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

Lin Ma †
National University of Singapore

Yang Tang ‡
Nanyang Technological University

November 15, 2016

Abstract

Larger cities are not always more unequal: while the 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 competitive 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. Our theoretical predictions are broadly supported by empirical tests using U.S. data.

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

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

*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.

†Email: ecsml@nus.edu.sg

‡Email: tangyang@ntu.edu.sg

1 Introduction

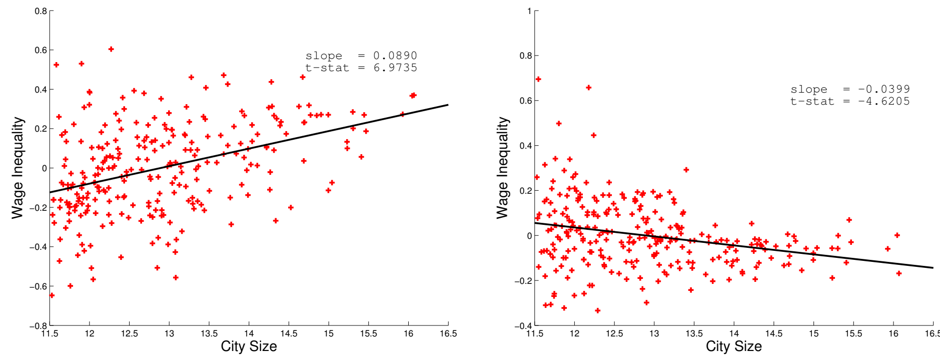
Are larger cities more unequal? Empirical evidence thus far seem to suggest a simple, positive answer. Baum-Snow and Pavan [2013] documents that a positive relationship between city size and wage inequality has developed from 1979 to 2007 in the U.S. Similarly, Behrens and Robert-Nicoud [2014a] also estimates that the size elasticity of income Gini coefficient to be positive at around 0.011.[1] However, inequality in itself is a multifaceted concept: oftentimes, inequality measured at different parts of the same distribution could behave differently, rendering simplistic answers insufficient. For example, studies in the macroeconomic literature have shown that in the U.S. the inequality measured at the right tail, such as the 90-to-50 ratios, have been steadily increasing over time, but measures taken at the left tail, such as the 50-to-10 ratios, have been stable or even slightly declining since the 1980s.[2] In this paper we show that the same complexity arises 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 looking at.

Figure 1 summarizes our empirical findings for 254 U.S. Metropolitan Statistical Areas (MSAs) in 2000. The left panel plots the 99-to-50 percentile wage ratio within the same city against the city size measured in population, while the right panel plots the 50-to-1 wage ratio in the same manner. Larger cities are more unequal, but only in the right tail: the size elasticity of the 99-to-50 ratio is indeed positive at around 0.0890. On the left tail larger cities are actually more equal: the size elasticity of the 50-to-1 ratio stands significantly negative at around -0.0399. This contrasting pattern is robust: it can be observed for other percentile-ratios of the wage, income, and earnings distributions, regardless of measures of city size.

Why do wage inequalities move in such a pattern with respect to city size? Exist-

[1] For more empirical literature on this, 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. 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.



(a) 99/50 Wage Ratio

(b) 50/01 Wage Ratio

Figure 1: Wage Inequality and City Size

Notes: The graphs plot the 99/50 and 50/1 wage ratio within each city against the city size measured in the logarithm of population in the U.S. The measures of inequality are the residuals after netting out the human capital measured as years of education, racial compositions, and state fixed effects. Each dot in the above graphs represents an MSA in IPUMS-USA, 2000. For more details, see Section 3.

ing theories provide few clue. Studies on inequality from the macroeconomic literature usually focus on the aggregate level and lack the geographic structure to compare inequality across regions. The "new economic geography" literature provides a natural setting in which one could tackle this issue. However, most of the theories there emphasize on the efficiency side of the model such as agglomeration and the productivity differences across cities and often overlook the distributional issues within cities. The small theoretical literature on within-city inequality, such as Behrens and Robert-Nicoud [2014b], Eeckhout et al. [2014], and Davis and Dingel [2012], all predict larger cities are always more unequal. One cannot easily adapt these models to explain the contrasting behaviors at the two tails simultaneously.

We propose a new framework to show that the empirical patterns documented above can naturally arise as a spatial equilibrium. In our model individuals with heterogeneous talents endogenously choose along three dimensions — location, industry, and occupation — to work in, creating firms and populating cities along the way. In equilibrium individuals will sort along all three dimensions. Along the location dimension, due to the existence of knowledge spillovers and congestion, individuals with similar talents tend to reside in the same city, and those with better talents form larger cities. We assume that industries differ in entry barriers to capture the differences in costs of education and training needed to join different industries. As a result, industry with higher entry barriers will pay a higher real wage, and attract individuals with better talents. Within industries, individuals further sort into occupations. Founding new firms incurs fixed costs, and thus only the most talented individuals will choose to do so and become entrepreneurs, while all the others will become workers that are employed by the former group. In the spatial equilibrium ex-ante identical locations will host cities different in output, population and industry compositions.

Multi-dimensional sorting generate different wage rates within each city-industry-occupation cell. Consistent with the empirical results in the literature, all the measures of wage in our model increase with city size. Beyond the literature we are the first to theoretically and empirically show that different skill premium shall increase with city size at different speeds, and thus lead to differences in inequality measured at different tails of the distribution.

The return to entrepreneurial talents benefits more from city size compared with the return to labor in our model. As the entrepreneurs are usually the top earners in each city, inequality measured at the right tail will be higher in larger cities. These are the results of 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 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]. Our paper abstracts away from the details of an executive compensation model, directly assumes the Roberts' Law, and studies its impact on inequality. The second mechanism is a natural result of the sorting pattern described before. Large cities are populated by large and productive firms, which push up the factor prices and push down product prices in equilibrium — making the market conditions tougher for smaller firms to operate. In order for the marginal entrepreneur to stay in the large city, the 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 need to compensate the marginal workers the increments in congestion disutility. As a result, the wage rate will increase with city size at much slower pace as opposed to firm size and entrepreneurial compensation. This mechanism is also supported by the empirical literature. Researchers have found the size elasticity of firm size is usually estimated to be much larger than that of the wage rate.[3]

Within the group of workers, the wage rate per unit of efficiency labor in all industries will be higher in larger cities to compensate higher congestion costs.[4] However, the wage rate of lower-paying industries must increase faster with city size, otherwise the workers in these industries, who have fewer units of efficiency labor to supply to the market, 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 products and services from these low-skill, low-paying industries — janitors, cashiers, or street vendors — to properly function, which implies that the market must compensate those working in such industries relatively more in equilibrium. As those working in the low-paying industries usually occupy a lower position in the wage distribution, the inequality measured at the left tail shall be smaller in large cities.

The seemingly opposite patterns of skill premium with respect to city size — that both the top entrepreneurial talents and the skills of working in bottom paying in-

[3]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 much smaller, 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.

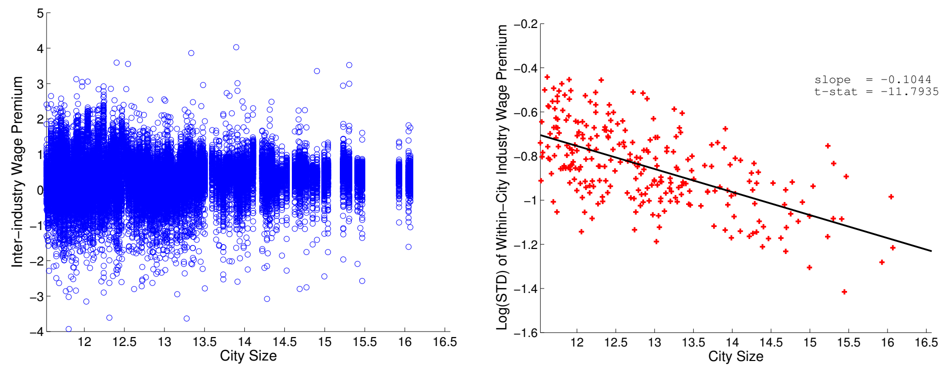
[4]One can interpret the higher congestion disutility as higher price levels in larger cities. Handbury and Weinstein [2015] documents that food price is lower in larger cities; however other price, especially housing price, is significantly higher in larger cities. Housing costs are also responsible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

dustries benefit more from city size than the "middle class" workers in relatively high paying industries — are actually the result of one simple equilibrium force. The market must compensate those with greater difficulties living in large cities relatively more in order to accommodate them in large cities. In addition to the congestion burden shared by every individual, entrepreneurs suffer from the competitive nature of the product market, while the workers in the bottom-paying industries lack sufficient supply of efficiency labor units.

Both our predictions on the city-size elasticity of skill premium are borne out in the data. We test these predictions with the IPUMS-2000 data in the U.S. with 254 MSAs and 201 industries. In our model all the measures of skill premium are based on wage rates per efficiency unit of labor supply, which is not the observed wage rate in the data. To make the measures comparable, we first net out the effects of observable individual characteristics, such as age, education, gender, and race, on wage rate, and measure skill premiums based on the residual wage in the data. We find that the skill premium of entrepreneurs increases faster with respect to city size as compared to that of the workers. On average, doubling the city size widens the wage gap between entrepreneurs and workers by around 2.1 to 2.3 percent, depending on the definitions of entrepreneurs.

To test our predictions on inter-industry skill premium, we first estimate these premium in each city with Monte-Carlo simulations[5]. The results are summarized in Figure 2, which plots the estimated industry wage premium and the logarithm of the standard deviation of industry wage premium within city against the city size. Our model predicts that the skill premium of low-paying industries increases faster with respect to city size. The direct implication is that the spread of inter-industry wage premium shall decrease with city size. The empirical results clearly support

[5] We employ Monte-Carlo simulation here because the sample sizes of each city-industry cell vary greatly within our samples, and on average industries in larger cities have more observations than their counter-parts in smaller cities. This mechanically leads to larger sample variances of inter-industry wage premium in smaller cities, which bias the results in our favor. We solve this problem by random re-sampling to ensure that the sample size used to estimate inter-industry wage premium is the same across cities.



(a) Industry Wage Premium by City Size (b) Log(STD) of Ind. Wage Premium by City Size

Figure 2: Spatial Variations of Inter-Industry Wage Premium

Notes: Panel (a) plots the industry wage premium estimated by Monte-Carlo simulations (1000 repetitions) in each city against the log of population. Panel (b) plots the log of standard deviation of industry wage premium within each city against the log of population. The slope and the t-statistics reported in Panel (b) are based on a simple linear regression between the two variables. For more details, please refer to Section 6.1.1. Data source: IPUMS-USA 2000 and Population Census 2000.

these predictions. The differences in spread between large and small cities are also sizable: doubling the city size decreases the standard deviation of the inter-industry wage premium within the city by 10.4 percent.

How important are the spatial variations of inter-industry wage premium in explaining the variations in within-city wage inequalities? We carry out a counterfactual exercise in which we shut down the spatial variations of inter-industry wage premium by forcing the distribution of inter-industry wage premium in every U.S. city to have the same standard deviation. We then re-construct the wage profile for each individual in our U.S. sample, and re-estimate the relationship between city size and the within-city wage inequality at the left tail in the counterfactual world. We compare the counterfactual results to a benchmark simulation designed to mimic the actual wage distribution in each city. We find that the size elasticities of the 50-to-10 wage ratios drops by 55 percent, and the size elasticities of the 50-to-01 wage ratios drop by around 36 percent in the counterfactual simulations. This indicates that a large proportion of the observed decline of wage inequalities in large cities can be explained by the spatial variations of inter-industry wage premium.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 documents the empirical pattern between wage inequality and city size. Section 4 presents the model, and Section 5 discusses the analytical results. Section 6 tests the key predictions of the model and perform counter-factual simulations, and Section 7 concludes.

2 Literature Review

Our paper is broadly related to several strands of existing literature. There is a small empirical literature on the relationship between inequality and city size. Baum-Snow and Pavan [2013] documents that wage inequality is higher in larger cities for both the 90-50 and the 50-10 percentile with several datasets without controlling for city level characteristics. Wheeler [2004] and Wheeler [2005] use U.S. census during 1950-1990 to show that the return to skills and wage inequality are both increasing with city size. Our results are different because we control for city level characteristics such as the average years of education and racial composition of the residents, and the state in which the MSA is located. Controlling these city-level characteristics is crucial to our exercise, as they within-city inequality through channels other than city size. The sample in Baum-Snow and Pavan [2013], Wheeler [2004] and Wheeler [2005] are all restricted to white male, and Wheeler [2005] only focuses on a selection of large individual cities. Our sample includes all male working population regardless of race in a large selection in MSAs. Glaeser et al. [2009] and Behrens and Robert-Nicoud [2014a] also document that inequalities are higher in larger cities with the Gini coefficient as measure.

