice $V_s$ is supermodular in manager skill $\alpha$ and city size $L$. Under regularity conditions so that the problem is concave and well defined, there is a unique city size that maximize utility,[16]

$$L_s(\alpha) = L_s(\alpha, M_s(\alpha)) = \arg\max_L V_s(\alpha, M_s(\alpha), L). \tag{1.10}$$

**Lemma 2.** *In equilibrium, $L_s(\alpha, M_s(\alpha))$ is non-decreasing in $\alpha$ within each sector. There is a*

[16]For the problem to be well-defined, the positive effects of agglomeration should not be too strong compared to the congestion forces. Later with functional form of $g(c_s, L)$, the condition is $c_s(\sigma_s - 1) - \phi < 0$.

*positive sorting between teams and city size.*

*Proof.* The results follow from the maximization problem. □

In equilibrium, workers of type $M_s(\alpha)$ are allocated towards managers' optimal location choice $L_s(\alpha)$. On one hand, it can be regarded as workers simply obeying the choices made by managers. On the other hand, the procedure can be viewed as a sequential game where managers move first therefore workers' payoff from locating in other cities is negative because of zero demand. Both views leads to co-location.[17]

Lemma 1 and 2 summarize the pattern of managers' decision conditional on fixed skill set in each occupation. Leveraging skill complementarity, high skill managers hire high skill workers and are able to afford higher cost in larger cities. Again, the optimal city size depends on both manager type and worker type, and the optimal worker type depends on the distribution and set of skills available.

### 1.2.5 Occupational Choice

I now endogenize occupational choices. In equilibrium, one skill type may be either managers or workers in a sector. Formally, utilities for a team of type $\{\alpha, M_s(\alpha)\}$ are

$$V_s(\alpha|\text{manager}) = \Lambda_s Q_s \left[\frac{f(\alpha, M_s(\alpha))g_s(L_s(\alpha, M_s(\alpha)))}{w_s(M_s(\alpha))}\right]^{\sigma_s - 1} - \bar{\gamma} L_s^{\phi}(\alpha) - \frac{\overline{V}_n \bar{n}}{1-\bar{n}}, \tag{1.11}$$

$$V_s(M_s(\alpha)|\text{worker}) = w_s(M_s(\alpha)) - \bar{n}\bar{\gamma} L_s^{\phi}(\alpha) - \frac{\overline{V}_n \bar{n}}{1-\bar{n}}. \tag{1.12}$$

**Selection into Non-Tradable Sector.** Recall that utilities in the non-tradable sectors are constant. Therefore, labor in non-tradable sectors is selected from the bottom of skill distribution. In equilibrium, each city has $\bar{n}$ share of population as non-tradable producers. However, there is

[17] In reality, there is also a scenario where workers have the strong power to shape the location of managers. But since the model is static, it does not feature this bargaining process.

an indeterminacy of how each tradable sector contributes to the non-tradable sector. For instance, in a simple world with two identical tradable sectors, as long as the sum of non-tradable producer is fixed, it is flexible in terms of the share contributed by each sector. Equilibrium in each sector thus varies by the quota. There is also a trade-off between total economic profit and equity. To simplify the case, the following assumption is made to pin down a unique composition of non-tradable labor without modeling trade-offs between sectors.[18]

**Assumption 2.** ***(Rules of Non-Tradable Labor Division)*** *The sectorial share of labor in non-tradable production equals the weight of sector population.*

The above rule of assembling non-tradable labor from various tradable sectors is equivalent to sector isolation. In other words, sectors feed on non-tradables produced by their own people. It can also be viewed as the result of repeated random draws from total population with intra-sector but not between-sector bargaining. Assumption 2 implies that for each sector, selection into non-tradable sector is governed by a cutoff $\underline{\alpha}_s$, which is the fixed percentile of sector skill distribution.

$$\overline{n} = \int_0^{\underline{\alpha}_s} \mu_s(\alpha)\,\mathrm{d}\alpha. \tag{1.13}$$

**Lemma 3.** *Under assumption 2, selection into non-tradable services is dictated by a cutoff $\underline{\alpha}_s$ such that individuals with skill below $\underline{\alpha}_s$ choose non-tradable services.*

*Proof.* The results follow from the assumption that non-tradable production could not leverage skills. Therefore high skill workers have comparative advantages in staying in tradable sectors. □

The model assumes a common $\overline{n}$ across sectors for a fact that there is no clear relationship between MSA size and the share of self-employed as documented in Behrens and Robert-Nicoud (2015) for US. It represents each person's consumption of non-tradable local services in the total

[18]This is not an issue in a single tradable sector scenario as in Davis and Dingel (2019).

output of non-tradable service. In addition, Eeckhout et al. (2014) documented thicker tails in skill distribution in larger US MSAs and use city-level extreme-skill complementarity to generate that prediction. Here I tailor to this particular feature through non-tradable sector. In equilibrium, each city will host some low skill people producing to support the city.

Given the lower bound of skill in tradable sectors, I now turn to occupation choice in tradable sectors. With single dimension in workers and managers skill, occupation selection is not obvious if trade-off between occupations also depends on other endogenous outcomes. In my case, the optimal location choices feed back to individual's utility through the cost of living and congestion. To establish clear comparative advantages, I assume the production function satisfies the following assumption.

**Assumption 3.** ***(Occupational Comparative Advantages)*** *For any vector* $(\alpha, z, \alpha')$ *in the domain of skill support,*

*1) $f$ satisfies* $\frac{f_1(\alpha,z)}{f(\alpha,z)} > \frac{f_2(\alpha',\alpha)}{f(\alpha',\alpha)}$.

*2) The necessity share $\overline{n}$ is sufficiently small. Specifically,* $\overline{n} < \frac{L^{\phi}(z)}{L^{\phi}(\alpha)}$.

The first part of Assumption 3 imposes sufficiently strong asymmetry between occupations rather than within occupations. This condition resembles Spence-Mirrlees condition that gives comparative advantages to skilled people and separates individuals into two convex sets if all individuals are located in the same city. But it is not sufficient to generate comparative advantages to skilled people when location choices are endogenous and differs on two sides of the cutoff. To that end, the second part is needed for a sufficient condition that support a unique threshold. Intuitively, it requires that the share of necessities should not exceed the relative cost of living between any two cities so that utility changes from relocating in different cities do not overwhelm the single-crossing condition.

**Lemma 4.** ***(Occupation Segregation)*** *Under Assumptions 3, occupation selection in each tradable sector is dictated by a cutoff $\hat{\alpha}_s$ such that individuals with skill above $\hat{\alpha}_s$ become managers and those with skill below $\hat{\alpha}_s$ become workers.*

*Proof.* See the appendix. $\square$

Lemma 4 is the results of Roy-model selection as in Lucas (1978). It should be noted that Assumption 3 is a strong assumption that are sufficient but globally not necessary for Lemma 4. Nevertheless, it is a reasonable assumption in practice accounting for organization hierarchy (Garicano and Rossi-Hansberg, 2015) and importance of management (Bloom et al. 2012). By sharing a unique global occupation cutoff in each tradable sector, Lemma 4 is consistent with recent arguments that selection into entrepreneurs are less evident conditional on sorting (Combes et al., 2012; Behrens et al., 2014).

Lemma 3 and lemma 4 together characterize the structure of occupational skill sets and provide boundary conditions for team matching: $M_s(\overline{\alpha}_s) = \hat{\alpha}_s$ and $M_s(\hat{\alpha}_s) = \underline{\alpha}_s$. The cutoffs then must satisfy indifference conditions between two occupations (for marginal type $\hat{\alpha}_s$),

$$\tilde{\pi}_s(\hat{\alpha}_s, \underline{\alpha}_s, L_s(\hat{\alpha}_s)) - \overline{n}p^n(L_s(\hat{\alpha}_s)) = w_s(\hat{\alpha}_s) - \overline{n}p^n(L_s(\overline{\alpha}_s)), \tag{1.14}$$

and between non-tradable and tradable sectors (for marginal type $\underline{\alpha}_s$),

$$p_j^n(L_j)(1-\overline{n}) - \gamma L_j^{\phi} = \overline{V}_n = w_s(\underline{\alpha}_s) - \overline{n}\gamma L_s^{\phi}(\hat{\alpha}_s) - \frac{\overline{V}_n \overline{n}}{1-\overline{n}}. \tag{1.15}$$

Equation (14) and (15) are important linkage in the system connecting cross-city prices, starting wages and occupational choices across sectors. These cutoffs reflect relative toughness of being a manager. They also determine the range of cities a sector appears. Intuitively, a high $\underline{\alpha}_s$ and $\hat{\alpha}_s$ increase the wage of entry level workers in the industry and the lower bound of city size this sector shows up. Similarly, $\hat{\alpha}_s$ and $\overline{\alpha}_s$ are related to the upper bound of city size of a sector.

These two equations are the crucial differences between this model and traditional occupation selection with no spatial structure. In particular, equation (14) implies indifference between being "the first in village" and "the second in Rome" where location and occupation are jointly determined. The framework deals with a more general case and links occupational selection and spatial sorting through assortative matching and comparative advantages.

To close the model, the equilibrium is finally pinned down by market clearing conditions. The aggregate supply of workers must equal to the aggregate mass of workers demanded by managers type by type. In aggregation, for any manager type $\alpha \in (\hat{\alpha}_s, \overline{\alpha}_s)$

$$\int_{\underline{\alpha}_s}^{M_s(\alpha)} \mu_s(z)\,\mathrm{d}z = \int_{\hat{\alpha}_s}^{\alpha} l_s(M_s(z);z)\mu_s(z)\,\mathrm{d}z,$$

where team size $l_s$ is also endogenously determined. Differentiate the equation with respect to $\alpha$ and plug in $l_s(M_\alpha;\alpha)$. The spacing of skill, i.e., $M_s'(\alpha)$, can be written as,

$$M_s'(\alpha) = \frac{(\sigma_s - 1)\Lambda_s Q_s}{w_s(M_s(\alpha))}\left[\frac{f(\alpha, M_s(\alpha))g_s(L_s(\alpha, M_s(\alpha)))}{w_s(M_s(\alpha))}\right]^{\sigma_s - 1}\frac{\mu_s(\alpha)}{\mu_s(M_\alpha)}. \tag{1.16}$$

The above system of equations (6)(10)(13)(14)(15)(16) and boundary conditions jointly characterize a unique equilibrium including the wage schedule, matching function, cutoffs, sorting outcomes and city size distribution in each sector.

Overall, the model brings the skill complementarity to a spatial setting where heterogenous individuals organize into production teams and sort across cities. In this long -run equilibrium, the spatial allocation of skills, the organization of skills, and the size distribution of cities are jointly determined rather than sequentially isolated. The teamwork reinforces each other's skills, enhances spatial sorting and boosts the capacity of cities, leading to greater productivity gaps.

### 1.2.6 Equilibrium

In this subsection, I define the equilibrium and characterize its properties.

**Definition 1.** *A spatial general equilibrium for a total population of size $\mathbb{L}$ and sector population $\mathbb{L}_s$ with respective skill distribution $\mu_s(\alpha), \alpha \in \mathbb{A}_s \equiv [0, \overline{\alpha}_s]$ on a set of city $\{\mathbb{C}\}$ is a set of price*

*$\{p_j^n\}$, population distribution $\{\mu_s(\alpha, j)\}$, matching function $\{M_s\}$, wage schedule $\{w_s\}$, cutoffs $\{\underline{\alpha}_s, \hat{\alpha}_s\}$ such that*

*1) Manager chooses the optimal location, labor types and quantity taking wage schedule as given;*

*2) Occupations are optimally chosen to maximize utility;*

*3) Matchings are optimally formed and consistent with each other's choice;*

*4) Labor and goods market clear for all tradable sectors;*

*5) Labor clears city by city for non-tradable sectors, in which utilities are equalized;*

*6) Total and sectorial population clear.*

That is, in addition to the system of equations described above, the following equations also hold.

$$\mu_s(\alpha) = \sum_j \mu_s(\alpha, j) \quad \forall \alpha, \tag{1.17}$$

$$L_j = \sum_s \mathbb{L}_s \int \mu_s(\alpha, j)\,\mathrm{d}\alpha \quad \forall j, \tag{1.18}$$

$$L_j^n = \sum_s \mathbb{L}_s \int_0^{\underline{\alpha}_s} \mu_s(\alpha, j)\,\mathrm{d}\alpha = L_j \bar{n} \quad \forall j. \tag{1.19}$$

The equilibrium is unique in terms of the city size distribution in each sector but says nothing about which sites become bigger cities as they are ex-ante identical. Without a coordination device providing optimal incentives to teams, there could be a coordination failure especially when the number of cities becomes discrete. In the appendix, I provide one such device through education. But there is still little to say about sectorial composition in each city.

**Within-Sector Property.** Since there is a positive assortative matching between manager and workers, observables of worker and managers are positively associated with city size.

**Proposition 1.** *In equilibrium within each sector, manager skill, worker skill, team productivity increases with city size. For any $L < L'$, take a manager $\alpha$ such that $L_s(\alpha) = L$ and a manager*

$\alpha'$ *such that* $L_s(\alpha') = L'$. *Then* $M_s(\alpha) < M_s(\alpha')$, $f(\alpha, M_s(\alpha)) < f(\alpha', M_s(\alpha'))$, $w_s(M_s(\alpha)) < w_s(M_s(\alpha'))$, $\pi_s(\alpha) < \pi(\alpha')$ *and* $p^n(L) < p^n(L')$.

*Proof.* See the appendix. □

Proposition 1 is consistent with stylized facts of agglomeration. In equilibrium, larger cities are more productive and more skilled on average. They also have higher average income and cost of living. Moreover, the team structure predicts co-location and co-work of skilled managers and skilled workers as teams. These predictions are also supported by empirical findings of assortative matching between workers and firms (Mion and Naticchioni, 2009; Card et al., 2013; Dauth et al., 2019). It should be noted that such prediction is not unique feature of my model. Other standard models in the urban literature produce same results.

For a given sector, there are three groups in a city $j$: managers of type $\alpha$, workers of type $M(\alpha)$ and non-tradable producers. Income of all groups increases with city size but the mechanism differs. For non-tradable producers, higher income is simply compensation for higher cost of living. For workers in tradable sectors, cross-city differences in wages come from the return to skill generated by assortative matching. It depends on the matching function and underlying skill distribution. There is no agglomeration externality on wages. For managers, income differences can be decomposed into three different sources. First, due to spatial sorting, managers in larger cities have higher skill that could generate higher incomes in any location with any partner. Second, since there is complementarity between two occupations, better managers benefit more from being able to work with better workers, i.e., assortative matching within a team. Third, firms sort across cities and benefit from agglomeration externalities. As the claimant of profit, better managers realize larger gains in larger cities.

**Skill Premium and Inequality.** The team structure sheds some light on skill premium and inequality across cities. I define skill premium in a city as the average income of tradable producers to the income of non-tradable producers. When the labor quantity $l(M_\alpha)$ is large, skill premium is approximated by wage-price ratio,

$$\frac{\frac{\sigma_s \pi(\alpha, M_\alpha, L)}{1+l}}{p^n(L)} \approx \frac{\sigma_s}{\sigma_s - 1} \frac{w(M_\alpha)}{p^n(L)}. \tag{1.20}$$

Therefore, cross-city variation in skill premium depends on the elasticity of wages to city size and the elasticity of price levels to city size. Intuitively, when the return to skill dominates increase in the cost of living with respect to city size, skill premium will rise. Due to the scarcity of skilled workers in the overall distribution, the skill premium is likely to be higher and more salient in larger cities as documented in Baum-Snow and Pavan (2013).

The team structure also links within-firm inequality to employment size as $\frac{\pi}{w} \propto l$. Theoretically, high skilled pairs do not necessarily lead to larger team size as both numerator and denominator are increasing with city size. But the empirical positive relationship between firm size and city size implies that within-firm inequality is higher in larger cities. For example, Mueller et al. (2017) find that within-firm pay inequality – wage differentials between top- and bottom-level jobs – increases with firm size in UK. In the next section, I also show that it is also true in my dataset. Using firm data from the US, Song et al. (2018) finds that between-firm wage inequality is driven by sorting of firms and workers, whereas within-firm inequality is mostly concentrated in larger firms. Through the team structure, my model features both within-firm and between-firm wage inequalities that are consistent with their findings

**Skill Shocks.** The analysis above highlights the importance of underlying skill distribution in determining matching patterns and equilibrium outcomes. The positive assortative matching have well-established properties that are useful to study the impact of skill shock on inequality. To illustrate, I consider comparative statics of a specific type of distribution shock in the short run.

**Proposition 2.** *(Short-Run Abundance in Extreme Skills) Consider a change in skill distribution* *of tradable sector from* $\mu_s$ *to* $\bar{\mu}_s$ *such that*

$$\frac{\bar{\mu}_{\alpha_2}}{\bar{\mu}_{\alpha_1}} \geq \frac{\mu_{\alpha_2}}{\mu_{\alpha_1}}, \ \forall \hat{\alpha}_s \leq \alpha_1 < \alpha_2 \leq \overline{\alpha}_s,$$

*and/or*

$$\frac{\bar{\mu}_{z_2}}{\bar{\mu}_{z_1}} \leq \frac{\mu_{z_2}}{\mu_{z_1}}, \ \forall \underline{\alpha}_s \leq z_1 < z_2 < \hat{\alpha}_s.$$

*That is, new distribution satisfies MLRP towards extreme skill, divided by* $\hat{\alpha}$*. Then in the short-run (no occupation change), the new matching function* $\overline{M}_s(\alpha) \leq M_s(\alpha)$*.*

*Proof.* See the appendix. □

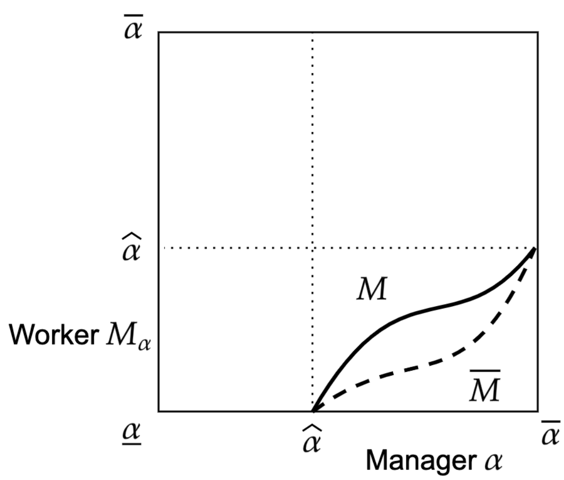



**Figure 1.1**: Change in Matching Function: Short-Run Abundance in Extreme Skills

The shock means relative demand for high skilled worker rises, and the relative supply of high skilled workers decreases. Then market clearing conditions require workers to move toward matching with more skilled managers. This implies worker downgrading from the view of managers and manager upgrading from workers' perspective.

In light of the properties of assortative matching, the model links interval-wise income inequality to the matching function. Since better partners increase the returns to skill within each occupation, inequality will increase when everyone in an interval gets a better match.

**Corollary 1.** *Let $M$ and $\overline{M}$ be two equilibrium matching functions such that all workers (managers) with skill $\alpha \in (\alpha_1, \alpha_2)$ are matched with better managers (workers) under $\overline{M}$ than those under $M$. Then income inequality over $(\alpha_1, \alpha_2)$ is higher under $\overline{M}$ than under $M$.*

*Proof.* I show the proof for workers. The relative wage of two workers can be expressed as

$$\log w(\alpha) - \log w(\alpha') = \int_{\alpha'}^{\alpha} \frac{f_2(M_z^{-1}, z)}{f(M_z^{-1}, z)} \mathrm{d}z,$$

If $\overline{M}^{-1} > M^{-1}$ for all $\alpha$ workers in the interval, the integral is strictly larger under $\overline{M}$ by log-supermodularity of $f$. Similar results hold for managers. □

Applying Corollary 1 to the extreme skill abundance shock, the inequality within workers increases and falls within managers in the economy. Since workers outnumber managers in most cases, the overall inequality in the economy is likely to increase. Similarly, the model predicts an increasing inequality as the economy experience fast expansion in higher education or structural transition toward urbanization, both leading to (relative) abundance in extreme skill in the short run. In terms of migration, workers are now matched with better managers in larger cities. Assuming that managers are not relocating in the short run, the shock implies internal migration toward larger cities. In the context of China, such phenomenon corresponds to the rapid urbanization in which workers move up the ladder of cities for better matches.

A similar effect can be created through changes in the necessities share $\overline{n}$. Changes in $\overline{n}$ can be viewed as either a preference shift, or productivity changes in non-tradable sector. All else equal, a smaller $\overline{n}$ releases more low skilled worker from non-tradable sector to tradable sectors, creating a short-run effect similar to abundance in extreme skill.

It is difficult to give analytical solutions to the long-run equilibrium or a general distribution shocks. Although the above example is also a special case in which the shock happens to be divided by $\hat{\alpha}$, it demonstrates the spatial sensitivity of teams to underlying distribution. This sensitivity echoes the "fast" location changes of industries across cities illustrated in Duranton

(2007).

**Cross-Sector Property.** Sectors may differ in production function $f_s$, the strength of agglomeration $c_s$ and exogenous skill distribution. Teams from different sectors thus locate in different sets of cities, and a city is a union of sectorial teams that choose the same sizes. I define the geographic distribution of firms as follows:

**Definition 2.** *The geographic distribution of firms in a sector is the probability that a team from the sector is in a city of size smaller than L,*

$$G(L|c_s,\mu_s,f_s) = Prob(a\ team\ from\ sector\ s\ is\ in\ a\ city\ of\ size\ smaller\ than\ L).$$

Since decisions are jointly determined, changes in parameters invoke not only relocation but also rematching. For instance, a right shift in the underlying skill distribution $\mu_s$ may lead to larger city sizes at the top tier. At the same, if the shift induces some types to switch from workers to managers, the bottom cities may become smaller as well. In other words, the geographic distribution of firms depends on the relative forces of sorting and rematching in shaping city sizes. Therefore, it is difficult to predict cross-sector comparative statics.

If I restrict to the cases where the indirect effect of rematching can be neglected, then it is possible to examine how the geographic distribution of teams in a sector vary with parameters holding all other sectorial characteristics constant, in particular the distribution of skill $\mu_s(\cdot)$.

**Proposition 3.** *When the indirect effect on matching is sufficiently small compared to changes in agglomeration forces, the geographic distribution $G_s$ of a high $c_s$ sector first-order stochastically dominates that of a low $c_{s'}$ sector, all else equal.*

*Proof.* When the indirect effect on matching is sufficiently small, the work type can be treated as fixed in each team and there is a one-to-one positive relationship between $c_s$ and optimal city size for every team. □

**City-Size Distribution.** In equilibrium, each sector spreads out over a range of cities and each city hosts multiple sectors and skills. However, the sector composition is not determined from the model. Under a continuous measure of cities, the mass of each city size is the sum of teams (managers and workers) from all sectors who sort into cities of size $L$, divided by $L$. Given skill distribution, changes in initial population size do not affect equilibrium, so the mass of cities increases but not the distribution of city size. Under certain conditions, the framework could generate Zipf's Law sector by sector.

**Proposition 4.** ***(Zipf's Law by Sector)*** *Provided that $\bar{n}$ is small and the gap between congestion force and agglomeration forces is small, the city size distribution converges to Zipf's Law sector by sector.*

*Proof.* See the appendix. □

It is known that the city size-rank in China does not follow Zipf's Law (Anderson and Ge, 2005). As shown in Figure 1.16, the log city size and log rank in my sample also rejects Zipf's Law. Proposition 4 provides an explanation that cities differ in their sector compositions therefore the overall size distribution of cities is a nonlinear combination of sectorial distributions. Industrial policies by central government could play a major role in shaping sector allocation and the city size in China.

**Social Welfare.** The equilibrium above is suboptimal because managers only consider the profit rather than the total output. Consider a social planner who maximize the overall output of individuals in a single-sector economy. Since the set of non-tradable producers does not change, I focus on the welfare of the tradable sector.

Planner's problem is to optimally choose – for each skill $\alpha$ – its occupation, team partners and location. In other words, the planner maximizes output from all permutations and combinations of team composition and location choices. It can be shown that the assortative matching conditional on occupation sets are efficient. However, the location choice made in a decentralized economy is suboptimal because it is optimized on managers short-sighted objective of their own profits rather than team output. Therefore, the social planner would develop larger cities compared to ones chosen by managers conditional on the same skill sets. However, when city size changes, it also changes the occupation cutoff, leading to different matching patterns, which is another source of distortion. The social optimal equilibrium then differs not only in the city-size distribution, but also in the team composition.

To induce such socially optimal equilibrium, a complicated incentive scheme is needed. More generally, in a multi-sector economy without exogenous division rules such as Assumption 2, the planner also needs to trade off between sectors to reach efficient industrial composition. Such analysis is important but beyond the scope of this paper.

### 1.2.7 Extensions

In this subsection, I relax the model assumption and show that the insight from model holds in the presence of migration barriers and idiosyncratic shocks. These extensions are important to bridge the model to estimation.

**Discrete Cities.** Conditional on a successful coordination, properties in the continuous case also apply to the discrete case. Better teams tend to locate in larger cities. Formally, let the equilibrium matching in a sector be $M(\alpha)$ with prevailing wage $w(\alpha)$ and boundaries (cutoffs) $(\underline{\alpha}, \hat{\alpha}, \overline{\alpha})$. Denote by $J$ the total number of cities.

Let there be a $2J$- partition of original skill support $\mathbb{A}$ such that

1) managers set $[\hat{\alpha}, \overline{\alpha}] = [\hat{\alpha}, \alpha_1] \cup [\alpha_1, \alpha_2] \cup \cdots \cup [\alpha_{J-1}, \overline{\alpha}]$;

2) workers set $[\underline{\alpha}, \hat{\alpha}] = [\underline{\alpha}, b_1] \cup [b_1, b_2] \cup \cdots \cup [b_{J-1}, \hat{\alpha}]$;

3) $M(\alpha_j) = b_j$.

In equilibrium, city $j \in J$ contains one subinterval of managers and corresponding subinterval of workers. That is, $\mathbb{A}_j = [\alpha_{j-1}, \alpha_j] \cup [b_{j-1}, b_j]$. There will be $J-1$ types of marginal workers, $J-1$ types of marginal managers who are indifferent between two cities respectively.

For them, the benefit of larger city's agglomeration externality are exactly offset by higher cost of living. This provides a set of conditions that pin down the cutoffs.

**Idiosyncratic Preferences**. With discrete cities, I introduce idiosyncratic shocks on managers after the matching but before the actual migration is completed. The purpose of this pause is to create an opportunity for managers to reconsider locational choices due to idiosyncratic preferences. Same as before, workers have no choices but to follow managers. They may sort according to their location preferences but in equilibrium the market always clears.

I separate the process into two stages. The first stage provides optimal matching and occupation outcomes. The second stage inherits the equilibrium outcomes (except optimal location choices) and embraces utility shocks. Formally, let starred variables be the solution to the model under perfect sorting. Instead of the optimal location $L^*$, each manager $i$ now chooses city size $\tilde{L}$ that maximizes utility equation (9) with idiosyncratic shocks, evaluated at the equilibrium wage schedule $w_s^*$ and matching function $M_s^*$. The solution can be expressed as

$$\tilde{L}_{i,s} = \arg\max_{L}\{V_s(\alpha, M_s^*(\alpha), L|\text{manager}^*, w_s^*)\exp(\varepsilon_{i,s,L})\},$$

where $V_s(\alpha, M_s^*(\alpha), L|\text{manager}^*, w_s^*)$ is the non-stochastic part of utility for managers of type $\alpha$. $\varepsilon_{i,s,L}$ are idiosyncratic shocks that are i.i.d. across cities and teams within sector $s$. I assume $\varepsilon_{i,s,L}$ is drawn from a Extreme Value Type I distribution with shape parameter $\nu_{\varepsilon,s}$ so that $\exp(\varepsilon_{i,s,L})$ follows Fréchet distribution with the same shape parameter. Standard procedures predict that the probability of team $\{\alpha, M_\alpha\}$ choosing a city of size $\tilde{L}$ is

$$\text{Prob}(\tilde{L}|\alpha, M_s^*(\alpha), w_s^*) = \frac{V_s(\alpha, M_s^*(\alpha), \tilde{L})^{\nu_{\varepsilon,s}}}{\sum_L V_s(\alpha, M_s^*(\alpha), L)^{\nu_{\varepsilon,s}}}. \tag{1.21}$$

Because $V_s(\alpha, M_s^*(\alpha), L)$ is concave and maximized at the original optimal city size $L_s^*(\alpha, M_s^*(\alpha))$, the mode of city size distribution among type $\alpha$ is thus $L_s^*(\alpha)$. The probability of locating in $L$ decreases as $L$ deviates away from $L^*$. By doing so, homogenous teams now spread across cities of different sizes. The the spread of sizes is governed by $\nu_{\varepsilon,s}$.

**Migration Barriers.** Classic migration cost can be easily embedded in the baseline model. Here I consider migration barriers in the context of China and focus on household registration system (hukou), which is also regarded as a source of misallocation (Tombe and Zhu, 2019).[19] In these studies, hukou retains people from moving towards more productive locations and distorts the allocation of skills.

In my model, hukou has another distinct implication for large cities. When hukou blocks low skill workers from following their managers to larger cities, it forces the matching to be made among high skills in the local labor market. In other words, it disconnects large cities from the global matching and creates local matching that makes teams more productive in larger cities. This segregation of the labor market increases the productivity of large cities and, in contrast to the migration barrier view, could enhance productivity gaps across cities. Therefore, the aggregate impact of hukou is ambiguous and difficult to model. I formalize the traditional view and leave the second channel for future study.

I make a simple extension to the model with birthplaces. After the general equilibrium is pinned down, managers choose between birthplaces with hukou and optimized locations without hukou. Similar to idiosyncratic shocks, I fix the equilibrium occupation and matching when managers make the final binary locational choices. Suppose that managers are born in different city sizes according to a prior population distribution bounded on $[\underline{L},\overline{L}]$ that is independent of skill $\alpha$. Once managers leave birthplace for a larger city, they lose a constant share $\tau$ of utility. Therefore, the value function of manager $i$ locating in a city of size $L$ can be written as

$$V_{i,s}(\alpha, M_s^*(\alpha), L|\text{manager}, w_s^*)[1-\tau\mathbb{1}(L > L_{\text{i, hukou}})],$$

[19]The system assigns a unique Hukou location (begins with birthplace) to each person regardless of future physical appearance in the labor market. The Hukou grants access to local public service and welfare such as children's education, medicare and social security etc. Transferring Hukou across cities are typically associated with stable jobs and housing purchases.

where $\mathbb{1}(L > L_{i,\text{hukou}})$ is an indicator for city sizes larger than $i$'s birthplace city size $L_{i,\text{hukou}}$. Notice that the value of hukou is higher for high-skill types who tend to choose larger cities. The presence of $\tau$ does not affect first order condition of matching therefore the team composition is the same as before. The difference is that instead of perfect sorting, managers may stick to their birthplaces. I now show that this stickiness is only binding to managers born in certain city sizes. Moreover, the range of binding cities shrinks as skills increase.

The additional problem for manager $i$ is to choose between two modes: with hukou at birthplace $L_{i,hukou}$ and without hukou at the optimal size $L^*$

$$\max_{\{L^*, L_{i,\text{hukou}}\}} \{V_s^*(\alpha, M_\alpha^*, L^* | \text{manager}, w_s^*)(1-\tau), V_{i,s}(\alpha, M_\alpha^*, L_{i,\text{hukou}} | \text{manager}, w_s^*)\}.$$

The probability of insisting on $L^*$ is then

$$\text{Prob}[V_{i,s}(L_{i,\text{hukou}}|\alpha) < V_s^*(L^*|\alpha)] = \text{Prob}[L_{i,\text{hukou}} < \hat{L}(\alpha)]$$

where $\hat{L}(\alpha) < L^*(\alpha)$ and solves $[\pi(\alpha, M_\alpha, L^*) - c(L^*)](1-\tau) = \pi(\alpha, M_\alpha, \hat{L}) - c(\hat{L})$.[20] In other words, for $\alpha$-individuals born in cities smaller than $\hat{L}(\alpha)$, the gains from migration is sufficiently high to forsake hukou at birthplace. In contrast, it is irrational to do so when the ideal city is only slightly bigger than birthplace. By implicit function theorem, $\hat{L}(\alpha)$ is increasing in $\alpha$. So high-skill people have higher geographic mobility regardless of hukou status. Unfortunately, the firm census I use does not contain information on employees' hukou status therefore it is not possible to calibrate migration barrier without additional dataset. I address this issue in several ways in the data and estimation.

There are also several interesting extensions to consider in the future. First, the free trade assumption greatly simplifies the computation process but are not appropriate for all sectors. I extend the model to costly trade in the appendix and show that the proposition still hold with

[20] I suppress the necessity cost in utility as it does not change the conclusion.

some modifications. There is a similar argument for amenity differences. But calibrating the extended model requires information on internal trade flows. Second, the model does not directly incorporate searching frictions and asymmetric information. For example, there might be a fixed cost to search for the right partners for managers and larger cities may serve as a more efficient matching platform (Dauth et al., 2019). Eeckhout (2018) and Eeckhout and Kircher (2018) provide more discussions with searching, endowment resources, moral hazard, etc. These extensions enrich the dimensions but lead to extra conditions for positive assortative matching and demanding information from data. Third, the model focuses on face-to-face interactions in production. It is of course possible to consider the fragmented cross-city production given the fast development of information and communication technology (Jiao and Tian, 2019). The model also describes a long-run equilibrium and abstracts away from potential gains from learning and experience.

## 1.3 Data and Empirical Evidence

In this section, I first introduce the data and definitions. I then present a set of cross-sectional facts that support model predictions. The results do not provide a causal interpretation.

### 1.3.1 Data

The main data are from the firm-level records of 2008 National Economic Census (NEC) in China conducted by National Bureau of Statistics (NBS).[21] The census is mandatory and covers all legal person units in the secondary and tertiary industries regardless of size, type and location.[22] In particular, I obtain a 20% sample of firms in manufacturing sectors and the full universe of firms in service sectors. The unit of observation is an enterprise (hereafter referred

[21] Survey data for self-employed individuals are not available. However, since most self-employed individuals provide small non-tradable local services, they are not critical to our main sorting results.

[22] Other commonly used firm survey covers only a subset of NEC. For example, Annual Survey of Industrial Firms (ASIF) only covers firms with annual sales above 20 million RMB and state-owned enterprises.

to as a firm or a team), which is a legal entity larger than an establishment but smaller than a consolidated corporation. They operate with high degree of independence and keep their own financial records for tax purposes.

The dataset includes information on basic characteristics of the legal entity units, four-digit industry code (Industrial classification GB/T 4754-2002), total employment, financial statement, production and operation etc. Most importantly, for service sectors, it reports firm employees' skill composition by five education levels. For the rest of the paper, I mainly focus on service sectors as skill composition information is critical to the identification of complementarity. Manufacturing sectors are included when outcomes are total employment, total labor payment, average wages and value-added. More details on manufacturing sectors and its identification are included in the appendix.

I make two restrictions to the data. First, I drop three sectors that are outside the scope of model analysis: mining, construction and real estate (0.42 million observations). Second, I exclude legal entities that are non-profit such as public institutions, NGOs and other social organizations (2 million observations). These units are often financed and supervised by local governments for public service. By doing so, I focus on firms that are relatively mobile and proactive for profit. The final dataset contains over 2.8 million firms, of which 86% are from service sectors. Based on the industrial code, I further categorize them into 45 sectors (27 in manufacturing and 18 in service) for structural estimation. The geographic distribution of the sample is shown in Table 1.4 in the appendix.

