A Tale of Two Tails: Wage Inequality and City Size *

Lin Ma
National University of Singapore

Yang Tang
Nanyang Technological University

January 8, 2019

Abstract

We document how within-city inequality vary across cities. While the residual wage gaps between the top and the median earners increase with the city size, the gaps between the median and the bottom earners shrink. We develop a spatial equilibrium model where heterogeneous individuals sort into entrepreneurs or workers in different industries and cities. Entrepreneurs benefit more than workers in larger cities due to knowledge spillovers, leading to higher top inequality. Bottom inequality shrinks in larger cities because low-income workers must be compensated to overcome the higher living costs. Empirical tests using U.S. data broadly support our theoretical predictions.

Keywords: wage inequality; sorting; wage distribution; city size; inter-industry wage premium.

JEL Classification: F12; J24; J31; R10; R23

*Email address: ecsml@nus.edu.sg, tangyang@ntu.edu.sg, respectively. The authors are grateful for stimulating discussions with Costas Arkolakis, Costas Azariadis, Davin Chor, Andrei Levecheko, Wen-Tai Hsu, and Ping Wang. We also greatly benefit from valuable suggestions and comments of various seminar and conferences at National University of Singapore, Nanyang Technological University, Singapore Management University, the Public Economic Theory Meeting 2015, the Society of Economics Theory Meeting 2015, Taipei International Conference on Growth, Dynamics, and Trade 2016, and the Asian Meeting of Econometric Society 2016. We are solely responsible for the remaining errors.

1 Introduction

Empirical studies often suggest that larger cities are unequal than smaller ones. Baum-Snow and Pavan (2013) document a positive relationship between city size and wage inequality in the U.S. during 1979-2007. Similarly, Behrens and Robert-Nicoud (2015) also estimate a positive city size elasticity of the income Gini coefficient at around 0.011.[1] However, inequality is a multifaceted concept; often, inequality measured at different parts of the distribution could behave differently. For example, the macroeconomic literature has shown that at the national level, the inequality measured at the right tail, such as the 90-50 ratios, have been steadily increasing over time, but measures took at the left tail, such as the 50-10 ratios, have been stable or even slightly declining since the 1980s.[2] In this paper, we document and rationalize the same complexity in the context of within-city inequality: larger cities are not always more unequal, and the exact answer depends on which part of the distribution we are examining.

We document a U-shape relationship between residual wage and city size in the United States: the residual wage rates at the top and the bottom percentiles increase faster with city size than those in the middle. Figure 1 summarizes the result: it plots the elasticity of residual wage rates with respect to city size at different percentiles in the U.S.[3] While the residual wage rates across the distribution are higher in larger cities, they increase with city size at different speeds. Particularly, the wage around the median increase at slower rates as compared to both the left and the right tail. For example, the city size elasticity around the median is 25 percent lower than that at the 5th percentile, and 33 percent lower than that at the 95th percentile.[4] This implies that while the right-tail inequality (95-50 gap) is higher in larger cities, the left-tail inequality (50-5 gap) is lower. We confirm that it is indeed the

[1]For more empirical literature on this issue, see Glaeser et al. (2009), Berube (2014), and Berube and Holmes (2015).

[2]For details, see Heathcote et al. (2010) for time trends of various measures of wage, earnings, and income inequality in the US. Piketty and Saez (2003) shows that even within the measures taken at the right tail, income shares at different percentiles exhibit different trends over time.

[3]Data source is Public Use Microdata Series (PUMS) in year 2000. City is defined as MSA. The residual wage rate is the observed wage rate with individual characteristics such as age, years of education, race, and marital status filtered out. See details in Section 2.

[4]The city size elasticity is 0.04, 0.03, and 0.045 at the 5th, 50th, and the 95th percentile as shown in Figure 1, respectively. The difference between the left (right) tail and the median is $1 - 0.03/0.04 = 0.25$ $(1 - 0.03/0.045 = 0.33)$.

case: while the size elasticity of the 95-50 gap is positive at around 0.026, the size elasticity of the 50-05 gap stands significantly negative at around -0.012. This contrasting pattern in city size inequality is also robust: it can be observed for a wide rage of percentile-ratios at both the left and the right tail; it is robust to different ways of estimating the residual wage as well as the unfiltered raw wage, and it is also robust across different measures of city size.

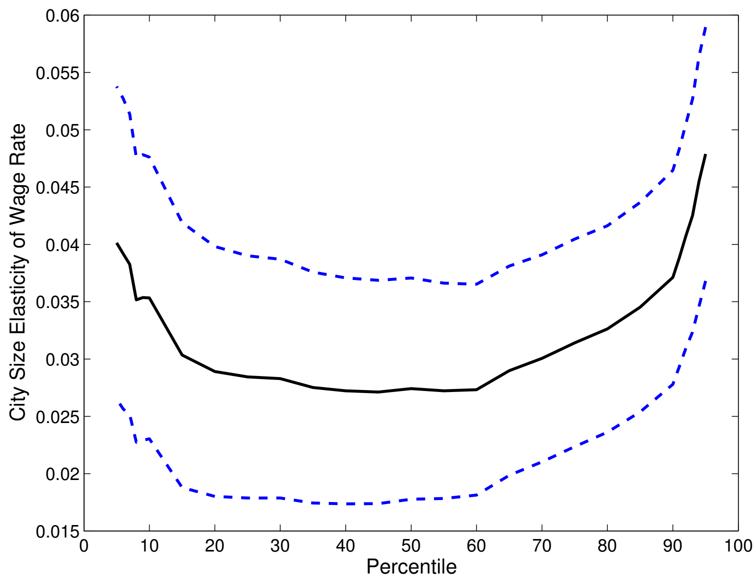


Figure 1: City Size Elasticity of Residual Wage at Different Percentiles

Note: This figure reports the city size elasticity of wage rate (α_1) at different percentiles from Equation (2). The dashed line is 95 percent confidence interval. See Section 2 for details. Data source: IPUMS-USA, 2000.

Two channels can potentially explain the observed pattern of within-city inequality: the variations in 1) inter-industry wage premium and 2) the entrepreneurship premium across cities. Firms within the same industries usually pay higher wage in larger cities in our data. However, we find that the wage rate in low-paying industries grows faster with city size as compared to high-paying industries, leading to smaller wage gaps between high- and low-paying industries in large cities. To the extent that workers are likely to occupy the middle and the lower segments of the income distribution, the spatial variations in inter-industry wage premium can potentially explain the compression of the left-tail inequality. On the other find, we document that the return to entrepreneurship raises with city size, leading to larger gaps between entrepreneurs and workers in larger cities. As entrepreneurs are also likely to be the top earners in each city, the spatial variations of entrepreneurship premium can potentially explain why the right-tail inequality is higher in larger cities.

To validate and quantify the relationship between inter-industry wage premium, entrepreneurship premium, and within-city inequality, we carry out two "counter-factual" econometric exercises. In the first exercise, we control for industry-city or occupation-city fixed effects when computing the residual wage to eliminate the spatial variations along these two dimensions. Once the industry-city fixed effects are controlled for, the pattern for left-tail inequality completely disappears, as larger cities no longer see lower inequality in the left-tail. Once we control for the occupation-city fixed effects, the positive relationship between city size and right-tail inequality drops by around 27 percent. In the second "counter-factual" exercise, we reconstruct the individual wage in data by eliminating the variance of inter-industry and entrepreneurship premium across cities. Once we remove the variance of the inter-industry wage premium across cities, the negative relationship between city size and inequality measured at the left tail weakens by around 50 percent. Similarly, once we eliminate the spatial variations of entrepreneurship, the city size elasticity at the right tail shrank by between 19 to 29 percent respectively.

In light of the empirical findings, we propose a simple spatial model in which the three pieces of empirical pattern emerge in equilibrium. In the model world, individuals with different levels of human capital choose a city to live in, and an industry and occupation cell to work in to maximize utility. We assume that industries and occupations vary in the entry barriers, and cities also differ in the congestion disutility that rises with an endogenously determined population. In equilibrium, location-industry-occupation choices with higher entry barriers might also offer higher return to human capital to attract people. As a result, assortative matching arises such that individuals with higher human capital sort into location-industry-occupation cells with higher entry barriers and higher returns. Across cities, assortative matching means that individuals with higher human capital reside in larger cities. Across industries, talented individuals join industries with higher entry barriers. Within industries, individuals can choose occupations: they can either create new firms by paying an extra fixed cost, or work for the existing firms. The additional barrier to entrepreneurship ensures that the most talented individuals choose to create new firms. In equilibrium, wage rates are endogenously determined in each city-industry-occupation cell, and the model delivers the three empirical patterns documented above in equilibrium.

Inequality measured at the right tail is higher in larger cities in our model. This is because the top earners in each city, the entrepreneurs, benefit more from city size as compared to the rest of the population. The spatial variation of the entrepreneurship premium depends on two mechanisms: 1) the income of the entrepreneurs positively depends on the size of the firm they manage, and 2) the average firm size increases faster with respect to city size as compared to the wage rates of workers. The first mechanism is our assumption based on both the empirical and the theoretical findings in the corporate finance literature.[5] The second mechanism is an equilibrium result derived from the assortative matching described before. Large cities are populated by large and productive firms which push up the factor prices and push down output prices in equilibrium, making them tougher for smaller firms to operate in. To retain the marginal entrepreneur in the large city, the factor market must compensate him not only the differences in congestion disutilities but also the differences in profit induced by market conditions between large and small cities. In contrast, the market in large cities only needs to compensate the marginal workers the increments in congestion disutility. As a result, the wage rate will increase with city size at a slower pace as opposed to firm size and entrepreneurial compensation. This mechanism is also supported by the empirical literature, which often found that the city size elasticity of firm size is higher than that of the wage rate.[6]

Meanwhile, the inequality measured at the left tail is smaller in large cities. Within the group of workers, the wage rates in all industries are higher in larger cities to compensate higher congestion costs.[7] However, the wage rate of lower-paying industries increase faster

[5]The positive relationship between entrepreneurial income and firm size has been extensively documented in the executive compensation literature since Roberts (1956). It is also the equilibrium compensation scheme found in a wide array of models of executive pay, such as in Gabaix and Landier (2008). We abstract away from the details of an executive compensation model, directly assume the Roberts' Law, and focus on its impact on inequality.

[6]For example, the city-size elasticity of firm employment is found to be between 0.5 and 0.7, such as in Glaeser and Kerr (2009) and Glaeser (2007). In contrast, the city-size elasticity of wage or earnings is usually between 0.05 and 0.1. See Roback (1982), Combes et al. (2008), Tabuchi and Yoshida (2000), Glaeser and Mare (2001) and Baum-Snow and Pavan (2012) for more details.

[7]One can interpret the higher congestion disutility as higher price levels in larger cities. Handbury and Weinstein (2015) documents that food price is lower in larger cities; however other price, especially housing price, is significantly higher in larger cities as documented in Moretti (2013). Housing costs are also responsible for a large fraction of household expenditure, and thus in the reality the aggregate price level in larger cities might still be higher than in smaller ones.

with city size in our model, the same as documented in the data; otherwise, the workers in these industries, who have fewer units of efficiency labor supply, will not have enough income to overcome the high living costs in large cities. Instead, they will either migrate to smaller cities or invest in education to join a higher-paying industry. However, every city needs the outputs from the low-skill, low-paying industries — janitors, cashiers, or street vendors — to function properly, which implies that the market must compensate those working in such industries relatively more in the equilibrium.

In addition to the theoretical results on inequality and industry-occupation premium, we also provide conditions under which a unique sorting equilibrium emerges in our model. Similar to many models in the economic geography literature, assortative matching in itself cannot pin down a unique equilibrium in our case. We prove that under a family of distributions of city size, which include power law and exponential distributions often documented in the empirical literature, the sorting patterns can be exactly characterized in our model. The key to our proof is the recurrence relations of order statistics in distributions well-studied in the statistics literature.[8] We show that the recurrence property can be used to reduce the number of potential sorting patterns in our model, and we believe our method and findings can be useful to other researchers in related fields.

Our paper is related to several strands of existing literature. First and foremost, we contribute to the empirical literature on within-city inequality. Baum-Snow and Pavan (2013) document that wage inequality is higher in larger cities for both the 90-50 and the 50-10 percentile. Wheeler (2004, 2005) show that the return to skills and wage inequality both increase with city size. Glaeser et al. (2009) and Behrens and Robert-Nicoud (2015) also document that the Gini coefficient is higher in larger cities. We highlight that the relationship between city size and within-city inequality is sensitive to the empirical specification and sample. Different from Baum-Snow and Pavan (2013), we control for city-level characteristics such as the average years of education, the racial composition, and the state in which the MSA is located in our estimation. We show that while the right-tail inequality is not sensitive to these control variables, the left-tail inequality is indeed sensitive. The samples in Baum-Snow and Pavan (2013) and Wheeler (2004, 2005) are all restricted to the white males,

[8]See Balakrishnan and Sultan (1998) for a survey of this topic.

and Wheeler (2005) only focuses on a selection of large cities. Our sample includes all male working population regardless of race in all MSAs with available data. Similar to our finding, Combes et al. (2012b) also find left-tail compression and right-tail dilution of within-city wage rate distribution in the French data.

On the theoretical side of the literature, Behrens et al. (2014), Behrens and Robert-Nicoud (2014), Davis and Dingel (2012), and Eeckhout et al. (2014) provide models in which within-city inequality vary across cities, and all the models predict higher inequality in larger cities. Our model instead allows for a flexible pattern of within-city inequality. Behrens et al. (forthcoming) provide a similar framework that embodies a Melitz-type trade model and entrepreneurship. While they focus on how market size affects entrepreneur entry, exit, and inequality, we focus on how market size affects inequality through industry and occupation premiums.

Our work is also related to the literature on the spatial variations of skill premium. Bacolod et al. (2009) find that large cities are more skilled than small cities in measures of skill other than education. Roback (1982), Combes et al. (2008), Davis and Dingel (2012) and Baum-Snow and Pavan (2012) provide the theoretical foundation and empirical support on the spatial sorting of skills and the city-size premium of skills. Hendricks (2011) show that larger cities tend to attract more skilled workers, and De la Roca and Puga (2017) show that the earnings premium in large cities is mainly due to learning effects. We extend this line of work by showing that the return to different skills increases with city size at different speeds, which in turn leads to rich patterns of inequality. A long tradition in the urban literature following Henderson (1974) emphasizes the role of specialization of industries across cities. Complementing to this literature, we show that the sorting of workers within the same industry can also explain a sizable proportion of the observed spatial variation of skill-premiums. Related to the skill premium, another strand of the literature, motivated by the works of Rosen (1987), Krueger and Summers (1988), and Katz et al. (1989), studies the determination of inter-industry wage premium. Our work is the first to explore the spatial dimension and provides a theoretical foundation on which inter-industry wage premium can vary systematically across cities.

Our paper is broadly related to the new economic geography literature, such as Combes et al. (2008), Gaubert (2017) and Tombe and Zhu (2015). We contribute to this literature by introducing a tractable framework with multi-dimensional sorting. We show that under reasonable assumptions, the sorting pattern and the equilibrium outcome are both unique, and thus the model can be used to study the distributional impacts of agglomeration, migration, and inter-city trade quantitatively in future works.

The rest of the paper is organized as follows. Section 2 documents the spatial variations of within-city inequality, inter-industry wage premium, and entrepreneurship premium. Section 3 presents the model, and Section 4 discusses the analytical results. Section 5 concludes.

2 Empirical Results

In this section, we first document that top and bottom wage inequality within a city moves in opposite directions with respect to city size. We then proceed to present two pieces of supporting evidence outlining the potential mechanisms: the spatial variations of inter-industry wage premium and entrepreneurial wage premium, and its relationship to within-city inequality. In the next section, we proceed to build the model that features these mechanisms and use them to explain the observed pattern of within-city inequality. In this sense, the supporting empirical evidence can also be interpreted as tests on our model mechanisms.

2.1 Data

All the empirical evidence is based on the individual level data from the Integrated Public Use Microdata Series (IPUMS) compiled by the University of Minnesota (Ruggles et al. (2010)). We use the 5 percent sample in the year 2000. We impose a few conventional sample restrictions. We drop the individuals 1) not in the labor force or unemployed, 2) working in government, military, religious and other non-profit entities, 3) the seasonal workers who work less than 10 weeks in the last year, and 4) those whose wage is lower than the federal minimum wage. Following Baum-Snow and Pavan (2013), we restrict our analysis to males, so the results are not compounded by gender wage premium. We interpret the Metropolitan Statistical Area (MSA) in which the individual works as the "city." We measure "raw wage" as the ratio between total wage income and usual hours worked.

Our final sample contains around 1.23 million individuals working in 254 MSAs.[9] The median "city" in our sample hosts 12,638 individuals, and the smallest "city" contains 445 individuals. More than 75 percent of our cities include at least 4,000 individuals. The large sample size in each city allows us to compute the percentile ratios within each city. We provide a detailed description of the data set and the variables constructed in Appendix B.

We use three measures of city size: 1) the private regional industry GDP, and 2) the total regional GDP, and 3) population. The first two measures are obtained from the Bureau of Economic Analysis (BEA) in 2001, and the last measure comes from the census in 2000. The results from all three measures only differ marginally, and therefore we only present and discuss the results based on private industry GDP in the main text. We report the robustness checks using the other two measures in the appendix.

2.2 City Size and Within-City Inequality

Consistent with our model presented later, we study residual wage inequality after controlling for individual characteristics with the following Mincer regression:

$$\ln(W_i) = \beta_0 + \beta_1 \cdot \mathbf{X_i} + \epsilon_i, \tag{1}$$

where W_i is the observed wage and $\mathbf{X_i}$ includes the years of education, age, marital status, and race of individual i. We use the residuals from the regression, $\widehat{\epsilon}_i$, to compute all the measures of inequality in the baseline results.

We first measure how income at different percentiles vary across cities with the following equation:

$$\hat{\epsilon}_q^j = \alpha_0 + \alpha_1 \ln Y^j + \mathbf{a} \cdot \mathbf{Z^j} + u^j, \tag{2}$$

where j indexes the city, Y^j is the measure of city size, and $\mathbf{Z}^j$ is a vector that controls for

[9]The number of MSAs in the sample increases to 264 if we use population as the measure of city size. The difference exists because the BEA only reports GDP at the more aggregated CMSA level, while the population data from the census are more dis-aggregated at the MSA level.

city-level characteristics, such as the average years of education, the share of the white population, and state dummy variables.[10] The left-hand-side variable, $\hat{\epsilon}_p^j$, is the $p = 5, 6, \cdots 95$-th percentile of residual wage in city j. The coefficient of interest is α_1 plotted in Figure 1 in the introduction. Residual wage at all percentiles increase with city size, but at different speeds. A U-shape emerges: the wage at the top and the bottom percentiles increases faster with city size than those in the middle. The graph suggests that the top inequalities such as the 95-50 and 90-50 gaps shall be higher, and the bottom inequalities such as the 50-10 and 50-5 gaps shall be lower in larger cities. At the same time, inequality measure such as the 75-to-25 or 90-to-10 ratios shall not vary across cities.

To confirm the pattern of inequality suggested in Figure 1, we first define tail inequality between the $q > p$-th percentile of residual wage within city j as:[11]

$$\text{Ineq}_{q,p}^j = \ln\left(\hat{\epsilon}_q^j - \hat{\epsilon}_p^j\right).$$

We then study how these measures of inequality vary with the city size with the following equation:

$$\text{Ineq}_{q,p}^j = \beta_0 + \beta_1 \ln Y^j + \mathbf{b} \cdot \mathbf{Z^j} + u^j, \tag{3}$$

where Y^j and $\mathbf{Z^j}$ are the same as defined in (2).

Table 1 reports the findings. The first two columns of the table report inequality measured at the left tail such as the 50-05 and 50-10 residual wage gaps and the last two columns report the right tail with the 95-50 and 90-50 gaps.[12] All the estimates of β_1 at the right tail are significantly greater than zero. This is consistent with the literature: the size of the city is positively correlated with the wage gap at the top of the distribution. On average a one percent increase in city size is associated with 0.017-0.026 percent *increase* in the 95-50 or

[10] Some MSAs span over multiple states. We assign the MSA to the state where most of its population lives in these cases.