The literature on the spatial variations of skill premium is relatively larger. Bacolod et al. [2009] measures skill by fundamental worker skills — including a wide range of cognitive, social and motor skills — rather on worker education. They find large cities are only slightly more skilled than are small cities. Roback [1982], Combes et al. [2008], Davis and Dingel [2012] and Baum-Snow and Pavan [2012] provide both 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 la Roca and Puga [2016] show that individual earnings are higher in larger cities mainly due to learning effects. We extend this line of work by showing, both theoretically and empirically, that the return to different skills shall increase with city size at different speeds, and how these lead to different patterns of inequality.

The theoretical studies on city size and inequality are relatively scarce. Behrens et al. [2014] provided a framework in which income distributions are not degenerate within a city. However, income distributions and inequality do not vary across cities in their model. Behrens and Robert-Nicoud [2014b], Davis and Dingel [2012], and Eeckhout et al. [2014] provide models in which inequality vary across cities. However, they predict inequality will always be higher in larger cities. Our model goes beyond this by allowing wage inequality measured at different tails to move in different directions.

Our paper is broadly related to the new economic geography literature, such as Combes et al. [2008], Gaubert [2014] and Tombe et al. [2015] etc. We contribute to this literature by introducing a tractable framework despite of the complexity induced by three dimensional sorting and endogenous distributions. We are able to show that under reasonable assumptions, the firm and city size distribution in our model follow Pareto distributions. This further uniquely pins down the sorting patterns along three dimensions, and thus an unique equilibrium. This also opens up the possibility to study the distributional impacts of agglomeration, migration, and inter-city trades for future quantitative works.

Lastly, our work is linked to the labor literature on inter-industry wage premium, motivated by the works of Rosen [1987], Krueger and Summers [1988], and Katz et al. [1989]. Previous studies have demonstrated how inter-industry wage premium can be explained with either worker-specific or firm-level characteristics. However, this literature focuses on the national level and overlooks the potential spatial dimension.

Our work is the first to provide a theoretical foundation on which inter-industry wage premium can vary by location. We are also the first to theoretically predict as well as empirically document that the relative wage premium in high-paying industries must increase slower with respect to city size.

3 Inequality and City Size: Empirical Patterns

In this section we document that top and bottom inequality within a city moves in opposite directions with respect to city size in the U.S. Our data come 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 year 2000 for our analysis. We impose a couple of conventional sample restrictions: we drop those 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". Within each MSA, we compute various measures of income, wage, and earnings inequality. We define income as the personal total income, earnings as total earned income, and wage as the ratio between total wage income and usual hours worked. We measure the size of the city using both the regional GDP estimates obtained from the Bureau of Economic Analysis (BEA) in year 2001, and the population reported by the census in 2000.

Our final sample contains around 1.23 million individuals working in 254 MSAs.[6] The median "city" in our sample hosts 12,638 individuals, and the smallest "city" contains 445 individuals. More than 75 percent of our "cities" contain at least 4,000 individuals. The large sample size in each city allows us to compute the percentile

[6]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.

ratios within each city. We provide a detailed description of the data set and the variables constructed in Appendix D.

We study how various measures of inequality within the city vary with the size of the city with the following equation:

$$\log(\text{Ineq}_i) = \beta_0 + \beta_1 \log Y_i + \mathbf{b} \cdot \mathbf{X_i} + \epsilon_i, \tag{1}$$

where i indexes the city. Ineq_i is a measure of inequality within city i, Y_i is a measure of the size of city i, and $\mathbf{X}_i$ is a vector that controls city-level characteristics, such as the average years of education among all the residents of the city, the standard deviation of years of education, state dummy variables, and racial compositions.[7] Our parameter of interest is β_1, the size elasticity of inequality.

VARIABLES	(1) 99/50 Ratio	(2) 99/50 Ratio	(3) 99/50 Ratio	(4) 50/01 Ratio	(5) 50/01 Ratio	(6) 50/01 Ratio
Log(Total GDP)	0.0854***			-0.0417***		
	(0.0197)			(0.0114)		
Log(Private Ind. GDP)		0.0791***			-0.0408***	
		(0.0183)			(0.0111)	
Log(Population)			0.0890***			-0.0399***
			(0.0241)			(0.0141)
Average Years of Edu.	2.014***	2.086***	2.388***	1.318***	1.332***	1.077***
	(0.467)	(0.457)	(0.506)	(0.423)	(0.431)	(0.350)
STD of Years of Edu.	0.581***	0.578***	0.604***	0.135	0.145	0.0969
	(0.138)	(0.142)	(0.150)	(0.105)	(0.106)	(0.104)
Share of White Population	0.260	0.213	0.159	-0.446**	-0.428**	-0.374*
	(0.350)	(0.348)	(0.366)	(0.206)	(0.200)	(0.210)
Constant	-3.374***	-3.495***	-4.569***	-0.763	-0.796	-0.120
	(0.974)	(0.962)	(1.018)	(0.911)	(0.925)	(0.705)
Observations	254	254	264	254	254	264
R-squared	0.556	0.553	0.548	0.324	0.325	0.295

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: 99/50 and 50/01 Wage Ratio by City Size, U.S. Data

Note: This table reports the results of estimating equation (1), while controlling for the average years of education, the standard deviation of years of education, and state dummies. The measure for inequality is the 99/50 and 50/01 ratios of wage ratio within an MSA. Data for income inequality come from 5 percent sample in year 2000 compiled by IPUMS. Data for GDP come from BEA regional GDP estimates. Data for population come from U.S. 2000 census.

[7] Some MSAs span over several states. In these cases, we assign the MSA to the state where most of its population lives.

Table 1 reports our main findings. The first three columns of the table report the estimation of equation (1) when we use the log of the wage ratio between the 99th and the 50th percentile of wage rate as measures of Ineq_i. All the estimates of β_1 are significantly larger than zero. This is consistent with the previous 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 total GDP is associated with $0.0791 - 0.0890$ percent *increase* in the 99-to-50 ratio, depending on the measure of city size. The opposite pattern emerges at the left tail: the wage ratio between the 50th and the 1st percentile decreases with city size. A one percent increase in total GDP is associated with a $0.0399 - 0.0417$ percent *decrease* in the 50-to-1 wage ratio.

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 various city level characteristics, while Baum-Snow and Pavan [2013] do not.[8] In the following, we highlight the our differences by first estimating equation (1) without any control, and then progressively add each control variable back. The results are reported in Table 2.

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.[9] The rest of the Table 2 shows that by adding back our control variables, the size elasticity gradually decreases from 0.021 to -0.015.[10]

[8] 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 here. 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 different city sizes.

[9] 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.

[10] Table 2 seems to suggest that the standard deviation of human capital distributions is the key variable behind the negative relationship, while the share of white population is not. However this is an illusion driven by the ordering of the variables. In the appendix, Table 7, we report similar regressions with all the possible combinations of control variables, and it shows that the relationship between city size and bottom inequality is always positive if we only control for one variable. We need to control for the average level of education and the racial compositions at the same time to generate the negative relationship.

VARIABLES	(1) 50/10 Ratio	(2) 50/10 Ratio	(3) 50/10 Ratio	(4) 50/10 Ratio	(5) 50/10 Ratio
Population	0.0212***	0.00947*	0.00184	-0.0112**	-0.0150**
	(0.00562)	(0.00497)	(0.00441)	(0.00507)	(0.00581)
Share of White Population			-0.252***	-0.341***	-0.318***
			(0.0640)	(0.0602)	(0.0591)
Average Years of Edu.				0.591***	0.728***
				(0.125)	(0.161)
STD of Years of Edu.					0.0525
					(0.0395)
Constant	0.553***	0.729***	0.772***	-0.609**	-0.815**
	(0.0732)	(0.0619)	(0.0536)	(0.298)	(0.333)
Observations	264	264	264	264	264
R-squared	0.047	0.617	0.652	0.687	0.691
State Dummies	No	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

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

Note: This table reports the results of estimating Equation (1) 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.

Shall we control for city level characteristics? The variables that we control for indeed affect wage distributions within the city in addition to city size. Cities with well-educated population or higher dispersion of human capital tend to be more unequal at both tails; the share of population that reports as "white" seems to reduce inequality at the lower end. The state in which the city locates in also matters, probably because minimum wage requirement varies from state to state. As these variables affect inequality 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. In our model and the empirical tests later in this paper, we also focus on the residual inequality — the inequality in wage rates after controlling for age, education, and years of experience. We focus on residual inequality mainly because most of the variations in wage cannot be explained by observable ([Mortensen, 2005]).

Can the empirical patterns documented in Table 1 be extended to other percentile ratios? We present a complete picture on the relationship between wage and city size in Figure 3. We estimate equation (1) with the level of wage at different percentiles at the left-hand-side, and report the estimated size elasticity (β_1) with 95 percent confidence

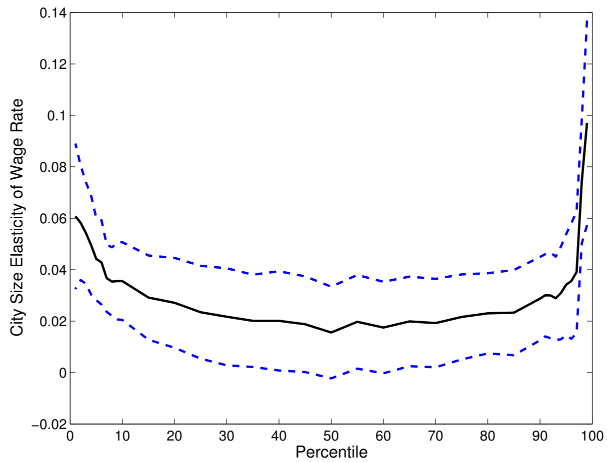


Figure 3: City Size Elasticity of Wage at Different Percentiles

Note: This figure reports the city size elasticity of wage rate (β_1) from Equation (1) 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.

interval in the graph. Wage rates at all percentiles increase with city size, but at different speed. A U-shape emerges: the wage at the top and the bottom percentiles increases faster with city size than those in the middle. Comparing any two points on the graph reveals how percentile ratios vary with city size. The graph indicates that the top inequalities such as the 95-to-50 ratios shall be higher and the bottom inequalities such as the 50-to-5 ratios shall be lower in larger cities, while inequality measured in the middle of the distribution, such as the 75-to-25 or 90-to-10 ratios, shall not vary across cities with different sizes at all. We report the estimation using (1) with various percentile ratios in Table 4 in the appendix, and they are consistent with the implications from the figure. Table 5 and Table 6 in appendix also report the same exercise with total personal income and labor earnings inequalities and the results are consistent with wage rates.

Autor and Dorn [2013] document the polarization of U.S. wage rates and argue that automation in manufacturing is the driver behind it. We show that similar polarization can be observed in the size elasticity of wages, which suggests that spatial agglomeration might have contributed to the polarization of wage rates as well. 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 larger, agglomeration will directly lead to the polarization of wage rates in addition to what happens in the technological front.

Why do inequality measured at different ends of the distribution tend to move in opposite directions with respect to city size? In the next section we present a general equilibrium model, and postulate that the differences in the city-size elasticities of skill premium might be able to explain this pattern.

4 Theoretical Framework

4.1 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 cumulative distribution function $G(x)$. x includes individual characteristics both observable and unobservable to an outside econometrician, but x is perfectly observable for all individuals inside the model.[11] Individuals can freely choose which city to live in.[12] Within each city, individuals can 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 finance, healthcare, and 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.[13]

Within the H sector, individuals can choose between two occupations — entrepreneurs

[11]This implies that the sorting in our model is based on both the observable and the unobservable. This introduces complexities when we test the predictions of the model later, and will be discussed in detail in the next section.

[12]Essentially we assume that there is no costs to migrate, and thus our equilibrium results do not depend on the initial population distribution.

[13]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 sections will still be true.

or workers — 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. Firm created by the entrepreneur with human capital level x in city j has the following production function:

$$Q_j(x) = A_j(x)(\ell - f),$$

where ℓ denotes efficiency labor input and f is fixed cost of production in units of efficiency labor. $A_j(x)$ is labor productivity of the firm:

$$A_j(x) = b_j(\bar{x}_j)e^x.$$

The above functional form follows Ma [2015], which implies that the firm productivity follows a Pareto distribution as long as human capital is exponential distributed — a property that we will exploit later to pin down the unique equilibrium. $b_j(\bar{x}_j)$ captures the average productivity in city j, and we assume it to be increasing in the average entrepreneurial talents in the city j, $\bar{x}_j$: an entrepreneur surrounded by other talented individuals will be more productive. This assumes positive human capital spillovers, and eventually leads to the productivity advantages of large cites. The income of the entrepreneur equals to the profit of the firm she owns.