I report summary of statistics in Table 1.1 for 18 service sectors. The largest service sector in terms of firm numbers is wholesale. I report the value-added (measure by profit from main business) and employment. The median number of employees in service sector is six. I also report the share of sector employment that have a four-year college degree. The most educated sector is software while the least one is hotels and catering. In the last column, I calculate the share of total employees that locate in large cities (over 4 million population defined below). The results suggest that labors are concentrating in larger cities especially for software, media and business services. Statistics for manufacturing are shown in Table 1.5. Comparing to service, manufacturing sectors have larger size and higher value added.

**Cities.** The geographic location of a firm can be identified precisely from the street address and area code. I aggregate areas to prefecture-level cities since there is no well-defined metropolitan statistical areas in China. A prefecture-level city is an administrative division below a province and above a county that includes both the urban and rural places. They form a complete partition of China like commuting zones in the United States but are twice larger on average. I exclude 53 prefecture-level areas and leagues that are vast but less dense areas.[23] The total number of cities is 287.

**City Size.** To measure the size of these cities, I use the number of working age population (between age 15 and 64) from 2010 population census, which is the closest population census to 2008 economic census. The population refers to the permanent residents including both registered population (hukou population) and migrant workers who had been living in the city for more than half a year. By including those migrant workers, the number to a large degree represents revealed preferences for locations in the labor market. Hukou may be a huge barrier for permanent rural-urban migration (Ngai et al., 2019) but it is less of a constraint for temporary cross-city migration especially for young cohorts. The share of residents without hukou in Beijing and Shanghai, two cities with the most stringent hukou policy, is 36% and 38% respectively in 2010, suggesting strong incentives to sort despite hukou concerns.

The average size of these cities is about 3 million, ranging from 0.18 million (Jiayuguan) to 19 million (Shanghai). I group the largest 74 cities (4 million above) that account for half of the above working age population as large cities. In the last column of Table 1.1, it can be seen that employment are concentrated in the large cities. In the estimation, I further rank and define

[23]These areas are mostly autonomous regions of minority groups. 39 of them are from Tibet, Xinjiang, Qinghai, Guizhou and Yunnan.

**Table 1.1**: Summary Statistics: Service Sectors

| | | log value-added | | | employment | | | % of total employees | |
|---|---|---|---|---|---|---|---|---|---|
| | N | mean | p25 | p75 | p25 | p50 | p75 | college | large cities |
| Transportation | 120,083 | 5.59 | 4.54 | 7.45 | 5 | 10 | 30 | 7.5 | 64.5 |
| Postal and warehousing | 20,878 | 5.43 | 4.62 | 7.51 | 5 | 12 | 29 | 9.5 | 62.9 |
| Telecommunication services | 17,065 | 4.84 | 3.78 | 6.8 | 3 | 7 | 20 | 32.0 | 60.2 |
| Computer services | 81,943 | 4.65 | 3.89 | 5.90 | 3 | 4 | 8 | 30.0 | 74.2 |
| Software | 38,628 | 4.48 | 3.40 | 6.78 | 3 | 7 | 16 | 62.3 | 91.5 |
| Wholesale | 826,403 | 5.56 | 4.62 | 7.04 | 4 | 6 | 11 | 14.4 | 67.3 |
| Retails | 524,938 | 5.17 | 4.17 | 6.54 | 3 | 5 | 10 | 8.5 | 63.4 |
| Hotels and catering | 133,460 | 5.37 | 4.80 | 6.52 | 8 | 16 | 36 | 1.1 | 65.6 |
| Finance | 7,913 | 5.17 | 4.33 | 7.03 | 4 | 6 | 10 | 30.0 | 56.6 |
| Leasing | 17,222 | 4.64 | 3.74 | 6.34 | 3 | 5 | 10 | 10.7 | 70.7 |
| Business services, non IT | 328,174 | 4.52 | 3.50 | 6.48 | 3 | 5 | 10 | 21.2 | 75.7 |
| Scientific service | 120,664 | 4.67 | 3.69 | 6.67 | 3 | 7 | 16 | 40.1 | 77.6 |
| Water, environment, utilities | 21,164 | 5.04 | 4.20 | 6.96 | 5 | 10 | 24 | 11.5 | 62.9 |
| Traditional service | 102,869 | 4.77 | 3.78 | 6.33 | 4 | 7 | 15 | 6.3 | 72.4 |
| Education | 20,774 | 4.96 | 4.01 | 6.61 | 4 | 9 | 20 | 25.8 | 66.6 |
| Health and welfare | 15,097 | 4.88 | 3.83 | 6.50 | 3 | 7 | 23 | 26.9 | 59.4 |
| Media | 7,797 | 4.96 | 3.99 | 7.17 | 4 | 10 | 25 | 37.9 | 83.1 |
| Sports and entertainment | 26,265 | 4.79 | 3.71 | 6.61 | 3 | 7 | 18 | 9.3 | 68.9 |
| Total | 2,431,337 | 5.16 | 4.22 | 6.73 | 3 | 6 | 12 | 14.6 | 68.1 |

Note: There are 18 service sectors in total. Value-added is measured in thousand RMB. The last two columns show the share of total employees with bachelor degrees or above, and the share of total employees in cities with over 4 million population.

four groups of cities each accounting for 25% of the total population. After normalizing city sizes by the smallest city size in the sample, the largest city size in four bins are 14, 22, 34 and 104 respectively.

I also collect prefectural-city level commercial housing prices from the China Real Estate Statistics Yearbook to approximate the congestion cost in the model. In the structural estimation below, I regress it on city total population in 2010 to calibrate parameters on congestion. Studies have shown that the value of public goods (priced and non-priced) is well capitalized in the housing market in China (Zheng and Kahn, 2008; Chen et al., 2019). Given that housing purchases are closely associated with hukou status, gains from local hukou are partly incorporated in congestion costs when measured by these housing prices.

### 1.3.2 Defining Skills

The dataset contains skill mix in each firm but not individual records such as experience, occupations and wages. Therefore, it is not possible to directly identify managers and workers in the model. Instead, following the comparative advantages of skilled persons in managerial position (Lucas, 1978), I approximate with the distribution of education. More specifically, I extract two tails of the education distribution in each firm and define the right tail as the higher level occupation, namely the “manager”, and the left tail as “workers”. The skill of each occupation are then measured by years of education.[24] Note that the skill mix differs across firms so any education level could be workers or managers. Therefore for each firm, I obtain two statistics that represent skill mix. Correlation of those two measures in the raw data are 0.4.

There are several caveats to this approximation. First, this occupation structure has a fixed two-layer hierarchies management in contrast to endogenous number of layers in Garicano (2000). However, the average number of education layers is 2.2 so a two-layer structure is a

[24]These five categories are: middle school and below, high school, two-year college, four-year college, graduate school and above. The corresponding years of education are set to 6, 12, 14, 16 and 20.

plausible choice. About 25% of firms in the data have only one education category, whereas 10% of firms have four or five education levels. Most importantly, as shown in Figure 1.15, there is no significant difference in skill dispersion across cities so it is unlikely to cause systematic bias due to the variable construction. As a robustness check, I also show results using 25th and 75th percentile of education distribution as alternative proxies. Second, the proxies capture the quality but not the quantity of human capital in each occupation. The skills of managers are the same for firms with one or ten college-educated managers. To mitigate such concern, I control for factors like total employment, skill dispersion and firm size in regressions.[25] Lastly, since the data has no personal information, I can not capture other unobserved ability and experience, as well as the composition differences in age, gender and hukou status. Ideally, a panel of employee-employer matched data with individual records would help address these concerns as in Dauth et al. (2019). To help reduce such bias, I also include four-digit sector fixed effect, firm age fixed effect, ownership fixed effect, etc. in regressions. These controls minimize the bias as worker characteristics become similar within a narrow group of similar firms.

### 1.3.3 Evidence on Assortative Matching

#### Stylized Facts

**Cross-City Productivity Difference.** I start with relationships between productivity and city size. I use the detailed three-digit industry codes and obtain 250 sectors that have more than 500 firm-level observations in manufacturing and service.[26] For each sector, I examine how city average firm value-added changes with city size. More specifically, I compute the average firm value-added for each city. I then regress the average value-added on city size, both in the log terms. The estimated coefficient is then the elasticity of average value-added to city size.

[25] Another concern is that for enterprises with only one person, there is no difference between manager and worker. However, only 4% of the observations have only one employee. All results holds if they are excluded.

[26] To avoid spurious results due to small sample, I restrict to sectors with sufficient observations. Results are similar with alternative restrictions.

Similarly, I also compute the elasticity of average employment to city size. Figure 1.2 plots these two elasticities for each sector.

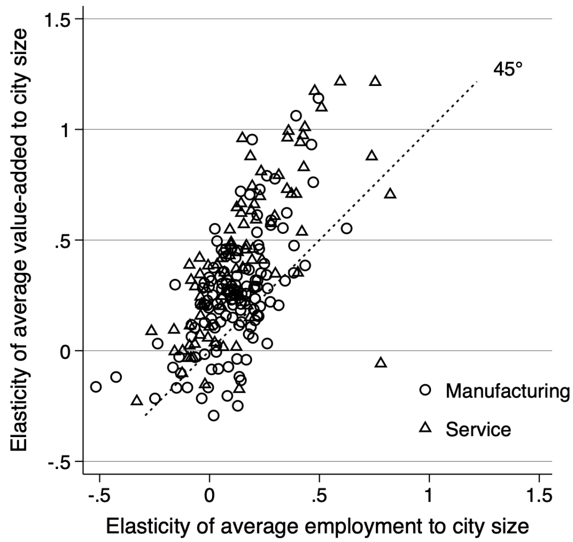



**Figure 1.2**: Elasticity of Mean Value-Added and Employment to City Size

Note: Elasticities are estimated from the 2008 firm census. Both manufacturing and service sectors are included. Each dot represents a sector $s$. Elasticity of variable $y$ to city size is $\hat{b}$ from $\log \text{mean } y(L_i) = a + b \log L_i + \varepsilon_i$, ran by each three-digit sector with more than 500 observations and weighted by the number of firms in city.

There is a positive correlation between firm productivities and city sizes. For 137 manufacturing industries, only 23 have negative elasticities in value-added and 27 in employment size. In 113 service sectors, those numbers are 10 and 32 respectively. In addition, most of the negative elasticities are small and insignificant. The elasticity of value-added is mostly higher than that of employment so output per worker is also increasing with city size. There is no significant difference between manufacturing and service sectors. In the appendix, I also show elasticities of two alternative measures with respect to city size: total revenue and total costs. As expected, they are positive and positively correlated. These relationships imply agglomeration economies.

**Skill Concentration.** Aggregating firm-level data, I show how employees with different levels of education are distributed across cities. Figure 1.3 plots the density for each of five education groups.

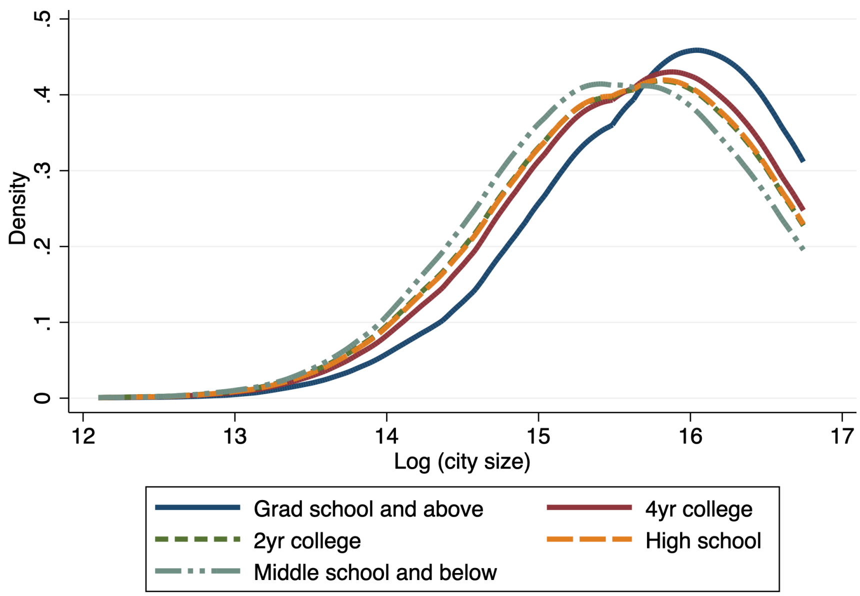



**Figure 1.3**: Geographic Distribution of Employees in Service Sectors, by Education

Note: This figure plots the distribution of firm employment across city sizes in China by education levels. County size are measured by the working age population. Employment data is aggregated from 2008 firm census in China. Service sectors only. Epanechnikov kernel, bandwidth=1.

The results show a clear first-order dominance relationship between different groups, suggesting a spatial sorting of skills across cities. One exception is that there is no significant difference between the distribution of two-year college and high school groups. In Figure 1.10, I also plot the density of firm-level weighted years of schooling from firm census data for all service sectors by a binary city size dummy. The results indicate that larger cities are on average more educated.

To complement firm-level data, I also draw from 2010 population census and plot the relationship between county size and weighted average years of education in Figure 1.4. Statistics from population census covers a wider range of employment, but also include people out of the labor force. Nevertheless, Figure 1.4 shows a strong positive correlation, suggesting that larger counties are more skilled.

Overall, these results suggest that there is spatial sorting of skills across cities in China,

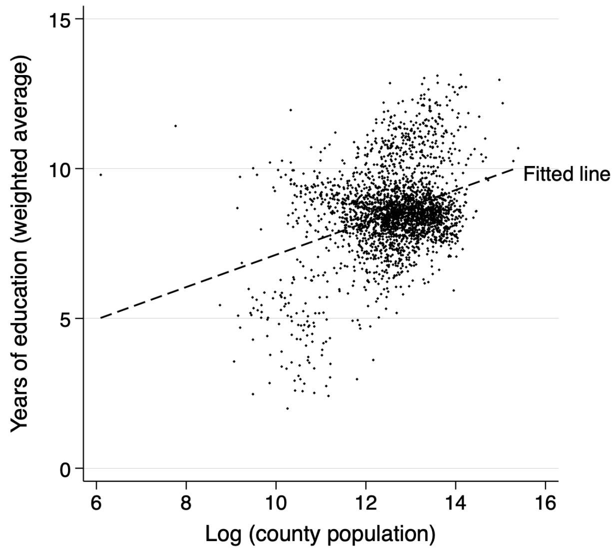



**Figure 1.4**: County Size and Education

Note: This figure plots the relationship between weight average years of education and county size in China. County size are measured by the population above school age. N=2533. Statistics are from 2010 population census in China. Dashed line is the fitted OLS, the slope is significantly positive at the 1% level.

and it is closely related to the observed agglomeration effect above. Therefore, their effects cannot be easily disentangled from the observed productivity gaps.

**Wage Inequality and City Size.** The model in general predicts higher inequality in larger cities. This relationship has been documented in many countries (e.g., Glaeser et al., 2009; Moretti, 2013). China is no exception. I calculate the average wage for each firm and the interquartile range of wage distribution in each city. Figure 1.5 shows how within-city wage inequality changes with city size using firms from both manufacturing and service sectors. The figure shows an increasing city wage inequality as city size gets larger. The results also complements with findings on China's inequality using household surveys (Xie and Zhou, 2014; Chen et al., 2018).

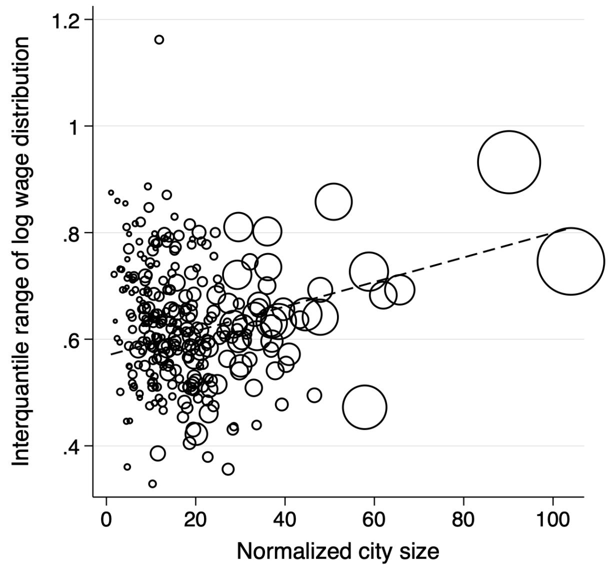



**Figure 1.5**: Within-City Inequality and City Size

Note: This figure plots the relationship between within-city inequality and city size. N=287. Statistics are calculated from 2008 firm census. Both manufacturing and service sectors are included. For each city, I first calculate average wage in each firm, and then calculate the interquartile range of log firm wages (75th - 25th percentile). The size of circle represents the number of firms in each city. Dashed line is the fitted line of weighted OLS, the slope is significantly positive a the 1% level.

### Evidence on Assortative Matching

In the model, assortative matching manifests in two dimensions: between cities and within cities. Between-city assortative matching is the result of spatial sorting where high-skill managers and high-skill workers co-locate in larger cities. Within-city co-work is the extension of assortative matching in the local labor market especially in the case of discrete city.

**Co-location.** For each sector, I compute two elasticities of average skills to city size. The elasticities are similarly estimated through city-level regressions. The only difference is that one elasticity uses the city average skills among the left-tail workers in each firm as dependent variable, whereas the other one draws on average skills among right-tail workers in each firm. These two elasticities capture how fast city skill mix changes with city size.

Figure 1.6 plots those estimated elasticities for 115 service sectors. The results show that

83 sectors, or 72% of 115 sectors, have positive elasticities of average education among right-tail workers. Meanwhile, 87% of sectors have positive elasticities among left-tail workers. Only six sectors have negative signs on both elasticities. In other words, for most of the sectors, the average education level at both tails increases with city size.

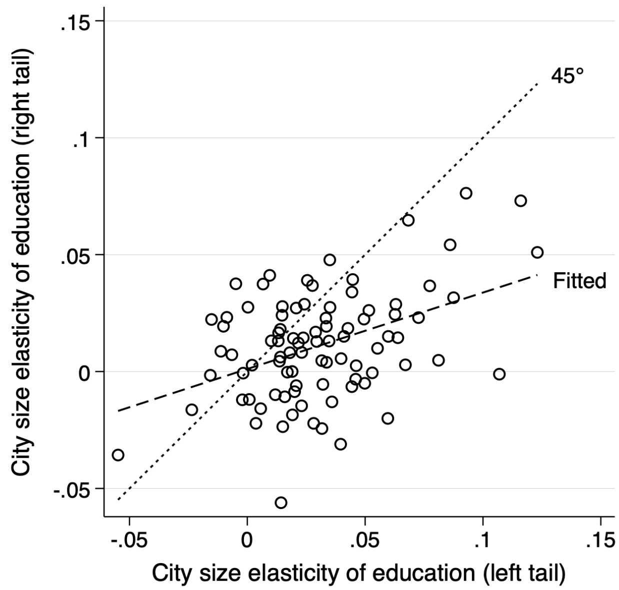



**Figure 1.6**: City Size Elasticity of Average Education at Tails: Service Sectors

Note: Elasticities are estimated from the 2008 firm census, service sectors only. Each dot represents a sector $s$. Elasticity of variable $y$ to city size is $\hat{b}$ from $\log \text{mean } y(L_i) = a + b \log L_i + \varepsilon_i$, ran by each three-digit service sector with more than 500 observations and weighted by the number of firms in city.

Two more comments are in order. First, there is a positive relationship between two elasticities. The slop of a simple OLS fitted line is 0.29 and significant at 1% level. The sensitivity of education in city size is correlated at two tails. Second, for most sectors, the elasticity among left-tail workers is higher than that among right-tail workers. In other words, the skill mix is shifting to the right with city size and the left-tail moves faster than the right tail. As a robustness check, I replicate Figure 1.6 with alternative definition of skill mix using the 25th and 75th percentile of employees' education. I also show the relationship using the most detailed four-digit sector code. The results are similar and included in the appendix. These two patterns can also be summarized by the elasticity of right-tail education to left-tail education at the city level. In the appendix, Figure 1.14 shows the distribution of elasticities for 115 service sectors. All elasticities are positive and less than 1, suggesting the same sensitivity implied by Figure 1.6.

**Co-work.** The cross-city evidence is suggestive but does not necessarily imply assortative matching between skills. There are other explanations that could generate the same sorting pattern. Imagine a case in which high cost of living in large cities crowd out low-skill workers as in Diamond (2016). This story generates the same co-location patterns without organizing them into teams. Workers of different skills may also operate and sort independently for agglomeration benefits among themselves. Besides, cities may function differently. For example, larger cities may be disproportionally occupied by headquarters and managerial units, shifting the overall skill distribution to the right. To further establish the fact on team structure and assortative matching, I use firm-level regressions controlling for other confounding factors.

If relatively more skilled people gather in large cities without team structure, there should be no systematic difference in skill composition across firms within a city. I investigate the education relationship between right-tail workers and left-tail workers in the same firm using the following regression:

$$yos^m_{ijs} = \alpha + \beta\, yos^w_{ijs} + \Gamma X_i + \theta_j + \zeta_s + \varepsilon_{ijs},$$

where the dependent variable $yos^m_{ijs}$ is the years of schooling of right-tail workers in firm $i$, city $j$, sector $s$. $yos^w_{ijs}$ is years of schooling of left-tail workers in the same firm. $X_i$ is a stack of firm level controls including number of establishments, open year, ownership type, number of education categories, value-added, total employment, assets and labor cost. These control variables address heterogeneity in firm age, types, size, capital intensity and skill dispersion. $\theta_j$ and $\zeta_s$ are city fixed effect and sector fixed effect respectively. Table 1.2 shows the results. The robust standard errors are clustered at the province level.

All columns show a significant correlation between top-worker education and bottom-worker education in the same firm. Column (2) to (4) include city size and gradually control for

**Table 1.2**: Education Correlation Between Left-Tail and Right-Tail Workers: Service Sectors

| | Years of education (right tail) | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Years of education (left tail) | .319$^{\dagger}$ | .314$^{\dagger}$ | .578$^{\dagger}$ | .567$^{\dagger}$ | .563$^{\dagger}$ | .559$^{\dagger}$ | .566$^{\dagger}$ |
| | [.016] | [.017] | [.015] | [.014] | [.014] | [.014] | [.014] |
| Log (city size) | | .160 | .064 | .049 | | | |
| | | [.080] | [.041] | [.035] | | | |
| Firm-level controls | | | Y | Y | Y | Y | Y |
| City FE | | | | | Y | | |
| County FE | | | | | | Y | |
| Four-digit sector FE | | | | Y | Y | Y | |
| City $\times$ two-digit sector FE | | | | | | | Y |
| $R^2$ | .163 | .165 | .835 | .837 | .839 | .841 | .840 |

Note: $^{\dagger}$ $p < 0.01$, $^{*}$ $p < 0.05$. N=2,394,832. Robust standard errors are in brackets, clustered at province level. There are 2415 counties in 287 cities and 310 four-digit sectors. Firm level controls include number of establishments, open year, ownership type, number of education categories, value-added, total employment, assets and labor cost.

more factors. While the coefficient on bottom workers' years of schooling is significant at 1% level, the city size effect is not significant and falling as more controls are included. Column (5) corresponds to the baseline specification above with city fixed effect and sector fixed effect. The results indicate that within a city and a sector, the skill of top workers is positively associated with the skill of bottom workers. Column (6) and (7) further enhance fixed effects to county level and city-sector cell. The results are similar to column (5), suggesting assortative matching at the firm level in each city. Since two tails of skill mix are increasing when city size increases, the entire skill mix also shift to the right. In Table 1.6, I replicate Table 1.2 using 25th and 75th percentile definition, the results are similar and robust.

Comparing to cross-city evidence, firm-level evidence reveals a consistent structure inside a city and a sector that can not be explained by other mechanisms alone at the city level (e.g., crowding). Many other stories at the firm-level are also ruled out by controlling for relevant factors in the regressions. Recall that firms in larger cities have more employees. So it is unlikely that firms are substituting away from low skilled workers to high skilled workers. These quality and quantity patterns can be easily reconciled using complementarity within firms.

In short, the regressions support my model predictions of positive assortative matching between skills. Better managers and better workers not only co-locate in larger cities but also co-work in teams. The evidence also implies that talents are on average more concentrated in larger cities and more productive firms as the coefficient is less than one.

## 1.4 Estimation

For the second half of the empirical investigation, I bring the model to data for structural estimation. After calibrating the model, I conduct two counterfactual analyses. First, I quantify the relative importance of firm-level skill complementarity by hypothetically shutting it down in the simulation. Second, I examine the general equilibrium effect of place-based policies subsidizing talented individuals in the second-tier cities in China.

I structurally estimate the model using the simulated method of moments (McFadden, 1989) separately for each sector. The procedure follows standard two steps as below. In the first step, I estimate three sets of parameters that can be inferred directly from the data, separately from the rest of the system. In the second step, I make parametric assumptions about the form of production function, matching function and underlying distributions in the extended model with discrete cities. I then simulate the decisions of firms and jointly calibrate the rest of parameters. I focus on tradable sectors and assume $\overline{n} = 0$ since self-employed sectors are not available in the data.[27] Here I show estimation for 18 service sectors. Results on 27 manufacturing sectors are included in the appendix.[28]

[27] Since the share of necessities $\overline{n}$ can not be backed out in the data, I assume for the convenience of estimation that $\overline{n}$ is sufficiently close to zero. Recall that in Assumption 3, $\overline{n}$ is assumed to be smaller than the relative cost of living. Using $\hat{\phi} = 0.2$, the upper bound of $\overline{n}$, using 18 million versus 0.18 million in the data, is around 0.4. Assumption 3 2) thus holds.

[28] Since skill composition information is missing in manufacturing firms in the data. Parameters are partially identified for manufacturing. See more details in the appendix.

### 1.4.1 Step One: External Parameters

I first calibrate $2S$ parameters directly from the census data using full sample. In each sector, the elasticity of substitution $\sigma_s$ in the CES demand function is calibrated to match the sector-level markup $\frac{\hat{\sigma}}{\hat{\sigma}_s - 1} = \text{mean}(\frac{\text{revenue}}{\text{total costs}})$, where total cost equals main business operating cost plus labor payment. The market size of each sector $Q_s$ is calibrated to the share of revenue from sector $s$ in total revenue.

For parameters $\phi$ and $\gamma$ in congestion cost, I regress prefecture-level commercial housing price $p_j^H$ in 2010 on total population for 287 prefectural cities.[29] The estimated elasticity $\hat{\phi}$ is 0.198. This magnitude is higher than what is commonly obtained in developed countries. Given China's land policies and regulations (Glaeser et al., 2017), this is not a surprising result. I calibrated $\hat{\phi} = 0.2$ and $\hat{\gamma} = 6.5$ for all sectors.

### 1.4.2 Model Specification

**Production Function.** I characterize and parameterize the production function in section 1.2.3 $f_s(\alpha, z)g(c_s, L)$ of a team of manager type $\alpha$, worker type $z$ and size $L$ as follows,

$$\log f_s(\alpha, z)g(c_s, L) = (1 + \log \alpha)[\delta_s + \lambda_s(1 + \log z)] + c_s \log(L/L_0), \tag{1.22}$$

where $\delta_s > 0$ controls the managerial technology and $\lambda_s > 0$ controls the degree of complementarity. It can be checked that this $f(\alpha, z)$ satisfies log-supermodular in $f$ and $g$ respectively. When $\lambda_s = 0$, the model reduces to the case with no complementarity. $c_s > 0$ is the sector-specific parameter that captures the sectorial strength of agglomeration force. $L_0$ is the smallest city size in the set of cities distribution. I denote $\tilde{L} = L/L_0$ the normalized city size.

The direct approach from here is to solve wage schedule, matching function, and oc-

[29] The elasticity of housing price to normalized total population is obtained by running $\log p_j^H = \gamma + \phi \log(L_j/L_0) + \varepsilon_j$.

cupation selection using the functional form of $f$ and skill distribution $\mu$. However, solving an assortative matching equilibrium with heterogeneous agents and endogenous occupation is technically difficult. I make two simplification of the model to ease the problem.

**Occupation.** Instead of simulating one skill distribution and allowing for occupation selection, I simulate skill distribution separately in each occupation so the distribution of manager skills and workers skills are independent with their own mean and variance. More specifically, I assume that manager skill $\alpha$ is distributed according to a log normal distribution with variance $\nu_{\alpha,s}$ and mean at the mode of normalized education in the data, truncated from left at zero. I also assume that the worker skill $z$ is log-normal with mean zero and variance $\nu_{z,s}$, truncated at zero as well. This specification allows imperfect occupation selection and the fact that manager skill distribution has a higher mean value.

**Matching Function.** In theory, the matching function is determined by the underlying skill distribution in each occupation. However, the assortative matching requires the skill set on both sides to be bounded, which introduces nuisance parameters into the model. In addition, the matching assignment is a mapping that can not be analytically expressed. Therefore, the wage schedule is also difficult to track. To reduce computational complexity, I approximate the matching function $M_\alpha$ by parameterizing the shape of matching function in a log linear form.

$$\log M_s(\alpha) = k_s \log \alpha, \tag{1.23}$$

where $k_s \leq 1$ is a parameter that governs the shape of matching. This form satisfies the following properties that are consistent with a general positive assortative matching function: $M(\infty) = \infty$, $M(1) = 1$, $M_\alpha < \alpha$ and $M' > 0$. Moreover, it is micro-founded through a Mincer equation that captures the second-order expression of wage function in log terms of skills. To see that, combine two functional forms on $f$ and $M_\alpha$ with the first order condition equation (6), and the wage for any worker type $z$ can be written as

$$\log w_s(z) = \int_1^z \frac{f_2(t^{1/k_s}, t)}{f(t^{1/k_s}, t)} \, \mathrm{d}t = \lambda_s \int_1^z \frac{\frac{\log t}{k_s} + 1}{t} \, \mathrm{d}t = \frac{\lambda_s}{2k_s} (\log z)^2 + \lambda_s \log z,$$

where $\log w_s(1)$ is normalized to 0. Since I do not distinguish experience and years of schooling, the wage function above to a large degree captures well the relationship between actual wage and skills. Intuitively, the coefficients stand for the return to education, which depends on the strength of skill complementarity and the shape of matching. Therefore, the matching function equation (23) is an appropriate approximation to the actual matching. Nevertheless, I plan to directly solve the assignment problem through simulating an agent-based economy in the future work.

Substitute $M_s(\alpha)$ and $w_s(M_s(\alpha))$ into production function, the problem can be expressed as

$$\max_{\tilde{L}} \{\Lambda_s Q_s \psi_s^{\sigma_s - 1} \tilde{L}^{c_s(\sigma_s - 1)} - \gamma \tilde{L}^{\phi}\}, \tag{1.24}$$

where $\psi_s(\alpha) = \frac{f(\alpha, M_\alpha)}{w(M_\alpha)}$ has the following structure

$$\begin{aligned} \log \psi_s &= (\lambda_s + \delta_s)(1 + \log \alpha) + \frac{\lambda_s k_s}{2} (\log \alpha)^2 \\ &= (\lambda_s + \delta_s)(1 + \frac{\log z}{k_s}) + \frac{\lambda_s}{2k_s} (\log z)^2. \end{aligned} \tag{1.25}$$

Because of the unique mapping of matching function, $\psi_s$ can be expressed entirely by manager skills or workers skills. Moreover, $\log \psi_s$ is a quadratic form of $\log \alpha$ or $\log z$, which means conditional on the distribution of $\psi_s, \alpha$ and $z$, parameters inside the team $\lambda_s, \delta_s$ and $k_s$ could be jointly identified.

**Traditional Density Agglomeration.** My model abstracts away from many other traditional sources of agglomeration externalities that do not interact with team productivity. To account for the density effect per se, I include a log linear term $L^{e_s}$ in the value function. Since

$L^{e_s}$ does not interact with team productivity, $e_s$ can be separately identified from $c_s$.

Finally, I introduce the idiosyncratic shocks described in the model extension. Combining all these considerations, the final maximization problem in the simulation is

$$\max_{\tilde{L}}\{(\Lambda_s Q_s \psi_s^{\sigma_s-1}\tilde{L}^{c_s(\sigma_s-1)} + \tilde{L}^{e_s} - \gamma\tilde{L}^{\phi})\exp(\varepsilon_{i,s,\tilde{L}})\}, \tag{1.26}$$

where $\varepsilon_{i,s,L}$ is the idiosyncratic productivity shock that are i.i.d. across teams and city size within a sector. I assume $\varepsilon_{i,s,L}$ is drawn independently from Extreme Value Type I distribution, with mean zero and shape parameter $\nu_{\varepsilon,s}$ for each sector $s$. Then $\exp(\varepsilon_{i,s,L})$ is drawn from Fréchet distribution with the same shape parameter.

### 1.4.3 Step Two: Identification and Targeted Moments

The remaining eight parameters, denoted by $\Theta_s = \{\delta_s, \lambda_s, k_s, c_s, e_s, \nu_{\alpha,s}, \nu_{z,s}, \nu_{\varepsilon,s}\}$ for each sector $s$ are jointly calibrated sector by sector. They are: team sorting agglomeration ($c$), non-sorting agglomeration externality ($e$); managerial technology ($\delta$), skill complementarity intensity ($\lambda$), matching parameter ($k$); variances of manager productivity ($\nu_\alpha$), worker productivity ($\nu_z$) and idiosyncratic shock ($\nu_\varepsilon$). I simulate the economy with 200 city size bins between the actual normalized city size 1 to 104. For each set of parameters, the model delivers a location choice for each firm.

I target six sets of moments (24 moments in total) from the data and choose the set of parameters that minimizes the distance between the model-implied moments and the data moments. I discuss the targeting moments and identification details one by one below. I use four city sizes bins defined in Section 3 with each bin contains a quarter of total population.

**Average Value-Added by City Bin.** For each city, I calculate the average firm value-added. I then compute log average city value-added for each of four city size bins. Finally, I normalized four bin value-added with their mean. These moments describe how average value-added increases with city size and capture how fast team productivity increases with city size. The information helps pin down the agglomeration parameters $c$ and $e$. Intuitively, both $c$ and $e$ govern the relationship between team productivity (thus value-added) and city size. But $c$ impacts more than $e$ because it entails the sorting of more productive teams into larger cities.

**Firm Value-added Distribution.** The second set of moments is the distribution of firm productivity measured by value-added. I use 25th, 50th, 75th and 90th percentile of the distribution of firms' value-added (normalized by median). These moments provide information on the dispersion of team productivity. Combined with other sets of moments below, it helps to identify $\nu_\alpha$, $\nu_z$ and $\nu_\varepsilon$.

**Firm Geographic Distribution.** To pin down the idiosyncratic shock variance $\nu_\varepsilon$, I use the 25th, 50th, 75th and 90th percentile of normalized firm city-size distribution. They describe the density of firms located in different city sizes. It contains information on the dispersion of team productivity and the idiosyncratic shock, conditional on other parameters. Together with the above set, they identify $\nu_\varepsilon$, therefore $\nu_\alpha$.

**Total Value Added Share by City Bin.** I also use the share of total value-added in each of four city bins to help calibrate above parameters. It summarizes the geographic distribution of economic outputs and helps identifying above parameters.

These four sets of moments are sufficient to identify parameters regarding the agglomeration externality and team productivity. To further identify the strength of skill complementarity and the shape of matching, I use skill distribution in each occupation.

**Average Skill by Occupation and City Bin.** I use average skill by city size bins in each occupation to identify $\delta$ and $\lambda$ and $k$. Conditional on agglomeration externality, the density of manager skill in each four city bins informs how skill $\alpha$ changes with city size. It corresponds to the first line of equation (25) where firm productivity is represented in terms of manager skills and parameters. Similarly, the distribution of worker skills in each of four city bins reveals the firm productivity in terms of worker skill $z$. They both come from the same distribution of firm productivity and the same parameters inside teams. Intuitively, a small difference in manager skill will lead to a large difference in location choice when skill complementarity is strong and the matching function is steep, and vice versa. Therefore the strength of complementarity parameter $\lambda$ and matching parameter $k$ are jointly identified from the perspective of two occupations because $\alpha$ and $z$ work through different combination of those factors but end with the same productivity. Together with other sets, the co-location pattern of two occupations identifies $\delta$, $\lambda$ and $k$ through "two sides of the same coin".