[11] A common practice in the literature is to use ratios between different percentiles of observed wage as measures of inequality. However, as residual wages contain negative values, taking the ratios is not informative. Therefore we take the differences between the percentiles to ensure that the measures are always positive, which in turn allows us to take logarithms and compute elasticities.

[12] We avoid using extreme percentiles at both ends such as the 1st or the 99th percentile as they are more likely to suffer from measurement errors, top or bottom coding, and other issues.

the 90-50 gap. The opposite pattern emerges at the left tail: a one percent increase in the city size is associated with a 0.011-0.012 percent *decrease* in the 50-05 and 50-10 wage gaps. Table A.5 in the appendix confirms that inequality measured toward the middle of the distribution such as 90-10 and 75-25 gaps do not vary across city size, as suggested by Figure 1.

Table 1: Left- and Right-tail inequality by City Size

LHS = Residual Wage Ineq.	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Private Ind. GDP)	-0.012***	-0.011**	0.026***	0.017***
	(0.005)	(0.005)	(0.005)	(0.004)
Average Years of Edu.	0.052***	0.050***	-0.002	-0.010
	(0.012)	(0.012)	(0.016)	(0.012)
Race == White	-0.424***	-0.408***	-0.277**	-0.231**
	(0.098)	(0.103)	(0.128)	(0.099)
N	254	254	254	254
R-squared	0.520	0.571	0.566	0.597
State FE	Yes	Yes	Yes	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the results of estimating equation (3), while controlling for the average years of education, share of white population, and state dummies. Data for income inequality come from 5 percent sample in year 2000 compiled by IPUMS. Data for GDP come from BEA regional GDP estimates.

Baum-Snow and Pavan (2013) reports that the 50-to-10 wage ratios are higher in larger cities (see Table 1 and Figure 2 in their paper). We find different results because we control for city-level characteristics, while Baum-Snow and Pavan (2013) do not.[13] We highlight the differences in Table 2 by first estimating equation (3) without any city-level control, and then progressively add back each control variable.

The first column of the table roughly replicates the findings in Baum-Snow and Pavan (2013) with our sample: inequality is indeed higher in larger cities at the left tail without any control at the city level.[14] The rest of Table 2 shows that by adding back our control

[13] The individual level regressions and counter-factual experiments in Baum-Snow and Pavan (2013) (Table 2 to 5 in their paper), in which they control for education and age of individuals, are not comparable with our estimations. Those tables study how 50-to-10 wage ratio at *national level* responds to the changes in city size, whereas we study how 50-to-10 wage ratio *within each city* changes with city size.

[14] This also suggests that whether to include all male working population or restrict to only white males, which is the main difference between our sample and sample in Baum-Snow and Pavan (2013), is not the cause of our differences.

Table 2: City Size and Wage Inequality, Effects of Education and Racial Composition

	LHS = 50-10 Wage Gap				
	(1)	(2)	(3)	(4)	(5)
Ln(Private Ind. GDP)	0.010**	0.006	0.000	0.001	-0.011**
	(0.005)	(0.006)	(0.007)	(0.005)	(0.005)
Average Years of Edu.			0.031***		0.050***
			(0.010)		(0.012)
Share of White Population				-0.299***	-0.408***
				(0.113)	(0.103)
N	254	254	254	254	254
R-squared	0.013	0.490	0.508	0.526	0.571
State FE	No	Yes	Yes	Yes	Yes

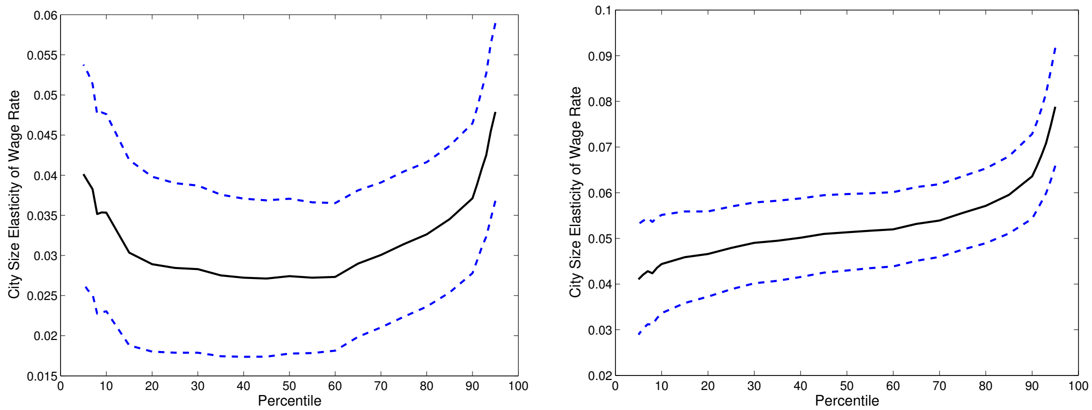
Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the results of estimating Equation (3) by progressively adding more city-level controls. Data source: IPUMS-USA, 2000.

variables, the size elasticity gradually decreases from 0.010 to -0.011. The control variables at the city level indeed affect within-city inequality in addition to city size: cities with more well-educated population tend to be more unequal at the left tail; the share of population that reports as "white" seems to reduce inequality at both tails. The state in which the city is located also matters, probably because minimum wage requirement varies from state to state.[15] As these variables probably affect inequality through channels such as institutional quality and mostly are correlated with city size as well, controlling for them will help us to single out the relationship between city size and inequality.

We also repeat the estimation of Equation 2 without any city-level controls and report the results in the right panel of Figure 2. Without the control variables wage level at all percentiles increases with city size, but those at the right tail see higher elasticity than those to the left, and thus inequality measured across the entire distribution shall increase with city size, a finding that resonates with Baum-Snow and Pavan (2013).

Autor and Dorn (2013) document the polarization of U.S. wage rates and argue that automation in the manufacturing industry is the driver behind it. We show that similar polarization can be observed in the size elasticity of wages, which suggests that the spatial

[15]We do not need to control for minimum wage at the city level since minimum wage within a state did not vary in 2000. The first city-level minimum wage was implemented in 2004 (Santa Fe, NM and San Francisco, CA) as documented in Schmitt and Rosnick (2011) and Vaghul and Zipperer (2016).



(a) With City Level Control

(b) No City-Level Control

Figure 2: City Size Elasticity of Wage at Different Percentiles

Note: This figure reports the city size elasticity of wage rate (β_1) from Equation 2. The dashed line is 95 percent confidence interval. The left panel the a replication of Figure 1 for comparison. Data source: IPUMS-USA, 2000.

agglomeration might have also contributed to the polarization of wage rates. Individuals at the top and the bottom of the wage distribution benefit more from living in large cities than the "middle-class." As cities grow, agglomeration can directly contribute to the polarization of wage rates in addition to the breakthroughs in the technology frontier.

We focus on percentile gaps as measures of inequality for a couple of reasons. Unlike scalar measures such as the Gini or the Theil index, percentile gaps provide a complete and flexible measure across the entire distribution. Higher 90-50 or 50-10 wage gaps can both lead to higher Gini coefficient, but the economic mechanism and policy implications can be different depending on whether higher inequality is driven by the movements at the left or the right tail. In the debate on whether urban inequality in itself is a problem to be concerned about, many authors such as Glaeser et al. (2008) pointed out that inequality is not necessarily undesirable. This is because the congregation of entrepreneurs and highly-skilled workers in large cities, which inevitably leads to higher inequality, is the precisely the advantage of large cities. At the same time, many voices in the public debate that label urban inequality as a "crisis" in American cities often attribute the undesirable outcomes such as higher crime rates and lower educational attainments to urban inequality as well. The debate on urban inequality would benefit from a clarification on what do we mean by "unequal." The benefits of urban inequality based on the agglomeration mechanisms are often linked to the right-tail inequality while many problems of inequality that are often inter-mingled with the problem of poverty, shall be analyzed with a focus on the left-tail.

Our message that large cities are only more unequal on the right-tail has a clear policy implication. Right-tail inequality in itself probably reflects the strength of large cities, and it is not a problem that the policymakers shall try to "fix". Expelling Google and Facebook from San Francisco might lower the right-tail inequality, but it is probably not a desirable policy for the majority of workers in the city. We also show that larger cities are more equal on the left-tail. In addition to complementing the existing literature on inequality which mainly focuses on right-tail measures, left-tail inequality in itself has distinct policy implications. The pattern on the left-tail implies that the income of the bottom earners drop fast as they move down the ladder of city size, and thus those in smaller cities are more likely to fall below the poverty line. This echoes the findings in Jargowsky (2015), which documents that concentrated poverty, a measure closely related to urban crime and social mobility, grows most in smaller cities such as Syracuse (NY) and Dayton (OH) in recent years. In the large cities such as New York City or Los Angeles, concentrated poverty has been declining. Our finding implies that while the ghettos in NYC or LA receive the most attention in the media, the policymakers at both the federal and the state level shall focus more on the poor neighborhoods in small cities. Not only the problems of poverty and left-tail inequality are more serious in smaller cities, the cities themselves probably lack the resource to solve the problems on their own.

Robustness Checks All the figures and tables for robustness checks are included in the appendix. Figure A.1 and Table A.1 report the robustness checks with two other measures of city size, the total regional GDP and population. Table A.2 adds in industry and occupation fixed effects in the Mincer equation (eq. 1) when computing the residual wage rates. Table A.3 reports the results using unfiltered raw wage. All the results in the robustness checks are qualitatively the same as those presented in the main text. Baum-Snow and Pavan (2013) measure city size using population, and we use private GDP in the comparison exercise reported in Table 2. Table A.4 repeats the same exercise with population as the measure of city size and the patterns remain the same as well.

2.3 Inter-Industry Wage Premium and Entrepreneurship Premium

In the rest of the section, we explore two potential channels that can potentially explain the patterns of within-city inequality: industry and occupational wage premium. We first show that both premiums vary systematically across cities in ways that are consistent with the patterns of inequality; we then quantify their importance in explaining the observed pattern of within-city inequality. These empirical patterns motivated the core mechanism of the theoretical model presented in Section 3.

Inter-Industry Wage Premium The literature has comprehensively documented that conditional on the observed individual characteristics, the residual wage varies substantially across industries, and thus the so-called "inter-industry wage premium" exists (Rosen, 1987; Krueger and Summers, 1988; Katz et al., 1989). We document that the inter-industry wage premium varies systematically across cities: they are higher in larger cities, but the premium in low-paying industries grows relatively faster with city size than the high-paying industries.

We first run the following regression to estimate the wage premium of each industry in every city j separately:

$$\ln\left(W_i^j\right) = \beta_0^j + \beta_1^j \mathbf{ind_i} + \beta_2^j \mathbf{occ}_i + \beta_3^j X_i + \epsilon_i^j, j = 1, 2, \cdots, J, \tag{4}$$

where W_i^j is the observed wage of individual i working in city j. $\mathbf{ind_i}$ is a vector of dummy variables to indicate the industry in which the individual works. Our industry classification comes from the 1990 census with 201 industries in the sample. We use "groceries" as the benchmark industry since it is present in all the cities. $\mathbf{occ}_i^j$ is another vector of dummy variables that control for the occupation of individual i.[16] X_i^j is a vector that controls the race, years of education, marital status, and age of the individual. We are interested in the vector β_1^j that measures the wage premium of each industry in city j relative to the benchmark industry. The above equation is estimated separately for each city and thus no

[16] We use the 1990 census occupational classification.

city fixed effects are needed.

We define the average premium of industry k as:

$$\bar{\gamma}(k) = \frac{\sum_{j=1}^{N(k)} \widehat{\beta_1^j}(k)}{N(k)},$$

which is the average of the k-th element of $\widehat{\beta_1^j}$ across all the $N(k)$ cities in which the industry is present. We then use the following equation across the entire sample to study the relationship between the average industry premium and city size:

$$\ln\left(W_i^j\right) = \beta_0 + \beta_1 \ln(Y^j) + \beta_2 \ln(Y^j) \times \bar{\gamma}(k(i)) + \beta_3 X_i + \mathbf{ind_i} + \mathbf{occ_i} + \epsilon_i, \qquad (5)$$

where Y^j is the size of the city j in which individual i lives, $k(i)$ is the industry in which the individual works, and all the other variables are the same as defined above. The coefficient of interest is β_2, the interaction between industry wage premium and city size.

Table 3 reports the results. The first column reports the regression without the interaction term and shows that across all the industries the wage premium is higher in larger cities, a result commonly found in the literature. The second column adds in the interaction term and shows that industries with higher average premium see their wage grow at a slower pace with city size. A 10-percent increase in the city size leads to a 0.45 percent increase in premium for industries at the 25th percentile wage premium and a 0.40 percent increase for industries at the 75th percentile. In the next column, we add in the city fixed effects to absorb the city size variable. The estimated relationship between the average industry premium and city size remains significantly negative. The last two columns report regressions using a binary definition of "high-premium industry", which is defined as those industries with average premium above the median of the distribution of $\bar{\gamma}(\cdot)$. The results are still similar: the city size elasticity of wage at high-premium industries is around 0.040, while the elasticity of the low-premium industries is about 10 percent higher at 0.044. Table A.7 reports robustness checks of the binary measure by defining "high-premium industry" based on the 25th or the 75th percentiles instead of the median, and we find similar results.

One potential explanation for the pattern documented above is the city specialization

Table 3: City Size and Inter-Industry Wage Premium

LHS = Ln(Wage)	Level Effect	Continuous Measure		Binary Measure	
	(1)	(2)	(3)	(4)	(5)
Ln(City Size)	0.042***	0.043***		0.044***	
	(0.000)	(0.000)		(0.001)	
(Avg. Ind. Premium) × Ln(City Size)		-0.018***	-0.012***		
		(0.003)	(0.003)		
(High-Premium Ind. Dummy) × Ln(City Size)				-0.004***	-0.003***
				(0.001)	(0.001)
Age	0.013***	0.013***	0.013***	0.013***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Edu.	0.049***	0.049***	0.047***	0.049***	0.048***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.194***	0.194***	0.197***	0.194***	0.198***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Race == White	0.128***	0.127***	0.132***	0.128***	0.133***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,215,182	1,215,182	1,237,695	1,237,695
R-squared	0.422	0.423	0.431	0.422	0.430
Industry FE	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes
City FE	No	No	Yes	No	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the estimation of equation (5) with variations. The left-hand-side variable is the logarithm of observed wage and "city size" is measured as private industry GDP. "Average Industry Premium" is a continuous measure computed from estimating equation (4), and "High-Premium Industry Dummy" is the associated binary measure in which the average premium is higher than the median. Data source: IPUMS-USA, 2000.

following the ideas of Henderson (1974). Certain industries are indeed concentrated in a small number of cities. For example, "metal mining" and "shoe repair shops" only exist in less than 50 cities, and 17 industries only in less than 100 cities. As a way to quantify the impact of city specialization, we repeat our exercise while restricting the sample to the industries that are present in more than 50, 100, or 200 cities. Table A.8 reports the results. The magnitude of the estimates barely changes if we drop the highly concentrated industries that only show up in less than 50 or 100 cities. The point estimates drop from -0.012 to -0.005 (still significant) if we restrict the sample to industries present in more than 200 cities for the continuous measure of industry premium, and similarly for the binary "high-paying" industry dummy. This implies that specialization can indeed explain a sizable fraction of the observed differences in city-size elasticities of industry premium. The model proposed in the paper does not feature specialization of cities, as all industries will be active in all cities. In this sense, the differences in the point estimate between specialized and non-specialized industries can be interpreted as a measure of the missing components of our model.

Table 4: City Size and Tradable Industry Wage Premium

LHS = Ln(Wage)	Benchmark Def.			Alternative Def.		
	(1)	(2)	(3)	(4)	(5)	(6)
Tradable Ind. Dummy	0.064***			0.074***		
	(0.001)			(0.001)		
Tradable Ind. Dummy × Ln(City Size)		-0.016***	-0.014***		-0.021***	-0.018***
		(0.001)	(0.001)		(0.001)	(0.001)
Ln(City Size)	0.043***	0.050***		0.043***	0.049***	
	(0.000)	(0.001)		(0.000)	(0.000)	
Age	0.014***	0.013***	0.013***	0.014***	0.013***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Edu.	0.054***	0.049***	0.047***	0.054***	0.049***	0.048***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.204***	0.194***	0.198***	0.204***	0.194***	0.198***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Race == White	0.132***	0.128***	0.133***	0.133***	0.128***	0.133***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695
R-squared	0.405	0.422	0.431	0.405	0.423	0.431
Industry FE	No	Yes	Yes	No	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	No	No	Yes	No	No	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the estimation of equation (5) with variations. "City size" is measured as private industry GDP. "Benchmark" definition of tradable industries include agriculture, manufacturing and construction industries in the 1990 industry classifications, and "alternative" definition classifies construction industries as non-tradable. Data source: IPUMS-USA, 2000.

A similar pattern can also be observed between tradable and non-tradable industries as well. We repeat the binary exercise above, replacing the "high-premium industry dummy" with a tradable industry dummy and report the results in Table 4. In the "benchmark" definition of tradable industries, we classify all the agriculture, manufacturing, and construction industries as tradable; in the "alternative" definition, we re-classify the construction industry to a non-tradable industry.[17] Similar to the results in Table 3, workers in tradable industries

[17]The benchmark classification of tradable industries include code 10 - 392 (excluding the agriculture

enjoy a 6.4 to 7.4 percent wage premium, but the premium decays with city size: doubling the city size reduces the tradable premium by around 1.4 and 1.8 percentage points.

Entrepreneurship Premium The entrepreneurship premium, on the other hand, increases with the city size, and thus can potentially explain the widened income gap at the right tail of the wage distribution. We follow the same binary estimation strategy outlined in Equation (5) and estimate the following equation:

$$\ln W_i^j = \beta_0 + \beta_1 E_i + \beta_2 \ln(Y^j) \times E_i + \beta_3 X_i + \mathbf{ind_i} + \mathbf{city_i} + \epsilon_i, \tag{6}$$

where i indexes individual, j indexes the city in which i lives, Y^j is the size of city j, and E_i is a binary variable that takes the value of 1 when the individual is defined as an entrepreneur. We use two definitions of "entrepreneurship". The benchmark defines "entrepreneur" as executives or as managers of finance, marketing, human resources in the 1990 occupational classification from the Bureau of Labor Statistics (see Table A.11 in the appendix for details). We also use an alternative definition of entrepreneurship that only includes "chief executives" as a robustness check. Similar to equation (5), $\mathbf{ind_i}$ and $\mathbf{city_i}$ are vectors of industry and city fixed effects, and X_i is a vector that controls for individual characteristics. The key parameter of interest is β_2, the city size elasticity of entrepreneurship premium.

Table 5 reports the results. The first column omits the interaction term and reports that entrepreneurs earn around 33 percent higher wage rate across all the cities. The second column controls for the interaction term and shows that the entrepreneurship premium is higher in larger cities: a 10-percent increase in the city size is associated with a 1.3 percent increase in entrepreneurship premium. In the third column, we additionally control for the occupational fixed effects which absorb the entrepreneurship dummy. The interaction with city size remains significant and positive, and the elasticity increases from 0.013 in the previous column to 0.021. The next three columns repeat the above exercise with the alternative definition of entrepreneurship. All the results remain the same qualitatively.

In the next section, we present a general equilibrium model that can generate the three patterns documented above with sorting of individuals across cities, industries, and occupa-

service industries 12, 20, and 30) and 500 - 561. The construction industry is code 60, "all constructions".

Table 5: U.S. Entrepreneurial Wage Premium and City Size.

LHS = Ln(Wage)	Benchmark Def.			Alternative Def.		
	(1)	(2)	(3)	(4)	(5)	(6)
Entrepreneur Dummy	0.327***	0.182***		0.621***	0.495***	
	(0.002)	(0.014)		(0.006)	(0.049)	
Entrepreneur Dummy × Ln(City Size)		0.013***	0.021***		0.011**	0.023***
		(0.001)	(0.001)		(0.004)	(0.004)
Age	0.014***	0.014***	0.013***	0.014***	0.014***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Edu.	0.068***	0.068***	0.047***	0.073***	0.073***	0.048***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.227***	0.227***	0.198***	0.236***	0.236***	0.198***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Race == White	0.169***	0.169***	0.132***	0.179***	0.179***	0.132***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695
R-squared	0.391	0.391	0.431	0.384	0.384	0.430
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	No	Yes	No	No	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Huber-White robust standard errors in parentheses. Results of estimating Equation (6). The left-hand-side variable is the logarithm of observed wage and "city size" is measured as private industry GDP. "Benchmark Def." includes all the occupations listed in Table A.11 as entrepreneurs, while "Alternative Def." only includes the "Chief executives and public administrators" (the 3rd row). Data source: IPUMS-USA, 2000.

tions.