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: 1) entrepreneurs wage increases with firm size, and 2) workers' wage is not directly linked to the firm they work for, are crucial in generating the patterns of inequality at the right tail. Both assumptions are based on the empirical findings in the literature of labor economics. The literature on executive compensation has extensively 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 [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 power function, the identity mapping, in our benchmark model.

Researchers have indeed documented the existence of positive firm-size-premium for workers as well (see Oi and Idson [1999] for summary). However, once individual characteristics have been controlled for, the size-premium usually shrinks significantly. For example, Abowd et al. [1999] documents that individual effects explain about 75 percent of the firm-size wage effect, while firm effects explain relative 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 summarized by x. 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 key results governing the patterns of inequality at the right tail, Proposition 8. 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 for workers as well. As long as the firm-size elasticity for entrepreneurs is higher than that of the workers — a simple assumption backed by most empirical studies on manager-to-worker pay ratios — our result survives. For concreteness of exposition we adopted some specific functional forms in this section, but our results are flexible enough to incorporate a wide range of variations.

We assume that the L goods are homogeneous and non-tradable across cities. The market is perfectly competitive, and individuals can produce without organizing into firms — this implies that there is no occupation choice in this 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 L sector with human capital x in city j is thus $z_j x$.

All goods in the H sector can be traded across cities, and trade incurs both fixed costs and variable costs. In order to export from city k to city j, the firm needs to pay f_{jk} in terms of labor in city k. Variable trade costs take the form of iceberg trade costs $\tau_{jk} > 1$: in order to sell one unit of goods in market j, the firm in city k needs to ship out τ_{jk} units of goods. We denote the price of variety x produced in city k and sold in city j as $p_{jk}(x)$.

Individuals in city j gain utilities from consumption of final goods, which is an aggregation of all the goods available in the city that she resides in:

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

where Λ_j is L goods input; $q(i)$ is the input of variety i of H goods, Ω_j is the set of available H goods in city j. α denotes the input share of H goods, and σ 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:

$$U = y_j - C(N_j)$$

where $C(N_j)$ captures congestion dis-utilities, which positively depends on the size of city j measured in population size. We assume that $C'(\cdot) > 0$ and $C''(\cdot) \geq 0$. Appendix B provides a simple extension of our model in which the congestion disutility is microfounded in a model with housing and land market. In the benchmark model we directly use the reduced functional from the extended model for the sake of simplicity.

The detailed solution of the model is provided in Appendix A.

5 Analytical Results

We define a **competitive spatial equilibrium** as a vector of location choices, a vector of industry choices, a vector of occupation choices, a vector of consumption choices, and a series of prices $\{w_j, z_j, p_{jk}(x)\}$ such that:

1. Given the prices, each individual maximizes her utility by choosing occupation, industry, location and consumption bundle; each firm maximizes its profit.
2. H and L goods markets clear in each city, and labor market clears in each city and sector
3. All cities are populated.

All the equilibria can be classified into two categories: symmetric or asymmetric. All the cities are ex-ante identical, and thus a symmetric equilibrium in which all the cities are identical always exists. However, throughout this paper, we are interested in asymmetric equilibrium, where cities are ex-post heterogeneous.

Proposition 1-4 show that all the asymmetric equilibria in the absence of migration cost are sorting equilibria: 1) within each city, individuals sort into different occupations and industries by human capital, and 2) within each occupation and industry, individuals also sort into cities with different sizes.[14] The first four propositions do not require restrictions on the human capital distribution, which leads to the potential of multiple equilibrium. In Proposition 5 and 6 we refine our results and show that under

[14]The above predictions do not depend on the initial distribution of population over cities. In general when migration cost is positive, the equilibrium sorting pattern will depend on the initial distribution and the migration costs. Everything else being equal, individuals with higher human capital endowment will tend to migrate to larger cities or higher-barrier occupations.

reasonable assumptions about the human capital distribution, a unique equilibrium will arise. All the proofs are provided in Appendix C.

Proposition 1 *In any asymmetric equilibrium:*

i) There exists a cutoff x_E such that individuals choose to become entrepreneur if and only if their human capital is higher than x_E.

ii) There exists a sequence of cutoffs $x_{E1} > x_{E2} > ...x_{EJ} = x_E$, such that entrepreneur with human capital $x \in [x_{Ej}, x_{Ej+1})$ will live in the same city. We name this city "j".

Proposition 1 is illustrated in Figure 4. It states that all the individuals with human capital above a certain threshold will choose to be entrepreneurs in the H sector, and they further sort into cities: the most talented entrepreneurs will group into city 1, followed by the less-talented entrepreneurs in city 2, and so on and so forth. The mechanism behind 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, cities with better entrepreneurs are also larger (Proposition 2), and thus post higher congestion disutility. In equilibrium, the trade-off between the benefit and cost of "better neighbors" pins down the sorting of entrepreneurs.

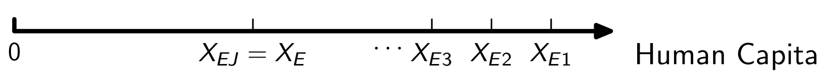


Figure 4: Sorting of Entrepreneurs

The sorting pattern of entrepreneurs lays the foundation of each city. The newly created firms, along with the demand for the products from the L sectors, attract individuals with human capital less than x_E to sort into workers in both the H and L sectors. Cities created by more-talented entrepreneurs will be larger:

Proposition 2 In any asymmetric equilibrium, if $j < i$, then $N_j > N_i$.

Within each city, individuals sort into different industry-occupation cells:

Proposition 3 *Let E_j, H_j, L_j be the set of human capital level (x) of individuals that choose to be entrepreneurs, workers in H sector, and workers in L sector in city j, respectively. In any asymmetric equilibrium: (i) wage rate is higher in H sector than in L sector in all cities: $w_j > z_j \ \forall j$, and (ii) within each city, individuals sort into occupations:*

$$\inf E_j \geq \sup H_j$$
$$\inf H_j \geq \sup L_j.$$

The individuals with highest human capital will be entrepreneurs in the H sector, followed by the workers in the H sector in each city. Those with the lowest human capital will be workers in the L sector. This is because of the differences in entry costs between industries and occupations. Working in the L sector requires zero entry cost; working in the H sector as worker requires an entry cost S; working in the H sector as entrepreneur requires the fixed cost f in addition to S. In equilibrium, jobs with higher entry barrier must pay more, otherwise no one will be willing to work. Moreover, it is relatively easier for individuals with higher human capital to overcome the higher entry cost and obtain the better-paying jobs — and thus the sorting pattern within each city arises.

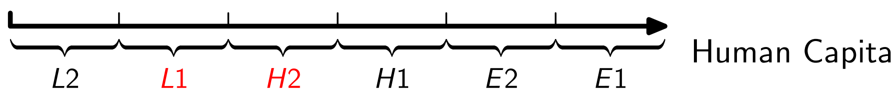
Workers in H and L sectors also sort across cities:

Proposition 4 *Let H_j, L_j be the set of human capital level (x) that individuals choose to become workers in H sector or L sector in city j, respectively. Within each occupa-*

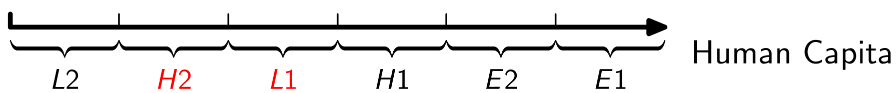
tion, individuals are sorted into cities. If $j < i$, then

$$\inf H_j \geq \sup H_i; \quad \inf L_j \geq \sup L_i.$$

The sorting pattern in Proposition 4 is weaker than that among the entrepreneurs: the union set of x of workers in either sector, $\cup_{j=1}^{J} H_j$ or $\cup_{j=1}^{J} L_j$, might not be a connected set on the real line, while the union of x of all the entrepreneurs, $\cup_{j=1}^{J} E_j$, is always a connected set. Due to this reason, Proposition 1-4 together cannot pin down a unique sorting pattern across both locations and occupations. For example, two potential sorting patterns may arise in the case of two cities:



(a) Type A Sorting



(b) Type B Sorting

Figure 5: Multiple Sorting Patterns with Two Cities

In the above example, Proposition 1 indicates that E_1 and E_2 must occupy the right end of the human capital distribution. Proposition 3 and 4 together imply that H_1 must immediately follow E_2, while L_2 must be at the left end of the human capital distribution. However, we do not know the relative positions between H_2 and L_1. In the first panel, individuals are first sorted into different occupations, and then within each occupation, they are further sorted in the cities. We call this type of sorting as "Type A" sorting. In this case, $\cup_{j=1}^{J} H_j$ and $\cup_{j=1}^{J} L_j$ are connected sets. Conversely in the second panel, workers are first sorted into different cities, and then into sectors — we name this type of sorting as "Type B". In this case, $\cup_{j=1}^{J} H_j$ are $\cup_{j=1}^{J} L_j$ are no longer connected sets.

Which sorting pattern will emerge as equilibrium depends on the value of S and $C(N_1) - C(N_2)$. Individuals will first sort by occupations if and only if $S > C(N_1) -$

$C(N_2)$, and they will first sort by cities if the reverse is true. We will formally prove the results in a general case of $J \geq 2$ cities, and here we first discuss the intuition. If the barriers between occupations are high relative to the barriers between cities, the marginal L worker with the highest possible human capital in a given city 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 $\cup_{j=1}^{J} L_j$ is a connected set and "Type A" sorting arises. If the barriers between cities are larger than S, the reverse will be true: those individuals will be indifferent between their current positions and switching into the H sector in the same city, before choosing to move to a larger city — in this case we have "Type B" sorting and all the sets $\{L_j\}$ and $\{H_j\}$ are interleaved.

In general, when there are J cities, the type of sorting pattern in equilibrium depends on the relationship between S and all the possible $C(N_j) - C(N_i)$. As $\{N_i\}$ 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. Luckily, under reasonable assumptions the city size, and thus the distribution of $C(N_j)$, approximates a power-law distribution in the limit:

Proposition 5 *If $G(x)$ follows an exponential distribution and $b_j = \exp(\bar{x}_j)$, then as $J \to \infty$:*

i) the firm employment distribution within each city follows a Power-law distribution, and

ii) city size measured in GDP follows Power-law distribution.

The first result on the firm size distribution is direct result of our assumption of human capital distribution and the functional form of $A(x)$, which follows Ma [2015]. The power law distribution of firm size is also well-documented in the literature (see Axtell [2001]). The second part of the proposition states that the city-size distribution will follow the firm size distribution, and be Pareto in the limit as well. Power-law distributions of city size are widely documented in the empirical literature, such as in Gabaix and Ioannides [2004]. It also arises in a wide array of theoretical models such as Behrens et al. [2014] and Gaubert [2014]. The power-law distribution of city size is theoretically important in our model, as it reduces the number of sorting patterns and pins down the sorting equilibrium:

Proposition 6 *If city size follows power law distribution and congestion utility is convex and monotonic in city size, then the sorting equilibrium is unique. Furthermore, there exists a j^* such that:*

$$\begin{aligned} C(N_J) + S &> C(N_{j*}) \\ C(N_J) + S &< C(N_{j*-1}) \end{aligned}$$

Individuals in cities $j > j^*$ follow Type-A sorting, and individuals living in cities $j <= j^*$ follow Type-B sorting, as shown in Figure 6.

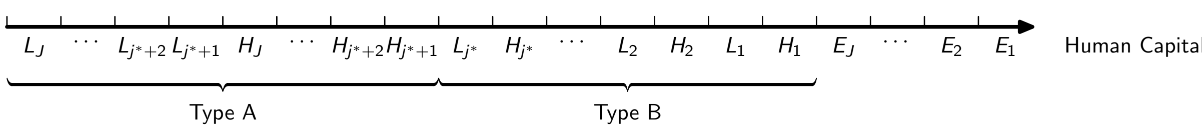


Figure 6: Sorting in N-city Case

The distribution of $\{N_i\}$ and the monotonicity and convexity of $C(\cdot)$ together imply that $C(N_j) - C(N_{j+1})$ is monotonically decreasing in j — the differences in congestion costs shrink as we move down the city size ladder. The monotonicity reduces the number of potential sorting patterns from the order of J-factorial to J. Furthermore, if we also know the tail index of the city size distribution — which, in the empirical literature, is widely estimated to be around 1 — we can directly pin down the entire sequence of the differential in city congestion costs, $\{C(N_j) - C(N_{j+1})\}$, using the log-rank property of the Pareto distribution. By comparing the value of S to the sequence of differences in congestion disutilities, we can pin down j^* as stated in Proposition 6, and thus the equilibrium.