This method also addresses some concerns about the hukou. The estimation targets moments of average skills in four broad city size bins. As long as the team locates inside the bin, it contributes the same set of information to estimation. For instance, whether a skilled manager locates in Beijing or Nanjing does not affect the average skills among the top city bin. Nevertheless, future work need to focus more on the role of trade cost and migration cost.

### 1.4.4 Implementation

Details on the procedures are included in the appendix. Here I provide a sketch of the procedure. I draw a sample of 100,000 managers for each sector and simulate their location choices among 200 bins along normalized city size from 1 to 104. I then draw $200\times 100{,}000$ idiosyncratic shocks for each team and city size bin and calculate manager utility for each city bin based on equation (26).[30] Managers then make discrete choices among 200 city size bins. I do the same for a sample of 100,000 workers to get the last four moments on worker skill above. Notice that these managers and workers are sampled from their own perspectives with a shadow partner linked through matching functions.

I then obtain the loss function using difference in 24 moments. The estimates $\hat{\Theta}_s$ minimizes

[30] Following Gaubert (2018), the shock is assumed to be bin-specific rather than city-specific.

the loss function

$$||m_s - \hat{m}_s(\Theta_s)||_{W_s} = (m_s - \hat{m}_s(\Theta_s))^T W_s (m_s - \hat{m}_s(\Theta_s))$$

where $m_s$ is the vector of moments computed from the raw data in sector $s$, and $\hat{m}_s(\Theta_s)$ is the vector of moments obtained from the simulated world using parameter $\Theta_s$. $W_s$ is the weighting matrix from variance covariance estimator.[31] I combine several global optimization algorithms (simulated annealing, surrogate optimization) and local minimization to search for the global optimized parameters in the eight-dimension space.

## 1.5 Estimation Results

### 1.5.1 Results and Model Fit

I present results on service sectors here and include results for manufacturing in the appendix. Table 1.3 shows the estimated parameters for 18 service sectors. Column (1) and (2) report the sorting agglomeration externality and traditional externality. Column (3) reports the variance of $\log \alpha$, the manager skills. Sectors with higher variance include retails and traditional service. In contrast, business service and scientific service have a variance close to zero, indicating homogeneity in manager skills in these sectors.

Turning to the internal structure, column (5) and (6) provide the relative importance of individual skill and teamwork. In particular, the larger $\lambda$ is, the stronger the importance of teamwork is. Interestingly, top sectors in this category include software, finance, business service and scientific service. When $\lambda$ gets close to zero, team structure breaks as workers and managers tend to work independently as in retails, hotels and catering and traditional services.

Finally, column (7) reports the matching shape parameters between two occupations. All

[31]The optimal weighting matrix is the inverse of the two-step variance covariance matrix, where the first step uses identity matrix as weight.

**Table 1.3**: Calibrated Parameters: Service Sectors

| | (1) $\hat{c}$ | (2) $\hat{e}$ | (3) $\hat{\nu}_\alpha$ | (4) $\hat{\nu}_\varepsilon$ | (5) $\hat{\delta}$ | (6) $\hat{\lambda}$ | (7) $\hat{k}$ | (8) $\hat{\nu}_z$ |
|---|---|---|---|---|---|---|---|---|
| Transportation | 0.202 | 0.006 | 0.642 | 1.969 | 1.085 | 0.382 | 0.029 | 0.038 |
| Postal and warehousing | 0.184 | 0.001 | 0.686 | 2.421 | 1.283 | 0.627 | 0.048 | 1.022 |
| Telecommunication services | 0.173 | 0.260 | 1.242 | 1.978 | 1.658 | 0.250 | 0.148 | 0.078 |
| Computer services | 0.304 | 0.169 | 1.079 | 1.672 | 0.986 | 0.044 | 0.078 | 1.343 |
| Software | 0.208 | 0.516 | 0.102 | 10.101 | 0.875 | 1.724 | 0.488 | 1.102 |
| Wholesale | 0.253 | 0.081 | 1.264 | 1.884 | 0.062 | 0.289 | 0.001 | 1.405 |
| Retails | 0.048 | 0.190 | 2.082 | 1.734 | 0.273 | 0.001 | 0.135 | 0.496 |
| Hotels and catering | 0.230 | 0.008 | 0.000 | 1.786 | 0.001 | 0.001 | 0.001 | 0.253 |
| Finance | 0.136 | 0.263 | 0.645 | 0.793 | 4.156 | 1.126 | 0.360 | 2.552 |
| Leasing | 0.349 | 0.018 | 0.137 | 2.391 | 6.325 | 0.253 | 0.039 | 2.321 |
| Business services, non IT | 0.398 | 0.075 | 0.000 | 6.166 | 6.426 | 1.018 | 0.098 | 1.648 |
| Scientific service | 0.267 | 0.306 | 0.000 | 9.853 | 1.326 | 1.815 | 0.275 | 0.783 |
| Water, environment management | 0.166 | 0.001 | 1.067 | 1.820 | 1.332 | 0.050 | 0.358 | 0.547 |
| Traditional service | 0.417 | 0.044 | 2.279 | 3.044 | 0.001 | 0.001 | 0.407 | 0.539 |
| Education | 0.238 | 0.001 | 1.311 | 2.224 | 1.413 | 0.012 | 0.006 | 3.113 |
| Health | 0.085 | 0.221 | 0.660 | 1.589 | 0.540 | 0.996 | 0.001 | 2.352 |
| Media | 0.355 | 0.155 | 0.760 | 9.941 | 2.080 | 0.321 | 0.244 | 0.820 |
| Sports and entertainment | 0.221 | 0.078 | 0.541 | 7.509 | 0.853 | 1.639 | 0.134 | 0.523 |

Note: This table contains calibrated values of parameters for 18 service sectors through the simulated method of moments. Estimation is conducted sector by sector. The standard errors are omitted. $c$ and $e$ are the team sorting agglomeration and non-sorting agglomeration externality; $\lambda$ is the complementarity intensity; $k$ is the matching parameter; $\delta$ is the managerial technology; $\nu_\alpha$ is the variance of manager productivity; $\nu_z$ is the variance of worker productivity; $\nu_\varepsilon$ is the variance of idiosyncratic shock.

estimates are smaller than one, which is consistent with the model specification. A larger $k$ implies higher concentration of skill in the sector. Leading sectors with high $k$ include software and traditional service. They both have smaller skill dispersion but likely at different levels. These results are broadly consistent with the intuition for most sectors.

In Figure 1.17 to 1.22, I report the model fit for all sets of moments. For each set of moments, the red solid lines connect four data moments. The blue dashed lines connect four simulated moments. For most sectors, the model tracks well with the data moments, especially for the first four sets of moments.[32]

I also show statistics that are not directly targeted – employment size. In particular, I

[32]For the moments on manager skill and workers skill, some sectors do not match well because the mean of distribution is fixed rather than sector specific to save dimensions of parameters.

show that the right tails of firm employment size distribution in the simulated equilibrium follow that of actual data. I focus on the right tails because the measurement error are small only when the labor quantity $l$ in the model is sufficiently larger. I provide detailed discussion on this issue in the appendix. Figure 1.23 show the actual distribution of employment using a random 10% sample from each sector. Figure 1.24 shows the histogram of simulated equilibrium employment truncated from the left at the median. The figure shows similar long right tails to the data. Overall, the results suggest a good fit of model to firms' economic activities.

### 1.5.2 Quantifying Skill Complementarity

Equipped with the estimated parameters, I quantify the relative importance of complementarity. Regressing the log of firm productivity on log of normalized city size yields a key measure of agglomeration: the elasticity of productivity to city size. To be consistent with Section 2, I measure productivity by simulated value-added and estimate the following regression.

$$\log \tilde{\psi}_{is} = \beta_0 + \beta_1 \log \tilde{L}_i + \xi_s + \varepsilon_i, \tag{1.27}$$

where $\tilde{\psi}_{is}$ is the simulated value-added of team $i$ in sector $s$. $\tilde{L}$ is the city size that the team chooses. $\xi_s$ is the sector fixed effects. The estimate $\beta_1$ is the elasticity of productivity to city size and a general measure of agglomeration. Using simulated data, the coefficient is 11.2%. The magnitude is higher than the range of existing measure of agglomeration externalities 3-8% (Rosenthal and Strange, 2004). The results suggest stronger agglomeration externalities among Chinese cities based on observed differences in productivity.

These productivity gains are driven by many sources. To measure their relative importance, I conduct counterfactual simulations shutting down different mechanisms one at a time. In particular, I first simulate the model with teams constrained to choose their city size as if they all had the average team productivity. In this case, the location choices are only driven by idiosyncratic shocks at the firm-level. The estimated elasticity drops to 7.5%, which means the overall firm sorting accounts for 33% of the aggregate productivity gaps. This share of systematic sorting is lower than the 50% found in Gaubert (2018) among French firms, probably due to higher migration barriers in China.

Within this 33% gaps explained by sorting, the key question of the paper is how much is caused by complementarity between managers and workers. To answer that question, I shut down complementarity while allowing heterogeneous managers to optimize their locations without partners. The city location choices are then driven by idiosyncratic shocks and heterogeneity of managers. As expected, the coefficient bounces back from 7.5% to 9.2%, meaning that the complementarity component is around 2 percent (11.2%-9.2%), which explains about 18% of the total agglomeration elasticity. It should be noted that the estimates also capture firm institutions that indirectly affect productivity through skill mix. The magnitude is not trivial compared to quantification of other factors. For example, Spanos (2019) estimates that difference in firm hierarchy accounts for 8.8%-22.4% of the productivity gaps in France. Tian (2019) estimates that labor division accounts for 15% of the gaps in Brazil.

Putting the magnitude of this effect into the sorting mechanism, then around 56% of the sorting mechanism comes from within-firm skill complementarity.[33] In other words, half of the sorting mechanism is attributed to the capability of better workers co-working and co-locating with better managers. To the best of my knowledge, this paper is the first one to structurally quantify the role of skill complementarity in explaining observed productivity gaps across cities.

## 1.6 Policy Analysis: Subsidizing Talents

The calibrated model is useful to conduct policy analysis through simulation. I apply it to study the trending place-based policies in China: subsidizing talents for residing in second-tier

[33] The share of complementarity within sorting is calculated by $(11.2\% - 9.2\%)/(11.2\% - 7.5\%) = 55.6\%$.

cities.

Recognizing the importance of human capital, dozens of cities in China have launched place-based policies to attract skilled individuals to their cities since 2017. By offering educated people income and housing subsidies, cash bonus and tax deductions, local governments attempt to influence the locational choice of individuals. For example, Changsha, the capital of Hunan province, offers housing and living subsidies of 6,000 to 15,000 yuan per year ($900 - $2200) to fresh college graduates moving to the city. Those with doctoral degrees can get 60,000 yuan in subsidies when they purchase their first house in the city. Hangzhou, the capital city of Zhejiang province, offers master's and doctoral graduates a lump-sum subsidy of 20,000 yuan and 30,000 yuan respectively, if they work or start up in Hangzhou within a year after graduation. The list continues and more cities are joining the "race for talent".

Unlike subsidizing talents at the country level, these policies are very local in terms of the amounts and forms of subsidies. In addition, the main targets of those policies are young cohorts who are well-educated and geographically mobile. Nevertheless, these policies could potentially affect millions of people. In 2019, 8.2 million college graduates entered the labor market and sorted across the cities. Local policymakers hope to sustain economic growth by attracting those skilled people to reside, invest and consume.

However, little is known about the general equilibrium effects of such policies on the overall economy and its efficiency. People on the margin would have different agglomeration gains at different places depending on the size of cities. Therefore, the aggregate effect on the overall economy is ambiguous and potentially heterogeneous across cities. Moreover, in the long run, whether the city could effectively retain the talents is unclear. With multiple cities completing for human capital at the same time, there is a risk of "race to the bottom" such that no place actually draws more talents in the new equilibrium. If not carefully designed, these policy may lead to undesirable effects and misallocation that outweighs the gains.

The calibrated model provides an opportunity to conduct counterfactual analysis of such policies. In particular, I consider policies issued in 32 second-tier cities in China that account for about 20% of the total population. The size of these cities ranges from 5 million (Kunming) to 7 million (Hangzhou). As a reference point, the baseline subsidy is set to 25% of individual income and the targeted education level is four-year college and above (16 or more years of schooling). I focus on the flow of new cohort rather than existing residents to match the intention of policies. I also assume that the cost of policy is relatively small comparing to the overall government budget so there is no direct burden on current residents in the city.

The simulation is a short-run counterfactual case where teams are relocated but the matching and occupation are fixed. I group cities into five quantiles where the second quantile (Q2) is treated. I focus on two outcomes for each quantiles: new employment and the number of incoming firms. For the entire cohort, I also consider the total output as a measure of welfare for the group because the total population are fixed. In the future work, I will consider the impact of such policies on existing firms and individuals.

**Local Effects.** In the short run, the policy benefits cities that subsidize talents. In targeted cities, the average number of firms attracted increases by 1.3%. In other words, compared to the case without subsidies, the number of incoming firms (managers) increases by 1.3% at the cost of other cities' losses. Figure 1.7 shows the percent changes in number of incoming firms by five quantiles, where the second quantile (Q2) is treated.

On average, larger cities (Q3-Q5) suffer more than smaller cities in terms of percentage losses. But since the changes are benchmarked to new cohorts (flow) rather than total populations (stock), the overall effect on number of firm may be smaller in larger cities. In Figure 1.29 and 1.30 in the appendix, I vary the level of subsidy to 10% and 50% respectively. The local impact on the number of incoming firms ranges from 0.55% to 2.7%.

Incentivizing managers not only relocate firms but also relocate matched workers. I calculate the equilibrium employment in each city size quantile with and without subsidies. Figure 1.8 shows the percent changes on new employment. The treated cities attract about 2.5%

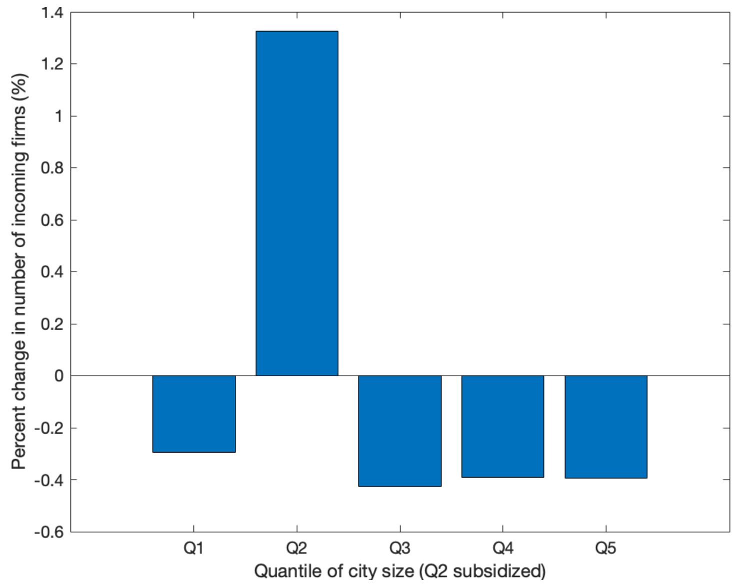



**Figure 1.7**: Effect of a 25% Subsidy on the Number of Incoming Firms by City Quantiles

Note: The figure shows percentage changes in number of incoming firms in each quantile of city size distribution. College graduates locating in Q2 are offered with a 25% income subsidy.

more employment with subsidies. As a back-of-the-envelope calculation, for a city of 6 million population that suppose to attract 0.5 million new employment before. The policy effect on total population is then 0.2%.

**Welfare.** By attracting skilled people to middle sized cities, the policy may create distortion and efficiency loss. I calculate the aggregate output by these new employment in the economy under different level of subsidies. Figure 1.9 shows the welfare loss in percentage as a function of levels of subsidies.

The simulation shows that these place-based policies lower welfare. Subsidies in second-tier cities that amounts to 25% of local income lead to a 0.77% loss in aggregate output produced by all new employment. Interestingly, the welfare loss is quite small at a lower subsidy level (<20%) presumably because the effect is dominated by agglomeration gains of relocating firms from smaller cities to larger ones. Recall that the competitive equilibrium is suboptimal. The output may actually increase if firms in small cities could relocate in larger ones. As subsidy

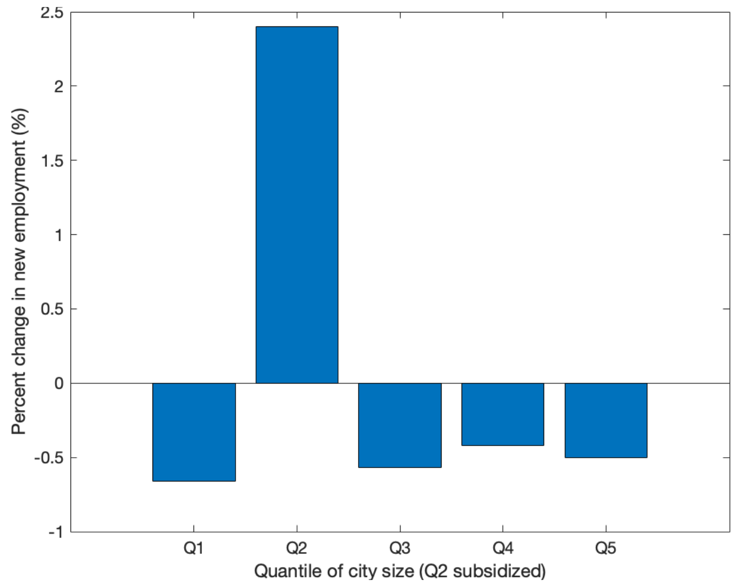



**Figure 1.8**: Effect of a 25% Subsidy on New Employment by City Quantiles

Note: The figure shows percentage changes in number of new employment in each quantile of city size distribution. College graduates locating in Q2 are offered with a 25% income subsidy.

increases, it starts to drag down firms in larger cities, which leads to welfare losses. The non-linear welfare loss suggests that an aggressive subsidy should be carefully examined. As migration barriers keep falling, the trade-off between regional equity and overall efficiency may become more intense.

Note that this welfare loss only comes from locational displacement of firms, and does not include the mismatching that may arise in local labor market. Moreover, with fixed number of cities, changes in population alter the city size thus agglomeration forces. Intuitively, as second-tier cities grow larger, they become more attractive to people who are eligible for the subsidy. Changes in the matching pattern may also lead to further relocation. The long-run aggregate impact of these policies gets complicated by the dynamics. In addition, other cities may respond with similar policies to counter the brain drains. The race for talents could lead to outcomes that may or may not be socially or politically desirable depending on the counterfactual state.

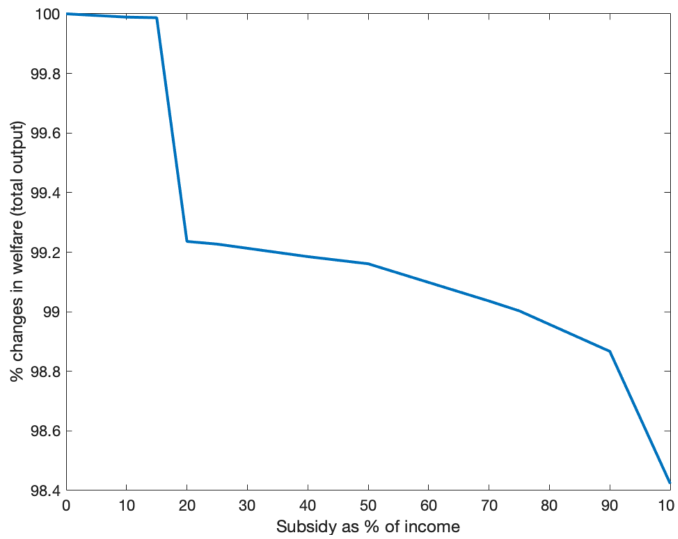



**Figure 1.9**: Effect of Income Subsidy on Aggregate Output by Subsidy Level

Note: The figure shows simulated percentage changes in welfare (output) at different levels of subsidy.

Overall, as a first attempt to evaluate the impact of those place-based policies in China, the results call for careful evaluation of the policies and more research on the optimal spatial designing at the top level. Such policies should be carefully evaluated and compared with other options such as improving public service, lowering congestion cost and reducing frictions in the labor market.

## 1.7 Conclusion

The clustering of people and economic activity is an important driver of development and inequality. I model a spatial economy with structures on internal organization that features a key nature of production: skill complementarity at the workplace, and I quantify its role in driving cross-city productivity gaps in China. I develop a spatial general equilibrium model with heterogeneous agents who simultaneously choose occupations, match with partners, and sort across cities in teams. The spatial allocation of skills, the organization of production, and the size distribution of cities are jointly determined in the equilibrium. The model highlights skill complementarity at the workplace as an important driver of spatial skill concentration that augments productivity gaps across cities with agglomeration economies.

Using firm census data that contains skill mixes at the firm-level, I find supporting evidence that skilled workers co-locate and co-work in larger cities. I structurally estimate the model through simulated method of moments and conduct counterfactual analysis. I find that skill complementarity accounts for a significant share of productivity gaps, about 18% of the overall agglomeration, or equivalently 56% of the systematic sorting of firms. The results provide the first structural estimate on the role of within-firm skill complementarity in spatial sorting and cross-city productivity gaps in China. Through the calibrated model, I evaluate the general equilibrium effect of recent policies among second-tier cities in China: subsidizing college graduates to attract talents. The simulation shows local gains in new employment at the expense of other cities and overall efficiency across the country, suggesting an equity-efficiency trade-off in a spatial economy.

Urbanization will continue to be one of the key themes for many countries. The locational choice of people and firms is going to shape the future of cities. For policymakers, it is crucial to understand the incentives behind sorting and how agents are economically organized and related in a spatial economy. Moreover, it is crucial to match population densities with appropriate skill composition, production structure and public service. The government should lower the barrier of forming efficient allocation of human capital so that the economic and social benefits of urbanization could be maximized and more widely shared.

## 1.8 Appendix

### 1.8.1 Figures and Tables

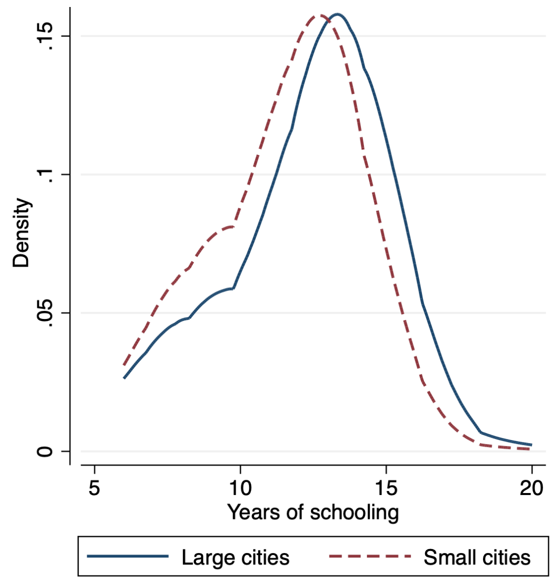



**Figure 1.10**: Distribution of Weighted Years of Education by City Size

Note: Kernel density of weighted years of schooling, service sectors only. Bandwidth =1. Larger cities are defined by over 4 million population.

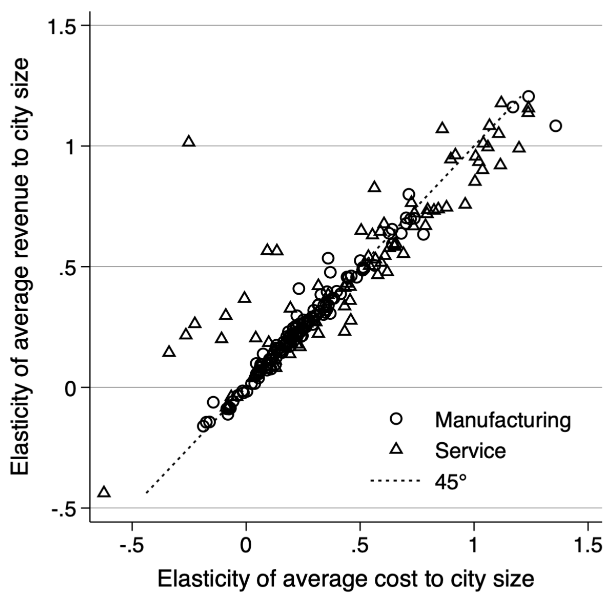



**Figure 1.11**: Elasticity of Mean Revenue and Cost to City Size

Note: Both manufacturing and service sectors are included. Each dot represents a sector $s$. Elasticity of variable $y$ to city size is $\hat{b}$ from $\log \text{mean} y(L_i) = a + b \log L_i + \varepsilon_i$, ran by each sector of three-digit industry code with more than 500 observations. Regressions are weighted by the number of firms in city-sector cell.

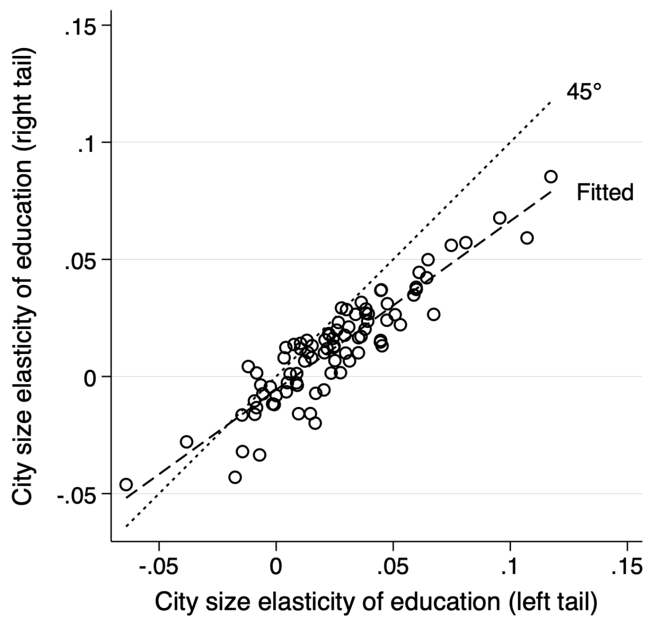



**Figure 1.12**: City Size Elasticity of Education at Two Tails: Alternative Proxies

Note: Service sectors only. Managers and workers are proxied by 25th and 75th percentiles of ranked employees in education respectively. Each dot represents a sector $s$. Elasticity of variable $y$ to city size is $\hat{b}$ from $\log \text{mean} y(L_i) = a + b \log L_i + \varepsilon_i$, ran by each service sector of 3-digits industry code with more than 500 observations. Regressions are weighted by the number of firms in city-sector cell. 28 of 115 sectors have negative elasticities for managers, and 18 have negative elasticities for workers.

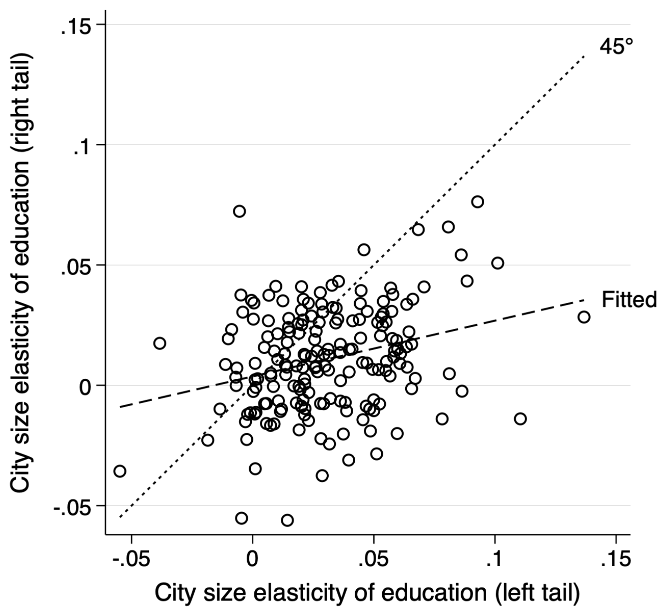



**Figure 1.13**: City Size Elasticity of Education at Two Tails: Four-Digit Sectors

Note: Service sectors only. Each dot represents a four-digit sector. Elasticity of variable $y$ to city size is $\hat{b}$ from $\log \text{mean}\, y(L_i) = a + b \log L_i + \varepsilon_i$, ran by each service sector of four-digit industry code with more than 500 observations. Regressions are weighted by the number of firms in city-sector cell. Of 215 sectors, 67 have negative elasticities for managers, 26 have negative elasticities for workers.

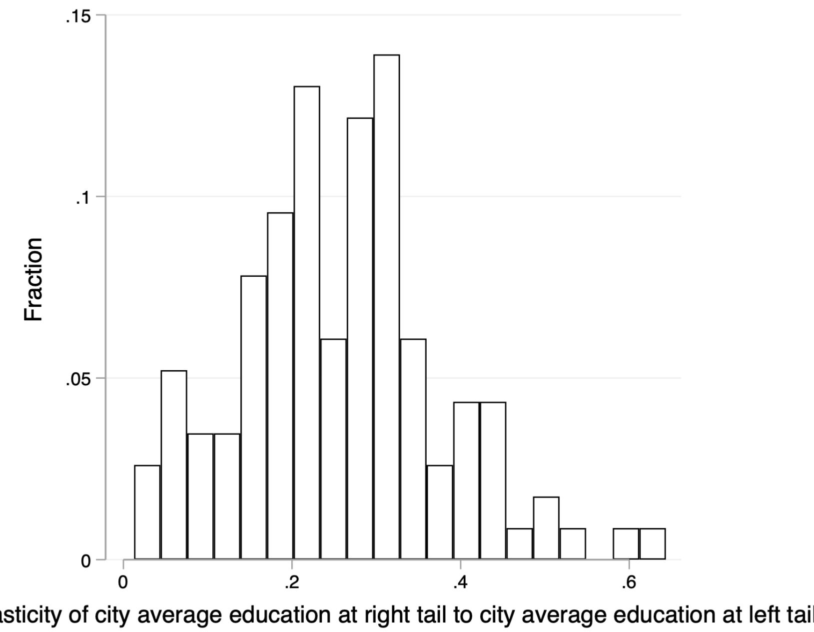



**Figure 1.14**: Histogram of Elasticity of City Average Right-Tail to Left-Tail Education

Note: Elasticity is obtained from $\hat{b}$ of regression: $\log \text{mean}\, m(L_i) = a + b \log \text{mean}\, w(L_i) + \varepsilon_i$, for each three-digit service sector with more than 500 firm observations.

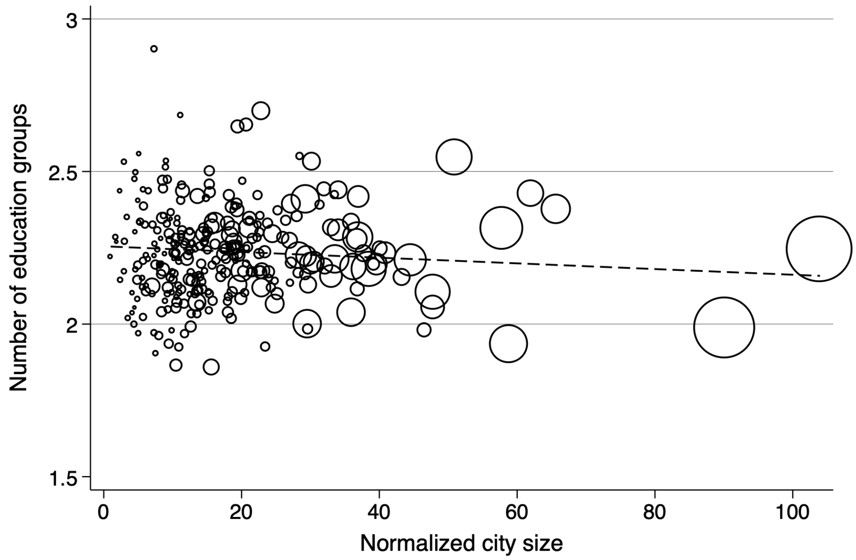



**Figure 1.15**: Skill Dispersion and City Size

Note: This figure plots the interquartile range of log average wages distribution for each city, N=287. The size of circle represents the relative number of firms in each city. Dashed line is the fitted OLS weighted by firm numbers. Both manufacturing and service sectors are included.

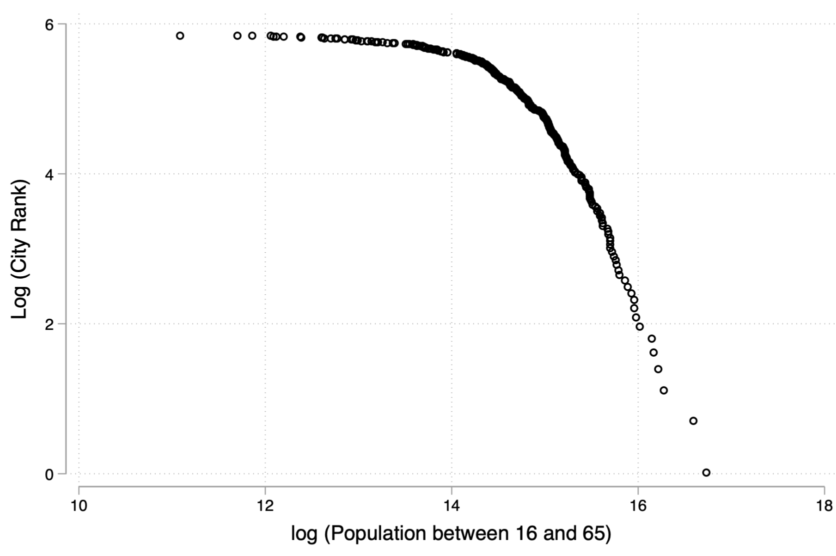



**Figure 1.16**: City Size and Rank

Note: This figure shows the city size (measured by working age population) and city rank for 287 cities in the data.

**Table 1.4**: Geographic Distribution of Firms: Service and Manufacturing

| Province | Observations | Percent |
|---|---|---|
| Guangdong | 310,846 | 11.04 |
| Jiangsu | 279,752 | 9.93 |
| Shandong | 249,297 | 8.85 |
| Shanghai | 235,832 | 8.37 |
| Zhejiang | 231,900 | 8.23 |
| Beijing | 204,991 | 7.28 |
| Liaoning | 144,682 | 5.14 |
| Hubei | 102,758 | 3.65 |
| Henan | 97,818 | 3.47 |
| Fujian | 90,790 | 3.22 |
| Sichuan | 87,084 | 3.09 |
| Hebei | 83,178 | 2.95 |
| Tianjin | 77,347 | 2.75 |
| Anhui | 69,511 | 2.47 |
| Hunan | 66,936 | 2.38 |
| Heilongjiang | 58,817 | 2.09 |
| Shaanxi | 51,779 | 1.84 |
| Shanxi | 49,862 | 1.77 |
| Jilin | 47,274 | 1.68 |
| Guangxi | 45,728 | 1.62 |
| Chongqing | 45,438 | 1.61 |
| Inner Mongolia | 39,300 | 1.40 |
| Jiangxi | 35,493 | 1.26 |
| Yunnan | 29,866 | 1.06 |
| Gansu | 22,985 | 0.82 |
| Xinjiang | 18,119 | 0.64 |
| Guizhou | 15,392 | 0.55 |
| Ningxia | 9,925 | 0.35 |
| Hainan | 9,559 | 0.34 |
| Qinghai | 3,623 | 0.13 |
| Tibet | 896 | 0.03 |
| Total | 2,816,778 | 100 |

Note: This table shows the distribution of firms in the sample by provinces.