Robustness Checks Table A.6, A.9, and A.10 report the robustness check using population as the measure of city size for all main results. Table A.7 experiments with different definitions of "high-premium industry", and Table A.8 drops the industries that are concentrated in a few cities.

2.4 Quantifying the importance of the mechanisms

In this section we quantify the importance of the two potential channels that we have discussed above. We carry out two exercises to do so. In the first exercise, we control for industry-city or occupation-city fixed effects when computing the residual wage to eliminate the spatial variations along these two dimensions. In the second exercise, we only eliminate the variance of inter-industry and entrepreneurship premiums across the cities. We repeat the earlier exercise on the simulated data and study how within-city inequality changes.

2.4.1 Industry-City and Occupation-City Fixed Effects

The measures of inequality in the previous parts are based on the residual wage from estimating equation 1. This equation does not control for the spatial variations of industry or occupation fixed effects.[18] To control for the spatial variation, we add in industry-by-city fixed effects in the Mincer regression to explain the patterns on the left tail, and similarly, occupation-by-city fixed effects to explain the right tail. In the validation exercise, instead of the equation (1), we estimate:

$$\ln(W_i) = \beta_0 + \beta_1 \cdot \mathbf{X_i} + \text{industry} \times \text{city}_i + \epsilon_i. \tag{7}$$

"$\text{industry} \times \text{city}_i$" is a dummy variable to indicate the industry and city in which individual i works, which allows the inter-industry wage premiums to vary across cities. We obtain

[18] Table A.2 reports the results that control for industry and occupation fixed effects, and the main results persist. This indicates that industry or occupation premium alone is not enough to explain the observed pattern of inequality. However, controlling for these fixed effects does not account for the *spatial variations* of these premiums, and we show in this section that it is the spatial variation of the premium that drives the the pattern of within-city inequality.

the residual wage from this equation, repeat the exercise in Section 2, and report the results of the left-tail inequality in the first two columns of Table 6. Once the industry-city fixed effects are controlled for, larger cities no longer see lower inequality on the left-tail. Instead, they are slightly more unequal than the smaller ones, a reversed pattern as compared to benchmark results. This indicates that the variations of inter-industry wage premium across cities are more than enough to explain the within-city inequality on the left.

Similarly, we estimate the following Mincer equation with occupation-city fixed effects:

$$\ln(W_i) = \beta_0 + \beta_1 \cdot \mathbf{X_i} + \text{occupation} \times \text{city}_i + \epsilon_i, \tag{8}$$

and repeat the exercise with the residual from the estimation. The results are reported in the last two columns in Table 6. Larger cities are still more unequal in the right-tail after controlling occupation-city fixed effects. However, the relationship becomes muted. In our benchmark results, the city size elasticity of the 95-50 gaps is 0.026; it drops by 27 percent to 0.019 in the counter-factual. Similarly, the size elasticity of the 90-50 gaps drop from 0.017 to 0.014. The variations in occupational premiums across cities can explain around one-third of the right-tail inequality. Unlike the left tail, the residual wage rates at the right tail depend more on unobservable individual characteristics and thus vary substantially more within a city-occupation cell. The variations of residual wage rates within all entrepreneurs in New York city are arguably larger than the variations across the residual wage rates janitors in the same city. This implies that any model that tries to explain the pattern on the right tail based on the observables will see their explanatory power dampened.

2.4.2 Eliminating Spatial Variance

In the second exercise, we only remove the differences in the second moment across cities, instead of industry-city and occupation-city fixed effects that eliminates all the spatial variations. We first directly estimate the premiums across cities and then eliminate the spatial variance of these premiums. We reconstruct the residual wage rates from the normalized industry and entrepreneurship premiums and then repeat the exercise in Section 2. The results are similar to the first set of counter-factual analysis.

Table 6: Counterfactual Regressions

LHS = Residual Wage Ineq.	Ind-Cty FE Filtered		Occ-Cty FE Filtered	
	50-05	50-10	95-50	90-50
Ln(Private Ind. GDP)	0.016***	0.015***	0.019***	0.014***
	(0.005)	(0.004)	(0.005)	(0.004)
Average Years of Edu.	0.026**	0.026**	0.023*	0.015
	(0.012)	(0.012)	(0.014)	(0.012)
Race == White	-0.299***	-0.310***	-0.328***	-0.282***
	(0.098)	(0.096)	(0.106)	(0.101)
N	254	254	254	254
R-squared	0.591	0.646	0.543	0.582
State FE	Yes	Yes	Yes	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the results of estimating equation (3). The measures of inequality are various percentile ratios of hourly wage within an MSA. Different from the results in Table 1, the measures of inequality are computed based on Eq 7 and 8 that control for industry-city or occupation-city fixed effects.

Inter-Industry Premium and the Left Tail Inequality We first compute a benchmark measure of wage that allows for the spatial variations of industry wage premium. We define the benchmark as the linear prediction based on the estimation of Equation (4):

$$\ln\left(\widehat{W_i^j}\right) = \widehat{\beta_0^j} + \widehat{\beta_1^j}\mathbf{ind_i} + \widehat{\beta_2^j}\mathbf{occ}_i + \widehat{\beta_3^j}X_i, j = 1, 2, \cdots, J,$$

where $\widehat{\beta_0^j}$, $\widehat{\beta_1^j}$, $\widehat{\beta_2^j}$, and $\widehat{\beta_3^j}$ are the OLS estimates of their counterparts in Equation (4). $\widehat{\beta_1^j}$ is the estimated industry wage premium in city j from the data, and its spread within city varies across different cities. To construct the counterfactual wage premium, $\widetilde{\beta_1^j}$, we eliminate the spatial variations in the spread of $\widehat{\beta_1^j}$:

$$\widetilde{\beta_1^j} = \bar{\sigma}\frac{\widehat{\beta_1^j}}{\sigma(\widehat{\beta_1^j})},$$

where $\sigma(\widehat{\beta_1^j})$ is the standard deviation of $\widehat{\beta_1^j}$ within city j, and $\bar{\sigma}$ is a scaling factor.[19] After the transformation $\widetilde{\beta_1^j}$ has the same standard deviation of $\bar{\sigma}$ across all cities. We then proceed

[19]The size of $\bar{\sigma}$ can be arbitrary. We define $\bar{\sigma}$ as the average $\sigma(\widehat{\beta_1^j})$ of the largest five cities. Changing $\bar{\sigma}$ moves all the $\widetilde{\beta_1^j}$ proportionally and thus will not affect the results.

to compute the counterfactual wage for each individual in our sample according to:

$$\ln\left(\widetilde{W}_i^j\right) = \widehat{\beta}_0^j + \widetilde{\beta}_1^j \mathbf{ind_i} + \widehat{\beta}_2^j \mathbf{occ}_i + \widehat{\beta}_3^j X_i, j = 1, 2, \cdots, J,$$

The counterfactual $\widetilde{W}_i^j$ is the same as the benchmark $\widehat{W}_i^j$ with $\widehat{\beta}_1^j$ replaced by $\widetilde{\beta}_1^j$.

We repeat the same exercise as in Section 2 using the benchmark wage $\widehat{W}_i^j$ and the counterfactual wage $\widetilde{W}_i^j$. We first compute residual wage inequality within each city and then estimate Equation (3) to study how within-city inequality varies across cities. The results are reported in the first panel of Table 7.

When we shut down the spatial variations of inter-industry wage premium, the negative relationship between city size and inequality measured at the left tail weakens significantly. The first two columns report the estimation of Equation (3) based on the benchmark $\widehat{W}_i^j$ and the next two columns report the results based on the counterfactual $\widetilde{W}_i^j$. The size elasticity of 50-05 and 50-10 wage gap is -0.068 and -0.045 in the benchmark case and drops to -0.040 and -0.022 in counter-factual case, respectively. This suggests that around half of the spatial variations of left-tail inequality reported in Section 2 can be explained by the two proposed mechanisms. Not surprisingly, the point estimates in the counterfactual case do not go to zero, suggesting that there exist other mechanisms in the data that can also explain the observed pattern. For example, Table A.8 in Section reports that the composition of industries across cities can potentially explain some of the observed variations in inter-industry wage premium, and thus can conceivably explain the left-tail inequality as well. Introducing the specialization of cities is beyond the scope of the current paper, and it could be a fruitful direction for future work.

Entrepreneurship Premium and the Right Tail Inequality To quantify the importance of entrepreneurship mechanism, we perform a similar counter-factual analysis by shutting down the spatial variations of entrepreneurship premium and then re-measure the right-tail inequality.

We start by computing the benchmark wage using the linear predictions similar to the

Table 7: Benchmark and Counterfactual Regressions

(a) Left Tail

	Baseline		Counter-Factual	
	50-05	50-10	50-05	50-10
Ln(Private Ind. GDP)	-0.068***	-0.045***	-0.040***	-0.022***
	(0.005)	(0.004)	(0.005)	(0.003)
Average Years of Edu.	0.050***	0.036***	0.036***	0.022***
	(0.010)	(0.009)	(0.010)	(0.008)
Share of White Population	-0.314***	-0.217***	-0.242**	-0.182**
	(0.095)	(0.077)	(0.104)	(0.077)
N	254	254	254	254
R-squared	0.554	0.470	0.392	0.374
State FE	Yes	Yes	Yes	Yes

Robust standard errors reported in parentheses
* p<0.10, ** p<0.05, *** p<0.01

(b) Right Tail

	Baseline		Counter-Factual	
	95-50	90-50	95-50	90-50
Ln(Private Ind. GDP)	0.014***	0.016***	0.010***	0.013***
	(0.004)	(0.004)	(0.003)	(0.004)
Average Years of Edu.	-0.020**	-0.032***	0.000	-0.019*
	(0.009)	(0.011)	(0.007)	(0.010)
Share of White Population	-0.035	-0.205***	-0.039	-0.205***
	(0.075)	(0.068)	(0.043)	(0.071)
N	254	254	254	254
R-squared	0.130	0.512	0.104	0.488
State FE	Yes	Yes	Yes	Yes

Robust standard errors reported in parentheses
* p<0.10, ** p<0.05, *** p<0.01

.

Note: The first panel of the table reports the estimation of Equation (3) with the benchmark $\widehat{W_i^j}$ and the counterfactual $\widetilde{W_i^j}$ instead of the data on the left tail. The second panel of the table reports the same estimation with right-tail inequality. All the other control variables are omitted in the tables. For more details, see the main text in Section 2 and 2.4

one from the previous section for each city separately:

$$\ln \widehat{W}_i^j = \widehat{\beta}_0^j + \widehat{\beta}_1^j E_i + \widehat{\beta}_2^j X_i + \widehat{\beta}_3^j \text{ind}_i, j = 1, 2, 3, \cdots J,$$

where the coefficients with hats are the estimates based on OLS and the $\widehat{W}_i^j$ is the predicted baseline income. We eliminate the spatial variation of $\widehat{\beta}_1^j$ across cities by taking the national average:

$$\widetilde{\beta}_1 = \frac{\sum_{j=1}^{J} \widehat{\beta}_1^j}{J},$$

and compute the counter-factual wage $\widetilde{W_i^j}$ as:

$$\ln \widetilde{W}_i^j = \widehat{\beta}_0^j + \widetilde{\beta}_1^j E_i + \widehat{\beta}_2^j X_i + \widehat{\beta}_3^j \text{ind}_i, j = 1, 2, 3, \cdots J.$$

Again, the only difference between $\widehat{W}_i^j$ and $\widetilde{W}_i^j$ is the coefficient on the entrepreneur dummy.

We then proceed to compute the right-tail inequality following the methods in the earlier parts of Section 2 and report the results in the second panel of Table 7. Using the benchmark predicted wage rates, the city size elasticity of the 95-50 and the 90-50 wage gaps is 0.014 and 0.016, and they drop to 0.010 and 0.013 respectively once we eliminate the spatial variations of entrepreneurship. It implies that between 18.8 and 28.6 percent of the observed variation on the right tail can be explained by the entrepreneurship premiums. The share explained by entrepreneurship premium is smaller as compared to the left tail because entrepreneurs only constitute a tiny fraction of the top 5 to 10 percent of the wage distribution. It is possible that workers within the top wage percentiles also earn higher premiums in larger cities, similar to the entrepreneurs. Nevertheless, the spatial variations of entrepreneurship premium are still responsible for a sizable part of the right-tail inequality.

3 Model

3.1 General Environment

The economy is geographically divided into $J \geq 1$ cities populated by a unit mass of individuals. Individuals are heterogeneous in their innate human capital endowments, x, which follows a continuous distribution with support on $\mathbb{R}^+$ and the cumulative distribution function $G(x)$. x includes individual characteristics that are both observable and unobservable to an outside econometrician, but x is perfectly observable to all individuals inside the model.

Individuals can freely choose which city to live in subject to no cost of migration. Within each city, individuals choose between two sectors, namely, high-entry-cost and low-entry-cost sector (thereafter referred as H and L sectors). H sectors are those that require certain certification or education to enter, such as manufacturing industries; L sectors are those with no entry barriers, such as low-end jobs in retailing and other service industries. In the model, individuals need to pay an entry cost $S > 0$ in unit of final consumption goods in order to work in the H industry.[20]

H Sector Individuals can choose between two occupations — entrepreneurs or workers — in the H sector following the occupational choice model in Lucas (1978). The market structure in the H sector is monopolistic competitive with differentiated products. To produce in the H sector, individuals need to first organize into firms. Any individual can choose to create a new firm, hire workers, and start production of a new variety. With a slight abuse of notation, we use x, the human capital of the entrepreneur, to index the variety she produces. The firm created by the entrepreneur with human capital level x in city j has the following production function:

$$Q_j(x) = b(\bar{x}_j)\psi(x)(\ell - f),$$

where ℓ denotes efficiency labor input. $\psi(x)$ is a strictly increasing and convex function that maps the human capital of the entrepreneur to the firm productivity. We assume $\psi(0) = 0$, $\lim_{x\to\infty}\psi(x) = \infty$, $\psi(x) > 0, \psi'(x) > 0$, and $\psi''(x) > 0$ for all x. f is the fixed cost of production in units of efficiency labor that is constant across cities and firms. The assumption of $\psi(0) = 0$ and $f > 0$ together imply that the individuals with $x = 0$ will never choose to be an entrepreneur in equilibrium. $b(\bar{x}_j)$ summarizes the city-level productivity. It is an increasing function of $\bar{x}_j$, the average level of entrepreneur talent in city j, captures the overall productivity in city j. The inter-linkage between the location-specific productivity and the average level of entrepreneurial talent is common in the literature of urban and economic geography. For example, it can be micro-founded in a model with

[20]The unit of S is not crucial for our results. If S is denoted in utility terms, or the unit of numeraire, all of our propositions later in the next section will still hold.

positive knowledge spillovers or return to face-to-face meetings. In this paper, we are abstract away from this micro-foundation by assuming the positive linkage, and study its impacts on within-city inequality. The income of the entrepreneur equals to the profit of the firm she owns.[21]

Individuals can also choose to work for an existing firm in the H sector. In this case, 1 unit of human capital directly translates into 1 unit of efficiency labor supply, and the income of a worker with human capital x is thus $w_j x$, where w_j is the wage rate per efficiency unit of labor in city j.

The above two assumptions are crucial in generating the patterns of inequality at the right tail: 1) entrepreneurs wage increases with firm size, and 2) workers' wage is not directly linked to the firm they work for. Both assumptions are broadly supported on the empirical findings in the literature. The literature on executive compensation has documented that the income of the top executives is proportional to a power function of the size of the firm she manages — a relationship also known as the "Roberts' law" (Roberts, 1956; Murphy, 1999). The positive correlation between entrepreneurial compensation and firm size is also rooted in many models of CEO pay, as long as in equilibrium the model sustains assortative matching between entrepreneurs and firms, such as in Gabaix and Landier (2008). We abstract away from most of the details of an executive compensation model and directly assumes equilibrium assortative matching as in Ma and Ruzic (2015) by linking the productivity of the firm with the human capital of the founder. For simplicity of exposition, we also assume the easiest form of the power function, the identity mapping, in our benchmark model.

In comparison to the executive compensation, the link between firm size and the average wage rates of workers appears to be much weaker. Researchers have indeed documented the existence of positive firm-size-premium for workers as well (Oi and Idson, 1999). However, once individual characteristics have been controlled for, the size-premium usually shrinks significantly. For example, Abowd et al. (1999) document that individual effects explain

[21] We assume that there is only one entrepreneur per firm in the model. This implies that only a small fraction of the working population see their wage rate increases with the size of the firm. This assumption is not critical for the theoretical results in the paper. The mechanism can be easily extended beyond the dichotomy between entrepreneurship and workers. In addition to entrepreneurs, a worker in any occupation or profession whose wage rate rises with the size of the firm shall benefit more from working in large cities. One can also directly scale up the measure of entrepreneurs per firm to account for multiple managers within the firm for quantitative purposes without affecting the main results of the paper.

about 75 percent of the firm-size wage effect, while firm effects explain relatively little. In our model, the wage rate w_j is not the observed wage rate in the data but the residual wage rate net of individual characteristics. For this reason, we assume it to be independent of firm sizes, and only determined in the local labor markets in a city.

The above two assumptions can be relaxed along several dimensions without affecting the main results (Proposition 2). The compensation function for entrepreneurs can be a power function or any function that is regularly-varying. We can also allow for positive firm-size elasticities of workers as well. As long as the firm-size elasticity of entrepreneurs is higher than that of the workers — a simple assumption backed by most empirical studies on manager-to-worker pay ratios — our results survive.

Trade Goods in the H sector can be traded across cities frictions. To export from city k to city j, firms need to incur iceberg trade costs so that in order to sell one unit of goods to market j, the firm in city k needs to ship out $\tau_{jk} > 1$ units of goods. We denote the price of variety x produced in city k and sold in city j as $p_{jk}(x)$.

L Sector L goods are homogeneous and non-tradable across cities. The market is perfectly competitive. Individuals can produce without organizing into firms, which imply that no occupation choice problem exists in the L sector. Production function in L sector is linear in labor supply, and 1 unit of labor input leads to 1 unit of output. Denote the price of L goods in city j as z_j. The income of an individual in the L sector with human capital x in city j is thus $z_j \times x$.

Preference Individuals in city j gain utilities from the consumption of final goods, which is an aggregation of all the goods available in the city they reside in:

$$y_j = \left[\frac{\int_{i \in \Omega_j} q(i)^{\frac{\sigma-1}{\sigma}} di}{\alpha} \right]^{\frac{\sigma\alpha}{\sigma-1}} \left[\frac{\Lambda_j}{1-\alpha} \right]^{1-\alpha}$$

where Λ_j is the consumption of L goods; $q(i)$ is the consumption of variety i of H goods, Ω_j is the set of available H goods in city j. α denotes the expenditure share of H goods, and $\sigma > 2$ is the elasticity of substitution between varieties.

Individual's utility function is linear in the consumption of final goods, and decreasing with congestion dis-utility. It is simply:

$$V = y_j - C(N_j),$$

if the individual works in the L sector, and

$$V = y_j - C(N_j) - S,$$

if the individual works in the H sector. $C(N_j)$ captures congestion dis-utilities, which positively depends on the population of city j. We assume that $C'(\cdot) > 0$ and $C''(\cdot) > 0$ so larger cities exert higher congestion dis-utility and the cost function is convex. Appendix C provides an extension of our model in which the congestion disutility is micro-founded in a model with the provision of public goods. In the benchmark model, we directly use the reduced functional from the extended model for the sake of simplicity.