Proposition 1- 6 together picture the sorting pattern in any asymmetric equilibrium. Next we show that in any sorting equilibrium, larger cities will have higher real GDP and consumption-equivalent utility level, attracts more productive firms. This finding is widely established empirically, such as in Combes et al. [2012] and Rosenthal and Strange [2004]. In terms of wage rate, we also prove results consistent with those in Davis and Dingel [2012] and Combes et al. [2008]: all wage rates in our model are higher in larger cities as well.

Proposition 7 *In any sorting equilibrium, if $j < i$, then the following properties hold:*

i) The real wage of H workers measured in unit cost of utility is higher in city j:

$$\frac{w_j}{P_j^{\alpha} z_j^{1-\alpha}} > \frac{w_i}{P_i^{\alpha} z_i^{1-\alpha}}.$$

ii) The real wages of workers in both sectors measured in term of H goods is higher in city j:

$$\frac{w_j}{P_j} > \frac{w_i}{P_i}, \quad \frac{z_j}{P_j} > \frac{z_i}{P_i}.$$

iii) The real income of any entrepreneur x measured in H goods is in city j is higher than that of any entrepreneur x' in city i:

$$\frac{\pi_j(x)}{P_j} > \frac{\pi_i(x')}{P_i}.$$

iv) Real GDP in city j is higher than city i:

$$\frac{R_j}{P_j} > \frac{R_i}{P_i}.$$

Lastly we show how wage inequality varies across cities:

Proposition 8 *In any sorting equilibrium, if $j < i$, then the following properties hold:*

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

$$\frac{1}{G(x \in E_j)} \int_{x \in E_j} \left(\frac{\pi(x)/x}{w} \right) dG(x) > \frac{1}{G(x \in E_i)} \int_{x \in E_i} \left(\frac{\pi(x)/x}{w} \right) dG(x).$$

ii The wage ratio between H and L sectors is smaller in city j:

$$\frac{w_i}{z_i} > \frac{w_j}{z_j}.$$

Proposition 8 is our key result. It states that entrepreneur's wage to H worker's wage ratio, which is the counter-part 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, $\pi(\cdot)$, 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. In equilibrium the entrepreneur's income equals to 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 in Glaeser and Kerr [2009], and around 0.7 for all firms in the U.S. data in 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[15], 0.05 in the French data (Combes et al. [2008]), and 0.1 in Japanese data (Tabuchi and Yoshida [2000])[16].

Our results on the right tail are also consistent with the findings in the literature that skill premium is higher in larger cities, such as in Davis and Dingel [2012], Davis and Dingel [2014] and Behrens and Robert-Nicoud [2014a]. 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 supply in the H sector. The spatial variations of entrepreneurial-skill premium is rooted in the fact that larger cities host disproportional more large firms in the equilibrium as stated in Proposition 1. Since the return to entrepreneurs' human capital is positively correlated with the size of the firm, whereas the return to workers' is purely city-specific and independent of firm size. This further implies that in the equilibrium, the difference in the rate of return to human capital between entrepreneurs and workers is greater in larger cities.

The second part of the proposition is a statement on the wage ratios between the 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 industries 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 median-to-bottom wage ratios, shall be smaller in larger cities.

[15]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.

[16]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 in Glaeser [2007] and Glaeser and Kerr [2009]. Baum-Snow and Pavan [2012] reports 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 most other estimates. Similar results can also be obtained in Glaeser and Mare [2001].

This result is driven by two forces. The first is the existence of the fixed cost of entry into H sector. Intuitively, when individuals are choosing between working in H v.s. 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 tends to be higher in larger cities, $w - z$ must be smaller in larger cities. If this condition is not met, the marginal individual between working in L and H sector in large cities will always find it better to switch to the H sector, a situation that cannot arise in equilibrium. In other words, in equilibrium 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 explains the pattern of wage inequality at the left tail in large cities. The second driving force is Proposition 7(ii), which states that z/P increases with the size of the city. This means the output price of L goods increases faster than that of the 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, in sorting equilibrium, w/z decreases with the size of the city.

The above two predictions are essentially statements on the city-size elasticity of wage rates. Proposition 7 states that the city-size elasticity of wage rates are all positive, and Proposition 8 further predicts that different wage rates shall have different size elasticities. Between different occupations, the city-size elasticity of entrepreneur compensation shall be higher than that of the workers' wage. Across different industries, the city-size elasticity of wages in low-paying industries shall be higher than that of the wages in high-paying industries. In the next section we are going to take these predictions to the data, and then further quantify their impacts on the observed patterns of within-city inequality documented in Section 3.

6 Empirical Analysis

In this section we empirically implement the model: we first test the model prediction on the differences in the city-size elasticities of wage rates, then we quantify the impacts of the differentials in city-size elasticity of wages on the spatial variation of wage inequality with counterfactual simulations.

6.1 Testing the Model Predictions

The main mechanism of our theory outlined in Proposition 7 and 8 can be summarized into the following hypothesis:

1. Between industries, the wage of working in high-paying industries shall increase with city size at slower speeds, as compared to the wage of working in low-paying industries. Therefore the spread of inter-industry wage premium shall decrease with city size.

2. Between occupations, entrepreneurs earn higher wage than workers: the relative wage premium of entrepreneurs increases with city size.

We test both hypotheses above using individual level IPUMS 2000, the same as the city-level analysis in Section 3. We impose the same data restrictions as in Section 3, and this leads us to a sample size of around 1.23 million individuals working in 201 industries [17] and 254 MSAs.

6.1.1 Inter-Industry Wage Premium and City Size

In order to test the first hypothesis, we first need to estimate the industry wage premium in each city. Recall that wage rate (w_j and z_j) in our theoretical framework are identical to all individuals in the same city: they are the residual wage after controlling for individual characteristics, summarized as x in our model. x include characteristics

[17] We use the 1990 Census Bureau industrial classification scheme as the definition of industry.

both observable and unobservable to the econometrician, but is perfectly observable for agents in the model. Therefore the sorting in our model is based on both observable and unobservable characteristics of the individuals. In the data we cannot observe the unobservable, and thus have to measure inter-industry wage premium as the residual premium by controlling for observable individual characteristics such as education attainments, age, and race. As the sorting is partially based on unobservable, our estimate for inter-industry wage premium is biased. However, the tests are not based on the level of the wage premium, but the difference of wage premium across cities. As the bias in our estimation is in the same direction for all cities, it will not affect our conclusion on the the differences in wage premiums across cities.

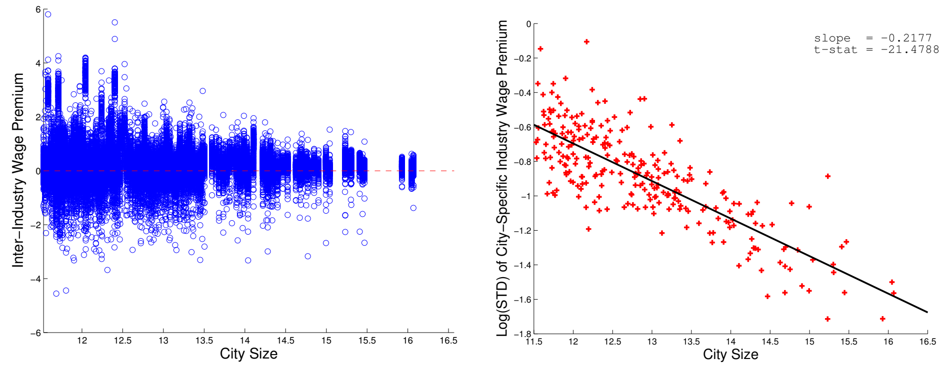
In every city j we run the following cross-sectional regression following the tradition in labor economics to estimate the wage premium of each industry relative to a benchmark industry, which we set it to be "agriculture production, crops".

$$\log W_i^j = \beta_0^j + \beta_1^j \mathbf{T_i^j} + \beta_2^j \mathbf{OC}_i^j + \gamma^j X_i^j + \epsilon_i^j, j = 1, 2, \cdots, J, \tag{2}$$

where i indexes individual, and j indexes the city in which individual i works. $\mathbf{T_i^j}$ is a vector of dummy variables to indicate the industry in which individual i works. If the individual works in industry k, then the kth element of $\mathbf{T_i^j}$ takes the value of 1, while all the other elements equal to 0. $\mathbf{OC}_i^j$ is another vector of dummy variables that controls for the occupation of individual i.[18] X_i^j is a vector that controls sex, race, years of education, marital status and age. We are interested in the vector β_1^j, which measures the wage premium of each industry in city j. Wage rate in the benchmark industry is normalized to be zero in each city, and we express the wage premium of the remaining 200 industries as percentage deviations from the benchmark industry. Our model predicts that the within-city wage premium shall have larger spreads in smaller cities.

The preliminary results of directly estimating equation (2) are reported in Figure

[18] Similar to the definition of industry, we use the 1990 census occupational classification here.



(a) Industry Wage Premium by City Size (b) Log(STD) of Ind. Wage Premium by City Size

Figure 7: Spatial Variations of Inter-Industry Wage Premium

Notes: Panel (a) plots the industry wage premium estimated in each city against the log of population. Panel (b) plots the log of standard deviation of industry wage premium within each city against the log of population. The slope and the t-statistics reported in Panel (b) are based on a simple linear regression between the two variables. Data source: IPUMS-USA 2000 and Population Census 2000.

7. The first panel of the figure plots all the estimated β_1^j against the size of city j measured in logarithm of population in year 2000. The figure confirms the prediction of our model: the wage gaps between high and low-paying industries shrink in larger cities. For cities below the median size, the industry wage premium at the top 10th percentile is on average around 0.77, while the industry wage premium at the bottom 10th percentile is around -0.29 — the spread between the two is at around 1.06. In contrast in cities with population above the median, the wage gaps are smaller by around a third: the wage premium at the top 10th percentile is at around 0.49, while it is -0.22 at the bottom 10th percentile, with a gap at around only 0.71.

The second panel of the same figure further measures the spread of industry wage premium within each city j using the logarithm of standard deviations of β_1^j. We plot the results against the the logarithm of population. The slope between the two variables stands significantly negative at around -0.217: on average, if city size doubles, the log standard deviation of industry wage premium will drop by around 21.7 percent.[19]

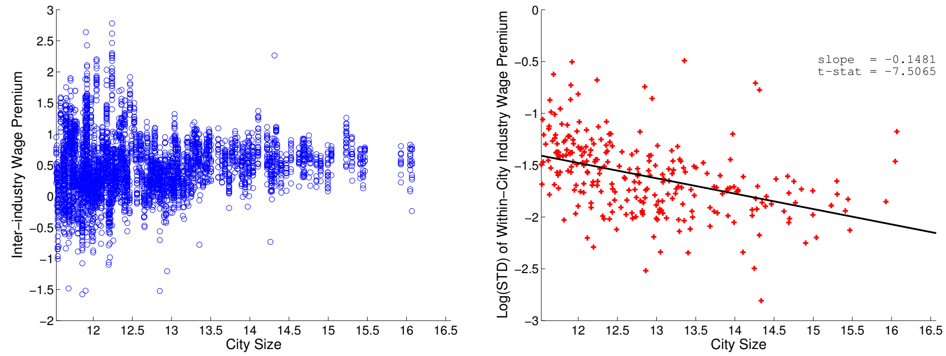
The above preliminary estimations of equation (2) suffer from sample size bias.

[19] Robustness checks using the GDP as measure of city size are reported in Figure 10 in the appendix. The results are essentially the same.

Naturally larger cities have more observations in the estimation of inter-industry wage premiums, and this could mechanically lead to lower within-city variances in larger cities — the premiums are simply more precisely measured in larger cities. We use Monte-Carlo re-sampling to correct for the differences in sample sizes. The smallest city in our estimation of equation (2) has 445 observations. We randomly draw 445 observations from each city, and re-estimate equation (2) based on this balanced dataset. We repeat the above re-sampling process 1,000 times, and approximate the industry-premium in each city with the sample average across all the 1,000 simulations. We report the results with the weighted average where the weight is the inverse of the standard errors of estimating β_1^j in each simulation. The results are reported in Figure 2 in the introduction. We report the robustness checks with simple averages in Figure 11.