**Table 1.5**: Summary Statistics: Manufacturing (20% sample)

| | | log value added | | | employment | | |
|---|---|---|---|---|---|---|---|
| | N | mean | p25 | p75 | median | p25 | p75 |
| Processing of agriculture products | 21,315 | 7.94 | 7.24 | 9.27 | 62 | 34 | 128 |
| Manufacture of food products | 7,690 | 8.13 | 7.34 | 9.39 | 85 | 45 | 183 |
| Manufacture of beverages and tobacco products | 5,176 | 8.58 | 7.55 | 9.78 | 85 | 42 | 186 |
| Manufacture of textiles | 32,293 | 7.37 | 6.77 | 8.68 | 85 | 43 | 180 |
| Manufacture of wearing apparel | 17,535 | 7.50 | 6.89 | 8.65 | 142 | 79 | 260 |
| Manufacture of leather goods and footwear, leather tanning | 8,365 | 7.74 | 6.99 | 8.83 | 125 | 65 | 278 |
| Manufacture and products of wood, except furniture | 9,782 | 7.59 | 6.95 | 8.81 | 78 | 45 | 140 |
| Manufacture of furniture | 5,182 | 7.68 | 7.06 | 8.87 | 89 | 50 | 182 |
| Manufacture of pulp, paper, and paper products | 9,641 | 7.47 | 6.81 | 8.83 | 72 | 40 | 140 |
| Publishing, printing, and reproduction of recorded media | 6,302 | 7.39 | 6.76 | 8.53 | 65 | 38 | 124 |
| Manufacture of arts and sports equipment | 4,699 | 7.54 | 6.95 | 8.69 | 110 | 58 | 247 |
| Petroleum refinery | 2,249 | 7.95 | 7.34 | 10.27 | 65 | 26 | 230 |
| Manufacture of chemicals and chemical products | 25,668 | 7.94 | 7.20 | 9.33 | 56 | 30 | 118 |
| Manufacture of pharmaceutical | 6,207 | 8.63 | 7.72 | 10.00 | 105 | 56 | 215 |
| Manufacture of chemical fiber, rubber and plastic products | 25,401 | 7.46 | 6.83 | 8.72 | 65 | 35 | 135 |
| Manufacture of glass, ceramic, brick, and cement products | 29,065 | 7.83 | 7.19 | 9.18 | 85 | 47 | 166 |
| Manufacture of basic and fabricated metals | 14,016 | 7.11 | 6.74 | 9.46 | 70 | 32 | 160 |
| Manufacture of metal products | 23,730 | 7.54 | 6.89 | 8.70 | 65 | 36 | 128 |
| Manufacture of general machinery | 35,894 | 7.76 | 6.97 | 8.86 | 62 | 35 | 120 |
| Manufacture of specialized machinery | 17,980 | 7.92 | 7.17 | 9.03 | 69 | 38 | 139 |
| Manufacture of transportation machinery | 17,840 | 7.92 | 7.11 | 9.19 | 85 | 45 | 185 |
| Manufacture of electrical machinery | 25,083 | 7.92 | 7.13 | 9.10 | 78 | 40 | 168 |
| Manufacture of telecommunications equipment | 12,872 | 7.74 | 7.22 | 9.46 | 128 | 60 | 320 |
| Manufacture of instrumentation and office machinery | 5,477 | 8.01 | 7.26 | 9.15 | 76 | 40 | 164 |
| Manufacture of artifacts and others | 6,937 | 7.64 | 6.94 | 8.69 | 90 | 48 | 186 |
| Recycling | 1,041 | 6.34 | 6.55 | 8.98 | 44 | 23 | 100 |
| Utilities | 8,001 | 6.23 | 7.19 | 9.75 | 135 | 53 | 335 |
| Total | 385,441 | 7.69 | 7.02 | 9.04 | 79 | 43 | 165 |

Note: This table summarizes value-added and total employment for 27 manufacturing sectors in the sample.

**Table 1.6**: Education Correlation Between Left and Right Tail Workers: Alternative Proxies

| | Years of education (right tail) | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Years of education (left tail) | .540$^{\dagger}$ | .537$^{\dagger}$ | .570$^{\dagger}$ | .536$^{\dagger}$ | .526$^{\dagger}$ | .515$^{\dagger}$ | .537$^{\dagger}$ |
| | [.017] | [.018] | [.017] | [.016] | [.016] | [.016] | [.016] |
| Log (city size) | | .100 | .110$^{*}$ | .071 | | | |
| | | [.052] | [.046] | [.044] | | | |
| Firm-level controls | | | Y | Y | Y | Y | Y |
| Four-digit sector FE | | | | Y | Y | Y | |
| Area FE | | | | | City | County | |
| City $\times$ 2-digit sector FE | | | | | | | Y |
| $R^2$ | .446 | .447 | .618 | .628 | .631 | .637 | .631 |

Note: Service sectors only. $^{\dagger}$ $p < 0.01$, $^{*}$ $p < 0.05$. N=2,394,832. Robust standard errors are in brackets, clustered at province level. Tails are proxied by 25th and 75th percentiles of education. There are 2415 counties in 287 cities and 310 four-digit sectors in 40 big sectors. Firm level controls include number of establishments, open year, ownership type, number of education categories, value-added, total employment, assets and labor cost.

### 1.8.2 Proof

**Proof of Lemma 1**

*Proof.* I first prove that conditioning on being in a city of size $L$, $M_s(\alpha|L)$ is increasing in $\alpha$. Since managers face same congestion and non-tradable prices, maximization is based on operating profit. Suppose two managers in the same sector with $\alpha_1 < \alpha_2$ have corresponding match $M_s(\alpha_1) = z_1$, $M_s(\alpha_2) = z_2$. The choices mean that for manager $\alpha_1$, worker of type $z_1$ is at least better than type $z_2$, $\pi(\alpha_1, z_1|L) > \pi(\alpha_1, z_2|L)$:

$$\frac{f(\alpha_1, z_1)}{f(\alpha_1, z_2)} \geq \frac{w_s(z_1)}{w_s(z_2)}.$$

Similarly, for the other manager

$$\frac{f(\alpha_2, z_2)}{f(\alpha_2, z_1)} \geq \frac{w_s(z_2)}{w_s(z_1)}.$$

Then if $z_2 < z_1$, log-supermudularity and above two inequalities imply

$$\frac{w_s(z_2)}{w_s(z_1)} \leq \frac{f(\alpha_2, z_2)}{f(\alpha_2, z_1)} < \frac{f(\alpha_1, z_2)}{f(\alpha_1, z_1)} \leq \frac{w_s(z_2)}{w_s(z_1)},$$

a contradiction. Therefore, $z_2 \geq z_1$. However, if $z_2 = z_1$, there will be mass points in skill distribution, which contradicts assumption that $\mu_s(\cdot)$ is well behaved. So $\alpha_2 > \alpha_1$. Similarly, there can not be multiple $\alpha$ that map to the same $z$.

To complete the proof, recall that matching schedule in equilibrium is independent from location choice, which means the ranking order of worker types are preserved no matter where managers locate in equilibrium. □

**Proof of Lemma 4**

*Proof.* Since both indirect utilities are continuous and occupation set is non-empty, there must be at least one type that are indifferent between two occupations. I drop subscript $s$ and the common constant terms in (11) and (12), which does not affect the results.

By envelope theorem, the marginal return of manager skill is

$$V'(\alpha|manager) = \pi'(\alpha) = (\sigma-1)\frac{f_1(\alpha,M_\alpha)}{f(\alpha,M_\alpha)}\Lambda Q\left[\frac{f(\alpha,M_\alpha)g(L(\alpha,M_\alpha))}{w(M_\alpha)}\right]^{\sigma-1} > 0.$$

Using FOC (4), the derivative of of worker's utility

$$V'(\alpha|\text{worker}) = \frac{f_2(\overline{M}_\alpha^{-1},\alpha)}{f(\overline{M}_\alpha^{-1},\alpha)}w(\alpha) - \overline{n}\overline{\gamma}\frac{\mathrm{d}L^\phi(\overline{M}_\alpha^{-1})}{\mathrm{d}\alpha},$$

where $\overline{M}^{-1}$ is the inverse of any arbitrary matching function. When type $\alpha$ is indifferent between two occupations, two occupation generates same utility $V(\alpha)$.

$$V'(\alpha|manager) = (\sigma-1)\frac{f_1(\alpha,M_\alpha)}{f(\alpha,M_\alpha)}\left[V+\overline{\gamma}L^\phi(\alpha)\right] = (\sigma-1)\frac{f_1}{f}V + \overline{\gamma}(\sigma-1)\frac{f_1}{f}L^\phi(\alpha),$$

$$V'(\alpha|\text{worker}) = \frac{f_2}{f}\left[V+\overline{n}\overline{\gamma}L^\phi(\overline{M}_\alpha^{-1})\right] - \overline{n}\overline{\gamma}\frac{\mathrm{d}L^\phi(\overline{M}_\alpha^{-1})}{\mathrm{d}\alpha} = \frac{f_2}{f}V + \overline{n}\overline{\gamma}\frac{f_2}{f}L^\phi(\overline{M}_\alpha^{-1}) - \overline{n}\overline{\gamma}\frac{\mathrm{d}L^\phi(\overline{M}_\alpha^{-1})}{\mathrm{d}\alpha}.$$

With assumption 3 1) and $\sigma > 2$, the first term $(\sigma-1)\frac{f_1}{f}V > \frac{f_2}{f}V$. To ensure single-crossing, a sufficient condition is

$$\overline{\gamma}(\sigma-1)\frac{f_1}{f}L^\phi(\alpha) > \overline{n}\overline{\gamma}\frac{f_2}{f}L^\phi(\overline{M}_\alpha^{-1}),$$

or equivalently

$$\overline{n} < \frac{L^\phi(\alpha)}{L^\phi(\overline{M}_\alpha^{-1})}.$$

By lemma 2, $L^\phi(\cdot)$ is strictly increasing. Since $\alpha$ is bounded, the right hand side is thus bounded. It then implies that a sufficient condition is

$$\overline{n} < \frac{L^\phi(\underline{\alpha})}{L^\phi(\overline{\alpha})}.$$

Together, assumption 3 guarantees single crossing of indirect utility function, and there can be only one such cutoff point in each sector. $\square$

### Proof of Proposition 1

*Proof.* The proof is straightforward following lemma 1 and lemma 2 using monotonicity of optimal choice function. Operating profit are higher for high skill managers because utility are increasing despite increasing land cost and cost of living. Therefore, it must increase with skill. $\square$

### Proof of Proposition 2

*Proof.* Since occupation does not change in the short run, cutoffs does not change. Matching function starts and ends at the same points as before. If the new matching curve $\overline{M}$ goes above $M$, there exists $\hat{\alpha}_s \leq \alpha_1 < \alpha_2 \leq \overline{\alpha}_s$ such that

a. $\overline{M}_{\alpha_1} = M_{\alpha_1} = z_1$ and $\overline{M}_{\alpha_2} = M_{\alpha_2} = z_2$;

b. $\overline{M}'_{\alpha_1} \geq M'_{\alpha_1}$ and $\overline{M}'_{\alpha_2} \leq M'_{\alpha_2}$;

c. For all $\alpha \in [\alpha_1, \alpha_2]$, $\overline{M}_\alpha \geq M_\alpha$.

Result *b* implies that

$$\frac{\overline{M}'_{\alpha_2}}{\overline{M}'_{\alpha_1}} \leq \frac{M'_{\alpha_2}}{M'_{\alpha_1}}.$$

By (15), it means

$$\frac{\bar{\mu}_{\alpha_2}\bar{\mu}_{z_1}}{\bar{\mu}_{\alpha_1}\bar{\mu}_{z_2}} \frac{\bar{w}_s^{\sigma_s}(\overline{M}_{\alpha_1})}{\bar{w}_s^{\sigma_s}(\overline{M}_{\alpha_2})} \leq \frac{\mu_{\alpha_2}\mu_{z_1}}{\mu_{\alpha_1}\mu_{z_2}} \frac{w_s^{\sigma_s}(M_{\alpha_1})}{w_s^{\sigma_s}(M_{\alpha_2})}.$$

By MLRP, this equation implies $\bar{w}(z_1)/\bar{w}(z_2) \leq w(z_1)/w(z_2)$. However, it can not happen by (5) and c. and strictly log super-modularity of $f$. Therefore $\overline{M}_\alpha$ must lie below the old one. $\square$

**Proof of Proposition 5**

*Proof.* I derive city size distribution for a single sector. The proof follows Behrens et al. (2014). Because population clears type by type, the mass of city chosen by a $\{\alpha, M_\alpha\}$ team is

$$n_s(\alpha) = \frac{\mathbb{L}_s[\mu_s(\alpha) + \mu_s(M_s(\alpha))]\,\mathrm{d}\alpha}{L_s(\alpha)} + \bar{n}.$$

Define $\psi_s(\alpha) = \frac{f(\alpha, M_\alpha)}{w_s(M_\alpha)}$. Notice that $\psi_s(\alpha)$ is strictly increasing in $\alpha$, so it is a one-to-one mapping from $\alpha$ to $\psi$, then I can construct a new probability distribution function $\hat{\mu}(\psi)$ and define the share of $\psi$-type team in sector population as $\hat{F}_s(\psi)$. So the mass of city for team $\psi$ is

$$n_s(\psi) = \frac{\mathbb{L}_s \hat{F}_s(\psi)}{L_s(\psi)} + \bar{n}.$$

Assume the same functional form as in estimation $g_s(L) = L^{c_s}$. The optimal city size in equilibrium is solved by

$$L_s(\psi) = \left[\frac{c_s(\sigma_s - 1)\Lambda_s Q_s}{\gamma\phi} \psi_s^{\sigma_s - 1}\right]^{\frac{1}{\phi - c_s(\sigma_s - 1)}}.$$

Invert the function and substitute $\psi$ with $L_s(\psi)$ in $n_s(\psi)$

$$\mathbb{L}_s \frac{F_s(\psi)}{L_s(\psi)} = \mathbb{L}_s \frac{\hat{F}_s(\hat{\Lambda}_s L_s(\psi)^{\hat{\phi}_s})}{L_s(\psi)},$$

with $\hat{\Lambda} \equiv \left[\frac{\gamma\phi}{c_s(\sigma_s - 1)\Lambda_s Q_s}\right]^{1-\sigma_s}$ and $\hat{\phi}_s = \frac{\phi}{\sigma_s - 1} - c_s$. Change of variables leads to the density of city size,

$$n_s(L) = \frac{\mathbb{L}_s \hat{\phi} \hat{\Lambda}_s}{C_s} \hat{F}_s(\hat{\Lambda}_s L^{\hat{\phi}_s}) L^{\hat{\phi}_s - 2} + \bar{n},$$

where $C_s$ is the total mass of cities in this sector, and the accumulative distribution of cities

$$N_s(L) = \frac{\mathbb{L}_s \hat{\phi} \hat{\Lambda}_s}{C_s} \int_0^L \hat{F}_s(\hat{\Lambda}_s q^{\hat{\phi}_s}) q^{\hat{\phi}_s - 2} \, \mathrm{d}q + L\bar{n}.$$

Following empirical evidence suggested in Behrens et al. (2014) that $\hat{\phi}_s$ is close to zero, a first-order approximation around $\hat{\phi}_s \approx 0$ is

$$n_s(L) = \frac{\mathbb{L}_s \hat{\phi}_s \hat{\Lambda}_s}{C_s} \hat{F}_s(\hat{\Lambda}_s) L^{-2} + \bar{n}.$$

When $\bar{n}$ is close to zero and agglomeration and congestion forces are close, the city size distribution follows a Pareto distribution with shape parameter 1. Using market clear condition, the equilibrium mass of cities can be solved,

$$C_s = \hat{\phi}_s \hat{\Lambda}_s f(\hat{\Lambda}_s) [\ln L(\bar{\phi}) - \ln L(\underline{\phi})] \mathbb{L}_s.$$

Finally, the aggregate city size distribution is a weighted sum of all sectors distribution. The weight is the sector share of population. The total mass of cities $C = \sum_s C_s$. □

### 1.8.3 Extensions of the Model

#### Costly Trade

In this extension, I consider costly trade. For simplicity, I illustrate with one sector and ignore $\bar{n}$. With trade cost, the price of variety produced $j$ at destination $k$ is further marked up by an iceberg cost $\tau_{jk}$, where $\tau_{jk} = 1$ when $j = k$. The local price index at destination $k$ is then

$$P_k = \left[ \int_{j \in C} \int_{\alpha \in \Omega(j)} \left( \frac{\tau_{jk} w(M_\alpha)}{f(\alpha, M_\alpha) g(L_j)} \right)^{1-\sigma} \tilde{\mu}(\alpha, j) \, \mathrm{d}\alpha \, \mathrm{d}j \right]^{\frac{1}{\sigma - 1}},$$

where $\Omega(j)$ is the set of managers (teams) that located in site $j$. $\tilde{\mu}(\alpha, j)$ is the probability density of team $\alpha$ in city $j$ redefined on managers set.

Let $\psi(\alpha, L_j) = \frac{f(\alpha, M_\alpha) g(L_j)}{w(M_\alpha)}$. Let city aggregate productivity

$$\Psi_j = \left[ \int_{\alpha \in \Omega(j)} \psi(\alpha, L_j)^{\sigma-1} \tilde{\mu}(\alpha, j) \, \mathrm{d}\alpha \right]^{\frac{1}{\sigma-1}}.$$

Then

$$P_k = \left[ \int_{j \in C} \left( \frac{\tau_{jk}}{\Psi_j} \right)^{1-\sigma} \mathrm{d}j \right]^{\frac{1}{\sigma-1}}.$$

From the perspective of a team of type $\alpha$ in site $k$, it has marginal cost $\frac{\tau_{kj}}{\psi(\alpha, L_k)}$ when selling to site $j$. Then the team's demand from city $j$ is

$$r_{ij}(\alpha) = \left( \frac{\tau_{kj}}{\psi(\alpha, L_k)} \right)^{1-\sigma} R_j P_j^{\sigma-1}$$

Total demand of city $j$ is $R_j$. Then the profit, if located in $k$, is

$$\pi(\alpha, k) = \frac{1}{\sigma} \int_{j \in C} \left( \frac{\tau_{kj}}{\psi(\alpha, L_k)} \right)^{1-\sigma} R_j P_j^{\sigma-1} \, \mathrm{d}j - C(L_k)$$

Let $k$'s market access be

$$MA_k = \int_{j \in C} \tau_{kj}^{1-\sigma} R_j P_j^{\sigma-1} \, \mathrm{d}j$$

Then

$$\pi(\alpha, k) = \frac{1}{\sigma} \psi(\alpha, L_k)^{\sigma-1} MA_k - C(L_k).$$

Recall that $\alpha$ and $L_k$ are separable in $\psi(\alpha, L_k)$ so that $\pi(\alpha, k)$ are supermodular in $\alpha$ and $L_k$. Therefore, in equilibrium with regularity conditions, more skilled teams locate in larger cities. Suppose not, there exist two teams $\alpha_1 < \alpha_2$ such that $\alpha_1$ chooses $j_1$ and $\alpha_2$

chooses $j_2$ and $L(j_1) > L(j_2)$. The revealed preference means $\pi(\alpha_1, L(j_1)) > \pi(\alpha_1, L(j_2))$ and $\pi(\alpha_2, L(j_2)) > \pi(\alpha_2, L(j_1))$. Notice that market access does not enter first order condition thus matching function, so teams can be index by manager skill $\alpha$.

By supermodularity of $\pi(\alpha, L_j)$ and sum of two inequalities above,

$$\pi(\alpha_2, L(j_1)) + \pi(\alpha_1, L(j_2)) > \pi(\alpha_1, L(j_1)) + \pi(\alpha_2, L(j_2)) > \pi(\alpha_2, L(j_1)) + \pi(\alpha_1, L(j_2)),$$

a contradiction. Thus, all lemmas and propositions still hold.

By introducing the trade cost, estimation requires additional trade information between cities. Even with the available data, it complicates the model with enormous parameters and abstract attentions from complementarity. Therefore, I choose to leave it out from estimation.

**Education as a Coordination Device**

Each potential city site has one city developer. They serve as the coordinator in the model that prevent individuals from coordination failure. Recall that in the baseline mode, I assume that individuals pay a sunk cost to get skill draws. I model such process in a two-generation style where the sunk cost of next generation is paid by incumbent residents in the city. With city developers, I delegate the education to each site where city developers announce the size of city as reserved seats for education and collect the total cost as a fixed share $F^e$ of local GDP. Then education budget differs across cities. Formally, the revenue from educating $L$ students is $F^e \bar{y} L$ where $F^e$ is the fixed share, $\bar{y}$ is the per capita output and $L$ is the city size. I assume such sunk cost does not enter the optimization decisions for the current generation.

Once the city developer claims the size of city, they also issue an ad-valorem subsidies at rate $D_s(L)$ to managers and workers coming to the city. The developers' objective is to maximize revenue net of subsidy paid out,

$$\begin{aligned}\Pi_L = F^e \bar{y} L - \sum_{s=1}^{S} \int_{\alpha} D_s(L) \tilde{\pi}_s(\alpha, M_\alpha, L) 1_s(\alpha, \text{manager}, L) \mu_s(\alpha) \mathbb{L}_s \,\mathrm{d}\alpha \\ - \sum_{s=1}^{S} \int_{\alpha} D_s(L) w_s(\alpha) 1_s(\alpha, \text{worker}, L) \mu_s(\alpha) \mathbb{L}_s \,\mathrm{d}\alpha,\end{aligned} \tag{1.28}$$

where $D_s(L)$ is the subsidy rate for city size $L$ that may or may not depend on $L$. $\tilde{\pi}_s$ and $w_s$ are operating profit and wages. I assume that all land owners are also compensated by the subsidy so that the combined subsidy $D_s(L)\pi_s(\alpha, M_\alpha, L) + D_s(L)c(L) = D_s(L)\tilde{\pi}_s(\alpha, M_\alpha, L)$. $1_s(\alpha, \text{worker}, L)$ and $1_s(\alpha, \text{manager}, L)$ are indicators of location and occupation of type $\alpha$. It equals to one if type $\alpha$ locates in city of size $L$ with corresponding occupation.

Since there is no difference in terms of which sector is drawn to a specific city, I show the case with one tradable sector and ignore non-tradable sector as the results still hold.

The monopolistic competitive market structure implies that the total wage payment is a fixed share of revenue. Therefore, the problem simplifies to

$$\Pi_L = F^e \bar{y} L - \int_{1_s(\alpha, \text{worker}, L)} D_s(L) \frac{\sigma_s w_s(\alpha) \tilde{l}_s(\alpha)}{\sigma_s - 1} \,\mathrm{d}\alpha = F^e \bar{y} L - D_s(L) \frac{\int_{1_s(\alpha, \text{worker}, L)} \frac{\sigma_s w_s(\alpha) \tilde{l}_s(\alpha)}{\sigma_s - 1} \,\mathrm{d}\alpha}{L} L,$$

where $\tilde{l}_s(\alpha)$ is the total number of $\alpha$ workers hired. There is perfect competition and free entry among city developers, which drives their profits to zero in equilibrium. So the only choice of subsidy rate to meet zero profit condition is $D_s(L) = F^e$, the share of education expenditure.

A city developer setting $D_s(L) > F^e$ will get negative profit, whereas setting $D_s(L) < F^e$ will attract no one. Note that since $D_s$ is irrelevant of $L$ or $\alpha$, the managers' problem does not change. The optimal matching, location and occupation remain the same as if there is no subsidy.

A city is built on open site when there is an incentive for city developers to do so. When there exists a set of firms and workers in city $L'$ who wish to locate in city size $L \neq L'$ that does not exist. A city developer will offer them a marginally lower $D' < D$ such that

$(1+D')\pi(\alpha,L') > (1+D)\pi(\alpha,L)$. Since $\pi(\alpha,L')$ is the optimal choice, there is always $D'$ that could steal attract these teams and make a positive profit for city developers. Since workers have no power to choose location, they follows to the new place as the demand outside city $L'$ is zero due to assortative matching. The construction of such process already demonstrate the stability of the equilibrium. The mass of cities adjusts so that each city has the right size in equilibrium.

### 1.8.4 Estimation

**Procedures**

Given the parameters separately calibrated from stage 1, the estimation follows the steps below:

1. Once and for all, I draw four sets of random seeds from uniform distribution on (0,1): one 100,000 for managers, 100,000 for workers, two 100,000 $\times$ 200 draws for firm-city bin idiosyncratic shocks for each occupation.

2. Given parameters, I transfer random seeds to the 100,000 realization of manager skill, 100,000 realization of worker skill and two 100,000$\times$200 idiosyncratic shocks for 200 bins for each occupations.

3. Given parameters, I calculate (26) based on realized productivities and shocks. The maximization delivers optimal choice of city size, and value added from the perspective of managers. It also gives optimal location of workers through the same matching parameters.

4. I compute 24 moments described above and the loss function compared to targeted moments, and use surrogate optimization to find the optimal parameters in the domain. To make sure the global minimum is achieved, I repeat this step for 50 times and find ones that has the smallest loss function.

5. Using the parameters estimated from above, I simulate another 50 random shocks to calculate the variance-covariance matrix of estimated moments. Then I use the inverse of this

$24 \times 24$ matrix as the new weighting matrix, instead of the identity matrix .

6. Repeat step 1-4 using the optimal weighting matrix until the optimal parameters are determined.

7. Calculate the standard errors, moments for figures and equilibrium using the optimal parameters.

Step 5 and 6 can be repeated many times until the weighting matrix converges to a stable matrix. However, such process requires enormous time and computing resources.

## Model Fit: Service Sectors


**Figure 1.17**: Share of Value-Added in City Bin

Note: Data: red solid line; Model: blue dashed line.

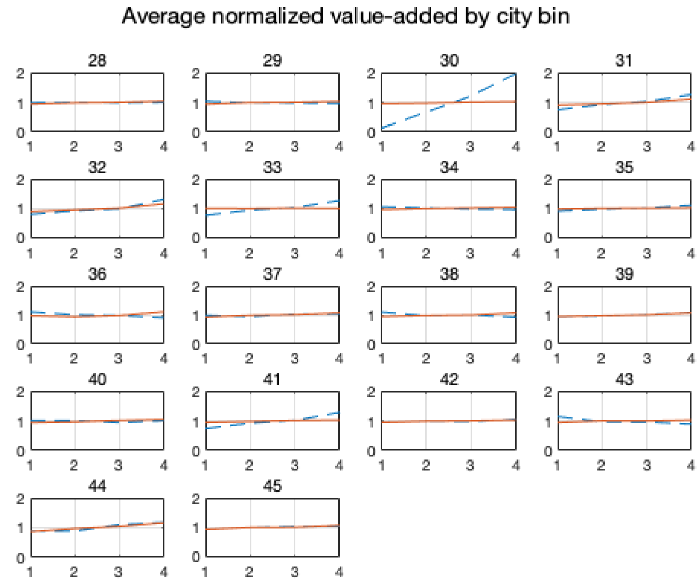



**Figure 1.18**: Average Normalized Value-Added by City Bin

Note: Data: red solid line; Model: blue dashed line.

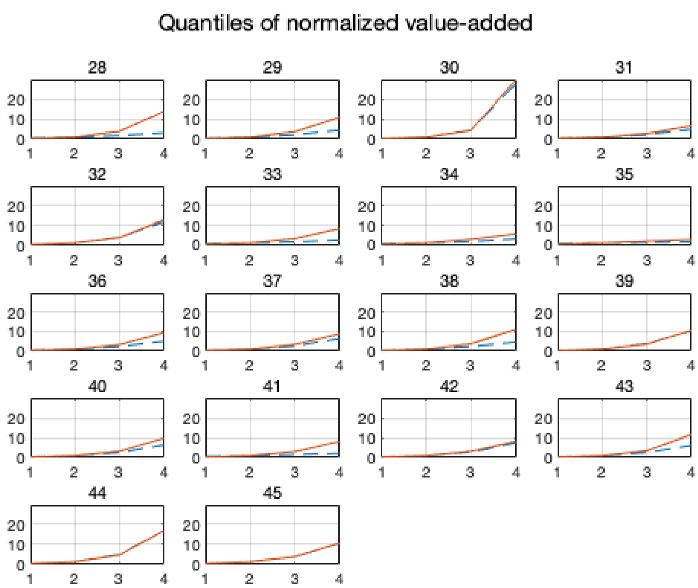



**Figure 1.19**: Quantiles of Normalized Value-Added

Note: Data: red solid line; Model: blue dashed line.

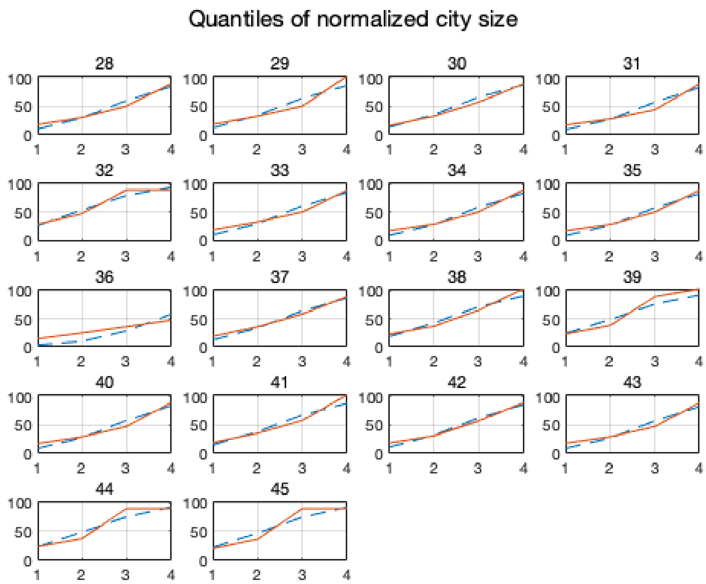



**Figure 1.20**: Quantiles of Normalized City Size

Note: Data: red solid line; Model: blue dashed line.

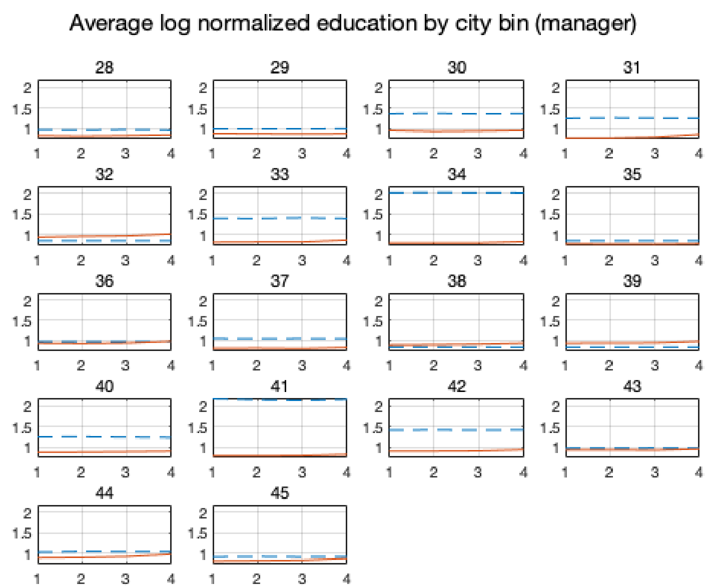



**Figure 1.21**: Average Manager Skill by City Bin

Note: Data: red solid line; Model: blue dashed line.

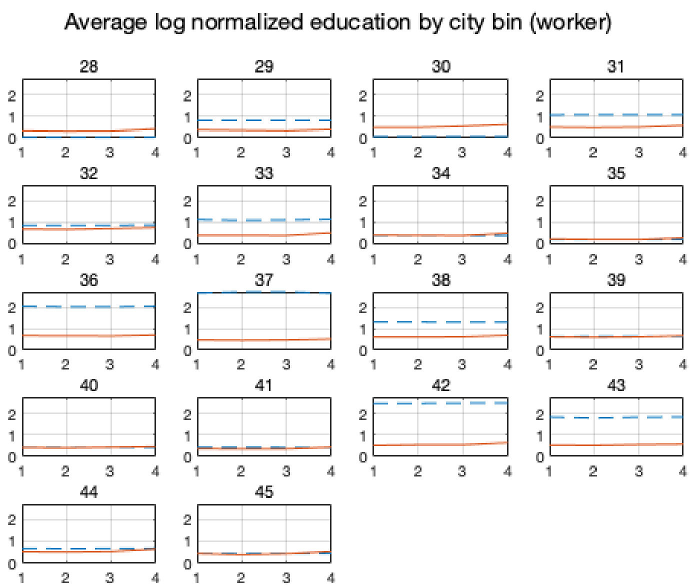



**Figure 1.22**: Average Worker Skill by City Bin

Note: Data: red solid line; Model: blue dashed line.

**Non-Targeted: Employment Size Distribution**

The above six moments are all directly targeted in the estimation. I now show one non-targeted statistics on firm employment.

Due to the simplification of model, there is a difference in the simulated employment and actual employment. This subsection provide technique note on their difference and show that only when $l$ is sufficiently large, the simulated distribution converges to actual distribution.

I drop sector index $s$. In estimation, two occupations are drawn from two distributions without specifying the exogenous size of occupation population. Let $K > 1$ be the relative measure of total workers to managers. Therefore the estimated version of equation (16) is

$$l = K \cdot M'(\alpha) \frac{\mu(M_\alpha)}{\mu(\alpha)}, \tag{1.29}$$

where $l(\alpha)$ is the number of workers in each firm. The right-hand side is computable using estimated parameters except for $K$. Therefore the simulated statistic is $l(\alpha)/K$. The actual firm size is $l(\alpha) + 1 + \Delta$, where $\Delta$ is the measure of all middle level workers that are missed in the model. Take log and express the measurement error as the ratio,

$$r = \frac{\log l - \log K}{\log(l + 1 + \Delta)} \tag{1.30}$$

Given K and $\Delta$, the ratio above is close to one only when $l(\alpha)$ is sufficiently large. Therefore, the simulated statistics are more accurate at the right tail.

In Figure 1.23 below, I take a 10% sample from each sector and plot the density of log employment. The density have a long right tail for most sectors. In Figure 1.24, I calculate the number of workers in each firm from simulation and plot the histogram of log employment truncated from the left at median. Their right tails are similar to the data. Notice there are visual difference between the histogram and density.

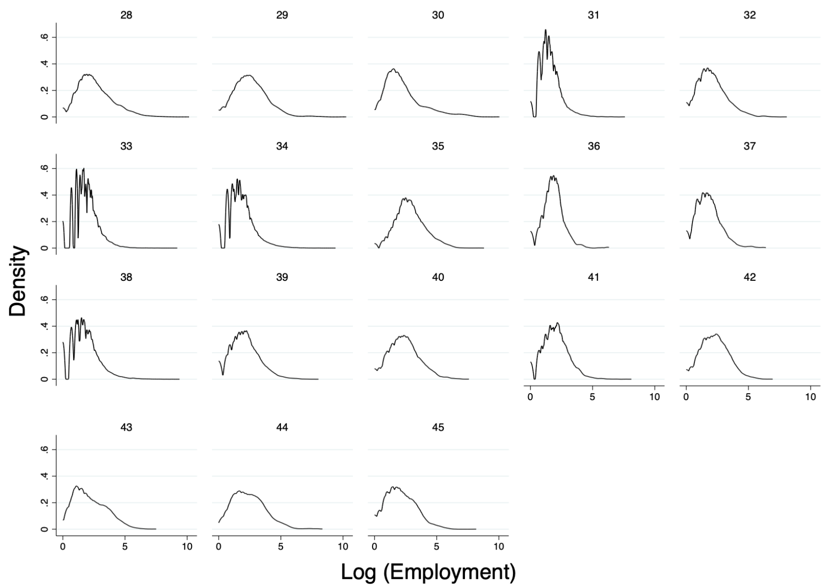



**Figure 1.23**: Density of Log (Employment): Data

Note: This panel shows density of firm employment based on 10% sampling of the full dataset by sectors.