4 Analytical Results

This section discusses the analytical results. We first define the equilibrium and the terminology that will be used throughout this section, and then proceed to characterize the central pattern of an asymmetric equilibrium: the assortative matching between individuals and location-industry choices. Based on the assortative matching, we show how within-city inequality vary across cities in the equilibrium. In the last part, we deal with the issue of multiple equilibrium, and provide conditions under which a unique equilibrium will emerge. Appendix D provides all the formal proofs, and we only discuss the intuition or the sketch of the proof in the main text.

4.1 Definition and Notation

We define a **spatial equilibrium** as a mapping from individual's human capital x to its corresponding location, industry, occupation and consumption choices, and a series of prices

$\{w_k,\, z_k,\, p_{kj}(x)\}$ such that:

1. Given the prices, each individual maximizes her utility by choosing the occupation, industry, location, and consumption bundle; each firm maximizes its profits.

2. H and L goods markets clear in each city and labor market clears in each city and sector.[22]

3. All cities are populated.

The potential equilibria can be either symmetric or asymmetric. As the cities are ex-ante identical, a symmetric equilibrium in which all the cities are identical in population, price, and sectoral composition always exists. However, throughout this paper, we are interested in the asymmetric equilibrium, where cities can be ex-post heterogeneous.

Before we delve into the analytical results, we first define the variables and notations that will be used in the rest of the section. The set of potential location choices is $k \in \{1, 2, \cdots, J\}$ and the set of occupation choices is $\omega \in \{E, H, L\}$, where E stands for the entrepreneur in the H sector, H and L stands for the worker in the H and L sectors, respectively. The individuals choose a location-sector combination in the the space that consists of $3J$ elements, and we refer the duplet (k, ω) as a **"choice"** of the individual.

The indirect utility function of individual x making the choice (k, ω) can be written as:

$$V_k^\omega(x) = A_k^\omega \cdot \phi^\omega(x) - B_k^\omega,$$

where A_k^ω is the return to human capital, B_k^ω is the costs of choosing (k, ω), and $\phi^\omega(x)$ is an occupation-specific function of human capital. For example, if x chooses to be an

[22] The market clearing condition in the H sector implies that trade must be balanced in each city. Combined with the non-tradability of the L sector, this implies that in equilibrium cities are not specialized in the sense that every single city must host both H and L sectors. This does not contradict the empirical pattern documented in Table A.8, in which many cities only host a selection of industries. The empirical pattern is based on a detailed classification of 201 industries, while in our model we adopt a broad definition of 2 industries. Whether we interpret the two industries as tradable/non-tradable or as high/low-paying industries, every city in the data host both types of the broadly-defined industries. However, as is shown in the empirical section, specialization of cities can indeed explain a sizable proportion of the observed variations of inter-industry wage premium across cities. We prioritize the simplicity of the model over empirical explanatory power in this paper and leave the extension of specialized industries to future works.

entrepreneur in city k, then:

$$V_k^E(x) = \underbrace{\frac{\pi_k}{P_k^\alpha z_k^{1-\alpha}}}_{A_k^E} \underbrace{(\psi(x))^{(\sigma-1)}}_{\phi^E(x)} - \underbrace{\left[C(N_k) + S + f \cdot \frac{w_k}{P_k^\alpha z_k^{1-\alpha}}\right]}_{B_k^E},$$

$$\pi_k = \frac{1}{\sigma}\sum_{j=1}^{J}\left[\alpha R_j P_j^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\frac{\tau_{jk} w_k}{b_k}\right)^{1-\sigma}\right].$$

The return to the choice, A_k^E depends on the profit that can be generated, π_k, and the ideal price index $P_k^\alpha z_k^{1-\alpha}$.[23] The barrier to (k, E) depends on the location-specific congestion, as well as the barrier into the industry, S, and the fixed costs of starting a firm in the units of utility, $f\frac{w_k}{P_k^\alpha z_k^{1-\alpha}}$.[24]

Similarly, the indirect utility functions for workers in the H or the L sector are, respectively:

$$V_k^H(x) = \underbrace{\frac{w_k}{P_k^\alpha z_k^{1-\alpha}}}_{A_k^H} \cdot x - \underbrace{[S + C(N_k)]}_{B_k^H},$$

$$V_k^L(x) = \underbrace{\frac{z_k}{P_k^\alpha z_k^{1-\alpha}}}_{A_k^L} \cdot x - \underbrace{C(N_k)}_{B_k^L}.$$

In the case of the workers, $\phi^H(x) = \phi^L(x) = x$.

Lastly, we denote the set of individuals with human capital x that optimally choose (k, ω) in an equilibrium as the set $\omega_k \subset \mathbb{R}^+$:

$$\omega_k = \left\{x \in \mathbb{R}^+ : V_k^\omega(x) \geq \max_{k'=\{1,2..J\}}\left[\max_{\omega'=\{E,H,L\}} V_{k'}^{\omega'}(x)\right]\right\}. \tag{9}$$

For example, the set of x that chooses to be entrepreneurs in city k is denoted as E_k, and

[23] In this section, we denote the city-level productivity, $b(\bar{x}_k)$ simply as b_k, and treat the variable as exogenously fixed. As individuals are atomistic in the model, they do not internalize their impact on the average entrepreneur talents in the city. As Corollary 1 shows, the sorting of the entrepreneurs in equilibrium indeed exhibit assortative matching when $b(\bar{x}_k)$ is taken as given. This implies that b_k is larger in larger cities, which is consistent with the prediction that both A_k^E and B_k^E are larger in larger cities.

[24] The fixed costs of starting a firm, $f \cdot w_k$, needs to be converted into utility terms, and therefore the ideal price index in the denominator.

for workers in the H and L sectors, as H_k and L_k. In any asymmetric equilibrium, we must have a positive measure of individuals in all city-sector-occupation cells, so $\omega_k \neq \emptyset, \forall \omega, k$.

4.2 Assortative Matching between Individuals and Choices

In this part, we first establish the assortative matching rule: individuals with higher x will be sorted into a choice (k, ω) that is more costly to enter. We then discuss the implications of assortative matching within each city and occupation. Our starting point is a lemma that will be used throughout the proofs: all the indirect utility functions only intersect at most once on $\mathbb{R}^+$:

Lemma 1. *For all $k, k' \in 1, 2, \cdots, J$ and $\omega, \omega' \in \{E, H, L\}$, there exists one and only one $x \in \mathbb{R}^+$ such that $V_k^{\omega}(x) = V_{k'}^{\omega'}(x)$.*

The single-crossing property simplifies the characterization of the equilibrium as it allows us to infer an individual's preference between two choices regardless of the monotonicity of $g(x) = V_k^{\omega}(x) - V_{k'}^{\omega'}(x)$. In the general case of non-monotonic $g(x)$, the Lemma above allows us to infer the sign of $g(x)$ by computing the sign of $g(0)$ and $\lim_{x \to \infty} g(x)$, which is straightforward to do in many cases without explicitly specifying the functional forms in the model. With the help of the single-crossing property, we can prove our main result of assortative matching:

Proposition 1. *Suppose there are two choices, (k, ω) and (k', ω'). If $x' \in \omega'_{k'}$, then $x' > \sup(\omega_k)$ if and only if $B_{k'}^{\omega'} > B_k^{\omega}$.*

Intuitively, the proposition states that more talented individuals will enter the cells with higher barriers. The proof is provided in the appendix, and we only discuss the sketch in the main text. Note that the difference between the two choices, $g(x) = V_{k'}^{\omega'}(x) - V_k^{\omega}(x)$, evaluated at $x = 0$ is simply $g(0) = B_k^{\omega} - B_{k'}^{\omega'}$. If $B_{k'}^{\omega'} < B_k^{\omega}$, then we can infer $g(0) > 0$. For any $x \in \omega^k$, we must have $g(x) < 0$ by revealed preference, and by the same logic, we must have $g(x') > 0$. However, as $0 < x < x'$ and $g(x)$ switches signs twice in the interval $(0, x')$, the continuity of $g(x)$ implies that there exists at least two values of $x_1, x_2 \in (0, x')$ such that $g(x_1) = g(x_2) = 0$. This contradicts with the single crossing property as stated in Lemma 1.

Proposition 1 is a useful result as it allows us to match individuals to choices by human capital and the entry barrier, B_k^{ω}. The assortative matching is an equilibrium result, and thus it takes all endogenous responses in general equilibrium, such as factor and output prices, as well as the distribution of population across space, into consideration, which in turn facilitates the characterization of the equilibrium without solving the model entirely.

The direct implication of the above proposition is that assortative matching shall arise along both the location and the industry dimension in our model. Our next corollary states that within a city, individuals sort across industries and occupations by talent and desirability; similarly, within an industry/occupation, individuals sort by talent and city size:

Corollary 1. *In any asymmetric equilibrium:*

(i) Within city k, individuals sort into occupations by entry barrier: $\inf E_k \geq \sup H_k$, *and* $\inf H_k \geq \sup L_k$.

(ii) Within each industry/sector ω, individuals sort by city size. If $N_{k'} > N_k$, then 1) $\inf E_{k'} \geq \sup E_k$, *2)* $\inf H_{k'} \geq \sup H_k$, *and 3)* $\inf L_{k'} \geq \sup L_k$.

(iii) There exists a cutoff x_E such that individuals choose to be an entrepreneur in some city k if and only if $x \geq x_E$.

We provide a sketch of the proof using the sorting of the entry barriers. Within each city k, all the occupations face the same congestion disutility; the entry barrier into entrepreneurship is the highest at $B_k^E = C(N_k) + S + f\frac{w_k}{P_k^{\alpha} z_k^{1-\alpha}}$, followed by the workers in the H sector, $B_k^H = C(N_k) + S$. The workers in the L sector do not face additional barriers other than the congestion disutility, so that $B_k^L = C(N_k)$. Proposition 1 implies that within the same city, the individuals with higher x sort into industry-occupation cells with higher entry barrier, and thus the within-city assortative matching.

Part (ii) of the proposition characterizes the assortative matching across city. Cities with higher population also have higher $C(N_k)$, and therefore attract individuals with higher human capital. For example, workers in the H sector face the entry barrier of $C(N_k) + S$, and thus individuals with higher human capital must be working in a larger city as implied by Proposition 1. The same logic applies to the workers in the L sector. For the entrepreneurs, the sorting of the entry barrier is less obvious as we would need cities with higher N_k to have higher $\frac{w_k}{P_k^{\alpha} z_k^{1-\alpha}}$ as well to ensure the assortative matching. Comparing the workers in the H sectors across cities, the cities with higher $B_k^H = S + C(N_k)$ must also have higher return to human capital, $A_k^H = \frac{w_k}{P_k^{\alpha} z_k^{1-\alpha}}$, otherwise the choice with higher barriers and lower return will be empty in equilibrium. From the entrepreneurs' perspective, this implies that workers are more expensive, and thus the fixed costs of entry are higher in larger cities as well. This implies that $C(N_k) + f \frac{w_k}{P_k^{\alpha} z_k^{1-\alpha}}$ must also be higher in larger cities.[25]

The third part of the Proposition 1 states that all the individuals with human capital above a certain threshold will choose to be entrepreneurs in some city. The grouping of the entrepreneurs at the top of the human capital distribution is driven by the differences in $\phi^{\omega}(x)$ function: while human capital translates into efficiency labor linearly in the case of workers, it adopts a convex functional form for entrepreneurs, and thus ensuring the cut-off. Combining this result with part (ii) of the same proposition, we can further infer that the union of x of all the entrepreneurs, $\cup_{k=1}^{J} E_k$, is always a connected set on the real line above x_E. Without the loss of generality, we define the city that host the most talented entrepreneurs as city 1, followed by the next group of most-talented entrepreneurs in city 2, and so on and so forth: $N_1 > N_2 > N_3, \cdots > N_J$. The mechanism behind the sorting is the trade-off between the human capital spillovers and competitiveness, which is consistent with the findings in the literature such as Glaeser et al. (2005) and Behrens et al. (2014). On the one hand, any entrepreneur prefers to work in the same city with highly-talented entrepreneurs due to the benefits of knowledge spillover. On the other hand, cities populated with talented entrepreneurs are also highly competitive: wage rates are high and the ideal price index low, which makes the city hard for the less-talented ones to survive. In addition,

[25] This sorting across implies that individuals with higher human capital are more likely to live in large cities, and the average level of human capital is higher in larger cities. These two predictions are strongly supported in the data: in our sample, individuals with college and above degrees are 44 - 53 percent more likely to live in cities with above-median GDP and similarly, the individuals in larger-than-median cities on average have 5.3 percent more years of education.Both statements are based on the sample in Section 2. The first statement is the result of the individual-level Logit regressions while controlling for individual age, marital status, and race; the second statement is based on a city-level OLS regression that controls for state dummies. Table A.12 in the appendix provides the detailed specification and results. Note that despite the positive correlation between city size and average level of human capital, the model-consistent way of estimating equation 3 still require both variables on the right hand side. This is because in equilibrium, the relationship between the two variables are neither linear nor log-linear, as shown in appendix E in a two-city case. Excluding the average human capital in Equation 3 leads to the omitted variable bias.

cities with better entrepreneurs are also larger, and thus post higher congestion dis-utility.[26] In equilibrium, the trade-off between the benefits and costs of "better neighbors" pins down the sorting of entrepreneurs. In comparison, $\cup_{k=1}^{J} H_k$ and $\cup_{k=1}^{J} L_k$ might not be connected sets and this leads to the issue of multiple equilibrium, which we will discuss at the last part of the section.

The above predictions on assortative matching do not depend on the initial distribution of population over cities, as we do not have migration frictions in the model. In general, when migration is costly, the equilibrium sorting pattern will depend on the initial distribution and the migration costs. Nevertheless, conditional on a specific location-industry cell, our model still predicts that individuals with higher human capital endowment will tend to migrate to larger cities or higher-barrier occupations.

Another implication of assortative matching is the advantage of the large cities often documented in the literature (Combes et al., 2012a; Rosenthal and Strange, 2004; Davis and Dingel, 2012; Combes et al., 2008). In the next corollary, we show that these patterns also arise in our model due to assortative matching: cities with larger population also have higher real GDP, consumption-equivalent utility, and real wage rates in all industry-occupation cells.

Corollary 2. *In any sorting equilibrium, if $N_{k'} > N_k$, then:*

(i) The real wage of H workers is higher in city k': $\frac{w_{k'}}{P_{k'}^{\alpha} z_{k'}^{1-\alpha}} > \frac{w_k}{P_k^{\alpha} z_k^{1-\alpha}}$.

(ii) The real wage rates of workers in both sectors measured in term of H goods are higher in city k': $\frac{w_{k'}}{P_{k'}} > \frac{w_k}{P_k}$, *and* $\frac{z_{k'}}{P_{k'}} > \frac{z_k}{P_k}$.

(iii) The return to entrepreneurship is higher in k': $\frac{\pi_{k'}}{P_{k'}} > \frac{\pi_k}{P_k}$.

(iv) Real GDP in city k' is higher: $\frac{R_{k'}}{P_{k'}^{\alpha} z_{k'}^{1-\alpha}} > \frac{R_k}{P_k^{\alpha} z_k^{1-\alpha}}$.

4.3 Within-City Inequality

We have described the assortative matching patterns in the equilibrium in the above section. In this section, we proceed to characterize the key result of the model, the within-city

[26] Our results do not imply that larger cities have higher or lower concentration of entrepreneurs, either in terms of absolute numbers or as a share of the population. Behrens et al. (forthcoming) provide a more detailed discussion on the indeterminacy of entrepreneurship concentrations across cities.

inequality:

Proposition 2. *In any sorting equilibrium, if $N_{k'} > N_k$, then :*

(i) the average ratio of profit per unit of efficiency labor supply to wage is higher in city j:

$$\frac{1}{G(x \in E_{k'})} \int_{x \in E_{k'}} \left(\frac{\pi_{k'}(\psi(x))^{(\sigma-1)}/x}{w_{k'}} \right) dG(x) > \frac{1}{G(x \in E_k)} \int_{x \in E_k} \left(\frac{\pi_k(\psi(x))^{(\sigma-1)}/x}{w_k} \right) dG(x).$$

(ii) the wage ratio between H and L sectors is lower in city k': $\frac{w_k}{z_k} > \frac{w_{k'}}{z_{k'}}$.

Proposition 2 is the main theoretical result. It states that entrepreneur's wage to H worker's wage ratio, which is the counterpart of the top-to-median wage ratio in the data, is increasing in city size; and the wage inequality measured at the left-tail — the wage gap between workers in the H and L — is decreasing with city size.

The first part of the proposition states that the wage gap measured at the right-tail of the distribution — the gap between the entrepreneurs and workers in the H sectors — widens in larger cities. We measure the wage rate of the entrepreneurs as their total income, $\pi_k(\cdot)$, divided by the level of their human capital x, which is the direct counter-part of wage rate per efficiency labor supply for the workers, w_k. The entrepreneur's income equals the profit of the firm, which is proportional to the sales of the firm. In light of this, the above proposition is the immediate implication that the average firm size increases faster with respect to city size relatively to wage rate in our model. Researchers have documented that both the average wage rate and the firm size increase with city size. However, the elasticity of city-size against average firm size is estimated to be much higher than against the average wage rate. For example, the city-size elasticity of firm employment is found to be around 0.5 for entering firms (Glaeser and Kerr, 2009), and around 0.7 for all firms in the U.S. (Glaeser, 2007). In contrast, the city-size elasticity of wage rate or earnings is significantly lower: around 0.046 in the U.S. earning data[27], 0.05 in the French data (Combes et al., 2008), and 0.1 in Japanese data (Tabuchi and Yoshida, 2000)[28].

[27] Roback (1982) reports that the coefficient on population of 98 cities is around 0.16E-7, and the average population in her sample is 2,866,958. This implies that the average size elasticity is around 0.046.

[28] Both Glaeser and Mare (2001) and Baum-Snow and Pavan (2012) report the city-size premium using dummy variables instead of elasticities, and thus their results are not directly comparable with those reported

Our results on the right tail are also consistent with the findings in the literature that skill premium is higher in larger cities (Davis and Dingel, 2012, 2014; Behrens and Robert-Nicoud, 2015). In the literature, the spatial variations of skill premium usually stem from mechanisms such as knowledge exchange, spatial sorting of individuals, or the uneven distribution of amenities. In our context, the skill premium refers to the premium of entrepreneurial skills v.s. labor in the H sector. The spatial variations of the entrepreneurial skill premium are rooted in the fact that larger cities host larger firms in equilibrium as stated in Corollary 1. As a result, the return to entrepreneurs' human capital is positively correlated with firm size, whereas the return to workers' is independent of it. This further implies that in the equilibrium, the difference in the rate of return to human capital between entrepreneurs and workers is higher in larger cities.

Our model can also shed light on the relationship between skilled-biased technological changes and the widening within-city income gap over time. Baum-Snow and Pavan (2013) document that within-city inequality was relatively constant across cities in the 1980s and larger cities only started to be more unequal in the recent decades. Through the lens of our model, one potential explanation is the advancements in information technology that expand the span of control of high-skilled individuals, allowing them to create larger firms in equilibrium. In our model, the effects of skill-biased technological change can be embedded in the functional form of $b(\bar{x})$, the knowledge spillover effects. Over time, if the $b(\cdot)$ function becomes steeper due to technological changes, entrepreneurs with the same x will be able to create larger firms. This implies higher within-city inequality, especially on the right-tail in later years.

The second part of the proposition is a statement on the wage ratios between workers in different industries, the w/z ratio. It shows that in our model: the relative wage premium of working in the high-paying industry decreases with the size of the city. In other words, the residual wage rate in low-paying sector must increase with city size at a faster speed, and therefore the wage inequality measured at the left-tail of the distribution, such as the

in Glaeser (2007) and Glaeser and Kerr (2009). Baum-Snow and Pavan (2012) report that the wage premium between large (with more than 1.5 million population) and small (with smaller than 0.25 million population) is at most 0.29. This roughly translates into an upper bound of the elasticity as $(0.29/(1.5/0.25)) \approx 0.0483$, which is in line with the other estimates. Similar results can also be obtained in Glaeser and Mare (2001).