The main message remains the same. With balanced sample sizes, the spread of industry-wage premiums is still significantly smaller in larger cities. The right panel also suggest that the differences in sample sizes indeed bias our preliminary estimates upwards. After we correct for the sample size bias, doubling the city size only leads to a 11.8 to 16.0 percentage point decreases in the standard deviations of within-city industry premiums, as compared to the 21.45 percentage points decrease in the preliminary estimations. Nevertheless, the negative slope is statistically significant and economically sizable.

Some inter-industry wage premium cannot be estimated in every city and every simulation in the results above. Due to small sample size, we might not be able to draw any individual working in a certain industry-city cell. This implies that while some points in the left panels of Figure 2 are computed from averages across 1,000 simulations, some other points are instead based on far smaller numbers of simulations. As the sampling bias could again, be more severe in smaller cities, our results provided in Figure 2 might still be biased in our favor. We therefore carry out another estimation of inter-industry wage premium based only on the industries-city cells that can be observed in all 1,000 simulations. The results are reported in Figure 8.



(a) Industry Wage Premium by City Size, Weighed Average

(b) Log(STD) of Ind. Wage Premium by City Size, Weighed Average

Figure 8: Spatial Variations of Inter-Industry Wage Premium, Monte-Carlo Simulation, Restricted Sample

Notes: We report the similar results as in Figure 2. The difference here is that we require all the points in the left panel are industry-city cells that show up in all 1,000 simulations. For more details, please refer to the note to Figure 2. Data source: IPUMS-USA 2000 and Population Census 2000.

The above restriction drops a number of industry-city cells in our scatter plots. Nevertheless, the main pattern – larger cities see smaller spread of inter-industry wage premium, emerges from the remaining sample. The right panels of the same figure also suggests that indeed, the differences in number of simulations in Figure 2 bias our estimates upwards. In the strictly balanced dataset, doubling the city size leads to a 7.5 to 9.2 percentage decrease in the standard deviations of within-city inter-industry wage premium.

6.1.2 Entrepreneurial Premium and City Size

Our model also predicts that entrepreneurs earn more than workers in the H sector, and the entrepreneurs' profit to workers' wage ratio should increase with city size. Similar to the previous set of tests, we estimate the following equation:

$$\log W_i = \beta_0 + \beta_1 E_i + \beta_2 E_i \cdot \log\left(Y_i^j\right) + \beta_3 T_i + \gamma X_i + \epsilon_i, \tag{3}$$

where i indexes individual, and j indexes the city in which the individual works. E_i is a binary variable that takes the value of 1 when the individual is broadly considered as an entrepreneur. In our benchmark regression we define entrepreneur based on the 1990 occupational classification from the Bureau of Labor Statistics. An individual is defined as an entrepreneur if she works as chief executives or as managers of finance, marketing, human resources, etc. We report the detailed definition of entrepreneurship in Table 10. Y_i^j measures the size of city j that individual i works in, T_i is the vector of dummy variables that controls for the industry in which individual i works, and X_i is a vector that controls for other individual characteristics. The key parameter of interest here is β_2. Our model 8 predicts that the entrepreneurial premium shall be higher in larger cities, and thus $\beta_2 > 0$.

The results are reported in the first panel of Table 8, and they are consistent with the predictions of our model. First, the wage and income of entrepreneurs is higher than those of the workers: both β_1 and β_2 are significantly positive. More importantly, the entrepreneurship premium *increases* with the size of the city. On average, doubling the city size is associated with a 1.3 percent increase in entrepreneurship premium.

This result is fairly robust to the changes in estimation specification. In the second panel of Table 8 we drop the dummy variable E_i and instead add in a vector of occupational dummies variables to control for the unobserved characteristics that vary with occupations. The estimated size elasticity is still significantly larger than zero. The magnitude of β_2 is larger in this case: doubling the city size now leads to a 2.1 percent increase in the entrepreneurship premium. Table 9 presents two other sets of robustness checks in which we vary the definition of entrepreneurs. The top panel uses a more restrictive definition: E_i only takes the value of 1 if the reported occupation is "chief executives". Under this coding, the population of entrepreneurs in the sample drops from 119,342 to only 15,482. We observe a higher entrepreneurship premium, and a positive size elasticity of the premium, though the standard error of the estimations increased significantly due to smaller sample of entrepreneurs.

We define the entrepreneurs to be managers. However, managers are not always the owners of the firm, but professionals hired by the owners to manage the on-going business. This definition is not the same as in our model, where the owners and the managers of the firm are the same. We use this definition mainly due to data restrictions: "managers" are clearly defined within the occupational classification codes, while owners of firms are usually defined as "self-employed", a category that pools the business owners in the H sector and the self-employed in the L sector together in our model. For our estimation, the differences between owners and top managers are not crucial either: both of them are residual claimants to the firm, and thus their income is likely to move together and scale with the size and performance of the firm they own or manage. Nevertheless, the distinction between the managers and owners introduce bias in our estimation of the size elasticity upward. As a robustness check, we use a broader definition of entrepreneurs to include the self-employed individuals. The inclusion of self-employed is likely to bias our estimation downwards, as many of the self-employed work in the L sectors. The results under the broad definition are reported in the second panel in Table 9. As expected, once self-employed individuals are classified as entrepreneurs, the coefficient on the entrepreneur dummy drops from 0.183 to around 0.074, indicating that the wage and income of the self-employed is lower as compared to the top managers in our benchmark sample. The coefficient on the interaction between entrepreneurship and city size, β_2, is roughly the same as compared to the benchmark estimation. This is probably due to the small sample size of self-employed individuals: there are only 60,553 individuals that are self-employed out of an estimation sample of 1.3 million.

6.2 Counterfactual Simulations

We proceed to quantify how much the observed spatial variations of within-city wage inequality can be explained by the spatial variations of inter-industry wage premium predicted by our model. To do this, we perform a counterfactual simulation, where we shut down the spatial variations of industry wage premium, and evaluate how city-level wage inequality may respond to the changes.

We first compute a benchmark measure of wage that incorporates the spatial variations of industry wage premium. We define the benchmark as the linear prediction based on the estimation of Equation (2):

$$\log \widehat{W_i^j} = \widehat{\beta}_0^j + \widehat{\beta}_1^j \mathbf{T_i^j} + \widehat{\beta}_2^j \mathbf{OC}_i^j + \widehat{\gamma}^j X_i^j, j = 1, 2, \cdots, J,$$

where $\widehat{\beta}_0^j$, $\widehat{\beta}_1^j$, $\widehat{\beta}_2^j$, and $\widehat{\gamma}^j$ are the OLS estimates of their counterparts in Equation (2). $\widehat{\beta_1^j}$ measures the industry wage premium in city j, and its spread within city varies across different cities. To eliminate the spatial variations in the spread of $\widehat{\beta_1^j}$, we compute counterfactual wage premium, $\widetilde{\beta_1^j}$, as follows:

$$\widetilde{\beta}_1^j = \bar{\sigma} \frac{\widehat{\beta_1^j}}{\sigma(\widehat{\beta_1^j})}.$$

In the above transformation, $\sigma(\widehat{\beta_1^j})$ is the standard deviation of $\widehat{\beta_1^j}$ within city j, and $\bar{\sigma}$ is a scaling factor that we define as the average $\sigma(\widehat{\beta_1^j})$ of the largest five cities.[20] After the transformation $\widetilde{\beta}_1^j$ will have the same standard deviation of $\bar{\sigma}$ across all cities.

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

$$\log \widetilde{W_i^j} = \widehat{\beta}_0^j + \widetilde{\beta}_1^j \mathbf{T_i^j} + \widehat{\beta}_2^j \mathbf{OC}_i^j + \widehat{\gamma}^j X_i^j, j = 1, 2, \cdots, J,$$

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

Given benchmark wage $\widehat{W}_i^j$ and the counterfactual wage $\widetilde{W_i^j}$, we repeat the same exercise as in Section 3 by first computing various measures of inequality within each city, then estimate Equation (1) to study how within-city inequality varies across cities.

[20]The size of $\bar{\sigma}$ can be arbitrary — we define $\bar{\sigma}$ in such way to create a counterfactual world where the spread of inter-industry wage premium in all cities are as small as those in the largest cities.

The results are reported in Table 3.

(a) Benchmark Estimation

VARIABLES	(1) 50/10 Ratio	(2) 50/10 Ratio	(3) 50/10 Ratio	(4) 50/01 Ratio	(5) 50/01 Ratio	(6) 50/01 Ratio
Total GDP	-0.0373***			-0.112***		
	(0.00565)			(0.0116)		
Private Ind. GDP		-0.0361***			-0.111***	
		(0.00546)			(0.0110)	
Population			-0.0398***			-0.124***
			(0.00589)			(0.0117)
Observations	254	254	264	254	254	264
R-squared	0.608	0.609	0.621	0.539	0.550	0.553

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) Counterfactual Estimation

VARIABLES	(1) 50/10 Ratio	(2) 50/10 Ratio	(3) 50/10 Ratio	(4) 50/01 Ratio	(5) 50/01 Ratio	(6) 50/01 Ratio
Total GDP	-0.0167***			-0.0773***		
	(0.00484)			(0.0122)		
Private Ind. GDP		-0.0167***			-0.0782***	
		(0.00465)			(0.0114)	
Population			-0.0182***			-0.0829***
			(0.00542)			(0.0115)
Observations	254	254	264	254	254	264
R-squared	0.663	0.664	0.677	0.491	0.504	0.534

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

.

Table 3: Benchmark and Counterfactual Regressions

Note: The first panel of the table reports the estimation of Equation (1) with the benchmark $\widehat{W_i^j}$. The second panel of the table reports the same estimation with the counterfactual $\widetilde{W_i^j}$ instead of the data. We control for the same variables as in Section 3. All the other control variables are omitted in the tables. For more details, see the main text in Section 3 and 6.

The first panel of the Table 3 reports the estimation of Equation (1) based on the benchmark $\widehat{W_i^j}$. The second panel reports the results based on the counterfactual $\widetilde{W_i^j}$. 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. For example, when we measure city size by total and private GDP, the size elasticity of 50/10 wage ratio is -0.037 and -0.036 in the benchmark case, respectively. In the counterfactual world, the size elasticities decrease by around 55 percent to -0.017. When we measure city size by population, the change is about the same magnitude:

the size elasticity decreases by 54 percent from -0.040 to -0.018. Similar results can be observed when we use 50/01 wage ratio. In this case, the size elasticity drops between 30 and 36 percent, from $[-0.124, -0.111]$ to $[-0.083, -0.078]$, varying slightly across different measures of city size. This suggests that between 1/3 and 1/2 of the spatial variations of within-city inequality reported in Section 3 can be explained by the spatial variations of inter-industry wage premium, as predicted by our model.

7 Conclusion

Our findings from the U.S. data suggest that top-to-median wage inequality tends to increase with city size, while the median-to-bottom inequality is decreasing with city size. We then develop a competitive spatial equilibrium framework with the feature that inequality measured at the top and bottom percentile can potentially move in opposite directions with respect to city size. Our model can also provide insights on the inter-industry wage premium puzzle by arguing that wage rate in low-paying industries needs to increase faster with city size than high-pay industries, otherwise the low talented individuals will not be able to survive in large cities due to the high living costs. The counter-factual simulation suggests that inter-industry premium can account for about half of observed spatial variations in within-city wage inequality.

Our model is rich yet highly tractable, and thus it admits the possibility for future quantitative work. The framework can be used to quantify the relative contributions of regional trade and migration barrier to wage inequalities and firm size distributions across cities. Moreover, for simplicity the industry entry barrier is exogenous in the current setup, and in reality this may correspond to search frictions and education cost etc. For future work, we may also consider to endogenize the entry barrier and enrich the labor market dynamics by allowing unemployment and job turnover etc.

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.

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820.

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

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. and F. Robert-Nicoud (2014a): "Agglomeration Theory with Heterogeneous Agents," Research Papers by the Institute of Economics and Econometrics, Geneva School of Economics and Management, University of Geneva 14096, Institut d'Economie et Economtrie, Universit de Genve.

——— (2014b): "Survival of the Fittest in Cities: Urbanisation and Inequality," *The Economic Journal*, 124, 1371–1400.

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 (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection," *Econometrica*, 80, 2543–2594.

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.

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. (2014): "Firm Sorting and Agglomeration," Working paper.

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. Resseger, and K. Tobio (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.