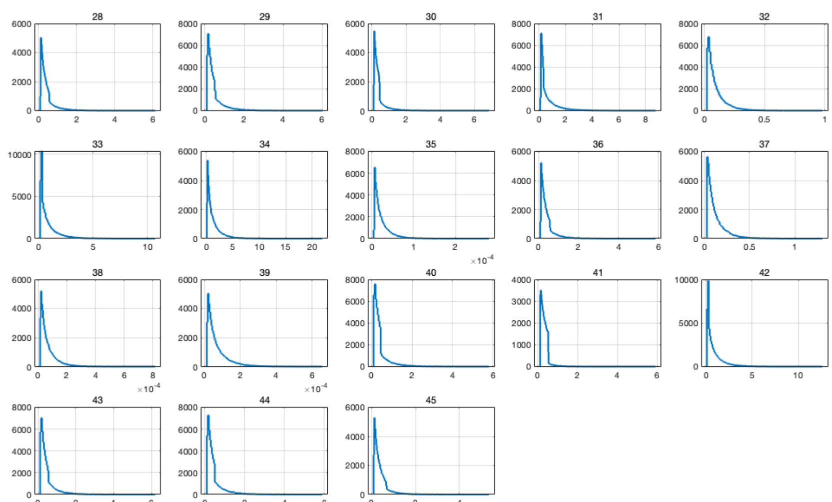


**Figure 1.24**: Distribution of Log (Employment): Simulated and Truncated

Note: This panel shows the distribution of simulated employment in equilibrium, truncated at the median. The magnitude of x-axis differs from actual distribution as sectors differs in their relative measure of total workers to managers, $K_s$.

### 1.8.5 Calibration: Manufacturing Sectors

For manufacturing sectors, I do not observe skill mix so only the first four sets of moments are available. Therefore, the complementarity $\lambda$, the matching parameter $k$, relative contribution $\delta$ and occupational productivity variance $\nu_z$ are not identified. Nevertheless, Table 1.7 shows the results.

Alternatively, manufacturing sectors can still be used to quantify the role of spatial sorting. Parameters at the city level – the variance of idiosyncratic shocks, sorting agglomeration $c$ and density agglomeration $e$ – are jointly estimated using the same procedures as above. Instead of modeling occupational skills, only the variance of firm productivity distribution is parameterized, which is identified with available moments. I leave it for future study.

Figure 1.25 to 1.28 below shows the model fit for 27 manufacturing sectors.


**Figure 1.25**: Share of Value-Added in City Bin: Manufacturing

Note: Data: red solid line; Model: blue dashed line.

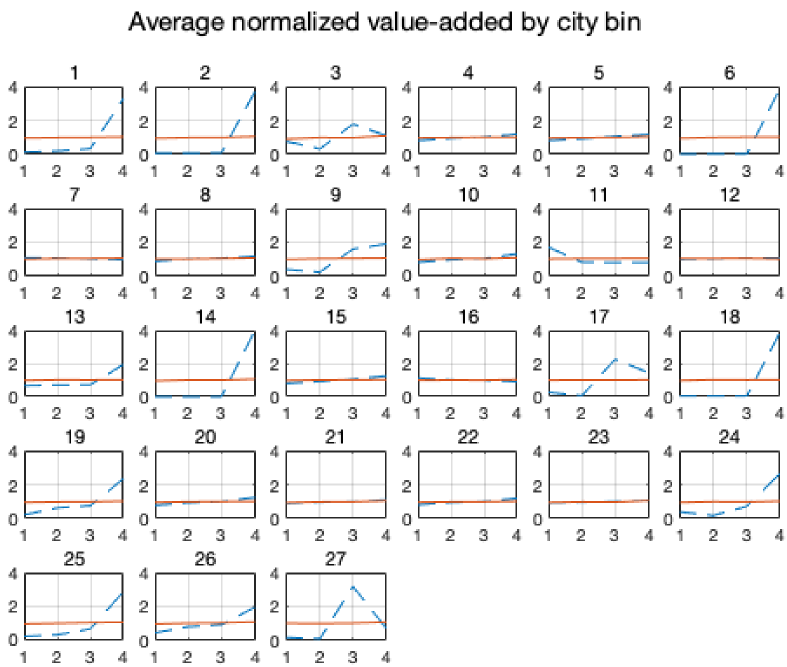



**Figure 1.26**: Average Normalized Value-Added by City Bin: Manufacturing

Note: Data: red solid line; Model: blue dashed line.

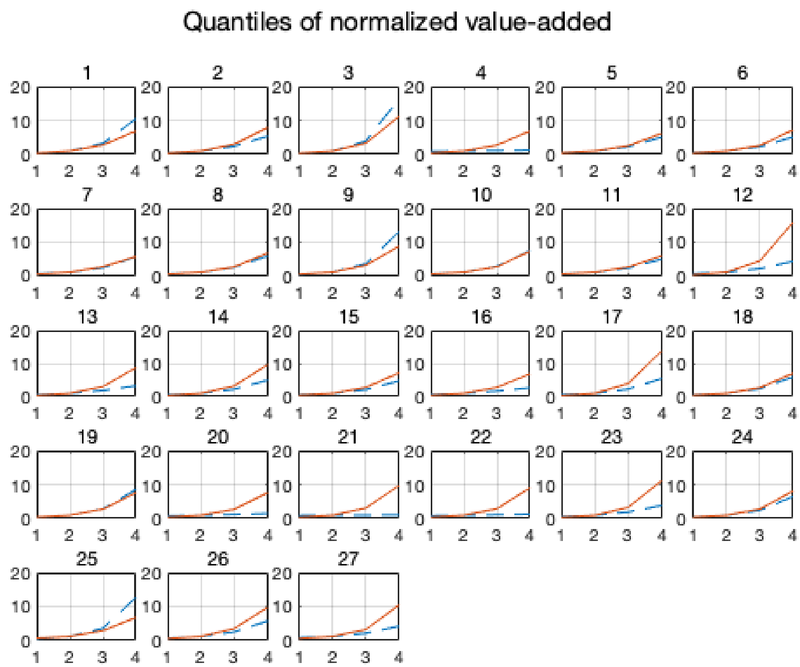



**Figure 1.27**: Quantiles of Normalized Value-Added: Manufacturing

Note: Data: red solid line; Model: blue dashed line.

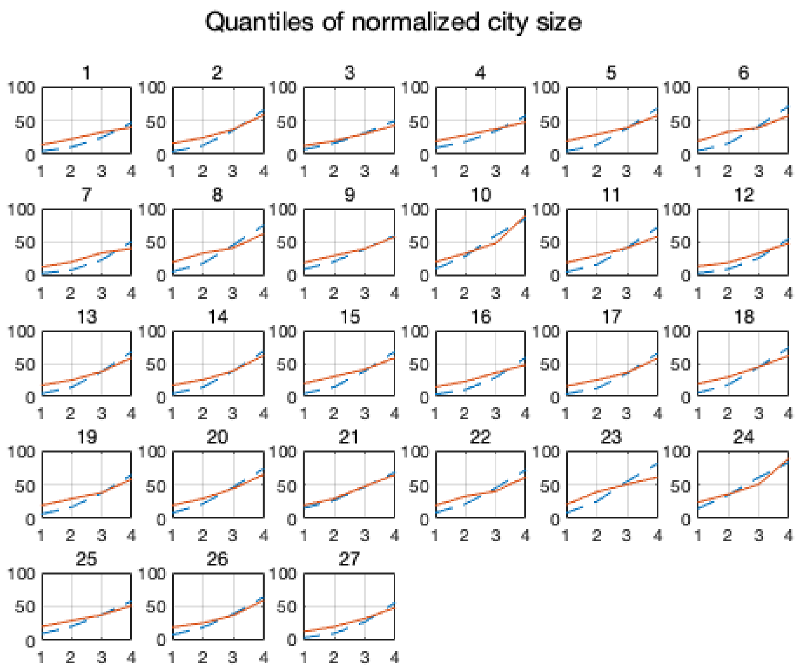



**Figure 1.28**: Quantiles of Normalized City Size: Manufacturing

Note: Data: red solid line; Model: blue dashed line.

**Table 1.7**: Estimated Parameters: Manufacturing

| | (1) $\hat{c}$ | (2) $\hat{e}$ | (3) $\hat{\nu}_\alpha$ | (4) $\hat{\nu}_\varepsilon$ | (5) $\hat{\delta}$ | (6) $\hat{\lambda}$ | (7) $\hat{k}$ | (8) $\hat{\nu}_z$ |
|---|---|---|---|---|---|---|---|---|
| Processing of agriculture products | 0.013 | 0.092 | 0.018 | 1.036 | 0.001 | 0.683 | 0.514 | 5.075 |
| Manufacture of food products | 0.044 | 0.039 | 0.740 | 1.044 | 0.361 | 0.001 | 0.010 | 3.527 |
| Manufacture of beverages and tobacco products | 0.010 | 0.346 | 0.011 | 1.340 | 1.162 | 0.246 | 0.010 | 1.665 |
| Manufacture of textiles | 0.010 | 0.072 | 0.010 | 0.010 | 0.000 | 0.473 | 0.764 | 0.000 |
| Manufacture of wearing apparel | 0.041 | 0.149 | 1.404 | 1.052 | 0.139 | 0.000 | 0.045 | 0.290 |
| Manufacture of leather goods and footwear, leather tanning | 0.089 | 0.191 | 1.823 | 1.090 | 0.210 | 0.000 | 0.010 | 0.102 |
| Manufacture and products of wood, except furniture | 0.017 | 0.198 | 1.452 | 0.762 | 0.149 | 0.006 | 0.013 | 0.018 |
| Manufacture of furniture | 0.035 | 0.247 | 0.967 | 1.090 | 0.219 | 0.000 | 0.641 | 1.612 |
| Manufacture of pulp, paper, and paper products | 0.010 | 0.012 | 0.010 | 1.235 | 0.685 | 0.000 | 0.134 | 1.701 |
| Publishing, printing, and reproduction of recorded media | 0.047 | 0.019 | 3.911 | 1.921 | 0.068 | 0.000 | 0.010 | 4.163 |
| Manufacture of arts and sports equipment | 0.017 | 0.235 | 0.342 | 1.042 | 0.020 | 0.178 | 0.963 | 4.806 |
| Petroleum refinery | 0.035 | 0.078 | 1.390 | 0.868 | 0.134 | 0.037 | 0.010 | 3.584 |
| Manufacture of chemicals and chemical products | 0.034 | 0.199 | 0.488 | 0.999 | 0.314 | 0.006 | 0.163 | 3.375 |
| Manufacture of pharmaceutical | 0.102 | 0.119 | 1.450 | 1.075 | 0.368 | 0.009 | 0.649 | 4.934 |
| Manufacture of chemical fiber, rubber and plastic products | 0.042 | 0.211 | 3.104 | 0.998 | 0.009 | 0.056 | 0.010 | 0.615 |
| Manufacture of glass, ceramic, brick, and cement products | 0.011 | 0.132 | 0.409 | 0.901 | 0.322 | 0.000 | 0.783 | 5.034 |
| Manufacture of basic and fabricated metals | 0.010 | 0.267 | 2.869 | 0.877 | 0.074 | 0.000 | 0.427 | 4.321 |
| Manufacture of metal products | 0.033 | 0.114 | 4.864 | 1.220 | 0.050 | 0.000 | 0.431 | 0.589 |
| Manufacture of general machinery | 0.012 | 0.353 | 0.011 | 0.998 | 0.000 | 0.660 | 0.065 | 4.226 |
| Manufacture of specialized machinery | 0.010 | 0.120 | 0.010 | 0.172 | 0.001 | 0.780 | 0.969 | 4.888 |
| Manufacture of transportation machinery | 0.010 | 0.010 | 0.010 | 0.010 | 0.400 | 0.391 | 0.010 | 2.048 |
| Manufacture of electrical machinery | 0.011 | 0.209 | 0.010 | 0.081 | 0.000 | 0.577 | 0.735 | 4.784 |
| Manufacture of telecommunications equipment | 0.041 | 0.090 | 1.965 | 1.610 | 0.115 | 0.000 | 0.305 | 4.423 |
| Manufacture of instrumentation and office machinery | 0.015 | 0.010 | 0.010 | 0.955 | 0.892 | 0.219 | 0.141 | 0.427 |
| Manufacture of artifacts and others | 0.010 | 0.203 | 0.009 | 1.197 | 0.731 | 0.001 | 0.410 | 3.003 |
| Recycling | 0.010 | 0.337 | 0.010 | 0.788 | 1.249 | 0.020 | 0.293 | 1.040 |
| Utilities | 0.038 | 0.237 | 4.432 | 0.784 | 0.162 | 0.000 | 0.320 | 1.020 |

Notes: $c$ and $e$ are the team sorting agglomeration and non-sorting agglomeration externality; $\lambda$ is the strength of complementarity intensity; $k$ is the matching parameter; $\delta$ is the relative importance of occupation; $\nu_\alpha$ is the variance of manager productivity; $\nu_z$ is the variance of worker productivity; $\nu_\varepsilon$ is the variance of idiosyncratic shock.

### 1.8.6 Alternative Subsidy Level

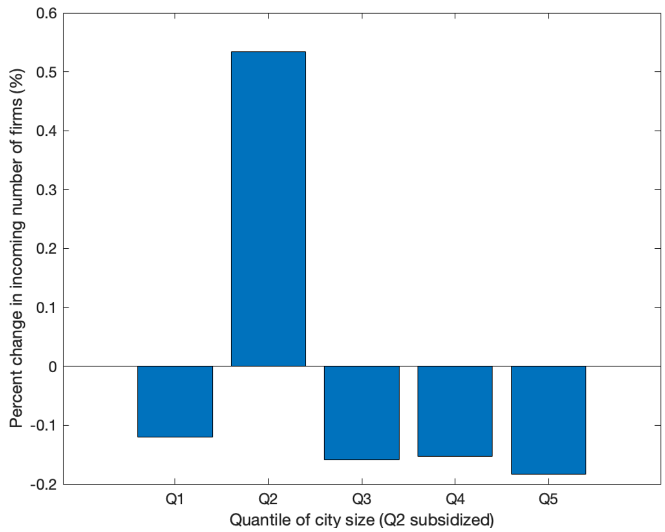



**Figure 1.29**: Effect of 10% Subsidy on the Number of Incoming Firms by City Quantile

Note: The figure shows percentage changes in each quantile of city size distribution when college graduates are offered with a 10% subsidy of original income if they locate in Q2.

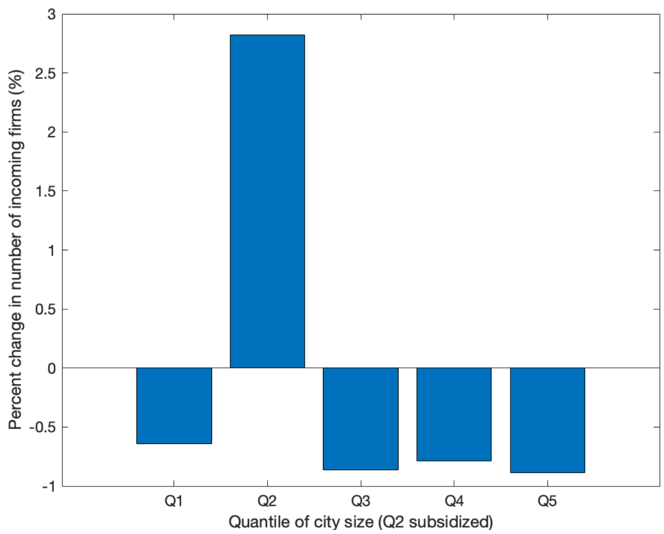



**Figure 1.30**: Effect of 50% Subsidy on the Number of Incoming Firms by City Quantile

Note: The figure shows percentage changes in each quantile of city size distribution when college graduates are offered with a 50% subsidy of original income if they locate in Q2.

# Chapter 2

# Talents on the Move: Income Gains from Internal Migration in China

## 2.1 Introduction

Income per capita across places within a country has led to increasing internal migration, especially in developing economies. The number of internal migrants, defined as movement across regional, district or municipal boundaries within a country, exceeds 740 million, three times larger than that of international immigrants (UNDP, 2009). For example, 326 million or 28.5% of the population in India are internal migrants (UNESCO, 2013). In China, the setting of this study, 247 million people (18% of the total population) are categorized as internal migrant workers in 2015.[1]

One of the foremost questions in migration is the return to migration in the labor market. However, estimating the causal impact of such geographic mobility has always been plagued by the selection into migration. Simply comparing outcomes between migrants and non-migrants can easily lead to biased conclusions as migrants are not randomly chosen from the population

[1] Report on China's Migrant Population Development, 2016

at the origin. While most studies focus on rural-urban migration, understanding the movement of young educated migrants has equal importance for regional development. The geographic distribution of young talents will large shape the landscape of local industries and economic growth in the future.

Moreover, how to interpret these gains has critical policy implications from the social planner's perspective. On one hand, it can be viewed simply as selection among workers with heterogeneous productivity. On the other hand, it may also imply misallocation, thus an opportunity to increase total outputs if one could be replaced from low-productivity places to high-productivity places. Quantifying the relative importance of each explanation would shed light on policies regarding population, labor market, and public spending, etc. For instance, due to limited public resources, several mega cities in China use Hukou policy to weaken the benefit attached to migration to control population.[2] Decisions on those man-made barriers (both extensive and intensive) should be carefully considered in conjunction with potential gains and trade-offs from migration. Despite such importance, gains from migration behavior and its underlying mechanism are still not well understood.

In this paper, I estimate the private income gains from internal migration among college graduates on the job market in China. For the main results, migration is defined as a residential movement across the provincial boarder. Using data from a unique cross-sectional survey conducted at the end of the job market season, I link detailed personal and household information to job market outcomes and construct a complete picture of home-education-work migration history. In general, I focus on the overall migration from the pre-college home place to the post-college job place. I start by showing a naive OLS regression as a benchmark, and control for detailed observed characteristics. I find that on average, there are 13%-18% gains from migration for a job. The results highlights the importance of location selection not only in ability (Dahl,

[2]Hukou is a nationwide registration system that tracks the belonging province of a person, usually birthplace or permanent workplace. It is often linked to access to public resources such as education, social security, and housing market, etc.

2002), but also in family background for young educated workers.

I then adopt propensity score matching methods to address the problem of selection into migration. I utilize previous migration behavior and three proxies for searching effort, risk preference, and network to predict the propensity to migrate. I then compare students who have similar propensity score but end up with different status. After correcting for selectivity, the estimated gap between migrants and non-migrants does not significantly change compared to OLS specification. Analysis on the determinants of migration suggest that selection indeed exists but have two opposite forces: the positive selection on ability is largely offset by the negative selection on family background, resulting in a similar effect to OLS estimation.

Given that the income gap can not be fully explained by observed difference, I further investigate the mechanism of wage gains from migration. I first examine the heterogeneity effect among individual variables. The results show that the gains from migration are not significantly different along the dimension of family background and personal ability, which draws the attention from individuals characteristics to the nature of location. I conduct several practices to test for the hypothesis that place per se affects migration gain. The cost of living contains information on place-specific factors. I adjust for the spatial difference in the cost of living across provinces by deflating the nominal wages by spatial inflation index, and find that the coefficient falls by at least 50%. To further shed light on the gains, I examine the correlation of migration with job type, firm type, and location, etc. I find that migrants are attracted to large firms and foreign-invested firms. Also, they are migrating toward east coast region, especially mega cities. Analysis from ideal job place and discrimination in job searching suggests that there may be an opportunity to improve efficiency by allocating skilled labor to those places. The results imply that misallocation could be the main driver of gains in migration. Policymakers might want to focus more on reducing frictions in the labor market.

This paper fills the gap in migration studies on internal migration among educated group in developing countries. Past literature in estimating private returns to migration mainly focuses on international migration. For example, Clemons et al. (2008) finds that the gain to migration for the Mexico-US case is between $6,700 to $8,000 after correcting for selection and cost of moving. Other studies also find varying amounts of gains to migration in many contexts and empirical strategies (Clemens, 2011; Clemens et al., 2009; Mckenzie et al., 2010; Rosenzweig, 2007; Walmsley and Winters, 2005). Most effort has been made to overcome selectivity issue stemmed from international migration.(Akee, 2010; Borjas, 1987; Chiquiar and Hanson, 2005).

In terms of internal migration, while many studies focus on the impacts and determinants of migration in general population(Greenwood, 1997; Molloy et al., 2011), only a few papers concentrate on the educated young group's migration decisions (Gottlieb and Joseph, 2006; Grogger and Hanson, 2015) and consequences (Kazakis and Faggian, 2016; Keith and McWilliams, 1999; Yankow, 2003). For example, Grogger and Hanson (2015) analyzes location choices of foreign-born PhDs in US universities and finds those stay in the US after graduation are positively selected. Yankow (2003) uses NLSY79 data and finds that highly educated young male workers have significant extended returns to migration. Studies in developing countries mostly focus on the effect of rural-urban migration on source areas such as consumption, poverty or inequality. (Beegle et al., 2011; Bryan et al., 2014) Moreover, few studies are able to connect migration behavior to immediate labor market performance. Using unique dataset and rich information, this study overcomes such problem and provides a reliable estimate of pecuniary returns right after migration.

This study also fits into the big picture of understanding spatial wage gaps. Many studies adopt a macro structural framework to understand the link between migration, productivity and the wage gap, which implicitly embed the gains from migration. (Lagakos and Waugh, 2013; Young, 2013). For example, Bryan and Morten (2019) studies the impact of migration on productivity difference cross Indonesia and suggests substantial gains from reducing migration costs. Hendricks and Schoellman (2016) quantifies the wage gains from migration to the US to reconsider human capital in development accounting. My paper adds to the micro evidence that even for educated group, there could be misallocation due to frictions in the labor market.

The rest of the paper is structured as follows. Section 2 introduces the survey data. Section 3 presents the empirical strategy. Section 4 shows the main findings and robustness checks. Section 5 further investigates the mechanism. Finally, section 6 concludes.

## 2.2 Data

### 2.2.1 China College Student Survey, 2010-2015

The main dataset used in this paper comes from China College Student Survey (CCSS) created by China Data Center at Tsinghua University. It is a cross-sectional survey on new graduating students conducted at the end of the job season in May and June each year from 2010 to 2015. The survey first chooses colleges according to geographic location and college type, and survey students on the job market.[3] The survey work in each college is managed by one to three college administrators in charge of teaching or student activities. These survey administrators are trained in Beijing in several days of intensive meetings. Then, they bring the questionnaires back to their campuses and randomly assign to graduates by their student ID at the end of the spring semester (end of job season). The timing provides the best up-to-date information on job market results for each graduate in that year. When a student finishes the questionnaire, the questionnaire would be sealed in coded envelopes for privacy concerns and mailed back to the data center by the survey administrators in that college. The survey and data entry process are all well-conducted and supervised with considerable care.

The survey starts with 19 colleges as a pilot in 2010. The number reaches its peak in 2013 with 65 colleges participated in that wave. Due to an unexpected budget cut, fewer colleges were surveyed in 2014 and 2015. Table 2.11 in the Appendix shows the number of colleges surveyed

[3]Areas include “Beijing, Shanghai, Tianjin”, Northeastern, Eastern, Central, and Western China. Three metropolises (Beijing, Shanghai, and Tianjin) are separated because they have extremely large concentrations of colleges, especially top universities.

in each wave by the school type. Since elite colleges are oversampled, I use the inverse of the proportion of selected colleges in each stratum as the weight in regressions when applicable. A total of 40,915 students in their graduating year are selected into the survey from 88 colleges across 26 provinces. Approximately 6 percent of students come from vocational colleges. 39 percent of students come from elite colleges (211 program schools).

The questionnaire is collaboratively designed with experts in other disciplines such as sociology and education. It collects information not only on demographics, family background, education history, and financial conditions but also on future plans, job offers and other characteristics. In particular, respondents were asked about the job searching process and labor market outcomes up to the time they were surveyed. For students who entered the job market and successfully received at least one job offer, they were asked to give the information on their best job offer including industry, location, starting salary, etc. Based on this information, I obtain detailed migration path from home to college, and then to the job place.

Table 2.1 summarizes the statistics of key variables pooling data from all years. About 66% of graduates choose to directly go to the job market, of which 76% received at least one job offer. About 70 percent of students claim that they already signed or will sign the contract on their reported best job offer. So the final sample size decreases to round 17,000. Therefore, the analysis applies to a particular selective sample of all college graduates.

The average nominal wage over the sample period is 2700 RMB. The survey also collected details about family background and performance in college. I use the score for the CEE exam as proxies for ability since the score is the main criterion for college admission and the only criterion for the majority of students, thus it is regarded as sufficient statistics for students' ability or IQ prior to college. I standardize the CEE score by year-province-track, so they are comparable across the nation.

**Table 2.1**: Descriptive Summary of Key Variables

| | Mean | S.D. | Min | Max | N |
|---|---|---|---|---|---|
| On Job Market | 0.66 | 0.47 | 0 | 1 | 39354 |
| No Job Offer | 0.24 | 0.43 | 0 | 1 | 24556 |
| Starting Salary (in RMB) | 2701.34 | 1201 | 600 | 8000.0 | 17489 |
| Migrate | 0.42 | 0.49 | 0 | 1 | 17467 |
| Age | 22.82 | 1.11 | 18 | 30 | 39759 |
| Female | 0.44 | 0.50 | 0 | 1 | 40728 |
| Married | 0.00 | 0.07 | 0 | 1 | 40716 |
| Minority | 0.08 | 0.26 | 0 | 1 | 40638 |
| Communist Party Member | 0.31 | 0.46 | 0 | 1 | 40553 |
| Family Size | 4.21 | 1.23 | 1 | 15.0 | 39696 |
| Has Gov Parent | 0.12 | 0.33 | 0 | 1 | 40898 |
| Has College Parent | 0.25 | 0.43 | 0.0 | 1 | 36973 |
| Log(Household Income) | 10.49 | 1.11 | 6.9 | 13.1 | 33933 |
| Log(Parent Income) | 9.79 | 2.37 | 0 | 13.1 | 34173 |
| Has Commercial Housing | 0.40 | 0.49 | 0 | 1 | 40462 |
| Home in City | 0.38 | 0.49 | 0 | 1 | 40575 |
| Elite University | 0.39 | 0.49 | 0 | 1 | 40898 |
| STEM | 0.70 | 0.46 | 0 | 1 | 40067 |
| CET4 Score | 462.82 | 62.79 | 100 | 700 | 29731 |
| Top 20% GPA | 0.46 | 0.50 | 0 | 1 | 38985 |
| Specialty certificates | 0.79 | 0.40 | 0 | 1 | 40544 |
| Student Union | 0.61 | 0.49 | 0 | 1 | 40898 |
| Provincial Awards | 0.12 | 0.32 | 0 | 1 | 40092 |
| Internship/Parttime | 0.72 | 0.45 | 0 | 1 | 40788 |
| CEE Chinese Z-score | 0.02 | 0.82 | -3 | 3 | 29523 |
| CEE Math Z-score | 0.02 | 0.86 | -3 | 3 | 29449 |
| CEE English Z-score | 0.02 | 0.85 | -3 | 2.9 | 29487 |
| CEE Comprehensive Z-score | 0.02 | 0.81 | -3 | 3 | 28181 |

### 2.2.2 Defining Migration

I construct migration path for each graduates from their home-college-job transition. Throughout the paper, I focus on the overall migration from home place to job place, and treat home-college as a temporary migration. In the main results, I use provincial borders as the boundary. By this definition, 42% of the students on the job market eventually work outside their home province, i.e., migrate for job.

Table 2.2 presents the mean comparison by migration status. On average, migrants earn more than those who stay at home province for a job. The difference is 443 RMB or 17%, and significant at 1% level. Other characteristics are also significantly different between the two

**Table 2.2**: Comparison of Mean by Migration Status

| | (1) Migrants | (2) Non-migrants | (3) Diff |
|---|---|---|---|
| Starting Salary (in RMB) | 2953.32 | 2510.24 | -443.08*** |
| | | | |
| Age | 23.03 | 22.92 | -0.12*** |
| Female | 0.35 | 0.50 | 0.15*** |
| Married | 0.00 | 0.00 | 0.00 |
| Minority | 0.08 | 0.06 | -0.02*** |
| Communist Party Member | 0.34 | 0.30 | -0.04*** |
| | | | |
| Family Size | 4.27 | 4.28 | 0.00 |
| Has Gov Parent | 0.09 | 0.10 | 0.01** |
| Has College Parent | 0.17 | 0.20 | 0.03*** |
| Log(Household Income) | 10.31 | 10.55 | 0.25*** |
| Log(Parent Income) | 9.51 | 9.85 | 0.33*** |
| Has Commercial Housing | 0.30 | 0.37 | 0.07*** |
| Home in City | 0.25 | 0.38 | 0.12*** |
| | | | |
| Elite University | 0.44 | 0.26 | -0.18*** |
| STEM | 0.79 | 0.67 | -0.12*** |
| CET4 Score | 460.96 | 451.61 | -9.35*** |
| Top 20% GPA | 0.41 | 0.43 | 0.02** |
| Specialty Certificates | 0.82 | 0.85 | 0.03*** |
| Student Union | 0.68 | 0.66 | -0.02*** |
| Provincial Awards | 0.12 | 0.10 | -0.02*** |
| Internship/Parttime | 0.85 | 0.83 | -0.02** |
| CEE Chinese Z-score | 0.02 | -0.04 | -0.06*** |
| CEE Math Z-score | 0.06 | -0.06 | -0.11*** |
| CEE English Z-score | 0.01 | -0.07 | -0.08*** |
| CEE Comprehensive Z-score | 0.08 | -0.06 | -0.14*** |
| N | 6964 | 9370 | |

groups except for marriage status and family size. There are several notable findings. First, female students are less likely to migrate. This is consistent with the risk preference that women are more risk-averse. Second, non-migrants are positively selected in terms of their family background. They are more likely to have rich, powerful and educated parents. In contrast, those who migrate are positively selected in terms of their ability or performance at college. These differences suggest that it is important to correct for selection into migration even for the educated.

## 2.3 Empirical Strategy

The primary interest of this paper is to quantify the private gains from migration in the labor market. Consider the following Mincer type equation that relates wages to migration and other covariates.

$$Y_i = \alpha + \beta M_i + \Gamma X_i + \kappa_h + \pi_s + \phi_m + \tau_d + \lambda_t + \varepsilon_i$$

where $Y_i$ is the log starting monthly salary in person $i$'s best job offer. $M_i$ is an indicator dummy for migrating for job. It equals 1 if the province of best offer job location is not the same as the home province for student $i$. In other words, I regard all the migration destination except home place as one alternative labor market. So the coefficient reflects national average gains for all migrants regardless of the destination. Vector $X_i$ denote a set of student attributes that affect earnings including personal characteristics, family background and college performance etc. In particular, it includes the CEE score as a proxy for ability.

To further account for the difference in earnings, I include the following fixed effects. Students in different provinces might have received differential exposure to the neighborhood and education prior to college (Chetty and Hendren, 2015), so I control for their home province fixed effect $\kappa_h$ to account for the difference in the average quality of primary and secondary education. School fixed effect $\pi_s$ reflects different qualities of education at each college. $\phi_m$ and $\tau_d$ represent major and industry fixed effects. $\lambda_t$ is the year fixed effect that picks up inflation and common wage growth. $\varepsilon_i$ represents the error term.

The parameter $\beta$ is the primary interest, which depicts the short-run monetary payoff to migrating for a job. Ideally, if the status of migration, $M_i$ is randomly assigned to two identical students so that one returns to the job market at home place, and the other one goes to an alternative market, then the difference of earnings will be a causal estimate on the impact of migrating for a job. However, as in the most literature among economics of migration, selection into migration emerges. Whether a student migrates for job is potentially highly selective as it is shown in Table 2.2. Moreover, the selection bias can not be easily signed. Depending on which underlying model of migration is used, the selection can be either positive, negative or both (Borjas, 1987; Grogger and Hanson, 2011).

To deal with the endogeneity of migration, I first use propensity score matching to compare students with same probability to migrate. Many attempts have been made to shed light on the theory of migration. Some simple facts were summarized as early as in the 19th-century (Ravenstein, 1885), where simple personal characteristics like gender are correlated with the distance of migration. Recent studies propose many factors that could influence migration choices, ranging from individual characteristics to macro opportunities, from economic variables to social connections (Bauernschuster et al., 2014; Borjas, 1989; Greenwood, 1997; Sjaastad, 1962).

The determinants of migration for college graduates are special. First, their network is relatively simple as they are young and mostly unmarried. The primary connection is their family. Second, the opportunity cost of moving is lower than aged workers since they have little fixed investment at the time of migration. Finally, their migration choice is exogenously determined in terms of the timing. Based on past findings and these specific contexts of graduate students, I estimate the following binary choice model:

$$\text{Prob}(M_i) = \alpha_0 + \Omega_0 Z_i + \Gamma_0 X_i + \kappa_h + \pi_s + \phi_m + \tau_d + \lambda_t + \nu_i$$

where $M_i$ is the dummy for migrating for job. $Z_i$ is a vector of variables that are related to migration but excluded from wage equation (e.g., migration cost). $X_i$ and other fixed effects are the same as before. $\nu_i$ is the error term.

Variables in $Z$ should reflect factors that influence migration choice but not wage offer. One would expect that migration is correlated with some intrinsic motivation. A natural proxy for that is searching effort (Chirinko, 1982), which is usually hard to measure. Fortunately, the survey has the number of CVs sent by the students, so I include it in $Z$ as a proxy for searching effort. I also use a dummy for participating in civil service exam as a proxy for the risk preference because the civil servants in China are relatively low-paid but very secure jobs. Students who participated in the exam are to some extent lean towards a lifelong secure job at least in the short run.[4]

Another aspect that is not included in the wage equation but relates to migration decision is the cost of migration. While some models assume that the cost is fixed (Borjas, 1987), others allow it to vary with education or network. In my case, most of the above variation in cost has been captured by personal and family characteristics in $X_i$, except for connection (Munshi, 2003). The survey asks whether students seek help from acquaintances such as relatives or friends during job hunting. Therefore in $Z$, I include a dummy for such help as well to represent how easy for students to access the labor market.

## 2.4 Estimation Results

### 2.4.1 A First Look from OLS

Table 2.3 presents the correlation between migrating for job and starting salary using OLS regressions. The dependent variable is the log of starting monthly salary for the best job offer. Migrate for job is a dummy that equals 1 if job province is different from home province. Column 1 includes basic demographics and year fixed effect. The coefficient on migration dummy is 0.16 and significant at 1%. This correlation means that migrating for a job is associated with a 16% increase in wage offer.

Column 2 and 3 add more control variables of family background and education performance in the regressions. The coefficient increases when family background is included, and drops to 12% when including education performance at school. This contrast indicates that while both family and education are positively correlated with wage, they have an opposite correlation

[4]For the 2011 survey, there was an additional module that directly measures risk preference. It can be used as an extension for that subset later.

**Table 2.3**: OLS Results of Migration Wage Premium

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Dependent Variable: Log (Starting Monthly Salary) | | | | | |
| Migrate for job | 0.16*** | 0.18*** | 0.12*** | 0.14*** | 0.15*** | 0.13*** |
| | (0.044) | (0.043) | (0.036) | (0.031) | (0.031) | (0.032) |
| Controls: | | | | | | |
| Demographics | Y | Y | Y | Y | Y | Y |
| Family background | | Y | Y | Y | Y | Y |
| College performance | | | Y | Y | Y | Y |
| CEE Z-score | | | | Y | Y | Y |
| Fixed effect | | | | | | |
| School& Home province | | | | | Y | Y |
| Major& Industry | | | | | | Y |
| Year | Y | Y | Y | Y | Y | Y |
| Observations | 16,034 | 13,035 | 10,757 | 10,171 | 10,171 | 10,079 |
| R-squared | 0.288 | 0.319 | 0.417 | 0.449 | 0.530 | 0.540 |

Note: Robust standard errors in parentheses, clustered by school. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. All regressions are weighted. Demographics: age, female, minority, marriage, political status; Family background: size, parent education, gov. official parent, household income, housing assets; Education: degree track, English, GPA, certificates, activities, awards, internship.

with migration behavior. When ability proxy, CEE score is included in column 4, the coefficient increases a little to 0.14, suggesting that college performance has captured a large portion of the unobserved ability.