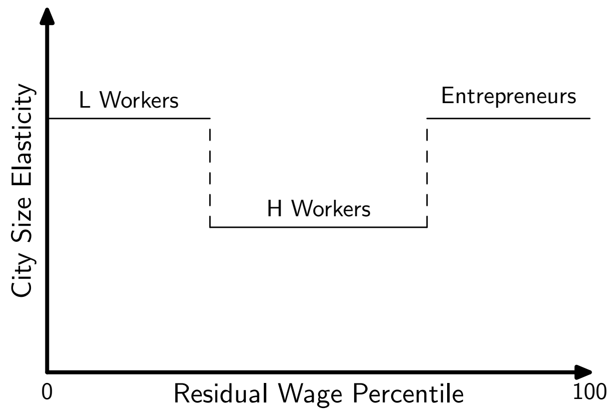


Figure 3: City Size Elasticity of Residual Wage in the Model

Note: this figure plots the implications of Proposition 2 and 2. We do not take a stand on whether the size elasticity of entrepreneurship is higher or lower than that of the workers in the L sector.

median-to-bottom wage ratios, shall be smaller in larger cities.

Two forces drive this result. The first is the existence of the fixed cost of entry into the H sector. Intuitively, when individuals are choosing between working in the H and L sector within a city, the choice boils down to the trade-off between the income premium of working in the H sector, $(w - z)x$, against the cost of entry, S. Since the average human capital is higher in larger cities, $w - z$ must be smaller in larger cities. Otherwise, the marginal worker between the L and H sector in large cities will always switch to the H sector, a situation that cannot arise in equilibrium. In other words, the market must compensate those working in L sectors relatively more in the large cities, otherwise, individuals will deviate from their equilibrium choice of location and industry. This channel is similar to the ideas in Roback (1982), in which the differences in housing market drive the pattern of wage inequality at the left tail in large cities. The second driving force z/P increases with the size of the city as stated in Proposition 2. This means the output price of L goods rises faster than that of the tradable H goods as the size of the city grows. The differences in output prices translate into the differences in factor prices, z/w, and as a result, w/z decreases with the size of the city.

The above two predictions are statements on the city-size elasticity of wage rates. Proposition 2 states that the city-size elasticity of wage rates are all positive, and Proposition 2 further predicts that wage rates differ in city-size elasticities. The city-size elasticity of entrepreneur compensation shall be higher than that of the workers' wage, and the city-size elasticity of wages in low-paying and non-tradable industries shall be higher than that of the wages in high-paying and tradable industries. We illustrate these predictions in Figure 3. The predicted pattern echoes the U-shaped pattern in Figure 1 in the introduction, and the predictions are also supported in data as shown in Section 2. These two propositions also imply that across cities, we shall expect the wage gap between the top earners in different cities to widen up, and the gap between median earners to narrow down, a finding similar to Giannone (2017).

Our results are based on the assumption that within a city, all the individuals face the same price index. In reality, this is not necessarily true: individuals with higher levels of education and income might choose more expensive housing and consumption bundles. Allowing for within-city variations of price index distorts our predictions on both tails asymmetrically. On the right-tail, it implies that the real wage inequality shall grow slower than what our model predicts, since the high-income individuals (entrepreneurs) might also face higher price index in larger cities, a finding similar to Moretti (2013). On the left tail, it implies that the real wage inequality between the median and the bottom earners might be wider than what our model predicts. Workers in the H sectors will not only experience slower growth in nominal wage rate but also suffer a higher price index as compared to the workers in the L sectors in large cities.

4.4 Multiple Sorting Patterns and Conditions for Uniqueness

In the previous parts, we have characterized the sorting and the within-city inequality in equilibrium. One remaining issue is the multiplicity of equilibria in our model. Models of economic geography often suffer the problem of multiple equilibria, and our model is no exception. In this section, we provide conditions under which a unique equilibrium exists.

The assortative matching pattern described in Proposition 1 cannot pin down a unique sorting pattern across all the (k, ω). Multiple sorting patterns could emerge among the workers because unlike the entrepreneurs, the union sets of the workers, $\cup_{k=1}^{J} H_k$ or $\cup_{k=1}^{J} L_k$, might not be a connected set on the real line. For example, two potential sorting patterns may arise in the case of two cities as illustrated in Figure 4. In the example, Proposition 1 indicates that E_1 and E_2 must occupy the right end of the human capital distribution. The proposition also implies that within the group of workers, the most talented workers must be in the H sector in large city, while the lest talented workers must be in the L sector in the small city, and thus pin down the set H_1 and L_2 on the real line. However, the relative positions between H_2 and L_1 cannot be determined. In the first panel, individuals first sort by industry, and then within each industry, they further sort into the cities. In this case, $\cup_{k=1}^{J} H_k$ and $\cup_{k=1}^{J} L_k$ are connected sets, and we call this sorting as **"industry-first" sorting**. Conversely, in the second panel, workers first sort into different cities, and then into sectors. In this case, $\cup_{k=1}^{J} H_k$ are $\cup_{k=1}^{J} L_k$ are no longer connected sets and we name this sorting as **"location-first" sorting**.

(a) Industry-First Sorting (b) Location-First Sorting

Figure 4: Multiple Sorting Patterns with Two Cities

Which sorting pattern will emerge in equilibrium depends on the barriers between industries, S, and cities, $C(N_k) - C(N_{k'})$. In the two-city case, industry-first sorting appears if $S > C(N_1) - C(N_2)$, and location-first sorting appears if the reverse is true. To see this, first assume that the barriers between industries are high relative to the barriers between cities. In this case, the marginal worker $x = \sup L_2$ shall be indifferent between moving into a larger city while staying in the L sector and his current position, rather than moving into the H sector in the same city. This implies that $\cup_{k=1}^{J} L_k$ is a connected set and industry-first sorting arises. On the other hand, If the barriers between cities are higher than S, the marginal worker x will be indifferent between his current positions and switching into the H sector in the same city, before choosing to move to a larger city. In this case, the sets $\{L_k\}$ and $\{H_k\}$ are interleaved, leading to location-first sorting. When $J > 2$, the sorting pattern depends on the relationship between S and all the possible combinations of $C(N_{k'}) - C(N_k)$. As $\{N_k\}$ is an endogenous object, the number of potential sorting patterns increases at the order of J-factorial, and it is impossible to push the results further without other assumptions.

To narrow down the potential types of sorting patterns, we need to reduce the potential combinations of $C(N_{k'}) - C(N_k)$ between any k and k'. In the next proposition, we show that if the population distribution across cities, N_k, belongs to a broad family of exponential-like distributions, then the sorting patterns and the equilibrium will be unique:

Proposition 3. *The asymmetric equilibrium is unique if the population distribution across cities is one of the following distributions: Pareto, exponential, Weibull, or Rayleigh*[29].

The common property we exploit to pin down the unique equilibrium from these family of distributions is the shrinking spacing between adjacent cities on the population ladder. It is well established in the statistics literature that if N_k follows one of the above distributions and $k' > k$, then $N_{k'} - N_{k'+1} < N_k - N_{k+1}$ (Kamps, 1991; Balakrishnan and Sultan, 1998). Together with the monotonicity and convexity of $C(\cdot)$, the shrinking spacing in population also implies that $C(N_k) - C(N_{k+1})$ is monotonically decreasing in k as well: the differences in congestion costs shrink as we move down the city size ladder. From the shrinking congestion costs, we can define a sequence of pivot cities, $k_1^*, k_2^*, \cdots$ such that

(1) $C(N_k) - C(N_{k+i}) > S$, for all $k < k_i^*$, and

(2) $C(N_k) - C(N_{k+i}) < S$, for all $k \geq k_i^*$.

Sorting patterns can be exactly characterized using the sequence, and we provide the details in the Appendix D.

The conditions on population distribution are broad enough to encompass both the empirical and the theoretical findings on city size distributions. In the data, the population distribution is often estimated to be Pareto (Gabaix and Ioannides, 2004); It also arises in an array of theoretical models such as Behrens et al. (2014) and Gaubert (2017). Our model allows for a wide range of potential distributions of population, depending on the distribution of human capital, $G(x)$. As our main predictions of within-city inequality is independent of the cross-city distribution of population, we refrain from imposing specific assumptions on

[29]We are not aware of any work that addresses the spacing of log-normal distributions. Unlike the other distributions in the above list, a simple expression does not exist because the CDF of log-normal distributions depends on Gauss error functions, which are not elementary. Balakrishnan and Chen (1999) provide numerical methods to compute the moments of log-normal order statistics, and Nadarajah (2008) discuss the explicit expression using hyper-geometric functions, which are not friendly to use analytically either. Nevertheless, all the simulation results indeed exhibit shrinking spacing among ordered samples from log-normal distributions, similar to the distributions listed above.

$G(x)$ in this paper, and leave the characterization of $G(x)$ and population distribution to future work.

5 Conclusion

We document a rich pattern of within-city inequality in the US: while top-to-median wage inequality tends to increase with city size, the median-to-bottom inequality decreases with city size instead. We then develop a spatial model to explain the pattern with individuals sorting along location, industry, and occupation. Our model delivers assortative matching between individuals and location-industry-occupation cells in equilibrium. It predicts that wage rate in low-paying industries needs to increase faster with city size than high-pay industries; otherwise, the workers in the low-paying industries cannot afford living in large cities. The validation test suggests that once the industry-city or occupation-city fixed effects are controlled for, larger cities no longer see lower inequality on the left-tail, and the relationship between city size and the inequality on the right-tail become muted. Once the spatial variations of inter-industry wage premium are removed, the within-city wage inequality has dropped by about half.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Autor, D. H. and D. Dorn (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103, 1553–97.

Bacolod, M., B. S. Blum, and W. C. Strange (2009): "Skills in the City," *Journal of Urban Economics*, 65, 136–153.

Balakrishnan, N. and W. Chen (1999): *Handbook of Tables for Order Statistics from Lognormal Distributions with Applications*, Springer.

Balakrishnan, N. and K. Sultan (1998): "Recurrence relations and identities for moments of order statistics," in *Order Statistics: Theory and Methods*, Elsevier, vol. 16 of *Handbook of Statistics*, 149 – 228.

Baum-Snow, N. and R. Pavan (2012): "Understanding the City Size Wage Gap," *The Review of Economic Studies*, 79, 88–127.

——— (2013): "Inequality and City Size," *The Review of Economics and Statistics*, 95, 1535–1548.

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

Behrens, K., D. Pokrovsky, and E. Zhelobodko (forthcoming): "Market size, occupational self-selection, sorting, and income inequality," *Journal of Regional Science*.

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

——— (2015): *Agglomeration Theory with Heterogeneous Agents*, Elsevier, vol. 5 of *Handbook of Regional and Urban Economics*, 171–245.

Berube, A. (2014): "All Cities Are Not Created Unequal," Tech. rep., Brookings Institution.

Berube, A. and N. Holmes (2015): "Some Cities Are Still More Unequal Than Others – an Update," Tech. rep., Brookings Institution.

Combes, P.-P., G. Duranton, and L. Gobillon (2008): "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, 63, 723–742.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012a): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection," *Econometrica*, 80, 2543–2594.

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux (2012b): "Sorting and local wage and skill distributions in France," *Regional Science and Urban Economics*, 42, 913 – 930, special issue In honor of Jacques Thisse.

Davis, D. R. and J. I. Dingel (2012): "A Spatial Knowledge Economy," NBER Working Papers 18188, National Bureau of Economic Research, Inc.

——— (2014): "The Comparative Advantage of Cities," NBER Working Papers 20602, National Bureau of Economic Research, Inc.

De la Roca, J. and D. Puga (2017): "Learning by Working in Big Cities," *Review of Economic Studies*, 84, 106–142.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014): "Spatial Sorting," *Journal of Political Economy*, 122, 554–620.

Gabaix, X. and Y. M. Ioannides (2004): "The Evolution of City Size Distributions," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson and J. F. Thisse, Elsevier, vol. 4 of *Handbook of Regional and Urban Economics*, chap. 53, 2341–2378.

Gabaix, X. and A. Landier (2008): “Why Has CEO Pay Increased So Much?” *The Quarterly Journal of Economics*, 123, 49–100.

Gaubert, C. (2017): “Firm Sorting and Agglomeration,” Mimeo.

Giannone, E. (2017): “Skilled-Biased Technical Change and Regional Convergence,” Mimeo.

Glaeser, E. L. (2007): “Entrepreneurship and the City,” NBER Working Papers 13551, National Bureau of Economic Research, Inc.

Glaeser, E. L., J. Gyourko, and R. E. Saks (2005): “Why Have Housing Prices Gone Up?” *American Economic Review*, 95, 329–333.

Glaeser, E. L. and W. R. Kerr (2009): “Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?” *Journal of Economics and Management Strategy*, 18, 623–663.

Glaeser, E. L. and D. C. Mare (2001): “Cities and Skills,” *Journal of Labor Economics*, 19, 316–42.

Glaeser, E. L., M. G. Resseger, and K. Tobio (2008): “Urban Inequality,” NBER Working Papers 14419, National Bureau of Economic Research, Inc.

——— (2009): “Inequality in Cities,” *Journal of Regional Science*, 49, 617–646.

Handbury, J. and D. E. Weinstein (2015): “Goods Prices and Availability in Cities,” *Review of Economic Studies*, 82, 258–296.

Heathcote, J., F. Perri, and G. L. Violante (2010): “Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006,” *Review of Economic Dynamics*, 13, 15–51.

Henderson, J. V. (1974): “The sizes and types of cities,” *The American Economic Review*, 640–656.

Hendricks, L. (2011): “The Skill Composition Of U.S. Cities,” *International Economic Review*, 52, 1–32.

Jargowsky, P. A. (2015): “Architecture of segregation: Civil unrest, the concentration of poverty, and public policy,” *The Century Foundation*, 7.

Kamps, U. (1991): “A general recurrence relation for moments of order statistics in a class of probability distributions and characterizations,” *Metrika*, 38, 215–225.

Katz, L. F., L. H. Summers, R. E. Hall, C. L. Schultze, and R. H. Topel (1989): “Industry Rents: Evidence and Implications,” *Brookings Papers on Economic Activity. Microeconomics*, 209–290.

Krueger, A. B. and L. H. Summers (1988): "Efficiency Wages and the Inter-industry Wage Structure," *Econometrica*, 56, 259–93.

Lucas, Jr, R. E. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–523.

Ma, L. and D. Ruzic (2015): "Globalization and Top Income Shares," Mimeo, National University of Singapore.

Moretti, E. (2013): "Real Wage Inequality," *American Economic Journal: Applied Economics*, 5, 65–103.

Murphy, K. J. (1999): "Executive Compensation," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter and D. Card, Elsevier, vol. 3, Part B, 2485 – 2563.

Nadarajah, S. (2008): "Explicit expressions for moments of order statistics," *Statistics & Probability Letters*, 78, 196 – 205.

Oi, W. Y. and T. L. Idson (1999): "Firm Size and Wages," in *Handbook of Labor Economics*, Elsevier, vol. 3, Part B, 2165 – 2214.

Piketty, T. and E. Saez (2003): "Income Inequality in the United States, 1913-1998," *The Quarterly Journal of Economics*, 118, 1–39.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

Roberts, D. R. (1956): "A General Theory of Executive Compensation Based on Statistically Tested Propositions," *The Quarterly Journal of Economics*, 70, 270–294.

Rosen, S. (1987): "The theory of equalizing differences," in *Handbook of Labor Economics*, ed. by A. O. and L. R., Elsevier, vol. 1, chap. 12, 641–692, 1 ed.

Rosenthal, S. S. and W. C. Strange (2004): "Evidence on the Nature and Sources of Agglomeration Economies," *Handbook of Regional and Urban Economics*, 4, 2119–2171.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010): "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Tech. rep., Minneapolis, MN: Minnesota Population Center [producer and distributor].

Schmitt, J. and D. Rosnick (2011): "The wage and employment impact of minimum-wage laws in three cities," *Center for Economic and Policy Research. Available at: http://www. cepr. net/documents/publications/min-wage-2011-03. pdf.*

Tabuchi, T. and A. Yoshida (2000): "Separating Urban Agglomeration Economies in Consumption and Production," *Journal of Urban Economics*, 48, 70 – 84.

Tombe, T. and X. Zhu (2015): "Trade, Migration and Productivity: A Quantitative Analysis of China," Mimeo.

Vaghul, K. and B. Zipperer (2016): “Historical State and Sub-State Minimum Wage Data,” *Washington Center for Equitable Growth.*

Wheeler, C. H. (2004): “On the Distributional Aspects of Urban Growth,” *Journal of Urban Economics*, 55, 371–397.

——— (2005): “Cities, Skills, and Inequality,” *Growth and Change*, 36, 329–353.

Online Appendix, Not for Publication

A Tables and Figures

Table A.1: City Size and Inequality: Alternative Measures of City Size

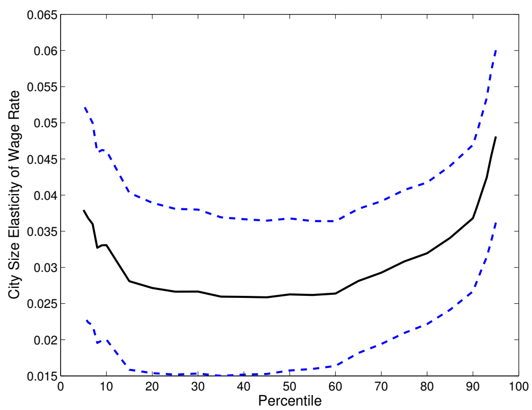
(a) City Size = Total GDP

LHS = Residual Wage Ineq.	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Total GDP)	-0.011**	-0.009*	0.028***	0.018***
	(0.005)	(0.005)	(0.005)	(0.004)
Average Years of Edu.	0.051***	0.048***	-0.004	-0.012
	(0.012)	(0.012)	(0.016)	(0.012)
Race == White	-0.418***	-0.399***	-0.258**	-0.217**
	(0.097)	(0.102)	(0.128)	(0.099)
N	254	254	254	254
R-squared	0.516	0.567	0.568	0.599
State FE	Yes	Yes	Yes	Yes

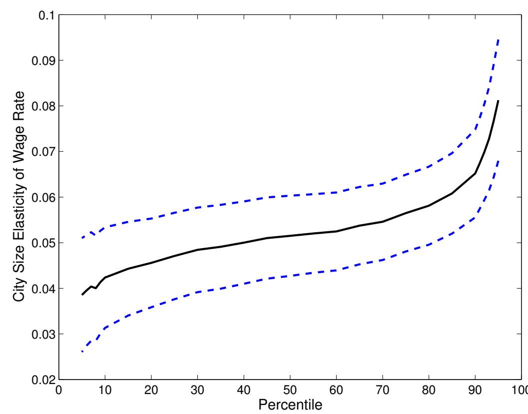
(b) City Size = Population

LHS = Residual Wage Ineq.	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Population)	-0.011**	-0.010*	0.022***	0.013*
	(0.005)	(0.005)	(0.007)	(0.007)
Average Years of Edu.	0.043***	0.040***	0.014	-0.001
	(0.010)	(0.010)	(0.018)	(0.015)
Race == White	-0.373***	-0.364***	-0.321**	-0.253**
	(0.098)	(0.099)	(0.153)	(0.122)
N	264	264	264	264
R-squared	0.557	0.587	0.543	0.556
State FE	Yes	Yes	Yes	Yes

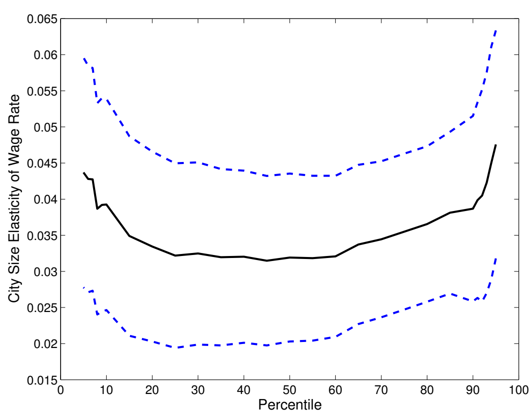
Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the results of estimating equation (3). The measures of inequality are various percentile ratios of hourly wage within an MSA. Population data come from U.S. 2000 census. For more details, see notes to Table 1.