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

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 Interindustry Wage Structure," *Econometrica*, 56, 259–93.

la Roca, J. D. and D. Puga (2016): "Learning By Working In Big Cities," *Review of Economic Studies*.

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

Ma, L. (2015): "Globalization and Top Income Shares," Working Paper.

Melitz, M. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

Mortensen, D. (2005): *Wage Dispersion: Why are Similar Workers Paid Differently?*, MIT press.

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.

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].

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

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

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.

Appendix:

A Solving the Model

A.1 Firm's Problem

The firm's problem can be solved as in a standard Melitz [2003] model. Each firm charges a constant markup over its marginal cost of production. Specifically, the price, quantity and profits for tradable goods produced in city k and exported to city j is:

$$\begin{aligned} p_{jk}(x) &= \frac{\sigma}{\sigma-1}\frac{\tau_{jk}w_k}{A_k(x)} \\ q_{jk}(x) &= \alpha R_j P_j^{\sigma-1}\left(\frac{\sigma-1}{\sigma}\frac{A_k(x)}{\tau_{jk}w_k}\right)^{\sigma} \\ \pi_{jk}(x) &= \left(\frac{\sigma}{\sigma-1}\frac{\tau_{jk}w_k}{A_k(x)}\right)q_{jk}(x) - f_{jk}w_k, \end{aligned}$$

where R_j is the total expenditure in city j, and P_j is the ideal price index of H goods in city j:

$$P_j = \left\{\sum_{k=1}^{J}\int_{x\in\Omega_{jk}} p_{jk}(x)^{1-\sigma}dG\left(x\right)\right\}^{\frac{1}{1-\sigma}}.$$

Ω_{jk} is the set of varieties exported from city k to city j. Firm x in city k will choose to export to city j if and only if $\pi_{jk}(x) \geq 0$, and therefore:

$$\Omega_{jk} = \left\{x|\pi_{jk}(x) \geq 0\right\}.$$

The total profit of firm x located in city j is thus:

$$\pi_j(x) = \sum_{k=1}^{J}\pi_{kj}(x)\cdot\mathbf{1}(x\in\Omega_{kj}).$$

A.2 Individual's Problem: Occupation, Industry, and Location Choices

Next we turn to the problem of each individual. Denote $V_j^E(x)$, $V_j^H(x)$ and $V_j^L(x)$ to be the indirect utility of an individual with human capital level x working as an entrepreneur, worker in H, and worker in L sector in city j, respectively:

$$\begin{aligned} V_j^E(x) &= \frac{\pi_j(x)}{P_j^\alpha z_j^{1-\alpha}} - S - C(N_j) \\ V_j^H(x) &= \frac{w_j \cdot x}{P_j^\alpha z_j^{1-\alpha}} - S - C(N_j) \\ V_j^L(x) &= \frac{z_j \cdot x}{P_j^\alpha z_j^{1-\alpha}} - C(N_j). \end{aligned}$$

The payoff for individual x in city j is thus denoted as:

$$V_j(x) = \max\left\{V_j^E(x), V_j^H(x), V_j^L(x)\right\}.$$

Individual will choose the city to maximize her utility:

$$\max_j \{V_j(x)\}_{j=1}^J.$$

B Congestion Disutility

We provide a very simple framework in this section to micro-found the assumption that individual's utility is decreasing with the population size of the city. Consider a representative worker in city j whose utilities positively depend on the consumption of goods and land. Specifically, the utility function takes the following form:

$$u(c_j, \ell_j) = \log(c_j) + \theta \log(\ell_j).$$

Denote wage rate in city j to be w_j. The budget constraint can thus be written as:

$$c_j + p_j \ell_j = w_j.$$

The demand for land is thus:

$$\ell_j = \frac{\theta w_j}{p_j(1+\theta)}.$$

Let the total land supply in city j be fixed at $\bar{L}_j$, and population size of city j be N_j. Land market clearing condition essentially implies that individual's utility equals:

$$U_j = \log\left(\frac{w_j}{1+\theta}\right) + \theta \log\left(\frac{\bar{L}_j}{N_j}\right).$$

It is straightforward to show individual's utilities increase with wage rate and decrease with population size of city j. Moreover, the land price can be pinned down from the land market clearing condition:

$$p_j = \frac{\theta w_j N_j}{(1+\theta)\bar{L}_j}.$$

C Proofs

C.1 Proposition 1

Proof (i) Wage income for an entrepreneur and a worker in any city k can be expressed as follows, respectively,:

$$\begin{aligned}\pi_k(x) &= \frac{1}{\sigma}\sum_{j=1}^{J}\left\{\left[\alpha R_j P_j^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\frac{\tau_{jk}w_k}{b_k e^x}\right)^{1-\sigma} - f_{jk}w_k\right]\cdot \mathbf{1}(x\in\Omega_{jk})\right\}\\ w_k(x) &= w_k x.\end{aligned}$$

Since $\pi_k(x)$ is increasing and convex in x, while $w_k(x)$ is linear in x, and thus it is straightforward to show

$$lim_{x\to\infty}\frac{\pi_k(x)}{w_k(x)} = \infty.$$

Therefore, individuals with higher human capital will choose to become an entrepreneur.

(ii) For any arbitrary index of cities, define $j^*(x)$ to be individual $x's$ most preferred city. Since entrepreneur's payoff is strictly increasing and convex in x, and this implies if individual of x prefers city $j^*(x)$ to other cities, then individuals with human capital level higher than x will all prefer city $j^*(x)$ to others.

Define x_{E1} to be the cutoff such that individuals with $x \geq x_{E1}$ have the same most preferred city, and we name this city "1". Similarly, denote x_{E2} to be the cutoff such that individuals with $x_{E1} > x \geq x_{E2}$ have the same most preferred city, and we name this city "2". Repeat the previous steps, we can find the sequence of $x_{E2} > ...x_{EJ} = x_E$, such that entrepreneur with human capital $x \in [x_{Ej}, x_{Ej+1})$ will live in the same city. We name this city "j". This completes the proof. ∎

C.2 Proposition 3

Proof See Figure 9 for intuition.

(i) In any city j, the indirect utility from being an H or L worker can be expressed as follows, respectively:

$$V_j^H(x) = w_j x(\frac{\alpha}{P_j})^{\alpha}(\frac{1-\alpha}{z_j})^{1-\alpha} - S - C(N_j)$$
$$V_j^L(x) = z_j x(\frac{\alpha}{P_j})^{\alpha}(\frac{1-\alpha}{z_j})^{1-\alpha} - C(N_j).$$

First, if $w_j > z_j$ does not hold, no individual prefers to work in the H sector, and this leads to an infinitely large labor demand for H workers, and a sufficiently high wage rate and a contradiction with our presumption. Therefore, $w_j > z_j$ must hold in any city j.

(ii) The indirect utility functions described above are increasing and linear in x, and thus these two lines only cross once. The single crossing property guarantees that individuals with higher x prefers to become H to L workers. We have also shown in Proposition 1 that most talented individuals will choose to become entrepreneur. Therefore, in any city j, individuals are sorted by their human capital levels into different occupations. ∎

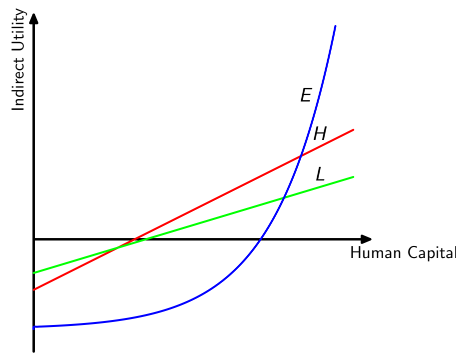


Figure 9: Sorting within a City

Note: This graph illustrates the results of Proposition 3. It plots the indirect utility (V) of individuals with different occupations in the same city against human capital.

C.3 Lemma

Proof We prove by contradiction. Suppose city j has smaller size than i, since congestion dis-utility positively depends on city size, and this implies there is a higher congestion dis-utility in city i than j. Moreover, the following has to hold in order for city i to be non-empty:

$$\frac{w_j}{P_j^\alpha z_j^{1-\alpha}} < \frac{w_i}{P_i^\alpha z_i^{1-\alpha}},$$

otherwise no individual will choose to be a worker in the H sectors of city i. Moreover, the fixed cost in utilities for entrepreneur to live in city i can be expressed as: $f\frac{w_i}{P_i^\alpha z_i^{1-\alpha}} + C(N_i)$. Our argument suggests that fixed cost for entrepreneurs living in city i is higher than city j. Therefore, entrepreneurs live in city i will have higher human capital level than j. But this contradicts with the prediction of Proposition 1, where city j is constructed to be a more talented city. Hence, we must have city j has a larger city size. ∎

C.4 Proposition 4

Proof $j < i$ implies city j has a larger city size, and thus a higher congestion dis-utility level: $C(N_j) > C(N_i)$ for all workers in the city. Therefore, within each occupation it must be those individuals with higher human capital choose to live in city j. Otherwise city j will be empty. This completes the proof. ∎

C.5 Proposition 5

Proof: When sorting equilibrium exists on a continuum of cities, by definition, there exists a one-to-one mapping between the "number" of cities, J, and the interval $[x_E, +\infty)$, where x_E is the cutoff of entrepreneurs as defined in Proposition 1. Under this condition, each city has only one type of entrepreneur x and $\bar{x} = x$. When $b(\cdot) = exp(\bar{x})$ holds, $A_j(x)$ follows a Pareto distribution:

$$\begin{aligned}
\Pr\left(A_j(x) \leq y\right) &= \Pr\left(e^{\bar{x}} e^{x} \leq y\right) \\
&= \Pr\left(e^{x} e^{x} \leq y\right) \\
&= \Pr\left(x \leq \frac{1}{2}\log y\right) \\
&= 1 - y^{-\frac{2}{\lambda}},
\end{aligned}$$

The last equation, which comes from the assumption that x follows an exponential distribution, is the CDF of a Pareto distribution with tail index $2/\lambda$.

Under the assumption of a continuum of cities, GDP in each city j can be expressed as:

$$\begin{aligned}
R_j &= \frac{1}{\alpha} P_j^{1-\sigma} \sum_k \alpha R_k P_k^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \frac{\tau_{kj} w_j}{b_j e^x P_j}\right)^{1-\sigma} \\
&= \frac{1}{\alpha} P_j^{1-\sigma} \sum_k \alpha R_k P_k^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \frac{\tau_{kj} w_j}{P_j}\right)^{1-\sigma} \cdot A_j(x)^{\sigma-1},
\end{aligned}$$

which is just the sales of the sole entrepreneur in city j. As it is standard in a Melitz model, H_j is proportional to the entrepreneur's productivity. This implies that R_j also follows a Pareto distribution with tail index equals to $2/[\lambda \cdot (\sigma - 1)]$. ∎

C.6 Proposition 6

Proof: Our proof takes the following three steps:

(i) if city size follows a Pareto distribution, then Pareto distribution properties guarantee the following relation between rank of city-size and city size:

$$log(i) - log(j) = \frac{1}{\alpha}\left[log(Size_j) - log(Size_i)\right]$$

where α denotes the tail-index of the Pareto distribution. This implies if $i < j$, then $Size_j < size_i$, moreover, $log(Size_i) - log(Size_{i+1})$ is decreasing in city rank

i. Without loss of generosity, we let congestion dis-utility function takes the following form:

$$B\,(size) = B * size^{\eta},\ \eta \geq 1$$

We can show $\{C(N_j) - C(N_{j+1})\}$ is decreasing in j since $C(N_j) - C(N_{j+1}) = B * size^{\eta}_{j+1}\left(\left(\frac{size_j}{size_{j+1}}\right)^{\eta} - 1\right)$. When $\eta \geq 1$, $\left(\frac{size_j}{size_{j+1}}\right)^{\eta} - 1 > 0$ holds. Moreover, both $\left(\frac{size_j}{size_{j+1}}\right)$ and $size_{j+1}$ are decreasing in j, and these together establish the argument.

(ii) Next, we show j^* defined in the proposition also satisfies:

$$\begin{array}{rcl} C(N_{j*}) - C(N_{j^*+1}) & < & S \\ C(N_{j*-1}) - C(N_{j^*}) & \geq & S \end{array}$$

Since $C(N_J) + S > C(N_{j*})$, this implies: $C(N_{j*}) < S$. Moreover,

$$\begin{array}{rcl} S & = & C(N_J) + S - C(N_J) \\ & = & (C(N_{J-1}) - C(N_J)) + (C(N_{J-2}) - C(N_{J-1})) + ... \\ & & \big(C(N_{j^*}) - C(N_{_{j^*+1}})\big) + (C(N_J) + S - C(N_{j^*})) \end{array}$$

Each of the element is non-negative, so $S > C(N_{j^*}) - C(N_{_{j^*+1}})$ holds. Furthermore,

$$C(N_{j*-1}) - C(N_{j^*}) > C(N_J) + S - C(N_{j^*}) > 0$$

This completes the arguments.