Studies have found that the quality of school affects the wage of their students in China (Li et al., 2012). I include school fixed effect to capture that. Besides that, I include home province fixed effect to account for the difference in primary education quality and neighborhood effect. Column 5 shows that although R-squared increases from 0.45 to 0.53, the coefficient stays relatively the same. Column 6 further includes major and job industry fixed effect. The coefficient remains at 0.13 and significant. For the rest of paper, I use the controls in column 6 as baseline. Unless otherwise specified, same fixed effects and covariates are all included.

### 2.4.2 Propensity Score Matching

The results above are potentially problematic as there might be other unobserved factors. A natural practice is to investigate factors that might affect migration status for jobs.

Table 2.4 shows the estimation results on the probability of migration for a job using a linear model. The first column regress migration for a job on previous migration for education and only controls for the same $X_i$ as above and home province fixed effect. The results show that if students experienced a temporary migration for education, their probability of working in a province other than home province is 36% higher. In other words, if they migrate before, they are less likely to return to work at home.

Column 2 further includes a school fixed effect. R-squared increased from 0.348 to 0.374, meaning that school also accounts for some variation in migrating for a job. Column 3 further controls for major and industry, the explanation power only increases a little.

**Table 2.4**: Linear Probability Model on Who Migrate for Job

| | Dependent variable: Migrate for job=1, otherwise 0 | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Migrate for College | 0.36*** | 0.31*** | 0.31*** | 0.31*** | 0.31*** |
| | (0.021) | (0.023) | (0.023) | (0.023) | (0.023) |
| Log(# of application) | | | | 0.0051 | 0.0049 |
| | | | | (0.0046) | (0.0046) |
| Took public servant exam | | | | -0.064*** | -0.063*** |
| | | | | (0.014) | (0.014) |
| Job search help | | | | | -0.036*** |
| | | | | | (0.0093) |
| Fixed Effect: | | | | | |
| Home province | Y | Y | Y | Y | Y |
| School | | Y | Y | Y | Y |
| Major&Industry | | | Y | Y | Y |
| Observations | 11,263 | 11,263 | 11,136 | 10,755 | 10,749 |
| R-squared | 0.348 | 0.374 | 0.382 | 0.384 | 0.385 |

Note: Robust standard errors in parentheses, clustered by school. * p<0.10, ** p<0.05, *** p<0.01. All columns control for demographics, family background and education.

Column 4 examines the impact of intrinsic motivation. The results show that the number of CVs sent has no significant impact on the probability of migrating for a job. Column 5 includes civil servant exam participation as a proxy for risk preference. For those who took the civil servant exam, the probability of migrating for job decreases by about 6.4%. Column 5 includes a dummy for the network. The results show that if students seek help from acquaintance during job hunting, they are significantly less likely to migrate for a job. On average, getting a connection decreases the chance of moving by 3.6%.

The effect of other control variables in $X_i$ is consistent with comparison in Table 2.2 across columns. Female students are significantly less likely to migrate. Families that purchase commercial housing and live in cities reduces the likelihood of students migration. Students who have higher GPAs, participate in student union and internship are more likely to migrate. All these results confirm that migrants are positively selected on ability but negatively selected on family background.

Based on the results in Column 5 of Table 2.4, I recalculate the propensity score using probit model. Figure 2.2 in appendix plots the density of predicted propensity score by actual migration status. To make the comparison group and treated group more similar, I use various ways to reduce the selection between them. Column 1 in Table 2.5 uses the sample that are common supported with a 20% trim. The second and third column uses one-to-one and five-nearest-neighbor matching respectively and finds similar results. Column 4-6 restrict sample with propensity score between different ranges. The coefficient of migrate for a job stays relatively stable between 0.13 and 0.15. All coefficients are significant at 1% level and are close to OLS estimates with rich controls.

In Table 2.3, the $R^2$ increases from 0.288 in column 1 to 0.540 with all controls in column 6, and even higher in column 6 of Table 2.5. Despite this large increase in explanatory power, the coefficient of interest remains relatively stable. In other words, adding a large number of observable wage determinants does not change the observed impact of migrating. Despite the effort to control for observed difference, there are still many unobserved factors that contribute to the outcome. How large is the omitted variable problem? Oster (2017) construct a statistic to

**Table 2.5**: Propensity Score Matching Results of Migration Wage Premium

| | (1) common support | (2) 1-1 match | (3) 1-5 match | (4) 1-5 matching, propensity score (0.2, 0.8) | (5) (0.3, 0.7) | (6) (0.4, 0.6) |
|---|---|---|---|---|---|---|
| Migrate for job | 0.13*** | 0.15*** | 0.13*** | 0.15*** | 0.14*** | 0.15*** |
| | (0.035) | (0.033) | (0.037) | (0.027) | (0.031) | (0.028) |
| Observations | 8862 | 5,979 | 7,320 | 5004 | 3570 | 1,978 |
| R-squared | 0.53 | 0.62 | 0.57 | 0.60 | 0.63 | 0.61 |

Note: Robust standard errors in parentheses, clustered by school. * p<0.10, ** p<0.05, *** p<0.01. All regressions are weighted, control for fixed effects and covariates as in column 6 of Table 2.3.

provide some insight on the issue. Following Oster (2017), the statistic is calculated as

$$\delta = \left(\frac{\beta_c}{\beta_u - \beta_c}\right) \times \left(\frac{R_c^2 - R_u^2}{0.3 \times R_c^2}\right),$$

where $\beta_c$ and $\beta_u$ are the estimates with and without controls, and $R_c^2$ and $R_u^2$ are the corresponding $R^2$. Using column 1 and 6 from Table 2.3, the result is $(\frac{.13}{.16-.13}) \times (\frac{.54-.29}{.3*.54}) = 6.7$. This means the selection on unobservables would have to be 6.7 times stronger than selection on observables to explain the results. With the rule of thumb cutoff for observational studies, which is 1, this provides some assurance that selection on unobservables is not severe. Nevertheless, the results could still be subject to omitted variable bias. Note that this matching further reduces the sample so the results apply to those group of students who have similar tendency to migrate. The effect of migration could be different at two ends of the score distribution.

### 2.4.3 Robustness Check

Although provincial boundary is a natural approach to define migration, one might worry that the distance of migration can be very short even it crosses provincial border. I use three different definitions of migration to see how the results vary. First, I use large regions that contain a few similar provinces to define migration. I divide all provinces into ten regions according to their similarity in language, culture, and climate. This is a more loose definition of migration that allows local cross-province movement.

Second, I define migration by the distance, where I choose 200 km to separate migrants and non-migrants. Finally, I restrict the definition of migration to those who have decided to accept the best job offer, so they are actually moving after graduation. Table 2.6 shows the results using OLS.

**Table 2.6**: Robustness Check on The Definition of Migration

| | (1) | (3) | (5) |
|---|---|---|---|
| | Log (starting salary) | | |
| Definition: | across region | over 200km | across province |
| | OLS | OLS | OLS |
| Migrate for job | 0.16*** | 0.13*** | 0.14*** |
| | (0.022) | (0.032) | (0.021) |
| Observations | 10079 | 10079 | 10079 |
| R-squared | 0.545 | 0.419 | 0.542 |

Note: Robust standard errors in parentheses, clustered by school. * p<0.10, ** p<0.05, *** p<0.01. All regressions are weighted and control for fixed effects and covariates as in column 6 of Table 2.3. Sample is restricted to students who have accepted an offer.

The overall results so far have important implications on the interpretation of wage gaps. First, there are positive selections for migrants according to ability, but the impact on wages is offset by negative selection on family background. Second, individual unobservables account for only a small portion of differences in wage gap since the matching estimates are similar to OLS results. It leads to the next question that if the selection does not explain the gap, what causes the gains at migration.

### 2.4.4 Heterogeneity

To shed light on the mechanism, I first investigate heterogeneous effects among different personal characteristics, family background, and education. Table 2.7 shows the results using

OLS. Matching produces similar results.

Panel A separates the sample by personal characteristics. The effect of migration is significantly larger for male students. The gap is smaller among the minority group. Communist Party membership enjoys the same migration gain as the rest.

Panel B examines the differential effects regarding family background. There are no significant differences across rich or poor families (separated by mean income), parental education or the size of the family (separated by mean family size). These results suggest that the gain from migration is unlikely to come from family background.

In Panel C, I estimate the gains by education. Column 1 and 2 compare the effect among students who pass the College English Test or not. Column 3 and 4 separate the sample by the initial college entrance exam score. Those who above the mean has similar gains from migration than those below. The last two columns examine the effect among different quality of schools. All columns show similar results that are not significantly different from each other, from 10% to 15%.

Panel B and Panel C together indicate that the gains from migration do not come from family background or education-related factors. This result is consistent with previous findings that the selection of heterogeneous students can not explain most of the gains from migration. So the next possibility is to investigate the difference between destination and origin places per se.

## 2.5 Into the Mechanism

In this section, I present evidence that links the migration gain with place-specific factors. There are many reasons for wage differences across places. First, it could be the case that purchasing power are different so higher wage only reflects a higher cost of living. Besides that, technology and capital stock may differ across places thus the same person may have difference productivities at different places. In additional, there could exist benefits from agglomeration

**Table 2.7**: Heterogeneity effect: OLS Estimation on Log(Starting Salary)

| Panel A | (1)<br>Male | (2)<br>Female | (3)<br>Minority | (4)<br>Non-minority | (5)<br>CCP | (6)<br>Non-CCP |
|---|---|---|---|---|---|---|
| Migrate for job | 0.17*** | 0.082 | 0.023 | 0.15*** | 0.12*** | 0.14*** |
| | (0.025) | (0.064) | (0.070) | (0.023) | (0.029) | (0.036) |
| Observations | 5,896 | 4,183 | 721 | 9,358 | 3,401 | 6,678 |
| R-squared | 0.480 | 0.552 | 0.701 | 0.550 | 0.522 | 0.558 |

| Panel B | (1)<br>Rich family | (2)<br>Poor family | (3)<br>College parent | (4)<br>Non-college | (5)<br>Large family | (6)<br>Small family |
|---|---|---|---|---|---|---|
| Migrate for job | 0.12*** | 0.15*** | 0.19*** | 0.12*** | 0.15*** | 0.13*** |
| | (0.016) | (0.028) | (0.027) | (0.035) | (0.031) | (0.035) |
| Observations | 3,472 | 6,607 | 1,725 | 8,354 | 3,625 | 6,454 |
| R-squared | 0.373 | 0.591 | 0.623 | 0.544 | 0.533 | 0.586 |

| Panel C | (1)<br>Pass CET4 | (2)<br>Not pass | (3)<br>High CEE | (4)<br>Low CEE | (5)<br>Elite Schools | (6)<br>Non-elite |
|---|---|---|---|---|---|---|
| Migrate for job | 0.14*** | 0.12*** | 0.15*** | 0.12*** | 0.10*** | 0.14*** |
| | (0.029) | (0.039) | (0.021) | (0.038) | (0.023) | (0.035) |
| Observations | 7,875 | 2,204 | 5,460 | 4,619 | 3,793 | 6,286 |
| R-squared | 0.574 | 0.497 | 0.528 | 0.517 | 0.413 | 0.536 |

Note: Robust standard errors in parentheses, clustered by school. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Controls and fixed effects are included.

externality, which further enhances the gap.

### 2.5.1 Adjusting for Cost of Living

The cost of living incorporates a lot of place-specific underlying factors. If I adjust for such difference and the gain from migration varies a lot, then it implies the gap can be attributed to local factors per se.

I borrow the calculation on spatial price differences in China from Brandt and Holz (2006). They construct a baseline basket of goods for urban areas and calculate the cost of basket for each province at a given time, then use urban CPI to figure out the cost changes over time. I normalize the cost of basket at the national level in 2010 to 100, and calculate the relative cost of all other provinces each year to get the deflator index. Finally, I deflate the wage by the index according to job location and year to get the real wage.

**Table 2.8**: Read Income Gains Adjusting for Cost of Living

| | (1) | (2) |
|---|---|---|
| | Log (Real Wage) | |
| | OLS | 1-5 Matching |
| Migrate for job | 0.064** | 0.056* |
| | (0.028) | (0.029) |
| Observations | 10079 | 7303 |
| R-squared | 0.468 | 0.514 |

Note: Robust standard errors in parentheses, clustered by school. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. All regressions are weighted and control for same fixed effects and covariates as in column 6 of Table 2.3.

Table 2.8 shows the results using OLS, matching. The magnitude of the coefficients all decreases. Compared to the estimated gap using nominal wages, at least 50% disappears with this simple deflation.

### 2.5.2 Job Characteristics

If the wage premium is associated with specific factors of the place, it must be reflected through the pattern of migration and the feature of firms. Therefore, I look at the correlation between migration and job characteristics. Table 2.9 shows the results.

**Table 2.9**: Migration on Job Benefits and Features: Linear probability OLS

| Panel A: Job Benefit | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Hukou | Health Insurance | Pension | Housing Fund |
| Migrate for job | 0.031 | 0.035* | 0.029 | 0.037** |
| **Panel B: Job Location** | (1) | (2) | (3) | (4) |
| | East Coast | West | Middle | Mega Cities |
| Migrate for job | 0.43*** | -0.19*** | -0.21*** | 0.26*** |
| **Panel C: Type of Job** | (1) | (2) | (3) | (4) |
| | Large Size Firms | State-owned Firms | Foreign Firms | Private Sector |
| Migrate for job | 0.13*** | -0.063** | 0.052*** | 0.047 |

Note: All regressions are weighted, clustered by school and control for fixed effects and covariates. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

In Panel A of Table 2.9, I test for non-pecuniary job benefits related to migration. The results show that migrants are more likely to get other bonus, in particular insurance and housing fund. Consistent with the process of Hukou policy, migration does not increase the change of getting a Hukou.

Panel B studies the location choice of migration. The results present strong migration pattern towards coastal area especially for metropolitan areas (Beijing, Shanghai, Guangzhou, and Shenzhen). On average, migrants are 26% more likely to end up with a job in megacities (column 4) compared to non-migrants. Accordingly, west and middle regions are experiencing brain drain as students are leaving those areas.

Panel C looks at firm size and type. Migrants are more related to a job in large firms, and they are moving away from state-owned firms towards foreign firms, which potentially has higher productivity and offers higher wage trajectory. This evidence indicates that human capital is concentrating towards developed and high productivity areas and sectors, thus suggesting the wage gap between migrants and non-migrants is highly likely due to the place-based factors such as capital complementarity, peer effect and environment of work.

### 2.5.3 Distribution of Migration Premium

So far, all the estimated gains are averaged all over the country. The next natural question is the distribution of migration premium across places. In particular, which province offers the highest migration premium to incoming migrants, and which province's students have the most attractive outside options?

To do that, I interact each destination province with the dummy of migration so that it captures migration premium associated with that specific destination. The coefficient on the new province-specific dummy then captures the wage difference between students who migrate to this province and the rest of the population (including local workers). The same can be done for each home province and the coefficients will capture migration gains associated with leaving specific source province. The former one describes the average migration gain on the new incoming migrants, thus I call it the inflow premium, whereas the latter one stands for the average premium for outflowing students from a province, or the outflow premium.

Figure 2.1 depicts both premiums for each province with the inflow premium on the y-axis and the outflow premium on the x-axis. The size of the circle represents the development of province measured by GDP per capita in 2014. Provinces above the horizontal zero line mean a positive inflow premium, i.e., students who migrate to those provinces receive a gain than the rest of the population. Therefore these places are more attractive. On the right of vertical zero line, provinces have a positive outflow premium, which means there are better outside options for students there.

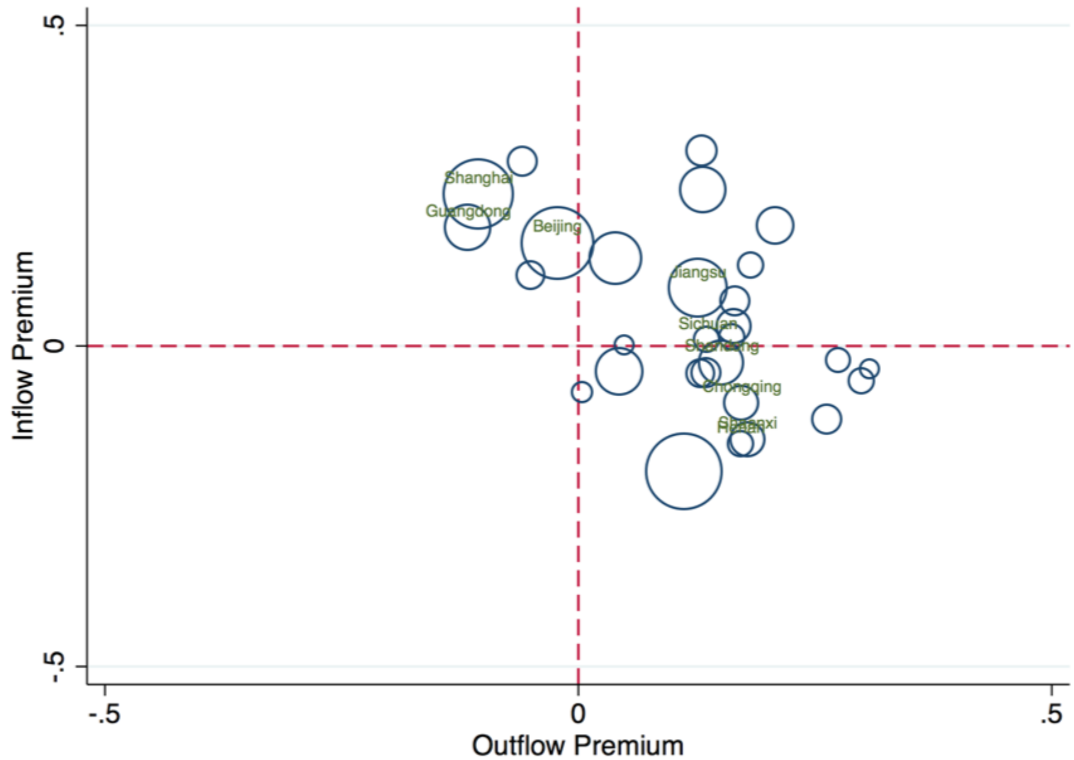



**Figure 2.1**: Inflow and Outflow Migration Premium

Note: The figure plots the inflow premium against outflow premium for each province. The inflow premium is the coefficient of the interaction term between migration for job and job province dummy. The outflow premium is the coefficient of the interaction term between migrate for job and home province dummy. All regressions use OLS on nominal wages and the same specification as column 6 of Table 3. The size of the circle represents the development of province measured by GDP per capita in 2014. Some coefficients are not significant.

I present the results using OLS estimation to show the relative position, some coefficients are not significantly different from zero. This is because the sample size can be small for some provinces based on several selected colleges. Nevertheless, it provides some insights into the productivity of provinces. First, developed provinces tend to have negative or small outflow premium, suggesting that they are more likely to retain talents. Second, apart from one province[5], other developed provinces have higher inflow premium, meaning they equip incoming migrants

[5]Tianjin, which has the highest GDP per capita but located at the bottom of the graph, has a special industrial zone that has only 15% of the population but accounts for over 50% of the GDP. Once adjusted, its per capita GDP is much lower.

with a higher premium. Specifically, three metropolitan areas, Beijing, Shanghai and Guangdong are all concentrated in the northwest corner with significantly larger inflow premium coefficients.

Overall, this exercise suggests that the wage gain from migration mainly comes from high-income provinces. Meanwhile, several megacities are able to retain their human capital. It should be noted that utility gains from amenity, idiosyncratic preference and other factors are not included here. If the advantages of amenity in large cities are considered as a premium, their position in Figure 1 would further shift towards the northwest, making them more attractive because of overall improvement.

### 2.5.4 Ideal Migration

In the survey, respondents were asked about their ideal job place. This information is especially useful in understanding the migration barrier or constraint faced by graduates. Overall, 60% of students have their actual job province as the ideal job province. But 40% of them are not having a job in their ideal province. This could either due to sorting that excludes them from the market, or it might stem from some misallocation that can actually be alleviated.

In Table 2.10, I regress a dummy for whether the job location is ideal on various factors. In particular, I focus on discrimination during job search and the physical distance from colleges to the ideal job market as a measure of searching cost.

In column 1, I include a dummy for whether the applicants felt being discriminated against because they are from rural areas. Similarly, column 2 and column 3 include dummies for being discriminated by Hukou and accent respectively. All three columns show a negative relation between discrimination and finding a job at the ideal province. In column 4, I also test the impact of physical distance from school to ideal job place. The result shows that the longer the distance (migration cost), the lower the chance of ending up with the ideal job province. The last column put all factors together, Hukou discrimination and distance still play significant roles in achieving ideal migration for a job. All personal, family and education controls including home place and

**Table 2.10**: Determinants of Landing at Ideal Job Location

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Land at Ideal Job Location | | | | |
| Feel Rural Discriminated | -0.036* | | | | 0.0048 |
| | (0.020) | | | | (0.023) |
| Feel Hukou Discriminated | | -0.069*** | | | -0.038*** |
| | | (0.0094) | | | (0.012) |
| Feel Accent Discriminated | | | -0.053*** | | 0.0024 |
| | | | (0.020) | | (0.023) |
| Log(distance from School to Ideal Job) | | | | -0.056*** | -0.056*** |
| | | | | (0.0044) | (0.0043) |
| Age | 0.0012 | 0.00062 | 0.0023 | -0.00031 | 0.0018 |
| | (0.0054) | (0.0048) | (0.0052) | (0.0041) | (0.0047) |
| Female | 0.082*** | 0.084*** | 0.084*** | 0.055*** | 0.057*** |
| | (0.012) | (0.011) | (0.012) | (0.0098) | (0.012) |
| Married | -0.078 | -0.025 | -0.080 | 0.036 | -0.017 |
| | (0.10) | (0.11) | (0.10) | (0.092) | (0.095) |
| Minority | 0.032 | 0.040* | 0.040** | 0.020 | 0.033 |
| | (0.021) | (0.021) | (0.020) | (0.018) | (0.021) |
| Communist Party Member | 0.019 | 0.015 | 0.018 | 0.0088 | 0.012 |
| | (0.011) | (0.012) | (0.011) | (0.012) | (0.011) |
| Controls | Y | Y | Y | Y | Y |
| Observations | 8,890 | 9,471 | 8,967 | 9,058 | 7,428 |
| R-squared | 0.213 | 0.213 | 0.213 | 0.312 | 0.319 |

Note: Robust standard errors in parentheses, clustered by school. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Controls include family background and college performance.

college fixed effects are included.

The simple correlations imply potential ways to reduce the barrier of migration. To improve the allocation of talents, the government should emphasize on eliminating discrimination in the job market because of Hukou and reducing migration cost such as traveling cost.

## 2.6 Conclusion

Geographic mobility of the educated labor group has significant meanings to economic and individual development. Not surprisingly, people migrate in the interest of economic profit especially for the educated people. But how much of the income difference across places is due to selection on individual characteristics? and how much is due to inherent productivity difference across places? This study finds evidence in favor of the latter hypothesis and quantifies the private gains at migration. I highlight the importance of educated migrants, as well as the difference in their migration selection.

While OLS estimates produce a 12%-18% gain from migration, results from the matching method reveal the importance of selectivity in both ability and family background. The magnitude of coefficients does not change not because there are no selections in observables, but because the two selections offset each other. The results are robust to various definitions of migration. Moreover, the heterogeneity test finds no significant difference in the gains from migration among different family and education groups.

The unexplained gap leads to an investigation on the other mechanism of the wage gap between migrants and non-migrants, the destination of migration. By adjusting the cost of living, I highlight the importance of mean wage difference across places. Over 50% of the observed gains can be attributed to differences in the cost of living, which incorporates information about local underlying factors. This finding means it is more likely to be the case that innovative and productive firms are offering wage premium to attract skilled worker from the outside labor market. Indeed, migrants are more likely to end up with large firms and foreign firm in east region. The pattern from inflow and outflow premium also support such argument. Finally, using ideal job place information, I look at the factors that prevent students from getting to their ideal job place. Two potential improvements are pointed out: reducing discrimination on Hukou and reducing physical migration cost.

There are still many issues worthy of investigation in the future. First, gathering more data on other educated labor force would help verify the external validity of this study. Second, what is the impact of migrants on the local labor market? Third, for provinces in different situations of human capital flows, what are effective policies that can improve the situation? For example, what is the role of college in retaining skilled labor? Finally, what are the impacts beyond the labor market, for instance, the marriage market, the housing market and long-term impact on social mobility. All these questions require much more effort and data on internal migrants in developing countries.

## 2.7 Appendix

**Table 2.11**: Number of Colleges Surveyed in CCSS, 2010-15

| | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|
| 211 program schools | 9 | 25 | 21 | 24 | 6 | 4 |
| Non-211 schools | 9 | 24 | 28 | 34 | 10 | 9 |
| Vocational schools | 1 | 1 | 1 | 7 | 1 | 0 |
| Total number | 19 | 50 | 50 | 65 | 17 | 13 |

Note: 211 program schools are elite 112 colleges. In the 1990s, the Chinese government put forward a proposal to enhance 100 colleges in the 21st century, which was later called the 211 Program. Although the proposal indicates only 100 colleges, in practice, 112 are covered by this program in 2010. Colleges covered by the Program have longer histories and offer high-quality education; more importantly, they also receive more financial support from the government.

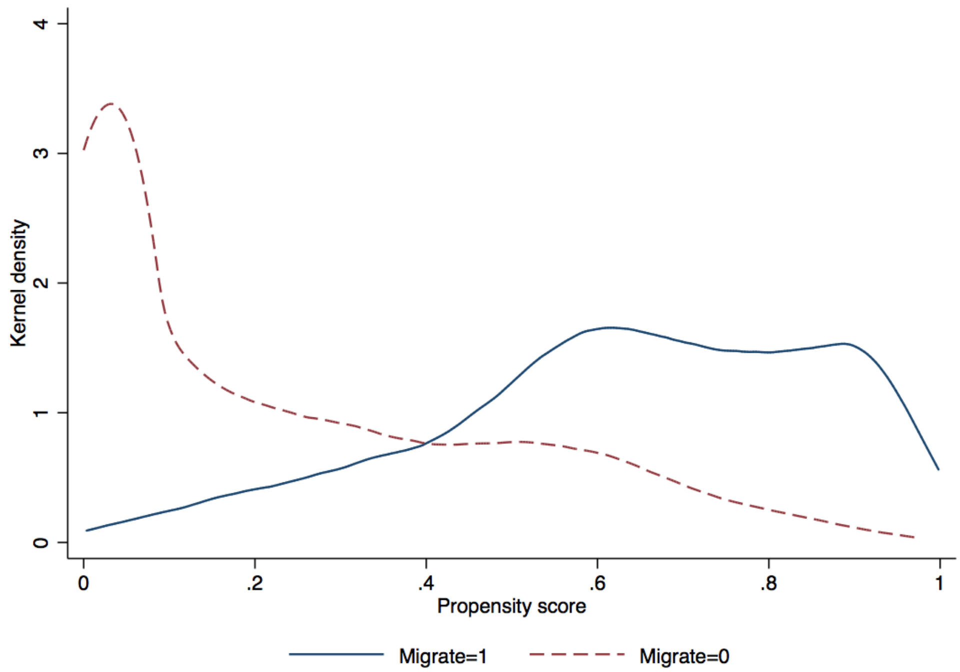



**Figure 2.2**: Density of Propensity Score by Migration Status

Note: The figure plots the density of predicted propensity score of migrating using Probit model. Same sample and control variables as in Table 2.3 column 6 are used.

# Chapter 3

# What We Learn from a Failed RD: Consequences of Retaking in High-Stakes Tests

## 3.1 Introduction

While exams and tests are commonly used to measure education output, scores and grades are often associated with further returns, e.g., honors, admission, license, scholarships. Many tests designed to measure achievement offer the opportunity to retake.[1] In contrast, some education systems rely heavily on one-time exam to assess students ability, thus a major impact on future outcomes (Jia and Li, 2017).[2]. In fact, a current debate in the education reform in China is whether such “one test for life" should be relaxed for more chances. However, the impacts of retake policy on the distribution of score, students' immediate efforts and long-term consequences are still a black box.

[1] Tests like SAT, ACT, GRE, GMAT, TOEFL, IELTS and GED all allow for retakes at varying fees. Individuals typically choose to only reveal the highest score to relevant parties.

[2] For example, college entrance exam in China and many other developing countries holds once a year. The opportunity cost of retaking is very high.

Retake is an option that reduces the adjustment cost. It improves the accuracy of individual scores by providing another opportunity to students who have “a bad day" during exams. It also allows for dynamic learning through feedback and inspires further efforts. These features are particularly salient in the presence of a threshold associated with certain standards and benefits. For students who are close to the cutoff, they have strong incentives to retake. However, there is relatively little evidence on the consequence of such retaking in the labor market especially in developing countries.

English, as a second language, is of great importance to college students in a globalized labor market. While learning English becomes extremely popular among Chinese students in the past two decades, there is no causal evidence that a higher English proficiency necessarily leads to a higher labor market outcome. Parents invest for their children in learning English mainly to help them gain access to high quality education resources and prepare for future study abroad. The total annual investment in teenager English training from Chinese parents is estimated to be 14 billion RMB in 2012. Given such a huge input in learning English, the returns in labor market remains unknown. It would be useful not only to have an estimate of returns to English but also to understand how English skills interact with labor markets–improvement in productivity or signaling.

In this paper, I study the consequence of retaking in an important exam of China, the College English Test (CET). The CET provides an official measure of English skills for nationwide college students to the labor market. A publicly know passing threshold and low-cost retaking policy makes such context an interesting case to investigate retaking behavior. Using a survey data on college students in China, I first estimate the return to English skills for the overall population. After controlling for rich characteristics, I find significant positive effects of English skills on wages. The return to English skill is 6% per hundred points measured in CET4 scores globally.

I then focus on the cutoff for passing CET4 at 425 points. With a perceived benefit from passing the threshold, retakes manipulate the score by bunching selective students right above the cutoff. I document massive retaking behavior in response to the presence of threshold due to low cost of retaking. I find that there is a 10% jump in starting salary after graduation for those who barely pass the test. Among students at risk of failing English test, retakers are positively selected in terms of abilities unrelated to English skills. I then analyze the underlying mechanism of such jump by exploiting the manipulation at the cutoff. I restrict the sample to multi-time retakers to adopt a regression discontinuity design and find little evidence of sheepskin effect for this subgroup. However, it should be noted that this does not identify the causal impact of simply crossing the threshold without retake. From the analysis in other job outcomes such as industries, location and firm types, I find that the wage gap is more likely to be associated with access to certain types of firms.

This paper is the first paper to estimate economic returns to English-language skills in China, and the first to causally document the retaking effect of CET4 in labor market. It connects to two major stream of literature. First, it adds evidence to the return to language skills as human capital. Studies trying to answer such questions are often set in developed countries or countries with linguistic diversity. Early studies focus on the wage gap between linguistic groups (Carliner, 1981; Grenier, 1984; McManus et al., 1983) and later on immigrants (Berman et al., 1994; Carnevale et al., 2001; Chiswick, 1991; Tainer, 1988). Recent work tries to address endogenous issue using IV (Bleakley and Chin, 2004; Chiswick and Miller, 1995; Shields and Price, 2002), panel data (Dustmann and van Soest, 2002; Saiz and Zoido, 2005) and exogenous shocks (Angrist et al., 2008; Angrist and Lavy, 1997). The estimates of returns to language skills vary across data, languages and methods, but most of them show a positive impact on wages. Nevertheless, few studies pay attentions to developing countries for the lack of data. A recent exception is Azam et al. (2013). They examine the returns to English in India, and find a 34% wage premium for male adults who speaks fluent English.

The paper also tackles on the role of retaking and provides a deeper understanding on the underlying mechanism of how English skills interacts with labor markets, which further guides the policy to improve efficiency in education reform. The retaking behavior have been studies recently mostly in a regression discontinuity framework that are closely related to the education outcomes in school and signaling effect of diploma or certificates (Clark and Martorell, 2014). For instance, Vigdor and Clotfelter (2003) finds that SAT retake policy places students with high test-taking costs at a disadvantage in admission.

One paper closely related to this one is Goodman et al. (2020) where they find retaking in SAT substantially improves scores and increases four-year college enrollment rates, particularly for low income and URM students. My paper suggests that retaking could have much longer impact even in the labor market. Similar to retaking, Diamond and Persson (2016) find bunching in Swedish math test score distributions under teachers grading discrepancy and they show that receiving a higher grade leads to far-reaching educational and earnings benefits. The manipulation stemmed from retaking also alerts the importance of assumptions in regression discontinuity. Unlike Urquiola and Verhoogen (2009) where manipulation is caused by market incentives, a simple threshold regulation in hiring could cause the same issue.

The rest of paper is organized in the following orders. Section 2 gives a review of institutional background and data. Section 3 shows the estimates on return to English skills. Section 4 focus on the role of retaking at the cutoff. Section 5 studies the mechanism of wage gap at cutoff by looking into other job outcomes. Finally, Section 6 concludes.

## 3.2 Background and Data

In this section, I describe the key features of the relevant institution, the College English Test (CET) in China. Since the dataset is the same as in Chapter 2, I only provide additional information on variables related to CET.

### 3.2.1 College English Test (CET) in China

College English Test (CET) is a nationwide English language test for current college students in China. The purpose of the test is to examine the English proficiency of undergraduate students in China and ensure students meet requirement of National College English Teaching Syllabus. It is administered by the National College English Testing Committee on behalf of the Ministry of Education. Since 1987, CET has been taken for almost thirty years. Table 3.7 shows the level of proficiency CET stands for.

CET consists of three bands: CET4 (typically taken by sophomore), CET6 (typically taken by junior students) and CET-Spoken English Test (CET-SET). In this paper, I focus on the written tests CET4.[3] Students majoring in any discipline except English are eligible for CET4. Over 9 million college students (40% of total registered college students) took CET4 and CET6 in June, 2017.

Both CET4 and CET6 lasts 140 minutes. There are four sections: listening, reading, comprehensive (cloze, proofreading or translation) and writing.[4] Each time the questions are designed by a committee of selected anonymous professors in English major to maintain a stable level of difficulty over time under the same guidance of "National College English Curriculum Requirements (after 2004)". Therefore, it serves as a consistent benchmark for English language proficiency all over the country.

CET exams are held twice a year, at the end of each semester (in June and December). At the designated weekend of the exams, students all over the country take the exams at the same time (usually at their own colleges). Only current registered college students are allowed to sign up for the test. After the test, all the answer sheets will be collected and delivered to grading

[3]CET-SET is available to undergraduates who have either passed the CET4 with a score of at least 550 or passed the CET6 with a score of at least 520. The number of CET-SET takers is much smaller compared to CET4 and CET6 takers.

[4]The test had one major reform in 2005 and some adjustment in 2012 in the structure and format. For example, after 2012, the length increases by 5 minutes and emphasize more on practical skills such as translation. But the score are overall comparable and consistent.

process organized by CET committee.

CET is a “criterion-related norm-referenced” exam. After 2005 reform, the full score is 710 and examinees’ original mark will be converted into the normal-distributed score with norm-referenced through scoring weighting and equating. A score report will be sent to every test taker. On this piece of paper, the total score, as well as sub-score is reported. The standard for passing is 425 or above in total score regardless of sub-score. Students are qualified for CET6 if they have passed CET4.

One important feature of the test is that students have no precise control over their score. There is no public rubric, indicator of exact points (even on the exam) and regrade policy. For subjective questions such as writing and translation, the score is also sensitive due to the noise from graders. Therefore, measurement error around true skill level happens all the time.