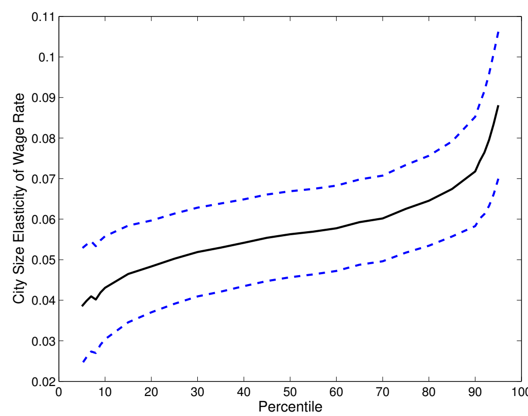
(a) Total GDP, with Control



(b) Total GDP, No Control



(c) Population, with Control



(d) Population, No Control

Figure A.1: City Size Elasticity of Wage at Different Percentiles, Robustness

Note: This figure reports the city size elasticity of wage rate (β_1) from Equation (3) with log(wage rates) at different percentiles within a city as dependent variable. The dashed line is 95 percent confidence interval. Data source: IPUMS-USA, 2000.

Table A.2: City Size and Inequality: alternative measures of residual wage

(a) City Size = Private GDP

LHS = Residual Wage Ineq. w/ FE Filtered	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Private Ind. GDP)	-0.014***	-0.011**	0.019***	0.007*
	(0.005)	(0.005)	(0.005)	(0.004)
Average Years of Edu.	0.041***	0.031**	0.018	0.015
	(0.013)	(0.014)	(0.013)	(0.012)
Race == White	-0.349***	-0.336***	-0.294***	-0.308***
	(0.109)	(0.104)	(0.105)	(0.091)
N	254	254	254	254
R-squared	0.466	0.535	0.560	0.544
State FE	Yes	Yes	Yes	Yes

(b) City Size = Total GDP

LHS = Residual Wage Ineq. w/ FE Filtered	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Total GDP)	-0.013**	-0.009*	0.021***	0.009**
	(0.005)	(0.005)	(0.005)	(0.004)
Average Years of Edu.	0.039***	0.029**	0.015	0.013
	(0.013)	(0.014)	(0.013)	(0.012)
Race == White	-0.346***	-0.327***	-0.272**	-0.295***
	(0.109)	(0.103)	(0.105)	(0.091)
N	254	254	254	254
R-squared	0.461	0.530	0.566	0.547
State FE	Yes	Yes	Yes	Yes

(c) City Size = Population

LHS = Residual Wage Ineq. w/ FE Filtered	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Population)	-0.015***	-0.012**	0.016**	0.005
	(0.005)	(0.005)	(0.006)	(0.006)
Average Years of Edu.	0.034***	0.023*	0.028**	0.018
	(0.012)	(0.013)	(0.014)	(0.013)
Race == White	-0.321***	-0.309***	-0.335***	-0.336***
	(0.109)	(0.109)	(0.121)	(0.110)
N	264	264	264	264
R-squared	0.497	0.555	0.560	0.533
State FE	Yes	Yes	Yes	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the results of estimating equation (3). The residual wage is computed from estimating equation 1 while controlling for industry and occupation fixed effects in addition. For more details, see notes to Table 1.

Table A.3: City Size and Inequality: Raw Wage

(a) City Size = Private GDP

LHS = Raw Wage Ineq.	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Private Ind. GDP)	-0.018***	-0.012**	0.047***	0.031***
	(0.007)	(0.005)	(0.008)	(0.005)
Average Years of Edu.	0.824***	0.688***	0.092	0.022
	(0.165)	(0.136)	(0.228)	(0.167)
Race == White	-0.495***	-0.386***	-0.236	-0.113
	(0.094)	(0.075)	(0.151)	(0.073)
N	254	254	254	254
R-squared	0.515	0.572	0.492	0.582
State FE	Yes	Yes	Yes	Yes

(b) City Size = Total GDP

LHS = Raw Wage Ineq.	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Total GDP)	-0.019***	-0.012**	0.050***	0.033***
	(0.007)	(0.005)	(0.008)	(0.005)
Average Years of Edu.	0.821***	0.687***	0.061	-0.003
	(0.168)	(0.138)	(0.230)	(0.166)
Race == White	-0.499***	-0.388***	-0.214	-0.097
	(0.094)	(0.075)	(0.150)	(0.073)
N	254	254	254	254
R-squared	0.513	0.571	0.494	0.585
State FE	Yes	Yes	Yes	Yes

(c) City Size = Population

LHS = Raw Wage Ineq.	Left Tail		Right Tail	
	50-05	50-10	95-50	90-50
Ln(Population)	-0.020***	-0.011**	0.036***	0.028***
	(0.007)	(0.005)	(0.011)	(0.007)
Average Years of Edu.	0.723***	0.591***	0.449	0.223
	(0.148)	(0.125)	(0.282)	(0.190)
Race == White	-0.453***	-0.341***	-0.329*	-0.165*
	(0.082)	(0.060)	(0.184)	(0.087)
N	264	264	264	264
R-squared	0.550	0.616	0.461	0.540
State FE	Yes	Yes	Yes	Yes

Note: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Huber-White robust standard errors in parentheses. This table reports the results of estimating equation (3) using observed wage instead of the residual wage. For more details, see notes to Table 1.

Table A.4: Population and Wage Inequality, Effects of Education and Racial Composition

	LHS = 50-10 Wage Gap				
	(1)	(2)	(3)	(4)	(5)
Ln(Population)	0.019***	0.008	0.003	0.002	-0.010*
	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)
Average Years of Edu.			0.026***		0.040***
			(0.009)		(0.010)
Share of White Population				-0.272**	-0.364***
				(0.108)	(0.099)
N	264	264	264	264	264
R-squared	0.036	0.529	0.541	0.556	0.587
State FE	No	Yes	Yes	Yes	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the results of estimating Equation (3) by progressively adding more controls. We only report the results with city size measured in the log of population for the sake of comparison with Baum-Snow and Pavan (2013). Data source: IPUMS-USA, 2000.

Table A.5: City Size and Inequality: Other Percentile Gaps

	Percentile Residual Wage Gap		
	95-05	90-10	75-25
Ln(Private Ind. GDP)	0.009	0.001	0.005
	(0.007)	(0.004)	(0.004)
Average Years of Edu.	0.049**	0.023**	0.011
	(0.019)	(0.011)	(0.011)
Race == White	-0.624***	-0.327***	-0.208**
	(0.160)	(0.094)	(0.096)
N	254	254	254
R-squared	0.627	0.660	0.632
State FE	Yes	Yes	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the results of estimating equation (3). The measures of inequality are various percentile ratios of hourly residual wage within an MSA. For more details, see notes to Table 1.

Table A.6: City Size and Inter-Industry Wage Premium, Robustness Check

(a) City Size = Population

LHS = Ln(Wage)	Level Effect	Continuous Measure		Binary Measure	
	(1)	(2)	(3)	(4)	(5)
Ln(City Size)	0.047***	0.049***		0.050***	
	(0.000)	(0.001)		(0.001)	
(Avg. Ind. Premium) × Ln(City Size)		-0.026***	-0.016***		
		(0.004)	(0.004)		
(High-Premium Ind. Dummy) × Ln(City Size)				-0.005***	-0.003***
				(0.001)	(0.001)
Age	0.013***	0.013***	0.013***	0.013***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Edu.	0.050***	0.050***	0.047***	0.050***	0.048***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.194***	0.194***	0.197***	0.194***	0.198***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Race == White	0.124***	0.124***	0.132***	0.124***	0.133***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,215,182	1,215,182	1,237,695	1,237,695
R-squared	0.420	0.421	0.431	0.420	0.431
Industry FE	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes
City FE	No	No	Yes	No	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the estimation of equation (5) with variations. The left-hand-side variable is the logarithm of observed wage and "city size" is measured as population. "Average Industry Premium" is a continuous measure computed from estimating equation (4), and "High-Premium Industry Dummy" is the associated binary measure in which the average premium is higher than the median. Data source: IPUMS-USA, 2000.

Table A.7: Robustness Check: Different Definition of "High-Premium Industries"

(a) City Size = GDP

LHS = Ln(Wage)	> 25%		> 75%	
	(1)	(2)	(3)	(4)
(High-Premium Ind. Dummy) × Ln(City Size)	-0.004***	-0.003***	-0.004***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,237,695	1,237,695	1,237,695
R-squared	0.422	0.430	0.422	0.430
Industry FE	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes
City FE	No	Yes	No	Yes

(b) City Size = Population

LHS = Ln(Wage)	> 25%		> 75%	
	(1)	(2)	(3)	(4)
(High-Premium Ind. Dummy) × Ln(City Size)	-0.005***	-0.003***	-0.007***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,237,695	1,237,695	1,237,695
R-squared	0.420	0.431	0.420	0.431
Industry FE	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes
City FE	No	Yes	No	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the estimation of Equation (5) with with binary measures of "high-premium industry". A high-paying industry is defined as those above the 25th or the 75th percentile among all the 201 industries. Table 3, the benchmark results, uses the 50th percentile as the cutoff. The left-hand-side variable is the logarithm of observed wage and "city size" is measured as private industry GDP or population. Data source: IPUMS-USA, 2000.

Table A.8: Robustness Check: the Effects of Specialization

(a) City Size = GDP

LHS = Ln(Wage)	≥ 50 Cities		≥ 100 Cities		≥ 200 Cities	
	(1)	(2)	(3)	(4)	(5)	(6)
(Avg. Ind. Premium)						
× Ln(City Size)	-0.012***		-0.012***		-0.005*	
	(0.003)		(0.003)		(0.003)	
(High-Premium Ind. Dummy)						
× Ln(City Size)		-0.003***		-0.003***		-0.002**
		(0.001)		(0.001)		(0.001)
N	1,214,781	1,237,294	1,208,792	1,231,305	1,062,965	1,085,478
R-squared	0.431	0.430	0.431	0.430	0.427	0.426
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

(b) City Size = Population

LHS = Ln(Wage)	≥ 50 Cities		≥ 100 Cities		≥ 200 Cities	
	(1)	(2)	(3)	(4)	(5)	(6)
(Avg. Ind. Premium)						
× Ln(City Size)	-0.016***		-0.016***		-0.010**	
	(0.004)		(0.004)		(0.004)	
(High-Premium Ind. Dummy)						
× Ln(City Size)		-0.003***		-0.003***		-0.002*
		(0.001)		(0.001)		(0.001)
N	1,214,781	1,237,294	1,209,590	1,232,103	1,097,275	1,119,788
R-squared	0.431	0.431	0.431	0.431	0.430	0.429
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the estimation of Equation (5) with all the fixed effects while dropping the industries that are concentrated in less than 50, 100, or 200 cities. Table 3, the benchmark results, uses all the industries. The left-hand-side variable is the logarithm of observed wage and “city size” is measured as private industry GDP or population. Data source: IPUMS-USA, 2000.

Table A.9: City Size and Tradable Industry Wage Premium, Robustness Check

LHS = Ln(Wage)	Benchmark Def.			Alternative Def.		
	(1)	(2)	(3)	(4)	(5)	(6)
Tradable Ind. Dummy	0.062***			0.072***		
	(0.001)			(0.001)		
Tradable Ind. Dummy × Ln(City Size)		-0.019***	-0.016***		-0.026***	-0.022***
		(0.001)	(0.001)		(0.001)	(0.001)
Ln(City Size)	0.049***	0.057***		0.049***	0.057***	
	(0.000)	(0.001)		(0.000)	(0.001)	
Age	0.014***	0.013***	0.013***	0.014***	0.013***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Edu.	0.055***	0.049***	0.047***	0.054***	0.049***	0.047***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.204***	0.194***	0.198***	0.203***	0.194***	0.198***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Race == White	0.128***	0.125***	0.133***	0.129***	0.124***	0.133***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695
R-squared	0.403	0.421	0.431	0.403	0.421	0.431
Industry FE	No	Yes	Yes	No	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	No	No	Yes	No	No	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. This table reports the estimation of equation (5) with variations. "City size" is measured as population. "Benchmark" definition of tradable industries include agriculture, manufacturing and construction industries in the 1990 industry classifications, and "alternative" definition classifies construction industries as non-tradable. Data source: IPUMS-USA, 2000.

Table A.10: Entrepreneurial Wage Premium and City Size, Robustness Check

LHS = Ln(Wage)	Benchmark Def.			Alternative Def.		
	(1)	(2)	(3)	(4)	(5)	(6)
Entrepreneur Dummy	0.327***	0.118***		0.620***	0.469***	
	(0.002)	(0.024)		(0.006)	(0.080)	
Entrepreneur Dummy × Ln(City Size)		0.014***	0.024***		0.010*	0.024***
		(0.002)	(0.002)		(0.005)	(0.005)
Age	0.014***	0.014***	0.013***	0.014***	0.014***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Edu.	0.068***	0.068***	0.047***	0.073***	0.073***	0.047***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.227***	0.227***	0.198***	0.236***	0.236***	0.198***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Race == White	0.170***	0.169***	0.132***	0.179***	0.179***	0.133***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695	1,237,695
R-squared	0.391	0.391	0.431	0.384	0.384	0.431
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	No	Yes	No	No	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: * p<0.10, ** p<0.05, *** p<0.01. Huber-White robust standard errors in parentheses. Results of estimating Equation (6). "City size" is measured as population. "Benchmark Def." includes all the occupations listed in Table A.11 as entrepreneurs, while "Alternative Def." only includes the "Chief executives and public administrators" (the 3rd row). Data source: IPUMS-USA, 2000.

Table A.11: Definition of Entrepreneurs, U.S.

Occupation, 1990 basis	No.	%
Managers and administrators, n.e.c.	69,863	58.54%
Managers and specialists in marketing, advertising, and public relations	18,669	15.64%
Chief executives and public administrators	15,482	12.97%
Financial managers	11,389	9.54%
Human resources and labor relations managers	3,939	3.30%
Total	119,342	100.00%

Note: This table reports the definition of entrepreneurs used in the estimation of Equation (6). In the benchmark regression, all the individuals with occupations listed above are defined as entrepreneurs. In the robustness checks with stricter definition of entrepreneurs, only those whose occupation is "Chief executives and public administrators" are defined as entrepreneurs. Data source: IPUMS-USA, 2000. Occupation definition follows the 1990 census standard.

Table A.12: Sorting of High-Skilled Individuals across Cities

	Pr(Large City), Logit		Pr(Entrepreneurs), Logit		Ln(Avg. Years of Edu.), OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
College and Above	1.443***	1.530***	4.454***	3.726***		
	(0.005)	(0.006)	(0.033)	(0.032)		
College and Above × Large City				1.366***		
				(0.010)		
Large City Dummy					0.042***	0.053***
					(0.007)	(0.010)
Age		1.002***	1.021***	1.021***		
		(0.000)	(0.000)	(0.000)		
Married		0.919***	1.723***	1.737***		
		(0.004)	(0.013)	(0.013)		
Race == White		0.553***	1.988***	2.052***		
		(0.003)	(0.020)	(0.020)		
Constant	0.863***	1.327***	0.007***	0.007***		
	(0.002)	(0.010)	(0.000)	(0.000)		
N	1,237,695	1,237,695	1,237,695	1,237,695	254	254
R-squared					0.068	0.342
State FE	N/A	N/A	N/A	N/A	No	Yes

Note: The first two columns report logit regressions in which we study the likelihood of highly-educated individuals living in larger cities. The odds ratios are reported in column 1 and 2. Column 3 and 4 report the OLS regression of average years of education on a large city dummy. "Large cities" refer to cities with higher-than-median private industry GDP. Point estimates are reported in Column 3 and 4.

B Data Description

The individual-level data come from IPUMS in 2000. The original dataset contains 6.44 million individuals. We restrict the sample to working population younger than 65. We also drop the individuals working in government, military, religious organizations and labor unions, or self-employed. We compute the hourly wage as total weekly wage income divided by usual hours worked in a week, and then drop the individuals whose hourly wage is smaller than the federal minimum wage of 7.5 dollars. These restrictions leave us with a sample of 1.23 million individuals to compute various measures of inequality. We define wage income as "total wage and salary"(variable name *incwage*) in the dataset.

We define the city to which an individual belongs as the metropolitan area in which the individual works (*pwmetro*). Under this definition, if an individual works in Detroit (*MSAcode 2160*) but lives in Ann Arbor (*MSAcode 440*), we treat this individual as a member of the city of Detroit. The sectoral level GDP for each metropolitan area come from the Bureau of Economic Analysis's(BEA) Regional GDP database in 2001. We match the metropolitan area in IPUMS and in the BEA dataset by name. The matched dataset contains 254 metropolitan areas. We use two definitions of GDP as measures of the economic size of a metropolitan area: total GDP and private industry GDP. The difference between the two is the government spending.

In our estimations reported in Section 2, we control for the state dummy variables and the racial compositions of each city. In most of the cases, a metropolitan area is located within a single state. In cases where a metropolitan area spans over two or three states, we use the state in which the majority of the population of the MSA is located in. Our results are robust if we randomize the state assignments. We control for the racial composition of each metropolitan area by the share of the white population, which is recorded in the dataset as *race* equals to 1.

C Congestion Disutility

In this appendix, we provide a simple framework to micro-found the assumption that part of individual's utility is decreasing with the size of the city due to congestion. Our baseline model assumes that the individual's utility function in city j is

$$V = y_j - C(R_j),$$

where y_j is the utility derived from consumption, and $C(R_j)$ is the congestion disutility.

Assume that instead of a reduced-form congestion dis-utility, individuals instead need to consume both the :

$$V = y_j + G_j,$$

where y denotes individual consumption of the tradable and the non-tradable goods other than housing, which is the same as in the main model. The consumption of y depends on the human capital and the income of the individual.

The utility also depends on the consumption of public goods, g. We assume that the supply of the public goods, $G(N)$, depends positively on the size of the city ($G'(N) \geq 0$).

This is to capture the idea that more populous cities have a higher budget to provide the infrastructure and other public services. We also assume that the public good is divisible, and all the residents equally share the consumption of the public goods. This implies that larger population crowd-out the consumption of the public good. In the equilibrium the following holds

$$g = \frac{G(N)}{N}.$$

The congestion disutility in the main model can in turn be expressed as

$$C(N) = -g = -\frac{G(N)}{N}.$$

We further assume that $G(N) - G'(N)N > 0$ so that $C'(N) > 0$ and also $G'(N) - G(N)/N > 1/2G''(N)N$ so that $C''(N) > 0$ as well.

D Proofs

D.1 Lemma 1

Proof. We separately discuss three potential cases:

1. If $\omega, \omega' \in \{H, L\}$, the single crossing condition is straightforward. This is because $V_k^H(x)$ and $V_k^L(x)$ functions are linear, and thus they can only intersect once at most on the entire real line. Therefore there exist at most one $x \in \mathbb{R}^+$ such that $V_k^\omega(x) = V_{k'}^{\omega'}(x)$, if $\omega, \omega' \in \{H, L\}$.

2. If $\omega = \omega' = E$, then the functions $V_k^E(x)$ and $V_{k'}^E(x)$ are both exponential functions. Define the difference between the two as

$$g(x) = V_{k'}^E(x) - V_k^E(x) = (A_{k'}^E - A_k^E)(\psi(x))^{\sigma-1} - (B_{k'}^E - B_k^E).$$

   The $g(x)$ function is monotonic due to the monotonicity of $(\psi(x))^{\sigma-1}$, and therefore there is exists only one point at which $g(x) = 0$. By the same logic, there exists at most one $x \in \mathbb{R}^+$ such that $V_k^E(x) = V_{k'}^E(x)$.

3. In the last case, we consider when $\omega \in \{H, L\}$ and $\omega' = E$. Again, define the difference as

$$g(x) = V_{k'}^E(x) - V_k^\omega(x) = A_{k'}^E (\psi(x))^{\sigma-1} - A_k^\omega x - (B_{k'}^E - B_k^\omega).$$

   This function is not necessarily monotonic. However, note that the second derivative of $g(x)$ is

$$g''(x) = A_k^E(\sigma - 1)\left[\psi(x)^{\sigma-3}\psi'(x) + \psi''(x)\psi(x)^{\sigma-2}\right]$$

   which is everywhere positive due to the monotonicity and the convexity of $\psi(x)$. This implies that the first derivative, $g'(x)$, is a monotonic function so there exists only one

x' such that $g'(x') = 0$. By Rolle's theorem, a unique solution of $g'(x)$ implies that at most there are two solutions of the $g(x)$ function on the entire real line.