(iii) The previous arguments enable us to rank the fixed cost to work in any city j as a worker in sector H or L as follows:

$$\begin{array}{rcl} C(N_J) & < & C(N_{J-1})...C(N_{j^*}) < C(N_J) + S < ...C(N_{j^*}) + S < ... \\ C(N_{j^*-1}) & < & C(N_{j^*-1}) + S < ...C(N_1) < C(N_1) + S \end{array}$$

It is straightforward to prove that if individual of human capital x prefers certain city/occupation to other combinations, then all individuals with higher human capital level will prefer the same. Therefore, in our non-empty city equilibrium, it must be the case that individuals with higher human capital levels choose to work in the city/occupation with higher fixed cost, otherwise certain city/occupation will be "empty". This completes proof. ∎

C.7 Proposition 7

Proof: i) For a worker with human capital level x, his payoff from staying in city i as a worker in H sector is given as:

$$V_i^W(x) - C(N_j) - S = w_i x(\frac{\alpha}{P_i})^{\alpha}(\frac{1-\alpha}{z_i})^{1-\alpha} - C(N_i) - S$$

In our constructed equilibrium, agents with higher x will prefer to work in city j to city i. According to the single-crossing property, this implies: $V_i^W(x) - C(N_i) - V_j^W(x) + C(N_j)$ is monotonically decreasing in x, that is:

$$w_i(\frac{\alpha}{P_i})^{\alpha}(\frac{1-\alpha}{z_i})^{1-\alpha} < w_j(\frac{\alpha}{P_j})^{\alpha}(\frac{1-\alpha}{z_j})^{1-\alpha}$$

The above is equivalent to:

$$\frac{W_j}{P_j^{\alpha} Z_j^{1-\alpha}} > \frac{W_i}{P_i^{\alpha} Z_i^{1-\alpha}} \tag{4}$$

ii) similarly, for an agent with human capital level x, his payoff from staying in city i as a worker in L sector is given as:

$$V_i^N(x) - C(N_j) = z_i x(\frac{\alpha}{P_i})^{\alpha}(\frac{1-\alpha}{z_i})^{1-\alpha} - C(N_i)$$

Again, in our constructed equilibrium, single-crossing property implies:

$$z_j(\frac{\alpha}{P_j})^{\alpha}(\frac{1-\alpha}{z_j})^{1-\alpha} > z_i(\frac{\alpha}{P_i})^{\alpha}(\frac{1-\alpha}{z_i})^{1-\alpha}$$

The inequality above is equivalent to:

$$\frac{z_j}{p_j} > \frac{z_i}{p_i} \tag{5}$$

Moreover, (4) and (5) also imply:

$$\frac{w_j}{P_j} > \frac{w_i}{P_i} \tag{6}$$

iii) When $j < i$, entrepreneur x chooses city j over i iff the following holds:

$$\frac{\pi_j(x)}{P_j^{\alpha} z_j^{1-\alpha}} > \frac{\pi_i(x)}{P_i^{\alpha} z_i^{1-\alpha}}$$

Moreover, from (ii) we have shown that $z_j/p_j > z_i/p_i$, and thus in order to have the inequality above hold, we need:

$$\frac{\pi_j(x)}{P_j} > \frac{\pi_i(x)}{P_i}$$

In our sorting equilibrium, individuals with higher x will be entrepreneurs in city j, so

$$\frac{\pi_j(x)}{P_j} > \frac{\pi_i(x')}{P_i}$$

This completes the proof.
iv) GDP in city j can be written as total sales of goods in H and L sector, that is:

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

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

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

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

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

So 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. $\blacksquare$.

C.8 Proposition 8

Proof: Define x_j to be the cutoff where agent is indifferent between becoming a worker in H and L sector in city j [21], this implies:

$$(\frac{\alpha}{P_j})^\alpha(\frac{1-\alpha}{z_j})^{1-\alpha} w_j x_j - S = (\frac{\alpha}{P_j})^\alpha(\frac{1-\alpha}{z_j})^{1-\alpha} z_j x_j$$

In the following we will prove if $j < i$, then $x_j > x_i$. Suppose not, then there exists some $x \in (x_j, x_i)$ and those individuals will prefer H_j to L_j,but L_i to H_i. We have shown in Proposition 2 that within each occupation individuals are sorted into cities. Therefore, for any individual of x who prefers L_i to H_i, we have $x < \inf H_i$. Moreover, the following also holds:

$$\inf H_i < \sup H_i < \inf H_j$$

[21] This is a hypothetical cutoff. In the equilibrium, individuals with x slightly below x_j may not work in L sector of city j

Therefore, individual of x will not work in city $j's$ H sector.

Similarly, for those who prefer H_j to L_j we have $x > \sup L_j$ and the following:

$$x > \sup L_j > \inf L_j > \sup L_i$$

So those individuals will not in city $i's$ L sector either. Hence, there is no city for individuals with $x \in (x_j, x_i)$ to work. This contradicts with our equilibrium definition. Therefore, the interval (x_j, x_i) does not exist, and $x_j > x_i$ holds.

We can then solve x_j as follows:

$$x_j = \frac{s}{(\frac{\alpha}{P_j})^{\alpha}(\frac{1-\alpha}{z_j})^{1-\alpha}(w_j - z_j)}$$

$x_j > x_i$ implies:

$$(\frac{w_j - z_j}{P_j})^{\alpha}(\frac{w_j - z_j}{z_j})^{1-\alpha} < (\frac{w_i - z_i}{P_i})^{\alpha}(\frac{w_i - z_i}{z_i})^{1-\alpha}$$

The inequality above is equivalent to:

$$(\frac{z_j}{P_j})^{\alpha}\frac{w_j - z_j}{z_j} < (\frac{z_i}{P_i})^{\alpha}\frac{w_i - z_i}{z_i}$$

we have shown $\frac{z_j}{P_j} > \frac{z_i}{P_i}$, so in order to have inequality above hold we need: $\frac{w_j - z_j}{z_j} < \frac{w_i - z_i}{z_i}$, which can be reduced to: $\frac{w_j}{z_j} < \frac{w_i}{z_i}$. This establishes the proof.

ii) The proof takes two steps: first, we show within any city $\pi_j(x)/x$ is increasing in x. Since the wage rate is fixed within each city, and thus if $\pi_j(x)/x$ is increasing in x, then so is $\frac{\pi(x)/x}{w}$. This implies in any city j the minimum of $\pi_j(x)/x$ should be equal $\pi_j\left(\underline{x}_j\right)/\underline{x}_j$, where $\underline{x}_j$ denotes the efficiency labor supply that the least talented entrepreneur has. Second, we show $\frac{\pi_j\left(\underline{x}_j\right)}{w_j \underline{x}_j} > \frac{\pi_{j+1}\left(\underline{x}_j\right)}{w_{j+1}\underline{x}_j}$. This statement implies the minimum value of $\frac{\pi_j\left(\underline{x}_j\right)}{w_j \underline{x}_j}$ is still larger than the maximum value of $\frac{\pi_{j+1}\left(\underline{x}_j\right)}{w_{j+1}\underline{x}_j}$ in city $j+1$.

First, we can express $\pi_j(x)/x$ in city j as follows:

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

To show $\pi_j(x)/x$ is increasing in x, we need to show $\frac{(b_j e^x)^{\sigma-1}}{x}$ is increasing in x.This is true when x is large enough. This condition is easily satisfied in our model, as the entrepreneurs are those with the highest efficiency labor supply, as stated in Preposition 1.

Second, to show $\frac{\pi_j(\underline{x}_j)}{w_j} > \frac{\pi_{j+1}(\underline{x}_j)}{w_{j+1}}$ is equivalent to show $\frac{\partial \frac{\pi(\underline{x})}{w}}{\partial \underline{x}} > 0$ Given $\bar{x}$, where a larger value of $\underline{x}$ implies a larger city. We express profit and real GDP in city j as follows:

$$\begin{aligned}
\pi_j(\underline{x}) &= \frac{1}{\sigma}\sum_k \alpha R_k P_k^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\frac{\tau_{kj}w_j}{b_j e^{\underline{x}}}\right)^{1-\sigma} - f w_j \\
R_j(x) &= \int_{x\in E_j}\sum_k \alpha R_k P_k^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\frac{\tau_{kj}w_j}{b_j e^{x}}\right)^{1-\sigma} dG(x)
\end{aligned}$$

Therefore, we have

$$\frac{\pi_j(\underline{x}) + f w_j}{R_j} = \frac{\left(b_j e^{\underline{x}}\right)^{\sigma-1}}{\int_{x\in E_j}\left(b_j e^{x}\right)^{\sigma-1} dG(x)}$$

That is,

$$\frac{\pi_j(\underline{x})}{P_j} = \frac{R_j}{P_j}\frac{\left(b_j e^{\underline{x}}\right)^{\sigma-1}}{\int_{x\in E_j}\left(b_j e^{x}\right)^{\sigma-1} dG(x)} - \frac{f w_j}{P_j}$$

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

$$\frac{w_j}{P_j} = \left(\int_{y\in E_j}\left(b_j e^{y}\right)^{\sigma-1} dG(y)\right)^{\frac{1}{\sigma-1}}\left(\frac{\sigma}{\sigma-1}\right)^{-1}$$

Therefore,

$$\frac{\pi_j(\underline{x})}{w_j} = \frac{\frac{\pi_j(\underline{x})}{P_j}}{\frac{w_j}{P_j}} = \frac{\frac{R_j}{P_j}\frac{\left(b_j e^{\underline{x}}\right)^{\sigma-1}}{\int_{x\in E_j}\left(b_j e^{x}\right)^{\sigma-1} dG(x)} - \frac{f w_j}{P_j}}{\left(\int_{y\in E_j}\left(b_j e^{y}\right)^{\sigma-1} dG(y)\right)^{\frac{1}{\sigma-1}}\left(\frac{\sigma}{\sigma-1}\right)^{-1}}$$

We need to find the condition such that $\left(b_j e^{\underline{x}}\right)^{\sigma-1}\left(\int_{y\in E_j}\left(b_j e^{y}\right)^{\sigma-1} dG(y)\right)^{\frac{\sigma}{1-\sigma}}$ is increasing in x, where

$$\begin{aligned}
&\int_{y\in E_j}\left(b_j e^{y}\right)^{\sigma-1} dG(y) \\
&= b_j^{\sigma-1}\lambda\left[\frac{e^{\bar{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda} - \frac{e^{\underline{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}\right]
\end{aligned}$$

So

$$\begin{aligned}
&\left(b_j e^{\underline{x}}\right)^{\sigma-1}\left(\int_{y\in E_j}\left(b_j e^{y}\right)^{\sigma-1} dG(y)\right)^{\frac{\sigma}{1-\sigma}} \\
&= b_j^{-1} e^{\underline{x}(\sigma-1)}\left(\lambda\left[\frac{e^{\bar{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda} - \frac{e^{\underline{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}\right]\right)^{\frac{\sigma}{1-\sigma}}
\end{aligned}$$

Differentiate the above equation with respect to $\underline{x}$:

$$\left(\lambda[\frac{e^{\bar{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}-\frac{e^{\underline{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}]\right)^{\frac{\sigma-1+\sigma}{1-\sigma}}\left\{-b_j^{-1}e^{\underline{x}(\sigma-1)}(\frac{\sigma}{1-\sigma})\lambda e^{\underline{x}(\sigma-1-\lambda)}+\right.$$
$$\left.(\sigma-1)b_j^{-1}e^{\underline{x}(\sigma-1)}\lambda[\frac{e^{\bar{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}-\frac{e^{\underline{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}]\right\}$$

Since $\left(\lambda\left[\frac{e^{\bar{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}-\frac{e^{\underline{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}\right]\right)^{\frac{\sigma-1+\sigma}{1-\sigma}}>0$ always holds, so we need to have

$$(\sigma-1)\,b_j^{-1}e^{\underline{x}(\sigma-1)}\lambda\left[\frac{e^{\bar{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}-\frac{e^{\underline{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}\right]>b_j^{-1}e^{\underline{x}(\sigma-1)}\left(\frac{\sigma}{1-\sigma}\right)\lambda e^{\underline{x}(\sigma-1-\lambda)}$$

That is,

$$(\sigma-1)\left[\frac{e^{\bar{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}-\frac{e^{\underline{x}(\sigma-1-\lambda)}}{\sigma-1-\lambda}\right]>\left(\frac{\sigma}{1-\sigma}\right)e^{\underline{x}(\sigma-1-\lambda)}$$

The above is always true when $\sigma>1$. $\blacksquare$

D Data Description

The individual level data come from IPUMS in 2000. In the original dataset we have 6.44 million individuals. We restrict our sample to working population younger than 65. We also drop the individuals working in government, military, religious organizations and labor unions, or self-employed from our sample. 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. Our measure of income refers to "total personal income"(*inctot*), earning refers to "total earned income"(*incearn*), wage income refers to "total wage and salary"(*incwage*).