College students can retake as many times as they want before graduation at the cost of $3-8 fees depending on college location. In theory, the maximum repeating times is eight. There is no adverse impact of retaking since past record (score and retake behavior) is not revealed in each report. In other words, each attempt is independent and only the maximum score matters. Due to low cost of retaking, it is a common practice to retake CET4 for most students till they choose to advance to CET6 or give up.

Firms and employers rely on the report as the proof of English skills and passing CET. Passing CET4 is one of the accomplishments in undergraduate study. It used to be the case that bachelor diploma requires the pass of CET4. Such policy was abandoned two decade ago, but the tradition of taking CET4 remains. In the job market, along with other international English tests, CET score is widely recognized by employers. Many firms list passing CET4 as one of the requirements in application.

### 3.2.2 China College Student Survey

The main dataset used in this paper comes from the same one in Chapter 2, China College Student Survey (CCSS). More details can be found in Section 2.2.1. In this study, I focus on the information on CET score in the survey.

In terms of language proficiency, students were asked to give their highest CET scores if they ever participated in CET. Both CET4 and CET6 scores are recorded if they ever participated the exams. About 70% students report CET4 records, and 40% provide scores on CET6.

After 2013, the number of attempts are also recorded, which provides direct information on retake behavior. For these years of survey, I leverage the retaking information and exploit the role of retaking at the cutoff. I drop 2012 wave because the question was rephrased to report their most recent CET score, which may provide additional information on English skill at the end of their college education. Therefore I separate it from the main sample (7710 observations). Table 3.8 in Appendix provides a more detailed information on the available variables from each year.

I also make the following selection on the sample. First, I exclude students from vocational college to include only four-year students. Second, I exclude students from English major and other language majors who are not subject to CET.[5] After these exclusions, it leaves 27929 observations (with 15171 from 2013-15). The outcome variable, starting monthly salary, is defined as the highest offer received at the time of the survey.

Table 3.1 presents summary statistics for the above sample. The average scores of CET4 and CET6 are around 462 and 441. Both mean values are higher than the cutoff value (425) for certificates. The pass rate for CET4 and CET6 in the sample is 80% and 62% respectively.

[5]Students majored in English take “Test for English major" (TEM) and other specialized tests.

**Table 3.1**: Statistical Summary of Key Variables

| | Mean | S.D. | Min | Max | count |
|---|---|---|---|---|---|
| Age | 22.90 | 1.09 | 18.0 | 30.0 | 27167 |
| Female | 0.41 | 0.49 | 0.0 | 1.0 | 27805 |
| Minority | 0.07 | 0.26 | 0.0 | 1.0 | 27723 |
| Communist Party Member | 0.32 | 0.47 | 0.0 | 1.0 | 27686 |
| | | | | | |
| Family Size | 4.23 | 1.24 | 1.0 | 15.0 | 27069 |
| Has Gov Parent | 0.12 | 0.33 | 0.0 | 1.0 | 27929 |
| Mom Schooling | 9.21 | 4.30 | 0.0 | 19.0 | 25585 |
| Dad Schooling | 10.39 | 3.79 | 0.0 | 19.0 | 25604 |
| Has College Parent | 0.24 | 0.43 | 0.0 | 1.0 | 25596 |
| Log(Household Income) | 10.47 | 1.11 | 6.9 | 13.1 | 23261 |
| | | | | | |
| Elite University | 0.40 | 0.49 | 0.0 | 1.0 | 27929 |
| STEM | 0.75 | 0.44 | 0.0 | 1.0 | 27288 |
| Top 20% GPA | 0.45 | 0.50 | 0.0 | 1.0 | 26749 |
| Student Union | 0.61 | 0.49 | 0.0 | 1.0 | 27929 |
| Provincial Awards | 0.12 | 0.33 | 0.0 | 1.0 | 27169 |
| Internship/Parttime | 0.71 | 0.45 | 0.0 | 1.0 | 27835 |
| Has Failed Courses | 0.41 | 0.49 | 0.0 | 1.0 | 26875 |
| | | | | | |
| CEE Chinese Z-score | 0.03 | 0.80 | -3.0 | 3.0 | 21149 |
| CEE Math Z-score | 0.07 | 0.83 | -3.0 | 3.0 | 21101 |
| CEE English Z-score | 0.05 | 0.83 | -3.0 | 2.9 | 21123 |
| CEE Comprehensive Z-score | 0.06 | 0.78 | -3.0 | 3.0 | 20390 |
| | | | | | |
| College English Test: | | | | | |
| CET4 Score | 462.00 | 60.92 | 100.0 | 700.0 | 21298 |
| CET4 Attempts | 1.93 | 1.39 | 0.0 | 8.0 | 13221 |
| CET6 Score | 440.85 | 65.08 | 100.0 | 700.0 | 11912 |
| CET6 Attempts | 1.98 | 1.24 | 0.0 | 8.0 | 8330 |
| | | | | | |
| Starting Salary (in RMB) | 2726.85 | 1168.97 | 600.0 | 8000.0 | 11917 |

Note: CEE score is the college entrance exam score standardized across year-province-track. The mean is slightly higher than zero as I excluded students from vocational colleges from the bottom of the distribution. CET attempts are available only for year 2013-2015.

## 3.3 Return to English Skills

### 3.3.1 OLS

A simple Mincer-type equation can be specified:

$$y_i = \alpha + \beta \text{ English}_i + \pi X_i + \lambda_t + \mu_s + \Delta_m + \varepsilon_i$$

where $y_i$ is the monthly starting wage of individual $i$. English is the proficiency of English measured in CET4 scores.$\lambda$, $\mu$ and $\Delta$ capture year fixed effect, school type and major fixed effects. $\beta$ is the coefficient of primary interest, the returns to English proficiency. $\varepsilon$ is the residual term.

$X$ is a rich set of controls. First, it includes demographic factors such as age, gender, and whether being a minority. Second, family background like parental income and education are included. To represents one's performance in colleges, dummy variables like whether being a Chinese Communist Party (CCP) member, whether joins in school student union or not, whether has an internship or not are added. To address the omitted variable problem, the most representative proxy for ability, the National College Entrance Examination (NCEE) score, as well as college academic ranking, is included.

The rich information from the unique dataset helps address potential biases due to omitted variables. First, since ability, the main omitted variable, is positively correlated with wage and language proficiency, $\beta$ is likely to be overestimated. With all the demographic information and academic performance prior to graduation, I am able to mitigate this bias using a compete set of control variables including important proxies for ability like National College Entrance Exam scores.

Second, unlike previous studies using self-reported proficiency levels, the CET score provides a nationwide standard measurement on English proficiency. This improvement gives us advantages not only in the accuracy of effect but also in creating regression discontinuity design. Thus the classic downward bias resulted from measurement error can be avoided.

The third benefit of this survey data is that college graduates have no formal working experience and their language skills are not developed for their career-related purpose, thus are less endogenous than experienced adults. In addition, college graduates have similar schooling years.

It should be noted that OLS estimation may not have a causal interpretation even after controlling for these variables. It is still possible that there are unaccounted omitted variables that simultaneously affect wage and English proficiency, such as the ability that are not captured by proxy variables. The results should be viewed as an upper bound of the true impact.

### 3.3.2 Results

Table 3.2 presents the results of OLS estimates on the full sample. In column 1, I only include CET4 score, demographics and year dummies. The coefficient of CET4 score is 0.23 and significant at the level of 1%. In other words, a hundred-point improvement in CET4 score is associated with 23% increase in monthly wage.

Column 2 adds individual characteristics including their political status (whether being a CCP member), social activity (whether serve in school student union), and academic performance (whether being top 20% in GPA). To address the potential problem of unobserved ability, standardized National College Entrance Exam score is also included as the proxy for students' ability. Unsurprisingly, the coefficient of CET4 score drops to 0.15 but is still significant at 1% level.

Column 3 further controls for family background. College parent is a dummy which equals one if either of the parents has a college degree. After controlling for parental income and education, the coefficient of CET4 decreases to 0.13, indicating that about half the effect of English proficiency on wage in column 1 can be attributed to ability and background.

In column 4, major fixed effect and school type fixed effect are included. The coefficient reduces to 0.065 with a significance level of 5%. Column 5 further adds school region fixed effect to account for job market environment. The coefficient slightly decreases to 0.06, and remains significant at 5% level. That means, on average, a student with a score of 600 in CET4 would have a 6% higher wage than one with 500 points given other factors fixed.

How difficult is it to improve CET score by a hundred points? Similar to GRE or SAT, official website has a reference table to exchange the points to the percentile ranking in norm-referenced test. For CET4, 410 points approximately corresponds to 11% percentile ranking. 510 corresponds to 55% and 610 corresponds to 95%. The quantile is similar for CET6. So a hundred

**Table 3.2**: OLS Estimation Results on Return to CET4

| Independent variables | Dependent variable: ln (Monthly wages) (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| CET4 score/100 | 0.23*** | 0.15*** | 0.13*** | 0.065** | 0.060** | | -0.062 |
| | (0.024) | (0.023) | (0.025) | (0.029) | (0.029) | | (0.053) |
| CET6 score/100 | | | | | | 0.071*** | 0.10** |
| | | | | | | (0.03) | (0.04) |
| Age | 0.0092 | 0.0012 | -0.0098 | -0.041*** | -0.039*** | -0.0084 | -0.0099 |
| | (0.007) | (0.006) | (0.007) | (0.012) | (0.012) | (0.016) | (0.015) |
| Female | -0.21*** | -0.20*** | -0.22*** | -0.11*** | -0.12*** | -0.11*** | -0.11*** |
| | (0.037) | (0.023) | (0.021) | (0.020) | (0.020) | (0.032) | (0.031) |
| Minority | 0.082 | -0.018 | 0.042 | -0.037 | -0.033 | 0.11* | 0.11* |
| | (0.079) | (0.076) | (0.120) | (0.110) | (0.120) | (0.063) | (0.066) |
| CCP member | | 0.11** | 0.13** | 0.068 | 0.074 | 0.014 | 0.014 |
| | | (0.045) | (0.062) | (0.046) | (0.047) | (0.023) | (0.023) |
| Student union | | 0.073 | 0.081 | 0.059 | 0.054 | 0.007 | 0.0026 |
| | | (0.059) | (0.056) | (0.044) | (0.044) | (0.056) | (0.056) |
| Internship | | -0.11** | -0.11*** | -0.092** | -0.098*** | -0.039 | -0.036 |
| | | (0.042) | (0.042) | (0.037) | (0.035) | (0.028) | (0.027) |
| Top 20% GPA | | -0.00016 | 0.0018 | 0.028 | 0.03 | 0.037 | 0.037 |
| | | (0.025) | (0.029) | (0.025) | (0.025) | (0.034) | (0.033) |
| NCEE score | | 0.11*** | 0.11*** | 0.061** | 0.061** | 0.0077 | 0.01 |
| | | (0.015) | (0.015) | (0.023) | (0.023) | (0.030) | (0.032) |
| College parents | | | 0.061* | 0.042 | 0.036 | 0.071*** | 0.075*** |
| | | | (0.036) | (0.029) | (0.028) | (0.025) | (0.028) |
| ln (parental income) | | | 0.0014 | -0.00021 | -0.0012 | 0.014** | 0.013** |
| | | | (0.0110) | (0.0092) | (0.0090) | (0.0063) | (0.0067) |
| Year FE | Y | Y | Y | Y | Y | Y | Y |
| Major FE | | | | Y | Y | Y | Y |
| School type FE | | | | Y | Y | Y | Y |
| School region FE | | | | | Y | Y | Y |
| Constant | 6.18*** | 6.82*** | 7.14*** | 7.82*** | 7.81*** | 7.01*** | 7.23*** |
| | (0.24) | (0.18) | (0.29) | (0.60) | (0.59) | (0.58) | (0.60) |
| Observations | 11612 | 7488 | 6520 | 6516 | 6516 | 3706 | 3625 |
| Adjusted $R^2$ | 0.24 | 0.31 | 0.31 | 0.37 | 0.38 | 0.35 | 0.36 |

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Each regression is sample weighted and clustered by school. School type includes 211, non 211 and vocational schools. Students are grouped into 13 majors.

points make a huge difference in English proficiency.

It is not surprising that the effect of CET4 keeps decreasing through column 1 to 5. About 75% of the effect that is originally attributed to language proficiency actually comes from difference in other factors like family background and ability. It is clear that I can not control for all the unobserved factors, but with a complete set of control variables including NCEE scores, I believe the bias should be minimized. In fact, a higher NCEE score has a positive impact on wage, which implies that NCEE score is indeed a good proxy for ability.

Students who pass CET4 (425 or above) are then eligible for taking CET6. Therefore, another measurement on English proficiency can be used if they take both tests. Column 6 uses CET6 score instead of CET4 score in regression with full specification. The result shows that the returns to English skill is about 7% for a 100-point improvement in CET6 score. Since CET6 is much more difficult in both vocabulary and comprehensive application, a larger effect is reasonable compared to the effect of CET4 in column 4.

In column 7, both CET4 and CET6 scores are included. Intuitively, when student acquires a good score on CET6, CET4 would become less useful in job market. The result confirms that there is a significant effect of CET6. The coefficient increases to 10% and is significant at 1% level. But the coefficient of CET4 is no longer significant.

The results should be interpreted carefully because the sample is now restricted to a smaller group. There are two kinds of selection. First, students who score below 425 is not eligible for CET6, thus are automatically dropped. Second, even when they pass CET4, students can choose not to take CET6. Those who take CET6 are likely to be more confident and skilled in their English. Both selections lead to a more capable group, i.e., high ability students, thus a nonrandom sample. We will address this issue soon.

### 3.3.3 Heterogeneity by CET status

It would be helpful to analyze the effect by proficiency group. According to cutoff of CET, students can be categorized into three groups, low skilled (students who do not pass CET4), medium skilled (pass CET4 but not CET6), and high skilled (pass CET6). Within each group, the effect of English skills may be different depending on which measurement is used. For low skilled group, only CET4 is available because they are not eligible for CET6. For medium skilled group, CET4 is available, but not all student has CET6 score. For high skilled group, both CET4 and CET6 can be applied. Table 3.3 presents the results within each group.

Column 1 in Table 3.3 shows the results using same OLS specification within low skilled group, i.e., students who do not pass CET4. The result shows that if the students do not pass CET4, CET4 score has no effect on wages, i.e., there is no marginal return when English skill is at a low level. Column 2 and 3 test the same idea on medium group. The results also suggest that within medium group. Neither CET4 and CET6 score matters.

Column 4 and 5 are similar to column 6 in Table 3.2. The difference is that the sample only includes students who score 425 or above in CET6, so the sample size decreases. The coefficient of CET4 is insignificant in column 4, which coincides with previous conclusion that for high ability student, CET4 is no longer effective. Column 5 shows that the effect of CET6 is about 9.4% for 100-point improvement above 425.

However, like I mentioned before, it is important that the selection into taking CET6 is properly addressed. Column 5 and 6 is subject to endogenous choice by students who passed CET4. Those who are eligible to take CET6 but choose not to is not the same as those participate in CET6. If people with high ability incline to further take CET6, then the estimates column 6 would be overestimated.

To mitigate this issue, I adopt propensity score matching to match students who have the same probability in taking CET6, but end up with different outcome. Probit model is used to calculate the score. To be robust, 20% trimming is also applied. The results are shown in

**Table 3.3**: OLS Estimation Results on Return to English by English Skill Group

| | Dependent variable: ln (Monthly wages) | | | | | |
|---|---|---|---|---|---|---|
| | Low Skilled | Medium Skilled | | High Skilled | | Matched |
| Independent variables | (1) | (2) | (3) | (4) | (5) | (6) |
| CET4 score/100 | 0.0038 | -0.015 | | 0.031 | | |
| | (0.024) | (0.039) | | (0.035) | | |
| CET6 score/100 | | | 0.055 | | 0.094 | 0.072** |
| | | | (0.080) | | (0.045) | (0.030) |
| Age | -0.069** | -0.016* | -0.0045 | -0.015 | -0.016 | -0.0053 |
| | (0.027) | (0.009) | (0.021) | (0.015) | (0.015) | (0.018) |
| Female | -0.11*** | -0.17*** | -0.14** | -0.091*** | -0.097*** | -0.095*** |
| | (0.027) | (0.040) | (0.058) | (0.024) | (0.024) | (0.034) |
| Minority | -0.11 | 0.062 | 0.17 | 0.078* | 0.075* | 0.12 |
| | (0.260) | (0.065) | (0.110) | (0.040) | (0.041) | (0.074) |
| CCP member | 0.23* | 0.029 | 0.023 | -0.0016 | -0.00026 | 0.013 |
| | (0.120) | (0.024) | (0.042) | (0.030) | (0.030) | (0.024) |
| Student union | 0.061 | 0.082** | 0.084 | -0.039 | -0.040 | -0.0040 |
| | (0.067) | (0.039) | (0.072) | (0.042) | (0.041) | (0.063) |
| Internship | -0.27*** | -0.01 | -0.068 | -0.025 | -0.022 | -0.047 |
| | (0.058) | (0.028) | (0.053) | (0.040) | (0.039) | (0.034) |
| Top 20% GPA | 0.0057 | 0.036 | 0.046 | 0.070* | 0.063* | 0.042 |
| | (0.015) | (0.049) | (0.048) | (0.037) | (0.037) | (0.036) |
| NCEE score | 0.080*** | 0.031 | -0.029 | 0.072*** | 0.067*** | 0.0026 |
| | (0.021) | (0.027) | (0.025) | (0.022) | (0.022) | (0.033) |
| College parents | -0.078* | 0.12** | 0.055 | 0.069** | 0.067** | 0.064 |
| | (0.043) | (0.047) | (0.056) | (0.030) | (0.030) | (0.033) |
| ln (Parental income) | -0.014 | 0.0086 | 0.0083 | 0.017* | 0.016* | 0.016** |
| | (0.014) | (0.006) | (0.011) | (0.009) | (0.009) | (0.0067) |
| Constant | 9.47*** | 7.96*** | 7.46*** | 7.25*** | 6.97*** | 7.29*** |
| | (0.790) | (0.250) | (0.540) | (0.510) | (0.550) | (0.47) |
| Observations | 1,276 | 3,099 | 1,484 | 2,113 | 2,113 | 2878 |
| Adjusted $R^2$ | 0.358 | 0.390 | 0.420 | 0.219 | 0.223 | 0.347 |

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Each regression is sample weighted and clustered by school. School type, region, year and major fixed effects are all included.

Column 6. After taking the selection into taking CET6 into account, the coefficient of CET6 score decreases as expected. The magnitude is now 7% and signifiant at 5% level. This estimate is more reliable than that in column 5, and is the same at column 6 in Table 3.2.

While estimates in high skilled group is consistent, the low skilled group results in Table 3.3 seems to be conflicted with the conclusion from Table 3.2. While CET4 shows a 6% return per 100 points in Table 3.2, it is not effective within each group in Table 3.3. It is less likely to be the case that the score is insensitive to English skills and fails to be a good measurement given the large dispersion in score. One possibility is that the effect of CET4, though represents skills, is not linear. In order to gain the benefit of learning English, it has to reach the threshold. The effect in Table 3.2 is merely a linear approximation, which neglects the underlying structure of effects. In light of this, further analysis is needed to reconcile the the previous two results.

## 3.4 Focusing on the Cutoff

In this section, I focus on CET4 and turn to the role of retake at the cutoff. Figure 3.1 depict the raw relationship between CET4 score and starting monthly wages for full sample.

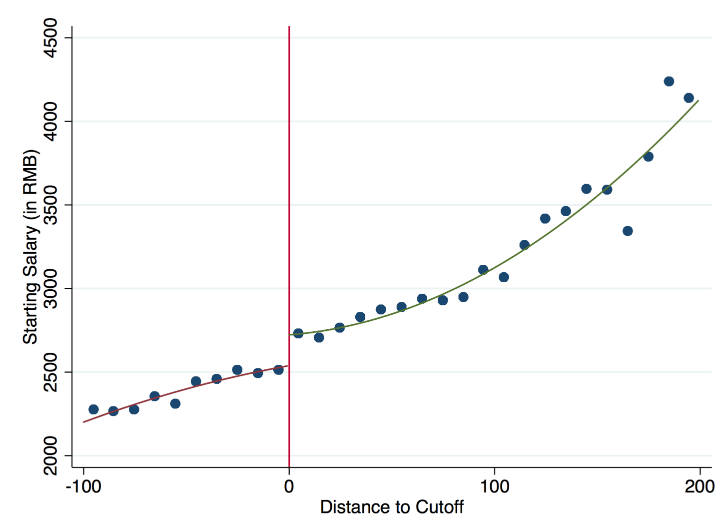



**Figure 3.1**: CET4 Score and Wages

There is a clear jump in the wages at the cutoff, which typically invokes a regression discontinuity design. However, due to retake, the distribution of maximum score violates the no manipulation assumption therefore invalidates the normal interpretation of the signaling effect. The main task for the rest of the paper is to explore rich information from the data and uncover the source of wage gap.

### 3.4.1 Who Are Retakers

Figure 3.2 (a) shows the raw distribution of the maximum CET4 scores centered to the 425 points cutoff. There are two dominant patterns from the score distribution. First, there are significant bunching at the threshold and the right of the it. Second, there are spikes in the density at 5 or 10 points. The later pattern can be smoothed as in Figure 3.2 (b) and (c) when excluding rounder and smoothing using 5-point bin respectively. The bunching behavior is clear in both figures that students are disproportionally distributed above the cutoff, a sign of manipulation. This feature is commonly found in exams with teacher discretion or retake (Diamond and Persson, 2016; Goodman et al., 2020).

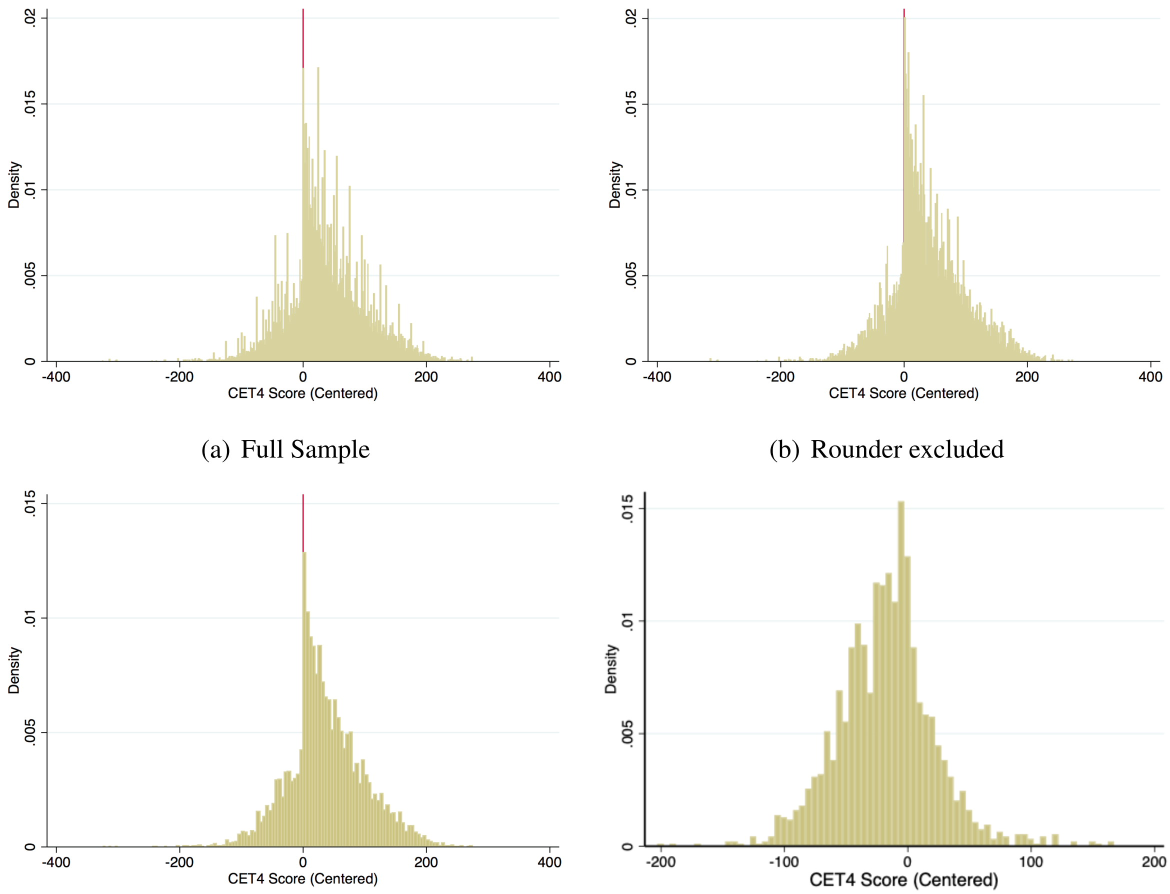
In terms of retaking, Table 3.4 below tabulates the number of attempts from 2013-15 sample. Out of 15171 observations, 12865 students participate CET4 at least once (85%), and most students take less than 5 times.[6]

**Table 3.4**: Number of Attempts, 2013-15

| Attempts | Frequency | Percent |
|---|---|---|
| 1 | 7194 | 55.92 |
| 2 | 1946 | 15.13 |
| 3 | 1720 | 13.37 |
| 4 | 1129 | 8.78 |
| 5 | 572 | 4.45 |
| 6 | 248 | 1.93 |
| 7 | 43 | 0.33 |
| 8 | 13 | 0.10 |
| Total | 12865 | |

To understand retake behavior along the maximum score distribution, I use the same sample and directly plot the average number of attempts by 5-point bin in Figure 3.3. The pattern is quite clear that there is a discontinuity in the retake behavior both in share and intensity. The retake behavior increases as the maximum score gets closer to the threshold. On average, students on the left of the cutoff take two more attempts than the ones on the right.

[6]The maximum number of attempts is 8. In practice, freshmen usually wait for a year to prepare and senior students may miss the last one because job season already start.




(a) Full Sample

(b) Rounder excluded

(c) Full sample, bin=5

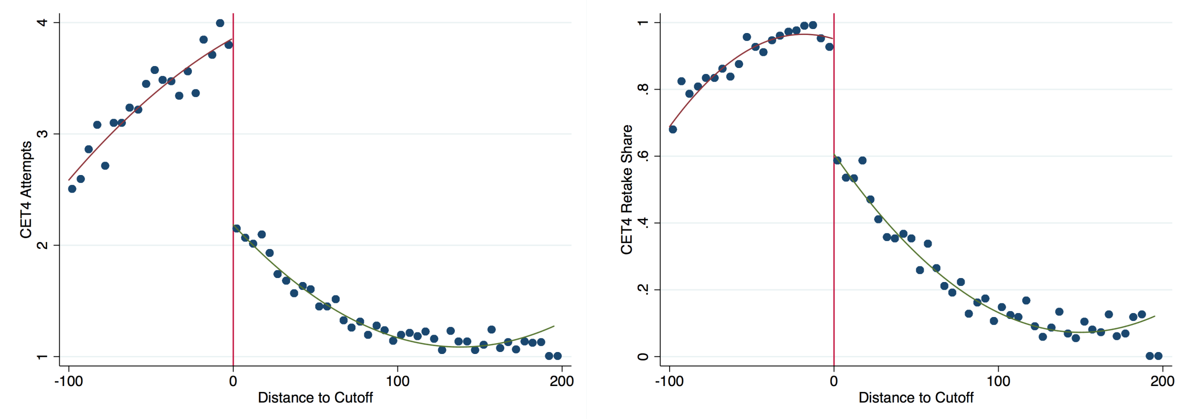
(d) Attempts>3, bin=5

**Figure 3.2**: Histogram of CET4 Maximum Score

Note: The figure plots the density of maximum CET4 score from 21298 observations in year 2010-11, 2013-15. Scores are centered at the 425 points cutoff (red line).

Visual inspection from Figure 3.3 and 3.4 indicates a strong incentive to retake CET4 to pass the threshold, and presumably students on each side of the cutoff are very different. The retaking behavior is increasing from below the threshold, and decreasing once the score exceeds the cutoff. There is a clear downward jump at the threshold for number of attempts and share of retakers.

To further analyze the characteristics of retakers, Table 3.5 takes a 50-point window on each side of the cutoff, and summarizes the mean characteristics in the first three column. In particular, column 2 and 3 separate students who take CET only once and pass from those who




(a) Average number of attempts (b) Average share of retakers

**Figure 3.3**: Retake behavior by 5-point bin, 2013-2015

Note: The figure plots the average of retake times and share by five points bin, 2013-15. Quadratic functions are fitted on either side of the threshold.

take multiple attempts to pass. Column 4 and 5 compare students who take advantages to pass (column 3) to the former two groups.

Column 4 shows that students who take advantages of retake have less educated parents and poor initial skills than those pass at first try. They are not only poor in English, but also other aspects like Math and Chinese. When compared with students who do not pass (include both single and multiple attempts), they are actually significant better in terms of college performance and initial ability. In addition, they are more likely to come from rich and large but not more educated families. The results indicate that students who pass through retake is positively selected in terms of their ability, performance and family income.

### 3.4.2 Can Observables Explain the Gap?

Given the above knowledge of selection, it is useful to quantify the gap controlling for the difference in observables. Formally, I present the results in Table 3.6. It should be noted that while the setup is similar to regression discontinuity design, it cannot be interpreted as the sheepskin effect from CET4 report.

**Table 3.5**: Mean Comparison Round Cutoff by Retake and Pass Status

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | within 50 points to cutoff | | | | |
| | No pass | Pass: once | Pass: retake | Diff: (3)-(2) | Diff: (3)-(1) |
| Age | 23.16 | 22.9 | 23.14 | 0.17*** | -0.02 |
| Female | 0.34 | 0.43 | 0.41 | -0.01 | 0.08*** |
| Minority | 0.11 | 0.06 | 0.08 | 0.02* | -0.02** |
| Communist Party Member | 0.22 | 0.32 | 0.28 | -0.04** | 0.06*** |
| Family Size | 4.34 | 4.42 | 4.49 | 0.07 | 0.15*** |
| Home in City | 0.32 | 0.31 | 0.31 | -0.01 | 0.00 |
| Has Gov Parent | 0.08 | 0.09 | 0.08 | -0.01 | 0.00 |
| Mom Schooling | 8.28 | 8.59 | 8.22 | -0.37** | -0.06 |
| Dad Schooling | 9.63 | 9.84 | 9.57 | -0.27* | -0.06 |
| Log(Household Income) | 10.41 | 10.64 | 10.60 | -0.05 | 0.19*** |
| Elite University | 0.29 | 0.30 | 0.28 | -0.02 | -0.01 |
| Top 20% GPA | 0.35 | 0.43 | 0.42 | -0.01 | 0.07*** |
| Student Union | 0.59 | 0.65 | 0.64 | -0.01 | 0.05** |
| Provincial Awards | 0.09 | 0.13 | 0.12 | -0.01 | 0.02** |
| Internship/Part time | 0.78 | 0.74 | 0.80 | 0.06*** | 0.02 |
| Has Failed Courses | 0.59 | 0.39 | 0.48 | 0.10*** | -0.10*** |
| CEE Chinese Z-score | -0.12 | 0.06 | -0.03 | -0.09*** | 0.10*** |
| CEE Math Z-score | -0.11 | 0.13 | -0.00 | -0.13*** | 0.11*** |
| CEE English Z-score | -0.33 | 0.10 | -0.09 | -0.19*** | 0.24*** |
| CEE Comprehensive Z-score | -0.04 | 0.05 | -0.01 | -0.06* | 0.03 |

Note: The first three columns summarize the mean statistics for three group round the 50 points window on each side of the cutoff. Number of observations varies across variables. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Panel A shows the results for CET4. The first two columns show the regression discontinuity estimates using parametric approach. A fourth order polynomial of the running variable is used on each side of the threshold. Column 1 presents the result with no additional control. It shows that students with a CET4 certificates has a benefit of additional 394 RMB compared to students who do not pass CET4, which equals to 17% increase in wage among students below the threshold.

Column 2 controls for some observed covariates that are prominent in the balance test. In particular, the standard NCEE score has a significant discontinuity at the cutoff, thus it is important to control for it. The other two covariates are dummies for college parents and 211 program schools, which represent family background and school quality respectively. The result shows a similar but less significant effect.

**Table 3.6**: Estimates of the Wage Jump at the Cutoff

| | Dependent variable: Monthly wage offer (RMB) | | | |
|---|---|---|---|---|
| | 4th order Polynomial | | Local linear regression | |
| | (1) | (2) | (3) | (4) |
| Panel A: CET4 | | | | |
| Pass CET4 | 394** | 379* | 493*** | 564*** |
| | (193.00) | (199.00) | (163.00) | (153.00) |
| Covariates | | Y | | Y |
| Observations | 2419 | 2419 | 2419 | 2419 |
| | | | | |
| Panel B: CET6 | | | | |
| Pass CET6 | 91 | 48 | -35 | -17 |
| | (292.00) | (263.00) | (214.00) | (302.00) |
| Covariates | | Y | | Y |
| Observations | 3162 | 3162 | 3162 | 3162 |

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. All regressions are weighted and clustered by school. Covariates includes standard NCEE score, dummy for 211 school and dummy for college parent. Local linear regression uses triangle kernel with a fixed bin of 10 points.

The last two columns in Table 3.6 use local linear regression to estimate the effect. It shows a 493-564 RMB (21%-25%) gap under two specifications. Both estimates are significant at 1% level and larger than parametric estimates in the first two columns. The estimates with covariates is similar to one without covariates in both approaches. The results suggest that the gap can not be simply explained by selection on observables like NCEE score or parental education. In fact, when I examine the at the cutoff for CET6 in panel B, it shows no significant wage gap at the threshold.

The difference in effects between CET4 and CET6 has interesting implications on the mechanism. Although both scores are positively correlated with English ability, labor markets treat CET4 differently from CET6 as passing CET4 is often listed as a minimum required. The only useful information conveyed in CET4 is whether the student pass the test or not. Firms use CET4 as a very basic screening mechanism. Detailed CET4 score is disregarded and the procedure is simplified only to eliminate unqualified candidates. This also explains why CET4 is insignificant once CET6 is included in Table 3.2 because having a CET6 score already demonstrates passing CET4.

In contrast, CET6 is often not required but optional by employers. CET6 score is more likely to represent actual proficiency in English skills as it is much more difficult compared to CET4. Therefore, the rest of the paper focuses on CET4 as it is more likely to be relevant to underlying retaking behavior.

However, it should be cautious that there might be unobservable qualities that contribute to the gap. For example, students who are persistent and motivated in learning would probably take advantages of retake, thus passing the threshold will be associated with personal traits that are unobserved but valuable in labor market.

### 3.4.3 Are Certificates Signaling

The full sample fails to meet the regression discontinuity design. However, for a subgroup, there is still a possibility to disentangle the signaling effect of CET4 certificates. In this subsection, I look for discontinuity in wages for students who have taken four times. By excluding students with low attempts, I focus on the group of retakers with similar level of struggle in CET4. Given that students have to pass CET4 to take CET6, it is important that each comparison should eliminate the influence of the other test. For CET4, I exclude students who have passed CET6 and have taken TOEFL or IELTS.

For these students on their last resort, students who are just below the threshold barely miss the treatment of passing CET4. The complicated rubric and grading system prevent anyone from choosing a score. Comparing students that is just below and above the cutoff therefore yields the estimates for having a piece of paper with score above 425.

To be more confident about above statement, I present two tests as the evidence. First, I check for the density of observations. Although the sample size reduces to about 15%, there is no longer bunching above the cutoff. Figure 3.2 (d) shows the distribution which satisfies McCrary test. Moreover there is no significant discontinuity in observables.