Note that we have assumed that f is large enough so that the individual $x = 0$ cannot be entrepreneurs. This means that $g(x) < 0, \forall (k', k, \omega)$. Also note that:

$$\lim_{x \to (-\infty)} g(x) = -A_k^\omega x - (B_{k'}^E - B_{k'}^\omega) = +\infty.$$

As the $g(x)$ function is continuous, it implies that there exists at least one solution for x between 0 and $-\infty$. As we have already shown that $g(x)$ can obtain at most two solutions on the entire real line, we can infer that there exists at most one solution of x between 0 and $+\infty$ such that $g(x) = 0$.

We have shown that in all cases, there exists at most one solution of x. Also note that we cannot have a case in which there exists no solution on $\{R\}^+$, as it would imply either $V_k^\omega(x) > V_{k'}^{\omega'}(x)$ or $V_k^\omega(x) < V_{k'}^{\omega'}(x)$. In either case, one of the sets ω_k or $\omega'_{k'}$, will be empty. Therefore we can conclude that there exists exactly one $x \in \mathbb{R}^+$ such that $V_k^\omega(x) = V_{k'}^{\omega'}(x)$. ∎

D.2 Proposition 1

Proof. Define the difference between the choices as:

$$g(x) = V_{k'}^{\omega'}(x) - V_k^\omega(x).$$

We first prove that if $x' \geq \sup \omega_k$, then $B_{k'}^{\omega'} > B_k^\omega$. As x' prefers the choice (k', ω'), it must be the case that $g(x') > 0$. We also know that $\forall x \in w_k$, $g(x) < 0$ as they prefer (k, ω) to (k', ω') by revealed preference. As Lemma 1 shows, there only exists one x^* such that $g(x^*) = 0$, we can infer that $\sup \omega_k \leq x^* \leq x'$. From this we can infer $g(0) < 0$ and $\lim_{x \to \infty} g(x) > 0$.

We now separately discuss the cases:

1. If the choice $\omega, \omega' \in \{H, L\}$, then $g(0) = -(B_{k'}^{\omega'} - B_k^\omega) < 0$ directly implies $B_{k'}^{\omega'} > B_k^\omega$, and thus the result.

2. If $\omega = \omega' = E$, then $\lim_{x \to \infty} g(x) = \lim_{x \to \infty} (A_{k'}^{\omega'} - A_k^\omega)(\psi(x))^{\sigma - 1}$. This can only be positive when $A_{k'}^{\omega'} > A_k^\omega$. Moreover, if $A_{k'}^{\omega'} > A_k^\omega$, then we must have $B_{k'}^{\omega'} > B_k^\omega$ as well. To see this, assume, $B_{k'}^{\omega'} < B_k^\omega$. In this case

$$g(x) = (A_{k'}^{\omega'} - A_k^\omega)(\psi(x))^{\sigma-1} - (B_{k'}^{\omega'} - B_k^\omega) > (A_{k'}^{\omega'} - A_k^\omega)(\psi(x))^{\sigma-1} > 0$$

for all x, which contradicts with $g(x) < 0, \forall x \in w^k$.

3. If $\omega' = E$ and $\omega \in \{H, L\}$, then $g(0) = A_{k'}^{\omega'} - (B_{k'}^{\omega'} - B_k^\omega)$. To have $g(0) < 0$, again we must have $-(B_{k'}^{\omega'} - B_k^\omega) < 0$ because $A_{k'}^{\omega'} > 0$. Therefore $B_{k'}^{\omega'} > B_k^\omega$ in this case as well.

4. Lastly, the case of $\omega = E$ and $\omega' \in \{H, L\}$. This case cannot happen because:

$$\lim_{x\to\infty} g(x) = \lim_{x\to\infty}\left[A_{k'}^{\omega'}x - A_k^E\left(\psi(x)\right)^{(\sigma-1)}\right]$$
$$= \lim_{x\to\infty}\left[A_{k'}^{\omega'} - (\sigma-1)A_k^E\left(\psi(x)\right)^{(\sigma-2)}\psi(x)'\right] = -\infty,$$

according to L'Hopital's rule and the assumptions that $\psi(x)' > 0$ and $\lim_{x\to\infty}\psi(x) = \infty$. This contradicts with $\lim_{x\to\infty} g(x) > 0$, and therefore cannot happen.

In all four cases, we have $B_{k'}^{\omega'} > B_k^{\omega}$, and therefore the result.

Next, we prove if $B_{k'}^{\omega'} > B_k^{\omega}$, then $x' > \sup \omega_k$. Let $x' \in \omega'_{k'}$ and $x \in \omega_k$. We know $g(x') > 0$ and $g(x) < 0$ by revealed preference. We also know that from Lemma 1 that there only exists one x^* such that $g(x^*) = 0$. To show $\inf \omega'_{k'} > \sup \omega_k$, we need to show either $g(0) < 0$ or $\lim_{x\to\infty} g(x) > 0$. Together with the single-crossing condition in Lemma 1, we can show that if any of the above inequality is true, then $\forall x \in [0, x^*), g(x) < 0$, and $\forall x \in (x^*, \infty), g(x) > 0$. This implies $x < x^*$ for all $x \in \omega_k$ and $x' > x^*$, for all $x' \in \omega'_{k'}$ and in turn, $x' > x^* > \sup \omega_k$.

Now we discuss again, case by case.

1. $\omega, \omega' \in \{H, L\}$. $g(0) = -(B_{k'}^{\omega'} - B_k^{\omega}) < 0$.

2. $\omega = \omega' = E$. $\lim_{x\infty} g(x) = \lim_{x\infty}(A_{k'}^{\omega'} - A_k^{\omega})\left(\psi(x)\right)^{\sigma-1}$. Using the same argument, we can infer $A_{k'}^{\omega'} > A_k^{\omega}$, otherwise $g(x) < 0, \forall x \in \omega'_{k'}$, which is a contradiction. Furthermore, $A_{k'}^{\omega'} > A_k^{\omega}$ implies $\lim_{x\infty} g(x) = \infty > 0$.

3. $\omega' = E$ and $\omega \in \{H, L\}$. In this case, our assumptions on $f > 0$ and $\psi(0) = 0$ imply $g(0) < 0$.

4. $\omega = E$ and $\omega' \in \{H, L\}$. This case cannot happen as shown above, $\lim_{x\to\infty} g(x) = -\infty$ regardless of the assumption on B_k^{ω} and $B_{k'}^{\omega'}$. By assumption we also have $g(0) < 0$. This means either $g(x) < 0$ for all x, or the single-crossing condition is violated. Neither can happen in the equilibrium.

As in all cases we have either $g(0) < 0$ or $\lim_{x\to\infty} g(x) > 0$, we have proved that $B_{k'}^{\omega'} > B_k^{\omega}$ implies $x' > \sup \omega_k$ for all $x' \in \omega'_{k'}$. ∎

D.3 Corollary 1

Proof. (i) Within any city k, we have:

$$B_k^E = C(N_k) + S + f\frac{w_k}{P_k^{\alpha} z_k^{1-\alpha}}$$
$$\geq B_k^H = C(N_k) + S$$
$$\geq B_k^L = C(N_k).$$

Proposition 1 then directly applies that $\inf E_k > \sup H_k$ and $\inf H_k > \sup L_k$.

(ii) First note that $N_{k'} > N_k$ implies $C(N_{k'}) > C(N_k)$ due to the monotonicity of $C(\cdot)$. To see the matching across cities, again we directly apply Proposition 1 by ranking the entry barriers. We discuss the case by case:

(a) $\omega = H$. We have $B^H_{k'} = C(N_{k'}) + S > C(N_k) + S = B^H_k$, and therefore $\inf H_{k'} > \sup H_k$.

(b) $\omega = L$. We have $B^L_{k'} = C(N_{k'}) > C(N_k) = B^L_k$, and therefore $\inf L_{k'} > \sup L_k$.

(c) $\omega = E$. Note that $B^H_{k'} > B^H_k$ implies $A^H_{k'} = \frac{w_{k'}}{P^\alpha_{k'} z^{1-\alpha}_{k'}} > \frac{w_k}{P^\alpha_k z^{1-\alpha}_k} = A^H_k$. If the inequality is not true, then

$$V^H_{k'} - V^H_k = (A^H_{k'} - A^H_k)x - (B^H_{k'} - B^H_k) < 0$$

for all x, which cannot happen in equilibrium. $\frac{w_{k'}}{P^\alpha_{k'} z^{1-\alpha}_{k'}} > \frac{w_k}{P^\alpha_k z^{1-\alpha}_k}$ implies:

$$\begin{aligned} B^E_{k'} &= C(N_{k'}) + S + f\frac{w_{k'}}{P^\alpha_{k'} z^{1-\alpha}_{k'}} \\ &\geq C(N_k) + S + f\frac{w_k}{P^\alpha_k z^{1-\alpha}_k} = B^E_k, \end{aligned}$$

which in turn implies $\inf E_{k'} > \sup E_k$.

(iii) Assume $\omega \in \{H, L\}$, and define:

$$g(x) = V^E_k(x) - V^\omega_{k'}(x)$$

as the premium of entrepreneurship in city k over all the other non-entrepreneurship jobs in any other city k'. Note that by L'Hopital's rule:

$$\lim_{x\to\infty} g(x) = \lim_{x\to\infty} \left[A^E_k (\psi(x))^{\sigma-1} - A^\omega_{k'} x\right] = \lim_{x\to\infty} (\sigma-1)A^E_k (\psi(x))^{\sigma-2} \psi(x)' = \infty > 0.$$

This implies that as x grows larger, entrepreneurship is always more attractive due to the convexity of the $\psi(x)$ function. Applying the single-crossing property in Lemma 1, we know that for all $(k', \omega), \omega \in \{H, L\}$ and (k, E), there exists a unique $x_{(k,E),(k',\omega)}$ such that $g(x_{(k,E),(k',\omega)}) = 0$ and $\inf E_k > x_{(k,E),(k',\omega)} > \sup \omega_{k'}$. This further implies that

$$\inf E_k > \sup(\cup^J_{k'=1} \cup_{\omega=\{H,L\}} \omega_{k'}).$$

As the inequality applies to all k, we can conclude:

$$\inf \cup^J_{k=1} E_k > \sup(\cup^J_{k'=1} \cup_{\omega=\{H,L\}} \omega_{k'}).$$

Now define

$$x_E = \inf \cup^J_{k=1} E_k.$$

If $x > x_E$, then $x > \sup \cup^J_{k'=1} \cup_{\omega=\{H,L\}}$, which in turn implies that $x \notin \cup^J_{k'=1} \cup_{\omega=\{H,L\}}$. We can then infer $x \in E_k$ for some k. The converse if also true. If $x \in \cup^J_{k=1} E_k$, then trivially $x > \inf \cup^J_{k=1} E_k = x_E$. ∎

D.4 Corollary 2

Proof. Note that if $N_{k'} > N_k$, then $C(N_{k'}) > C(N_k)$ by the monotonicity of the $C(\cdot)$ function.

(i) $C(N_{k'}) > C(N_k)$ implies $B^H_{k'} > B^H_k$. In this case, we must have $A^H_{k'} = \frac{w_{k'}}{P^{\alpha}_{k'} z^{1-\alpha}_{k'}} > \frac{w_k}{P^{\alpha}_k z^{1-\alpha}_k} = A^H_k$. If the inequality is not true, then

$$V^H_{k'} - V^H_k = (A^H_{k'} - A^H_k)x - (B^H_{k'} - B^H_k) < 0$$

for all x, which cannot happen in equilibrium.

(ii) Apply the same logic above to the L sector, we know $A^L_{k'} = \frac{z_{k'}}{P^{\alpha}_{k'} z^{1-\alpha}_{k'}} > \frac{z_k}{P^{\alpha}_k z^{1-\alpha}_k}$. This expression simplifies to $\frac{z_{k'}}{P_{k'}} > \frac{z_k}{P_k}$. Conditional on this, the result in part (i) then simplifies to $\frac{w_{k'}}{P_{k'}} > \frac{w_k}{P_k}$.

(iii) Apply the same logic to the entrepreneurs, we also have $A^E_{k'} = \frac{\pi_{k'}}{P^{\alpha}_{k'} z^{1-\alpha}_{k'}} > A^E_k = \frac{\pi_k}{P^{\alpha}_k z^{1-\alpha}_k}$. Plug in the result that $\frac{z_{k'}}{P_{k'}} > \frac{z_k}{P_k}$, we get $\frac{\pi_{k'}}{P_{k'}} > \frac{\pi_k}{P_k}$.

(iv) GDP in city j can be written as the total income earned by entrepreneurs, together with workers from H and L sector:

$$R_j = \int_{x\in E_j} \sum_k \alpha R_k P_k^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \frac{\tau_{kj} w_j}{b_j \psi(x)} \right)^{1-\sigma} dG(x) + z_j \int_{x\in N_j} x^{\gamma} dG(x).$$

Since $z_j \int_{x\in N_j} x^{\gamma} dG(x) = (1-\alpha) R_j$, the above is equivalent to:

$$\alpha R_j = \int_{x\in E_j} \sum_k \alpha R_k P_k^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \frac{\tau_{kj} w_j}{b_j \psi(x)} \right)^{1-\sigma} dG(x).$$

Profit for entrepreneur of x in city j is given as:

$$\pi_j(x) = \frac{1}{\sigma} \sum_k \alpha R_k P_k^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \frac{\tau_{kj} w_j}{b_j \psi(x)} \right)^{1-\sigma} - f w_j.$$

Therefore, the real GDP in city j can be written as:

$$\frac{\alpha R_j}{P_j} = \int_{x\in E_j} \frac{(\pi_j(x) + f w_j)\sigma}{P_j} dG(x).$$

Given results in (ii) and (iii), we have both $\pi_j(x)/P_j > \pi_i(x)/P_i$ and $w_j(x)/P_j > w_i(x)/P_i$ hold. This completes the proof. ∎

D.5 Proposition 2

Proof. (i) Define x_k and $x_{k'}$ to be the cutoff where individual is indifferent between becoming a worker in H and L sector of city k and k', respectively:

$$V_k^H(x_k) = V_k^L(x_k)$$
$$V_{k'}^H(x_{k'}) = V_{k'}^L(x_{k'})$$

We first show that if $N_{k'} > N_k$, then $x_{k'} > x_k$.

The ranking of the A terms can be one of the two possibilities:

$$A_{k'}^H > A_{k'}^L > A_k^H > A_k^L$$

or

$$A_{k'}^H > A_k^H > A_{k'}^L > A_k^L$$

We first focus on the first case. Define x^* as the solution to $V_{k'}^L(x^*) = V_{k'}^L(x^*)$, the individual indifferent between working in the L sector in the two cities. First note we must have $x_{k'} > x^*$. If this is not the case and $x_{k'} < x^*$, then all the $x < x^*$ prefer L_k to $L_{k'}$; all the $x > x^* > x_{k'}$ prefer $H_{k'}$ to $L_{k'}$, and thus no one works in $L_{k'}$. This implies $x_{k'} > x^*$. We must also have $x_k < x^*$, otherwise all the $x < x_k$ prefer L_k to H_k; all the $x > x_k > x^*$ prefer $L_{k'}$ to H_k because $A_{k'}^L > A_k^H$, and thus no worker would be employed in H_k. This proves that $x_{k'} > x_k$. A similar argument carries through in the second case.

We can then solve $x_k, x_{k'}$ as follows:

$$x_k = \frac{S}{\left(\frac{\alpha}{P_k}\right)^{\alpha}\left(\frac{1-\alpha}{z_k}\right)^{1-\alpha}(w_k - z_k)}$$
$$x_{k'} = \frac{S}{\left(\frac{\alpha}{P_{k'}}\right)^{\alpha}\left(\frac{1-\alpha}{z_{k'}}\right)^{1-\alpha}(w_{k'} - z_{k'})}$$

$x_{k'} > x_k$ implies:

$$\left(\frac{z_{k'}}{P_{k'}}\right)^{\alpha}\frac{w_{k'} - z_{k'}}{z_{k'}} < \left(\frac{z_k}{P_k}\right)^{\alpha}\frac{w_k - z_k}{z_k}$$

From Proposition 2, we know $\frac{z_{k'}}{P_{k'}} > \frac{z_k}{P_k}$, so the above inequality reduces to:

$$\frac{w_k - z_k}{z_k} > \frac{w_{k'} - z_{k'}}{z_{k'}}$$
$$\frac{w_k}{z_k} > \frac{w_{k'}}{z_{k'}}.$$

This establishes the proof.

(ii) First note that within any city, $\psi(x)^{\sigma-1}/x$ is increasing in x as long as $\sigma > 2$:

$$\begin{aligned}\frac{\partial\left(\psi(x)^{(\sigma-1)}/x\right)}{\partial x} &= \psi(x)^{(\sigma-2)}\frac{1}{x}\left[(\sigma-1)\,\psi'(x)-\frac{\psi(x)}{x}\right],\\ &= \psi(x)^{(\sigma-2)}\frac{1}{x}\left[(\sigma-1)\,\psi'(x)-\psi'(\zeta)\right],\ \zeta\in(0,x)\,,\end{aligned}$$

where we apply the mean value theorem for the last step. Since $\psi''(x) > 0$, so the above is positive if $\sigma > 2$. $\pi_k\psi(x)^{(\sigma-1)}/x$ is increasing in x implies that $\frac{\pi_k\psi(x)^{\sigma-1}/x}{w_k}$ increases with x as well. Therefore in any city k, the minimum of $\pi_k\psi(x)^{\sigma-1}/x/(w_k x)$ should be equal $\pi_k\psi(x_k)^{\sigma-1}/x/(w_k x_k)$, where x_k denotes the efficiency labor supply that the least talented entrepreneur has.

Second, we show

$$\frac{\pi_{k'}\psi(x_{k'})^{\sigma-1}}{(w_{k'}x_{k'})} > \frac{\pi_k\psi(x_k)^{\sigma-1}}{(w_k x_k)} \tag{10}$$

Note that we can express profit and real GDP in city j as follows:

$$\begin{aligned}\pi_k\psi(x_k)^{\sigma-1} &= \frac{1}{\sigma}\sum_{j=1}^{J}\alpha R_j P_j^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\frac{\tau_{jk}w_k}{b_k}\right)^{1-\sigma}\psi(x_k)^{\sigma-1}-fw_k\\ R_k &= \int_{x\in E_k}\sum_{j=1}^{J}\alpha R_j P_j^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\frac{\tau_{jk}w_k}{b_k\psi(x_k)}\right)^{1-\sigma}dG\left(x\right)\end{aligned}$$

Therefore, we have

$$\begin{aligned}\frac{\pi_k\psi(x_k)^{\sigma-1}+fw_k}{R_k} &= \frac{1}{\sigma}\frac{(b_k\psi(x_k))^{\sigma-1}}{\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}\,dG\left(x\right)}\\ \frac{\pi_k\psi(x_k)^{\sigma-1}}{P_k} &= \frac{R_k}{P_k}\frac{1}{\sigma}\frac{(b_k\psi(x_k))^{\sigma-1}}{\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}\,dG\left(x\right)}-\frac{fw_k}{P_k}\end{aligned}$$

Moreover, given the definition of price index, we can write real wage $\frac{w_k}{P_k}$ into the following:

$$\frac{w_k}{P_k} = \left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}\,dG\left(x\right)\right)^{\frac{1}{\sigma-1}}\left(\frac{\sigma}{\sigma-1}\right)^{-1}$$

Combine the two parts:

$$\frac{\pi_k\psi(x_k)^{\sigma-1}}{w_k} = \frac{\frac{\pi_k\psi(x_k)^{\sigma-1}}{P_k}}{\frac{w_k}{P_k}} = \frac{\frac{R_k}{P_k}\frac{1}{\sigma}\frac{(b_k\psi(x_k))^{\sigma-1}}{\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)}}{\left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}\,dG\left(x\right)\right)^{\frac{1}{\sigma-1}}\left(\frac{\sigma}{\sigma-1}\right)^{-1}}-f$$

Since $R_{k'}/P_{k'} > R_k/P_k$ as shown in Corollary 2, we only need to show

$$(b_{k'}\psi(x_{k'}))^{\sigma-1}\left(\int_{x\in E_{k'}}(b_{k'}\psi(x))^{\sigma-1}\,dG\left(x\right)\right)^{\frac{\sigma}{1-\sigma}} > (b_k\psi(x_k))^{\sigma-1}\left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}\,dG\left(x\right)\right)^{\frac{\sigma}{1-\sigma}} \tag{11}$$

to have $\frac{\pi_{k'}\psi(x_{k'})^{\sigma-1}}{w_{k'}} > \frac{\pi_k\psi(x_k)^{\sigma-1}}{w_k}$.

Differentiate $(b_k\psi(x_k))^{\sigma-1}\left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)\right)^{\frac{\sigma}{1-\sigma}}$ with respect to x_k:

$$\begin{aligned}
& (\sigma-1)(b_k\psi(x_k))^{\sigma-2}b_k\psi'(x_k)\left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)\right)^{\frac{\sigma}{1-\sigma}} \\
& -(b_k\psi(x_k))^{\sigma-1}\frac{\sigma}{1-\sigma}\left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)\right)^{\frac{2\sigma-1}{1-\sigma}}(b_k\psi(x_k))^{\sigma-1}g(x_k) \\
= \; & (b_k\psi(x_k))^{\sigma-1}\left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)\right)^{\frac{\sigma}{1-\sigma}}\left[\frac{(\sigma-1)}{b_k\psi(x_k)}-\frac{\sigma}{1-\sigma}\frac{g(x_k)}{\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)}\right] \\
= \; & (b_k\psi(x_k))^{\sigma-1}\left(\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)\right)^{\frac{\sigma}{1-\sigma}}\left[\frac{(\sigma-1)}{b_k\psi(x_k)}+\frac{g(x_k)\left(\frac{1}{\sigma-1}+1\right)}{\int_{x\in E_k}(b_k\psi(x))^{\sigma-1}dG(x)}\right]
\end{aligned}$$

when $\sigma > 2$, it is straightforward to show the above is positive. Proposition 1 states that larger cities have better entrepreneurs, which means $x_{k'} > x_k$, which in turn implies that the inequality in Eq. (11) is true, and thus $\frac{\pi_{k'}\psi(x_{k'})^{(\sigma-1)}}{w_{k'}} > \frac{\pi_k\psi(x_k)^{(\sigma-1)}}{w_k}$.
∎

D.6 Proposition 3

Proof: Our proof takes four steps:

(i) First, we show that if we rank the cities by the size so that $N_1 > N_2 > N_3 > \cdots, N_J$, and the distribution of city size follows one of the following specific distributions, then the expected one-step spacing in population, $E(N_k) - E(N_{k+1})$, is decreasing in k.

We prove the result on the pattern of shrinking spacing separately for each distribution.

(1) **Exponential Distribution.** Kamps (1991) provided a general formula to explicitly characterize any moment of 1-step spacing between exponentially distributed order statistics. If we apply his formula to the first moment and fit it into our notations, it is straight forward to show that the expected spacing between two cities is

$$E(N_k) - E(N_{k+1}) = \frac{1}{J-(J-k)+1} = \frac{1}{k+1}.$$

which is clearly decreasing in k.

(2) **Pareto Distribution.** Kamps (1991) also provided a general formula for the Pareto distribution, which we manipulate similarly as in the case of exponential

distribution. For any integer $q < 0$:

$$E(N_k) - E(N_{k+1}) = \frac{J!(J-(J-k)+q)!}{(J+q)!(J-(J-k)+1)!}$$
$$= \frac{J!}{(J+q)!}\frac{(k+q)!}{(k+1)!}$$
$$= \frac{J!}{(J+q)!}\frac{1}{\prod_{i=q+1}^{1}(k+i)},$$

which is again, clearly decreasing in k.

(3) **Weibull and Rayleigh Distribution** Balakrishnan and Sultan (1998) showed that for these two distributions, the expected spacing can be expressed as

$$E(N_k) - E(N_{k+1}) = \frac{\beta}{d(J-(J-k)+1)} = \frac{\beta}{d(k+1)},$$

which is decreasing in k. The constants β and d are both positive. We reach our formula by setting $\alpha = 0$ in Eq.(21.6) in Balakrishnan and Sultan (1998).

Given the above distributions, we can infer that the spacing in the expected population distribution must be decreasing in the city index. In the rest of the proof, we omit the expectation operator for notational ease.

$$N_{k'} - N_{k'+1} < N_k - N_{k+1} \tag{12}$$

(ii) In the second step, we show that if the spacing between population distribution is shrinking over city index k, then the spacing between the congestion costs must be shrinking as well.

We first show that the 1-step spacing must be smaller if k is higher. In the model we assume that the $C(\cdot)$ is monotone and convex. The definition of convex function implies:

$$\frac{C(N_k) - C(N_{k+1})}{N_k - N_{k+1}} \geq \frac{C(N_{k'}) - C(N_{k'+1})}{N_{k'} - N_{k'+1}}.$$

Combine this with inequality in Eq (12), we immediately have:

$$\delta_k(1) = C(N_k) - C(N_{k+1}) > C(N_{k'}) - C(N_{k'+1}) = \delta_{k'}(1).$$

To ease notations, we define the i-step spacing between the congestion costs in city k and $k+i$ as:

$$\delta_k(i) \equiv C(N_k) - C(N_{k+i}) = \sum_{k=0}^{i-1}(C(N_{k+k}) - C(N_{k+k+1})) = \sum_{k=0}^{i-1}\delta_{k+k}(1).$$

Note that the i-step spacing is equivalent to the summation of i 1-step spacing, as shown in the last two parts of the equation above.

It is then straight forward to see that the i-step spacing shrink over city index as well, that is, $\delta_{k'}(i) < \delta_k(i)$:

$$\delta_{k'}(i) = \sum_{j=0}^{i-1} \delta_{k'+j}(1) < \sum_{j=0}^{i-1} \delta_{k+j}(1) = \delta_k(i)$$

(iii) As the sequence of $\delta_k(i)$ are declining in city size, there exists a city index $\{k_i^*\}$ such that S is sandwiched between $\delta_{k_i^*}(i)$ and $\delta_{k_i^*-1}(i)$:

$$\delta_{J-i}(i) < \delta_{J-i+1}(i)... < \delta_{k_i^*}(i) < S < \delta_{k_i^*-1}(i)... < \delta_1(i),$$

or formally, we define $k_i^* \in \{1, 2, \cdots, J\}$ as the city that the following two conditions are both satisfied:

(1) $\delta_k(i) > S$, for all $k < k_i^*$, and

(2) $\delta_k(i) < S$, for all $k \geq k_i^*$.

In the case of $S > \delta_1(i)$, we define $k_i^* = 1$; in the case of $S < \delta_{J-i}(i)$, we define $k_i^* = J$.

The last result we need before describing the sorting pattern is that the sequence of $\{k_i^*\}$ is weakly increasing in i:

$$k_1^* \leq k_2^*, \cdots k_{J-1}^* \leq k_J^*.$$

We prove this result by contradiction. Suppose we have $i' > i$ and $k_{i'}^* < k_i^*$. By the definition of k_i^*, we know that $\delta_k(i) > S$ for all $k < k_i^*$. By the definition of $k_{i'}^*$, we also know $\delta_{k_{i'}^*}(i') < S$. Moreover,

$$\delta_{k_{i'}^*}(i') = \delta_{k_{i'}^*}(i) + \sum_{j=0}^{i'-i-1} \delta_{k_{i'}^*+i+j}(1) > \delta_{k_{i'}^*}(i).$$

and therefore we can infer $\delta_{k_{i'}^*}(i) < S$ as well. However, this contradicts with $\delta_k(i) > S$ for all $k < k_i^*$, and therefore we must have

$$i' > i \implies k_{i'}^* \geq k_i^*.$$

(iv) Now we are ready to characterize the sorting pattern. Corollary 1 states that all the $\{E_k\}$ occupy the right end of the real line, and therefore we only need to characterize the distribution of workers across the cities.

We characterize the distribution city by city. Given the order of k_i^*, we start from the largest cities with $k < k_1^*$. Corollary 1 also implies that within the set of workers, the most talented workers are in H_1, which is the H sector in the largest city. The next group of workers either works in L_1 or in H_2, the next two potential groups with lower entry barriers. In this case, since $k < k_1^*$, we know $\delta_1(1) = C(N_1) - C(N_2) > S$, and thus $C(N_1) > C(N_2) + S$. This implies that it is harder to choose L_1 than H_2, and thus the next group of workers will be employed in L_1. The group of workers following

L_1 face a choice between H_2 and L_2, in which the barriers to H_2 will always be higher ($C(N_2) + S > C(N_2)$), and thus they will be employed in H_2. The same logic can be applied to all the cities with $k \leq k_1^*$ and therefore we know among cities $k < k_1^*$, the workers will be sorted, by x from low to high, as

$$L_{k_1^*-1} H_{k_1^*-1} L_{k_1^*-2} H_{k_1^*-2} \cdots L_2 H_2 L_1 H_1$$

Now we consider the next group of cities with $k_1^* \leq k < k_2^*$. Within this group of cities, we always have $\delta_k(1) < S$ and $\delta_k(2) > S$. Again, the most talented workers will choose the job with the highest entry barrier, $H_{k_1^*}$. The next group of workers, facing a choice between $H_{k_1^*+1}$ and $L_{k_1^*}$, however, since $\delta_{k_1^*}(1) = C(N_{k_1^*}) - C(N_{k_1^*+1}) < S$, and thus they will choose to work in $H_{k_1^*+1}$. The subsequent group face a choice now between $H_{k_1^*+2}$ and $L_{k_1^*}$. As we know $\delta_{k_1^*}(2) = C(N_{k_1^*}) - C(N_{k_1^*+2}) > S$, this group of workers will choose $L_{k_1^*}$. The next group of workers then face the choice between $H_{k_1^*+2}$ and $L_{k_1^*+1}$. Following similar logic, we can infer that the workers will choose $H_{k_1^*+2}$. We can apply this logic all the way till $L_{k_2^*-2}$. The group of workers following $L_{k_2^*-2}$ face a choice between $L_{k_2^*-1}$ and $H_{k_2^*}$ and as $\delta_{k_2^*-1}(1) < S$. It follows that $H_{k_2^*}$ will be preferred. The next group, who choose between $H_{k_2^*+1}$ and $L_{k_2^*-1}$, prefer $L_{k_2^*-1}$ as $\delta_{k_2^*-1}(2) > S$. Taking stock, the sorting pattern so far is:

$$L_{k_2^*-1} H_{k_2^*} L_{k_2^*-2} H_{k_2^*-1} \cdots L_{k_1^*+2} H_{k_1^*+3} L_{k_1^*+1} H_{k_1^*+2} L_{k_1^*} H_{k_1^*+1} H_{k_1^*}.$$

Note that the sorting pattern is characterized by two consecutive H sectors at the beginning, and then again, H and L sectors taking turns with a 2-step spacing between them, e.g. L_k followed by H_{k+2}.

The same pattern naturally extend to $k_2^* \leq k < k_3^*$. It is straightforward to verify that the following sorting pattern holds:

$$H_{k_2^*+4} L_{k_2^*+1} H_{k_2^*+3} L_{k_2^*} H_{k_2^*+2} H_{k_2^*+1}.$$

Again, we observe two consecutive H sectors at the beginning, followed by inter-locking L and H sectors, this time with a 3-step spacing.

We can continue this pattern until H_J is reached. At this point, there will be a group of L_k choices left un-filled, and the remaining workers will directly sort into these L sectors by the prediction of corollary 1. The number of cities with un-filled L sector can be determined as the smallest i such that:

$$\bar{i} = \text{argmin}_i C(N_{J-i}) - C(N_J) > S$$

In the end, the left-tail of workers will be sorted as

$$L_J L_{J-1} \cdots L_{J-\bar{i}} H_J L_{J-\bar{i}-1} H_{J-1} \cdots$$

In the end, the unique sorting pattern of individuals over (k, ω) on $\mathbb{R}^+$ that emerges under the above-mentioned population distribution is summarized in Figure D.2:

As shown by the arguments above, this sorting pattern is unique. ∎

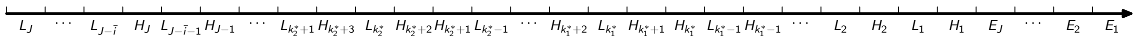


Figure D.2: Sorting with Multiple Cities

E City Size and Human Capital Stock

We consider a case of two-city and focus on a "location-first" sorting pattern. Without loss of generality, we assume the human capital distribution follows the exponential distribution. In the equilibrium, the most talented agents become entrepreneurs in city 1; less talented agents are first sorted into H and L sector in city 1, followed by the least talented ones working in city 2. The cutoff human capital of being entrepreneur in city 1 is determined as:

$$A_1 e^{x_{E1}(\sigma-1)} - B_1 = A_2 e^{x_{E1}(\sigma-1)} - B_2.$$

The cutoff human capital between H and L sector within each city is determined as:

$$\begin{aligned} A_1^H x_{H1} - B_1^H &= A_1^L x_{H1} - B_1^H, \\ A_2^H x_{H2} - B_2^H &= A_2^L x_{H2} - B_2^L. \end{aligned}$$

Moreover, the cutoff human capital between being an entrepreneur in city 2 and a worker of H sector in city 1 is given as:

$$A_1^H x_E - B_1^H = A_2 \left(x_E\right) e^{x_E(\sigma-1)} - B_2.$$

Finally, the cutoff human capital between being in L sector of city 1 and H sector of city 2 is:

$$A_1^L x_1 - B_1^H = A_2^H x_1 - B_2^H.$$

Denote $R_i (i = 1, 2)$ as the total expenditure in city i. Given the Cobb-Douglas utility function specification, αR_i is the sales revenue from H sector. The CES preference over composite variety from H sector implies a fraction $(\sigma - 1)/\sigma$ of the total sales revenue are spent on wages paid to the workers in the production. Labor market clearing condition from H sector in each city is:

$$\begin{aligned} \frac{\sigma-1}{\sigma} \alpha R_1 &= w_1 \int_{x_{H1}}^{x_E} x dG\left(x\right), \\ \frac{\sigma-1}{\sigma} \alpha R_2 &= w_2 \int_{x_{H2}}^{x_1} x dG\left(x\right). \end{aligned}$$

The remaining a fraction $1 - \alpha$ of the total expenditure are devoted to the non-tradable goods in L sector. Market clearing condition of L sector in each city is thus:

$$\begin{aligned} \left(1-\alpha\right) R_1 &= z_1 \int_{x_1}^{x_{H1}} x dG\left(x\right), \\ \left(1-\alpha\right) R_2 &= z_2 \int_{\underline{x}}^{x_{H2}} x dG\left(x\right). \end{aligned}$$

$R_j\,(j=1,2)$ denotes the total expenditure in city j, which in turn equals to the total income earned by all the agents :

$$\int_x \sigma A_j e^{x(\sigma-1)} P_j^{\alpha} z_j^{1-\alpha} dG(x) + (1-\alpha) R_j = R_j,\ j=1,2.$$

The first part in above represents the profits and wage income earned by all the entrepreneurs and workers from the H sector. The second part is the income earned by workers from L sector.

Rearranging the expressions above into the following:

$$\begin{aligned} R_1 &= \frac{1}{\alpha}\int_{x_{E1}} \sigma A_1 e^{x(\sigma-1)} P_1^{\alpha} z_1^{1-\alpha} dG(x), \\ &= \frac{\sigma}{\alpha} A_1 \frac{\lambda e^{x_{E1}(\sigma-1-\lambda)}}{1+\lambda-\sigma} P_1^{\alpha} z_1^{1-\alpha}. \end{aligned}$$

$$\begin{aligned} R_2 &= \frac{1}{\alpha}\int_{x_E}^{x_{E1}} \sigma A_2 e^{x(\sigma-1)} P_2^{\alpha} z_2^{1-\alpha} dG(x), \\ &= \frac{\sigma}{\alpha} A_2 \lambda \left(\frac{e^{x_E(\sigma-1-\lambda)}}{1+\lambda-\sigma} - \frac{e^{x_{E1}(\sigma-1-\lambda)}}{1+\lambda-\sigma}\right) P_2^{\alpha} z_2^{1-\alpha}. \end{aligned}$$

To establish the relation between city size and human capital stock, we compute the human capital stock in the following.

$$\begin{aligned} H_1 &= \int_{x_{E1}} x\lambda e^{-\lambda x} dx + \int_{x_{H1}}^{x_E} x\lambda e^{-\lambda x} dx + \int_{x_1}^{x_{H1}} x\lambda e^{-\lambda x} dx \\ &= e^{-\lambda x_{E1}} \left(\frac{\lambda x_{E1}+1}{\lambda}\right) + e^{-\lambda x_E}\left(\frac{-\lambda x_E - 1}{\lambda}\right) - e^{-\lambda x_1}\left(\frac{-\lambda x_1 - 1}{\lambda}\right). \end{aligned}$$

Human capital stock in city 2 is:

$$\begin{aligned} H_2 &= \int_{x_E}^{x_{E1}} x\lambda e^{-\lambda x} dx + \int_{x_{H2}}^{x_1} x\lambda e^{-\lambda x} dx + \int_{\underline{x}}^{x_{H2}} x\lambda e^{-\lambda x} dx \\ &= e^{-\lambda x_{E1}}\left(\frac{-\lambda x_{E1}-1}{\lambda}\right) - e^{-\lambda x_E}\left(\frac{-\lambda x_E - 1}{\lambda}\right) + e^{-\lambda x_1}\left(\frac{-\lambda x_1 - 1}{\lambda}\right) - e^{-\lambda \underline{x}}\left(\frac{-\lambda \underline{x} - 1}{\lambda}\right) \end{aligned}$$

Human capital stock in city 1 depends on x_{E1}, x_E and x_1, while the city size of city 1 only depends on x_{E1}.

It is more straightforward to consider an autarky case, where the wage rate in both cities are normalized to be 1. The total income at both cities can be obtained using the labor market clearing condition in H sector:

$$\begin{aligned} R_1 &= \frac{\sigma}{(\sigma-1)\alpha}\left[e^{-\lambda x_E}\left(\frac{-\lambda x_E - 1}{\lambda}\right) - e^{-\lambda x_{H1}}\left(\frac{-\lambda x_{H1} - 1}{\lambda}\right)\right] \\ R_2 &= \frac{\sigma}{(\sigma-1)\alpha}\left[e^{-\lambda x_1}\left(\frac{-\lambda x_1 - 1}{\lambda}\right) - e^{-\lambda x_{H2}}\left(\frac{-\lambda x_{H2} - 1}{\lambda}\right)\right] \end{aligned}$$

We can further establish a relation between R_1 and H_1 as follows:

$$R_1 \frac{(\sigma-1)\,\alpha}{\sigma} = H_1 - e^{-\lambda x_{E1}} \left(\frac{\lambda x_{E1}+1}{\lambda} \right) + e^{-\lambda x_1} \left(\frac{-\lambda x_1 - 1}{\lambda} \right) - e^{-\lambda x_{H1}} \left(\frac{-\lambda x_{H1} - 1}{\lambda} \right)$$

and

$$\begin{aligned} R_2 \frac{(\sigma-1)\,\alpha}{\sigma} = H_2 - e^{-\lambda x_{E1}} \left(\frac{-\lambda x_{E1} - 1}{\lambda} \right) + e^{-\lambda x_E} \left(\frac{-\lambda x_E - 1}{\lambda} \right) \\ + e^{-\lambda \underline{x}} \left(\frac{-\lambda \underline{x} - 1}{\lambda} \right) - e^{-\lambda x_{H2}} \left(\frac{-\lambda x_{H2} - 1}{\lambda} \right) \end{aligned}$$

As shown in the above two equations, the relationship between R_i and H_i is neither linear nor log-linear; instead, higher order terms exist and thus controlling for one variable cannot fully encapsulate the other variable in a linear regression setup. Instead, both variables need to be controlled for on the right-hand-side of the equation, otherwise the estimation might suffer from missing variable bias.