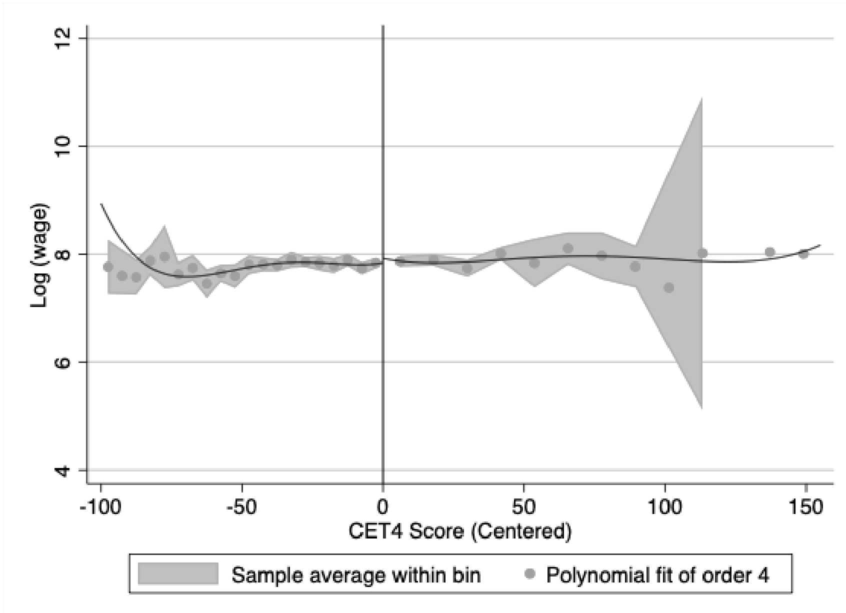



**Figure 3.4**: CET4 Score and Wages: Multi-time Retakers

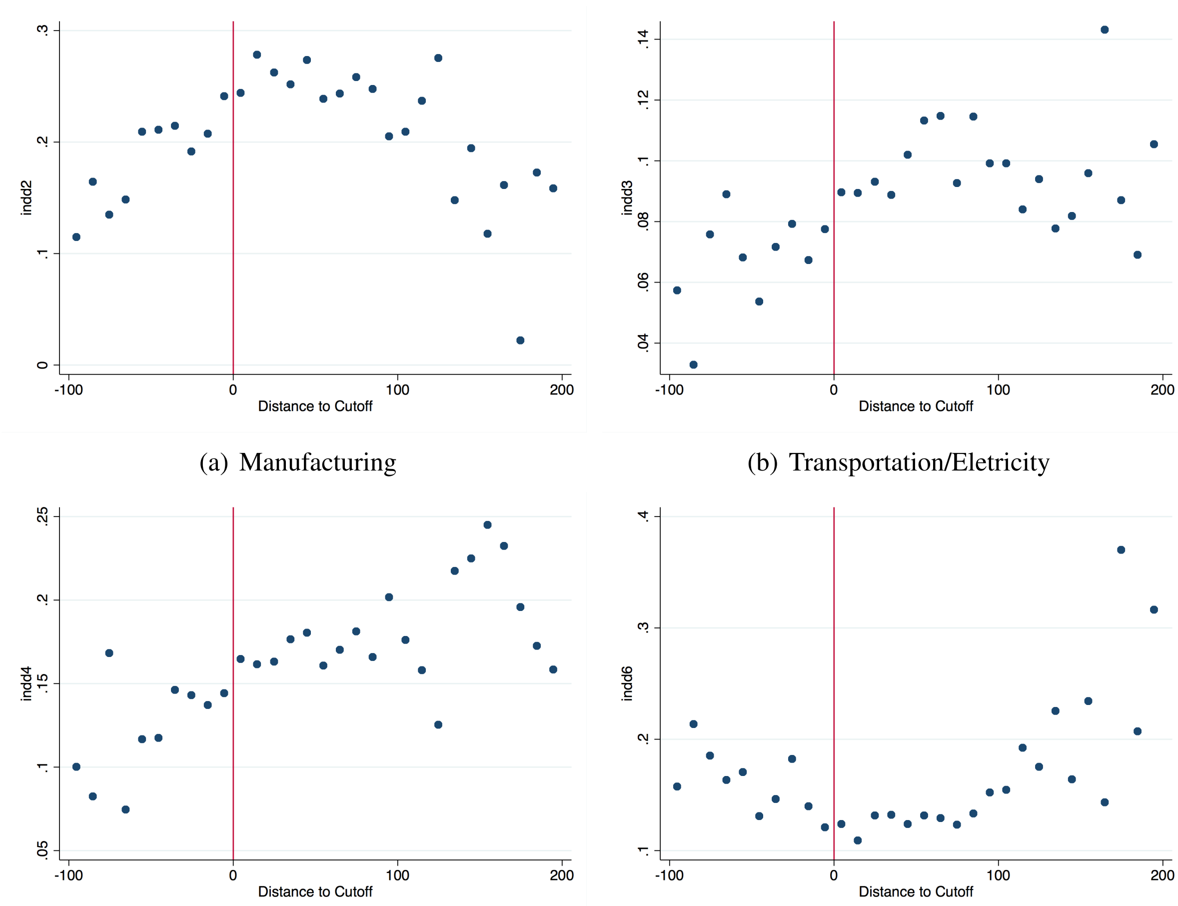
Note: The figure plots the average of wages within bins for CET4 retakers (n>3). The fitted line is estimated with 4th order polynomial.

Figure 3.4 shows the RD estimation of wages by the score of multi-time retakers. The figure shows no significant gap at the cutoff. For this particular group of students – multi-time retakers, there is no sheepskin effect from passing CET4. The results imply that the initial gap observed in the full sample has little to do with the signaling value of passing CET4. The gap is mainly caused by a compositional change across the cutoff and suggests strong sorting according to some unobserved quality such as resilience or learning skills. The retaking provide the opportunity to sort themselves to the right of the cutoff.

## 3.5 Other Job Outcomes

I have presented two main results relating the CET4 scores to wages. OLS estimation shows that CET4 score is overall positively correlated with wages. While there is a wage gap at the cutoff, regression discontinuity shows that just passing CET4 alone has no return for multi-time retakers, which suggests that the gap may come from students who are inherently different from retakers. This should show up in other dimensions of outcomes.

Figure 3.5 shows the average share of students within 10-point bin in four major industries. The figure shows that while there are selection into industries at difference scores, no discontinuity is found at the cutoff.




(a) Manufacturing

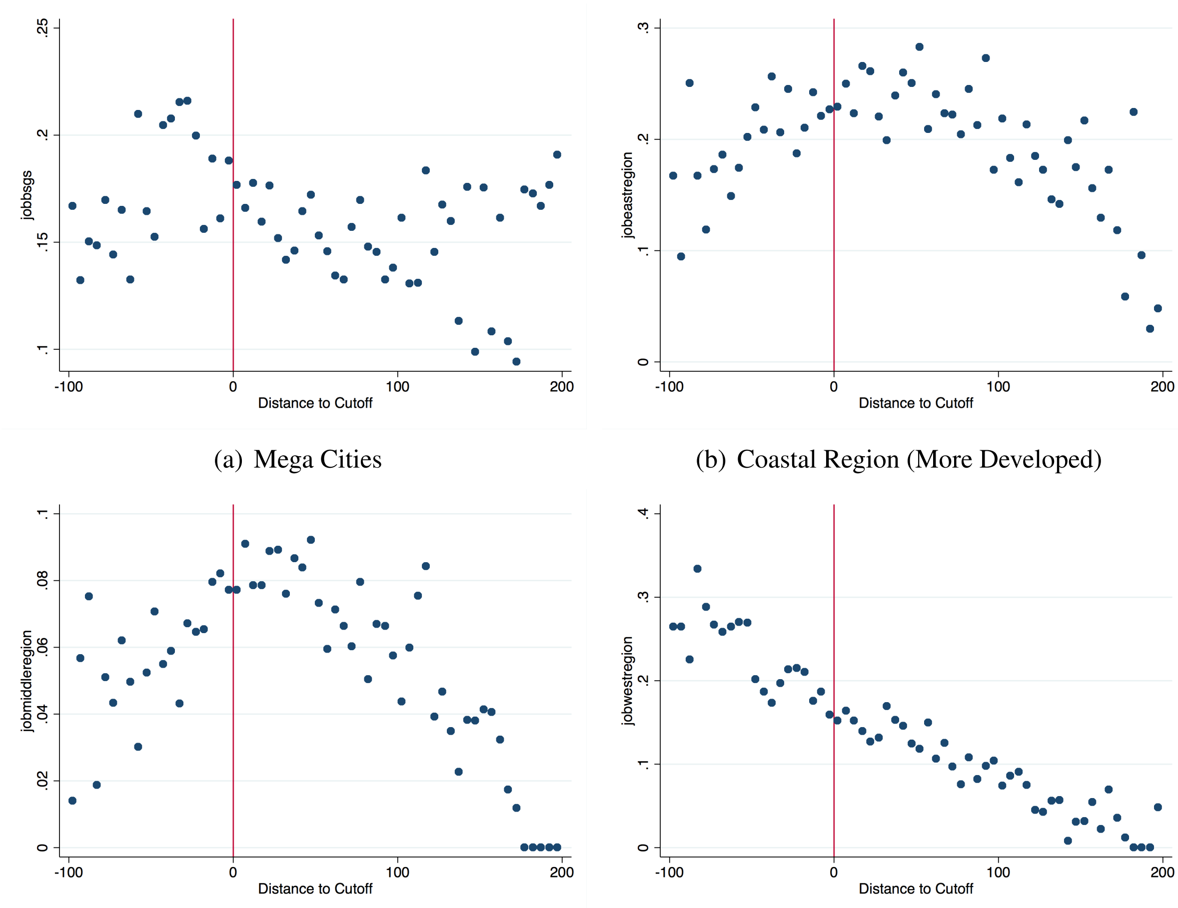
(b) Transportation/Eletricity

(c) IT

(d) Finance/Real Estate

**Figure 3.5**: Share of Jobs in Industries

Figure 3.6 examines the share of jobs in different locations. There is no significant discontinuity at the cutoff in each figure.




(a) Mega Cities

(b) Coastal Region (More Developed)

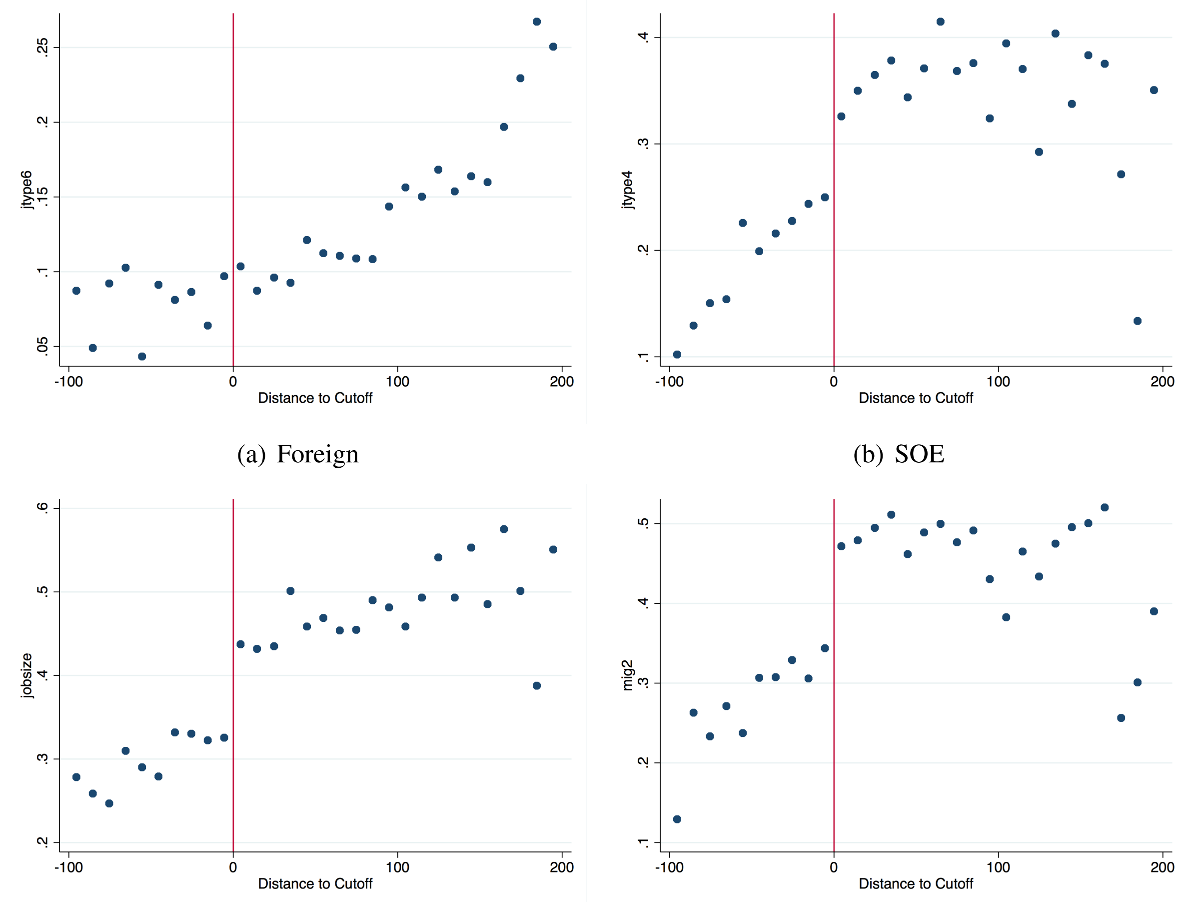
(c) Center

(d) West Region

**Figure 3.6**: Share of Jobs in Regions

Figure 3.7 examines the characteristics of firms they work for. While there is no significant discontinuity in the share of jobs in foreign-owned firms. Figure (b) shows a significant jump in the share of jobs in state-owned firms. Figure (c) also indicates that those who pass CET4 are more likely to work in firms with larger size. This is consistent with the fact that CET4 is used as an initial screening SOE in hiring.

Figure (d) demonstrates another impact of passing CET4. It allows candidates to increase the probability to land in other job markets. To summarize, passing CET4 under retaking policy




(a) Foreign (b) SOE

(c) Large Size (d) Migrate for Jobs

**Figure 3.7**: Share of Jobs by Firm Type

increases the chance to enter SOE, to large firm and geographic mobility. The above results implies that the wage gap at the cutoff may reflect selection into firm types.

## 3.6 Conclusion

What is the role of English skills and test in the labor market? This paper provides evidence on the labor market return to English skills in China and examines the consequence of retaking policy. I exploit a nationwide English test in China and analyze the wage gap at the passing threshold. After controlling for rich characteristics, I find significant positive effects of English skills on wages. The return to English skill is 6% per hundred points measured in CET4 scores globally. When focusing on the retaking behavior at the cutoff for passing CET4. I document massive retaking behavior in response to the presence of threshold due to low cost of retaking. There is a 10% jump in starting salary after graduation for those who barely pass the test. Among students at risk of failing English test, retakers are positively selected in terms of abilities unrelated to English skills. I then analyze the underlying mechanism of such wage jump by exploiting the manipulation at the cutoff. I restrict the sample to multi-time retakers to adopt a regression discontinuity design and find little evidence of sheepskin effect for this subgroup. From the analysis in other job outcomes such as industries, location and firm types, I find that the wage gap is more likely to be associated with access to certain types of firms.

These findings provide deeper understanding than previous work where all the language premium is attributed to skill. The results show that under low-cost retaking policy, the threshold serves as a screening mechanism for certain firms. The mechanism is of great importance because it leads to different policies in improving efficiency. On one hand, if English skills work as merely a screening device for unobserved ability, it is not necessarily the most efficient one and may be revised or replaced. Students who may be talented and suitable for certain jobs should not be constrained by language skills. On the other hand, if the effect comes directly from the improvement in productivity, then government should emphasize on the general access to learning English and leave the decision to individuals through markets. For this study, the former one applies to lows skilled students in CET4 and the latter one applies to high skilled students.

Policymakers should pay more attentions to the actual demand of skills, which may not directly related to English testing. It is important to specialize the exam for the right purpose. Heavy reliance on English tests urges government to reduce asymmetric information in the labor market. Potential research in the future includes non-market benefits from English skills, long-term career impacts and overall social efficiency.

## 3.7 Appendix

**Table 3.7**: College English Test (CET) Proficiency at Each Level

| CET Level | Proficiency |
|---|---|
| 1 | High School Completion |
| 2 | Junior College Entrance |
| 3 | Junior College Completion |
| 4 | College Completion |
| 5 | Grad School Entrance |
| 6 | Grad School Completion |

Note: Only CET4 and CET6 are tested nation wide.

**Table 3.8**: Information on Students' CET4 for Each Survey Year

| | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|
| Number of Attempts | | | ✓ | ✓ | ✓ | ✓ |
| Maximum Score | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Year of Maximum Score | | | | ✓ | ✓ | ✓ |
| Most Recent Score (and Year Taken) | | | ✓ | | | |

# Bibliography


Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6):2127–2189.

Akee, R. (2010). Who Leaves? Deciphering Immigrant Self-Selection from a Developing Country. *Economic Development and Cultural Change*, 58(2):323–344.

Allen, T. and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Anderson, G. and Ge, Y. (2005). The Size Distribution of Chinese Cities. *Regional Science and Urban Economics*, 35(6):756–776.

Andersson, F., Burgess, S., and Lane, J. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61:112–128.

Angrist, J. D., Chin, A., and Godoy, R. (2008). Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap? *Journal of Development Economics*, 85(1-2):105–128.

Angrist, J. D. and Lavy, V. (1997). The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco. *Journal of Labor Economics*, 15(1):S48–S76.

Antràs, P., Garicano, L., and Rossi-Hansberg, E. (2006). Offshoring in a Knowledge Economy. *The Quarterly Journal of Economics*, 121(1):31–77.

Arzaghi, M. and Henderson, J. V. (2008). Networking off Madison Avenue. *The Review of Economic Studies*, 75(4):1011–1038.

Au, C.-C. and Henderson, J. V. (2006a). Are Chinese Cities Too Small? *The Review of Economic Studies*, 73(3):549–576.

Au, C.-C. and Henderson, J. V. (2006b). How Migration Restrictions Limit Agglomeration and Productivity in China. *Journal of development economics*, 80(2):350–388.

Autor, D. (2019). Work of the Past, Work of the Future. *AER: Papers and Proceeding*, 109(5):5.

Azam, M., Chin, A., and Prakash, N. (2013). The Returns to English-Language Skills in India. *Economic Development and Cultural Change*, 61(2):335–367.

Baldwin, R. E. and Okubo, T. (2006). Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting. *Journal of Economic Geography*, 6(3):323–346.

Bauernschuster, S., Falck, O., Heblich, S., Suedekum, J., and Lameli, A. (2014). Why are educated and risk-loving persons more mobile across regions? *Journal of Economic Behavior and Organization*, 98:56–69.

Baum-Snow, N., Freedman, M., and Pavan, R. (2018). Why Has Urban Inequality Increased? *American Economic Journal: Applied Economics*, 10(4):1–42.

Baum-Snow, N. and Pavan, R. (2013). Inequality and City Size. *Review of Economics and Statistics*, 95(5):1535–1548.

Becker, G. S. (1973). A Theory of Marriage: Part I. *The Journal of Political Economy*.

Beegle, K., De Weerdt, J., and Dercon, S. (2011). Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey. *Review of Economics and Statistics*, 93(3):1010–1033.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive Cities: Sorting, Selection, And Agglomeration. *The Journal of Political Economy*, 122(3):507–553.

Behrens, K. and Robert-Nicoud, F. (2014). Survival of the Fittest in Cities: Urbanisation and Inequality. *The Economic Journal*, 124(581):1371–1400.

Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration Theory with Heterogeneous Agents. *Handbook of Regional and Urban Economics*, 5:171–245.

Berman, E., Bound, J., and Griliches, Z. (1994). Changes in the Demand for Skilled Labor within U. S. Manufacturing: Evidence from the Annual Survey of Manufactures. *The Quarterly Journal of Economics*, 109(2):367–397.

Bleakley, H. and Chin, A. (2004). Language Skills and Earnings: Evidence from Childhood Immigrants. *The Review of Economics and Statistics*, 86(2):481–496.

Bloom, N., Eifert, B., Mahajan, A., Mckenzie, D. J., and Roberts, J. (2012). Does Management Matter? Evidence from India. *The Quarterly Journal of Economics*, 128(1):1–51.

Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *The American Economic Review*, 77(4):531–553.

Borjas, G. J. (1989). Economic Theory and International Migration. *The International Migration Review*, 23(3):457–485.

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2012). Relaxing Hukou: Increased labor mobility and China's economic geography. *Journal of Urban Economics*, 73(2-3):252–266.

Brandt, L. and Holz, C. A. (2006). Spatial Price Differences in China: Estimates and Implications. *Economic Development and Cultural Change*, 55(1):43–86.

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, 82(5):1671–1748.

Bryan, G., Glaeser, E., and Tsivanidis, N. (2019). Cities in the Developing World.

Bryan, G. and Morten, M. (2019). The Aggregate Productivity Effects of Internal Migration: Evidence From Indonesia. *The Journal of Political Economy*.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Carliner, G. (1981). Wage Differences by Language Group and the Market for Language Skills in Canada. *Journal of Human Resources*, 16(3):384–399.

Carnevale, A. P., Fry, R. A., and Lowell, B. L. (2001). Understanding, Speaking, Reading, Writing, and Earnings in the Immigrant Labor Market. *The American Economic Review*, 91(2):159–163.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. pages 679–741. Elsevier.

Chade, H., Eeckhout, J., and Smith, L. (2017). Sorting through Search and Matching Models in Economics. *Journal of Economic Literature*, 55(2):493–544.

Chauvin, J. P., Glaeser, E. L., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98:17–49.

Chen, B., Liu, D., and Lu, M. (2018). City size, Migration and Urban Inequality in China. *China Economic Review*, 51:42–58.

Chen, Y., Shi, S., and Tang, Y. (2019). Valuing the Urban Hukou in China: Evidence From a Regression Discontinuity Design for Housing Prices. *Journal of Development Economics*.

Chetty, R. and Hendren, N. (2015). The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates. *Harvard University and NBER*.

Chiquiar, D. and Hanson, G. H. (2005). International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113(2):239–281.

Chirinko, R. S. (1982). An empirical investigation of the returns to job search. *The American Economic Review*.

Chiswick, B. R. (1991). Speaking, Reading, and Earnings among Low-Skilled Immigrants. *Journal of Labor Economics*, 9(2):149–170.

Chiswick, B. R. and Miller, P. W. (1995). The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics*, 13(2):246–288.

Clark, D. and Martorell, P. (2014). The Signaling Value of a High School Diploma. *The Journal of Political Economy*, 122(2):282–318.

Clemens, M. A. (2011). Economics and Emigration: Trillion-Dollar Bills on the Sidewalk? *Journal of Economic Perspectives*, 25(3):83–106.

Clemens, M. A., Montenegro, C. E., and Pritchett, L. (2009). The Place Premium: Wage Differences for Identical Workers Across the U.S. Border. *SSRN Electronic Journal*.

Clemons, M., Montenegro, C., and Pritchett, L. (2008). *The great discrimination: borders as a labor market barrier*. Mimeogr.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial Wage Disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. *Handbook of Regional and Urban Economics*, 5:247–348.

Cornelissen, T., Dustmann, C., and Schönberg, U. (2017). Peer Effects in the Workplace. *The American Economic Review*, 107(2):425–56.

Costinot, A. (2009). On the Origins of Comparative Advantage. *Journal of International Economics*, 77(2):255–264.

Costinot, A. and Vogel, J. (2010). Matching and Inequality in the World Economy. *The Journal of Political Economy*, 118(4):747–786.

Costinot, A. and Vogel, J. (2015). Beyond Ricardo: Assignment Models in International Trade. *Annual Review of Economics*, 7(1):31–62.

Dahl, G. B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, 70(6):2367–2420.

Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2019). Matching in Cities.

Davis, D. and Dingel, J. (2017). The Comparative Advantage of Cities.

Davis, D. R. and Dingel, J. I. (2019). A Spatial Knowledge Economy. *The American Economic Review*, 109(1):153–170.

de la Roca, J. (2017). Selection in Initial and Return Migration: Evidence From Moves Across Spanish Cities. *Journal of Urban Economics*, 100:33–53.

de la Roca, J. and Puga, D. (2017). Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1):106–142.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2018). The Geography of Development. *The Journal of Political Economy*, 126(3):903–983.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *The American Economic Review*, 106(3):479–524.

Diamond, R. and Persson, P. (2016). The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. *NBER Working Paper*.

Dingel, J. I., Miscio, A., and Davis, D. R. (2019). Cities, Lights, and Skills in Developing Economies. *Journal of Urban Economics*.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *The American Economic Review*, 108(4-5):899–934.

Duranton, G. (2007). Urban Evolutions: The Fast, the Slow, and the Still. *American Economic Review*, 97(1):197–221.

Duranton, G. and Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. *Handbook of Regional and Urban Economics*.

Dustmann, C. and van Soest, A. (2002). Language and the Earnings of Immigrants. *Industrial & Labor Relations Review*, 55(3):473–492.

Eckert, F., Ganapati, S., and Conor, W. (2019a). Skilled Tradable Services: The Transformation of US High-Skill Labor Markets.

Eckert, F., Hejlesen, M., and Walsh, C. (2019b). The Return to Big City Experience: Evidence from Danish Refugees.

Eeckhout, J. (2018). Sorting in the Labor Market. *Annual Review of Economics*, 10(1):1–29.

Eeckhout, J. and Jovanovic, B. (2012). Occupational Choice and Development. *Journal of Economic Theory*, 147(2):657–683.

Eeckhout, J. and Kircher, P. (2018). Assortative Matching With Large Firms. *Econometrica*, 86(1):85–132.

Eeckhout, J., Pinheiro, R., and Schmidheiny, K. (2014). Spatial Sorting. *The Journal of Political Economy*, 122(3):554–620.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100(3):1195–1213.

Fan, J. (2019). Internal Geography, Labor Mobility, and the Distributional Impacts of Trade. *American Economic Journal: Macroeconomics*, 11(3):252–88.

Fleisher, B., Li, H., and Zhao, M. Q. (2010). Human Capital, Economic Growth, and Regional Inequality in China. *Journal of development economics*, 92(2):215–231.

Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. *The Journal of Political Economy*, 108(5):874–904.

Garicano, L. and Rossi-Hansberg, E. (2006). Organization and Inequality in a Knowledge Economy. *The Quarterly Journal of Economics*, 121(4):1383–1435.

Garicano, L. and Rossi-Hansberg, E. (2015). Knowledge-Based Hierarchies: Using Organizations to Understand the Economy. *Annual Review of Economics*, 7(1):1–30.

Gaubert, C. (2018). Firm Sorting and Agglomeration. *The American Economic Review*, 108(11):3117–53.

Glaeser, E., Huang, W., Ma, Y., and Shleifer, A. (2017). A Real Estate Boom with Chinese Characteristics. *Journal of Economic Perspectives*, 31(1):93–116.

Glaeser, E. L., Kolko, J., and Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, 1(1):27–50.

Glaeser, E. L. and Lu, M. (2018). Human-Capital Externalities in China. *NBER Working Paper*.

Glaeser, E. L. and Maré, D. C. (2015). Cities and Skills. *Journal of Labor Economics*, 19(2):316–342.

Glaeser, E. L., Resseger, M., and Tobio, K. (2009). Inequality in Cities. *Journal of Regional Science*, 49(4):617–646.

Goodman, J., Gurantz, O., and Smith, J. (2020). Take Two! SAT Retaking and College Enrollment Gaps. *American Economic Journal: Economic Policy*.

Gottlieb, P. D. and Joseph, G. (2006). College-to-Work Migration of Technology Graduates and Holders of Doctorates Within the United States. *Journal of Regional Science*, 46(4):627–659.

Gould, E. D. (2007). Cities, Workers, And Wages: A Structural Analysis of the Urban Wage Premium. *The Review of Economic Studies*, 74:477–506.

Greenwood, M. J. (1997). Chapter 12 Internal migration in developed countries. In *Handbook of Population and Family Economics Volume 1*, pages 647–720. Elsevier.

Grenier, G. (1984). The Effects of Language Characteristics on the Wages of Hispanic-American Males. *Journal of Human Resources*, 19(1):35–52.

Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of development economics*, 95(1):42–57.

Grogger, J. and Hanson, G. H. (2015). Attracting Talent: Location Choices of Foreign-Born PhDs in the United States. *Journal of Labor Economics*, 33(S1):S5–S38.

Grossman, G. M., Helpman, E., and Kircher, P. (2016). Matching, Sorting, and the Distributional Effects of International Trade. *The Journal of Political Economy*, 125(1):224–264.

Hendricks, L. and Schoellman, T. (2016). Human Capital and Development Accounting: New Evidence from Wage Gains at Migration.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The Allocation of Talent and U.S. Economic Growth. *Econometrica*, 87:1439–1474.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Hsieh, C.-T. and Moretti, E. (2019). Housing Constraints and Spatial Misallocation. *American Economic Journal: Macroeconomics*, 11(2):1–39.

Iranzo, S., Schivardi, F., and , E. T. (2008). Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data. *Journal of Labor Economics*, 26:247–285.

Jarosch, G., Oberfield, E., and Rossi-Hansberg, E. (2019). Learning from Coworkers. *NBER Working Paper*.

Jia, R. and Li, H. (2017). Access to Elite Education, Wage Premium, and Social Mobility: Evidence from China's College Entrance Exam.

Jiao, Y. and Tian, L. (2019). Geographic Fragmentation in a Knowledge Economy.

Jovanovic, B. (1994). Firm Formation with Heterogeneous Management and Labor Skills. *Small Business Economics*, 6(3):185–191.

Kazakis, P. and Faggian, A. (2016). Mobility, education and labor market outcomes for U.S. graduates: Is selectivity important? *The Annals of Regional Science*, pages 1–28.

Keith, K. and McWilliams, A. (1999). The Returns to Mobility and Job Search by Gender. *Industrial & Labor Relations Review*, 52(3):460–477.

Kremer, M. (1993). The O-Ring Theory of Economic Development. *The Quarterly Journal of Economics*, 108(3):551–575.

Lagakos, D., Mobarak, A. M., and Waugh, M. (2018). The Welfare Effects of Encouraging Rural-Urban Migration.

Lagakos, D. and Waugh, M. E. (2013). *Selection, agriculture, and cross-country productivity differences*.

Li, H., Meng, L., Shi, X., and Wu, B. (2012). Does attending elite colleges pay in China? *Journal of comparative economics*, 40(1):78–88.

Liang, W. and Lu, M. (2017). Growth Led by Human Capital in Big Cities: Exploring Complementarities and Spatial Agglomeration of the Workforce with Various Skills. *China Economic Review*, page 101113.

Lucas, R. E. (1978). On the Size Distribution of Business Firms. *The Bell Journal of Economics*, 9(2):508.

Ma, L. and Tang, Y. (2019). Geography, Trade, and Internal Migration in China. *Journal of Urban Economics*.

Marshall, A. (1890). Principles of economics. Vol. 1.

Mas, A. and Moretti, E. (2009). Peers at Work. *The American Economic Review*, 99(1):112–45.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57:995–1026.

Mckenzie, D., Gibson, J., and Stillman, S. (2010). How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration. *Journal of the European Economic Association*, 8(4):913–945.

McManus, W. S., Gould, W., and Welch, F. R. (1983). Earnings of Hispanic Men: The Role of English Language Proficiency. *Journal of Labor Economics*, 1(2):101–130.

Mion, G. and Naticchioni, P. (2009). The Spatial Sorting and Matching of Skills and Firms. *The Canadian Journal of Economics*, 42(1):28–55.

Molloy, R., Smith, C. L., and Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, 25(3):173–196.

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *The American Economic Review*, 94(3):656–690.

Moretti, E. (2013). Real Wage Inequality. *American Economic Journal: Applied Economics*, 5(1):65–103.

Mueller, H. M., Ouimet, P. P., and Simintzi, E. (2017). Wage Inequality and Firm Growth. *The American Economic Review*, 107(5):379–83.

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics*, 118(2):549–599.

Neffke, F. (2017). Coworker Complementarity. *Working Papers*.

Ngai, L. R., Pissarides, C. A., and Wang, J. (2019). China's Mobility Barriers and Employment Allocations. *Journal of the European Economic Association*, 17(5).

Oster, E. (2017). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Rauch, J. E. (1993). Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics*, 34(3):380–400.

Ravenstein, E. G. (1885). The Laws of Migration. *Journal of the Statistical Society of London*, 48(2):167.

Restuccia, D. and Rogerson, R. (2017). The Causes and Costs of Misallocation. *Journal of Economic Perspectives*, 31(3):151–174.

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and Aggregate Productivity: a Quantitative Cross-Country Analysis. *Journal of Monetary Economics*, 55(2):234–250.

Rice, P., Venables, A. J., and Patacchini, E. (2006). Spatial Determinants of Productivity: Analysis for the Regions of Great Britain. *Regional Science and Urban Economics*, 36(6):727–752.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. *Handbook of Regional and Urban Economics*, 4:2119–2171.

Rosenzweig, M. (2007). Education and migration: A global perspective. *Yale University*.

Rossi-Hansberg, E., Sarte, P.-D., and Owens, R. (2009). Firm Fragmentation and Urban Patterns. *International Economic Review*.

Saiz, A. and Zoido, E. (2005). Listening to What the World Says: Bilingualism and Earnings in the United States. *The Review of Economics and Statistics*, 87(3):523–538.

Sampson, T. (2014). Selection into Trade and Wage Inequality. *American Economic Journal: Microeconomics*, 6(3):157–202.

Santamaría, C. (2019). Small Teams in Big Cities: Inequality, City Size, and the Organization of Production.

Sattinger, M. (1975). Comparative Advantage and the Distributions of Earnings and Abilities. *Econometrica*.

Serafinelli, M. (2019). “Good” Firms, Worker Flows, and Local Productivity. *Journal of Labor Economics*, 37(3):747–792.

Shields, M. A. and Price, S. W. (2002). The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas. *Journal of Population Economics*, 15(1):137–160.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70(5):80–93.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and von Wachter, T. (2018). Firming Up Inequality. *The Quarterly Journal of Economics*, 134(1):1–50.

Spanos, G. (2019). Firm Organization and Productivity Across Locations. *Journal of Urban Economics*.

Tainer, E. (1988). English Language Proficiency and the Determination of Earnings among Foreign-Born Men. *Journal of Human Resources*, 23(1):108–122.

Tian, L. (2019). Division of Labor and Extent of the Market: Theory and Evidence from Brazil.

Tombe, T. and Zhu, X. (2019). Trade, Migration, and Productivity: A Quantitative Analysis of China. *The American Economic Review*, 109(5):1843–72.

Topkis, D. M. (1978). Minimizing a Submodular Function on a Lattice. *Operations Research*, 26(2):305–321.

United Nations (2018). World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). Technical report, New York: United Nations, Department of Economic and Social Affairs, Population Division.

Urquiola, M. and Verhoogen, E. (2009). Class-Size Caps, Sorting, and the Regression-Discontinuity Design. *The American Economic Review*, 99(1):179–215.

Vigdor, J. L. and Clotfelter, C. T. (2003). Retaking the SAT. *Journal of Human Resources*, XXXVIII(1):1–33.

Vollrath, D. (2014). The Efficiency of Human Capital Allocations in Developing Countries. *Journal of Development Economics*, 108:106–118.

Walmsley, T. L. and Winters, L. A. (2005). Relaxing the Restrictions on the Temporary Movement of Natural Persons: A Simulation Analysis. *Journal of Economic Integration*, 20(4):688–726.

Xie, Y. and Zhou, X. (2014). Income inequality in today's china. *Proceedings of the National Academy of Sciences*, 111(19):6928–6933.

Yankow, J. J. (2003). Migration, Job Change, and Wage Growth: A New Perspective on the Pecuniary Return to Geographic Mobility. *Journal of Regional Science*, 43(3):483–516.

Young, A. (2013). Inequality, the Urban-Rural Gap, and Migration. *The Quarterly Journal of Economics*, 128(4):1727–1785.

Zheng, S. and Kahn, M. E. (2008). Land and Residential Property Markets in a Booming Economy: New Evidence from Beijing. *Journal of Urban Economics*, 63(2):743–757.