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 in 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. In our matched data set we have 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 3, 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 several cases, a metropolitan area might span over two or three states. In this case we assign the state dummy randomly. Our result are robust by varying the state assignments for these cities. We control for the racial composition of each metropolitan area by the share of white population, which is recorded in the dataset as *race* equals to 1.

E Tables and Figures

VARIABLES	(1) 95/50 Ratio	(2) 95/50 Ratio	(3) 95/50 Ratio	(4) 50/05 Ratio	(5) 50/05 Ratio	(6) 50/05 Ratio
Log(Total GDP)	0.0228**			-0.0266***		
	(0.00899)			(0.00803)		
Log(Private Ind. GDP)		0.0203**			-0.0269***	
		(0.00839)			(0.00815)	
Log(Population)			0.00638			-0.0269***
			(0.0171)			(0.00846)
Average Years of Edu.	1.187***	1.225***	1.508***	1.156***	1.188***	0.960***
	(0.440)	(0.442)	(0.537)	(0.190)	(0.186)	(0.141)
STD of Years of Edu.	0.375***	0.378***	0.407***	0.112*	0.121**	0.0910*
	(0.0662)	(0.0665)	(0.0825)	(0.0569)	(0.0560)	(0.0500)
Share of White Population	-0.0607	-0.0753	-0.148	-0.453***	-0.444***	-0.413***
	(0.139)	(0.138)	(0.175)	(0.111)	(0.112)	(0.0993)
Constant	-1.579	-1.646	-2.217*	-1.353***	-1.418***	-0.808***
	(1.029)	(1.039)	(1.112)	(0.401)	(0.396)	(0.295)
Observations	254	254	264	254	254	264
R-squared	0.668	0.666	0.653	0.615	0.618	0.641

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) 90/50 Ratio	(2) 90/50 Ratio	(3) 90/50 Ratio	(4) 50/10 Ratio	(5) 50/10 Ratio	(6) 50/10 Ratio
Log(Total GDP)	0.0162**			-0.0174***		
	(0.00629)			(0.00531)		
Log(Private Ind. GDP)		0.0141**			-0.0175***	
		(0.00613)			(0.00563)	
Log(Population)			0.00834			-0.0150**
			(0.0116)			(0.00634)
Average Years of Edu.	0.718**	0.751**	0.927**	0.921***	0.939***	0.728***
	(0.323)	(0.327)	(0.380)	(0.188)	(0.196)	(0.172)
STD of Years of Edu.	0.241***	0.243***	0.271***	0.0778**	0.0838**	0.0525
	(0.0376)	(0.0376)	(0.0483)	(0.0370)	(0.0380)	(0.0384)
Share of White Population	0.00180	-0.00917	-0.0448	-0.357***	-0.350***	-0.318***
	(0.0754)	(0.0743)	(0.0821)	(0.0795)	(0.0803)	(0.0679)
Constant	-0.827	-0.886	-1.261	-1.286***	-1.324***	-0.815*
	(0.769)	(0.780)	(0.816)	(0.466)	(0.478)	(0.413)
Observations	254	254	264	254	254	264
R-squared	0.730	0.728	0.708	0.659	0.661	0.691

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) 75/50 Ratio	(2) 75/50 Ratio	(3) 75/50 Ratio	(4) 50/25 Ratio	(5) 50/25 Ratio	(6) 50/25 Ratio
Log(Total GDP)	0.00791**			-0.00601*		
	(0.00334)			(0.00315)		
Log(Private Ind. GDP)		0.00675*			-0.00546*	
		(0.00351)			(0.00302)	
Log(Population)			0.00384			-0.00497
			(0.00523)			(0.00354)
Average Years of Edu.	0.355**	0.375**	0.456**	0.480***	0.472***	0.403***
	(0.161)	(0.167)	(0.176)	(0.107)	(0.109)	(0.106)
STD of Years of Edu.	0.104***	0.106***	0.113***	0.0751***	0.0749***	0.0657**
	(0.0197)	(0.0199)	(0.0233)	(0.0248)	(0.0257)	(0.0294)
Share of White Population	-0.0492*	-0.0549**	-0.0682**	-0.156***	-0.152***	-0.146***
	(0.0263)	(0.0264)	(0.0279)	(0.0435)	(0.0430)	(0.0361)
Constant	-0.437	-0.473	-0.654*	-0.619**	-0.606**	-0.434*
	(0.382)	(0.396)	(0.386)	(0.266)	(0.268)	(0.250)
Observations	254	254	264	254	254	264
R-squared	0.687	0.685	0.689	0.659	0.659	0.671

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Wage Ratios by City Size, U.S. Data

Note: This table reports the results of estimating equation (1). The measures of inequality are various percentile ratios of hourly wage within an MSA. For more details, see notes to Table 1.

VARIABLES	(1) 99/50 Ratio	(2) 99/50 Ratio	(3) 99/50 Ratio	(4) 50/01 Ratio	(5) 50/01 Ratio	(6) 50/01 Ratio
Log(Total GDP)	0.0847***			-0.0260***		
	(0.0231)			(0.00894)		
Log(Private Ind. GDP)		0.0800***			-0.0260***	
		(0.0215)			(0.00848)	
Log(Population)			0.0843***			-0.0232**
			(0.0277)			(0.0106)
Average Years of Edu.	1.905***	1.939***	2.258***	1.264***	1.288***	1.078***
	(0.656)	(0.625)	(0.681)	(0.388)	(0.390)	(0.292)
STD of Years of Edu.	0.582***	0.573***	0.618***	0.0756	0.0838	0.0421
	(0.159)	(0.162)	(0.166)	(0.0939)	(0.0952)	(0.0838)
Share of White Population	0.487	0.444	0.371	-0.510***	-0.500***	-0.442***
	(0.377)	(0.379)	(0.381)	(0.162)	(0.159)	(0.155)
Constant	-2.836*	-2.886**	-3.903**	-1.009	-1.058	-0.548
	(1.425)	(1.370)	(1.458)	(0.881)	(0.885)	(0.683)
Observations	254	254	264	254	254	264
R-squared	0.500	0.499	0.475	0.392	0.393	0.372

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) 95/50 Ratio	(2) 95/50 Ratio	(3) 95/50 Ratio	(4) 50/05 Ratio	(5) 50/05 Ratio	(6) 50/05 Ratio
Log(Total GDP)	0.0204*			-0.0242***		
	(0.0111)			(0.00633)		
Log(Private Ind. GDP)		0.0188*			-0.0250***	
		(0.0102)			(0.00643)	
Log(Population)			0.00497			-0.0220***
			(0.0181)			(0.00707)
Average Years of Edu.	1.754***	1.773***	2.084***	1.200***	1.240***	0.957***
	(0.440)	(0.434)	(0.532)	(0.175)	(0.175)	(0.138)
STD of Years of Edu.	0.471***	0.471***	0.496***	0.0900*	0.101*	0.0537
	(0.0831)	(0.0829)	(0.0909)	(0.0526)	(0.0521)	(0.0454)
Share of White Population	-0.164	-0.176	-0.272	-0.459***	-0.452***	-0.387***
	(0.170)	(0.168)	(0.208)	(0.0932)	(0.0939)	(0.0721)
Constant	-2.782***	-2.815***	-3.469***	-1.569***	-1.648***	-0.963***
	(1.025)	(1.021)	(1.106)	(0.385)	(0.387)	(0.285)
Observations	254	254	264	254	254	264
R-squared	0.667	0.666	0.679	0.551	0.555	0.604

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) 90/50 Ratio	(2) 90/50 Ratio	(3) 90/50 Ratio	(4) 50/10 Ratio	(5) 50/10 Ratio	(6) 50/10 Ratio
Log(Total GDP)	0.0137**			-0.0142***		
	(0.00631)			(0.00496)		
Log(Private Ind. GDP)		0.0119**			-0.0148***	
		(0.00591)			(0.00538)	
Log(Population)			0.00425			-0.0106
			(0.0135)			(0.00651)
Average Years of Edu.	1.019***	1.049***	1.257***	1.106***	1.134***	0.892***
	(0.292)	(0.292)	(0.419)	(0.157)	(0.161)	(0.151)
STD of Years of Edu.	0.302***	0.305***	0.331***	0.0847**	0.0915**	0.0574
	(0.0449)	(0.0450)	(0.0608)	(0.0403)	(0.0416)	(0.0414)
Share of White Population	-0.0436	-0.0532	-0.112	-0.348***	-0.344***	-0.299***
	(0.0794)	(0.0782)	(0.0932)	(0.0799)	(0.0801)	(0.0721)
Constant	-1.419*	-1.473**	-1.909**	-1.787***	-1.842***	-1.286***
	(0.712)	(0.716)	(0.890)	(0.366)	(0.371)	(0.331)
Observations	254	254	264	254	254	264
R-squared	0.736	0.734	0.708	0.640	0.642	0.656

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Income Ratios by City Size, U.S. Data

Note: This table reports the results of estimating equation (1). The measure for inequality is the 90/50 and 50/10 ratios of total income within an MSA. For more details, see notes to Table 1

VARIABLES	(1) 99/50 Ratio	(2) 99/50 Ratio	(3) 99/50 Ratio	(4) 50/01 Ratio	(5) 50/01 Ratio	(6) 50/01 Ratio
Log(Total GDP)	0.0932***			-0.0359***		
	(0.0297)			(0.0110)		
Log(Private Ind. GDP)		0.0882***			-0.0340***	
		(0.0265)			(0.0100)	
Log(Population)			0.103***			-0.0286**
			(0.0353)			(0.0141)
Average Years of Edu.	2.088**	2.124***	2.377***	1.248***	1.236***	0.952***
	(0.824)	(0.767)	(0.819)	(0.394)	(0.396)	(0.322)
STD of Years of Edu.	0.645***	0.635***	0.652***	0.119	0.123	0.0734
	(0.207)	(0.206)	(0.214)	(0.105)	(0.108)	(0.0998)
Share of White Population	0.506	0.458	0.411	-0.585***	-0.567***	-0.508**
	(0.444)	(0.444)	(0.459)	(0.207)	(0.200)	(0.213)
Constant	-3.242*	-3.294**	-4.339**	-0.628	-0.612	0.0650
	(1.682)	(1.582)	(1.636)	(0.851)	(0.851)	(0.676)
Observations	254	254	264	254	254	264
R-squared	0.492	0.491	0.470	0.370	0.369	0.341

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) 95/50 Ratio	(2) 95/50 Ratio	(3) 95/50 Ratio	(4) 50/05 Ratio	(5) 50/05 Ratio	(6) 50/05 Ratio
Log(Total GDP)	0.0273**			-0.0288***		
	(0.0106)			(0.00732)		
Log(Private Ind. GDP)		0.0250**			-0.0291***	
		(0.00977)			(0.00741)	
Log(Population)			0.0119			-0.0267***
			(0.0179)			(0.00717)
Average Years of Edu.	1.384**	1.414**	1.595***	1.317***	1.348***	1.079***
	(0.554)	(0.552)	(0.586)	(0.179)	(0.184)	(0.135)
STD of Years of Edu.	0.435***	0.435***	0.450***	0.102*	0.112**	0.0722
	(0.0973)	(0.0972)	(0.100)	(0.0535)	(0.0536)	(0.0467)
Share of White Population	-0.106	-0.122	-0.185	-0.439***	-0.429***	-0.360***
	(0.186)	(0.186)	(0.209)	(0.0989)	(0.0983)	(0.0770)
Constant	-2.046	-2.099	-2.453*	-1.749***	-1.813***	-1.129***
	(1.302)	(1.307)	(1.248)	(0.384)	(0.396)	(0.292)
Observations	254	254	264	254	254	264
R-squared	0.626	0.624	0.641	0.569	0.573	0.605

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) 90/50 Ratio	(2) 90/50 Ratio	(3) 90/50 Ratio	(4) 50/10 Ratio	(5) 50/10 Ratio	(6) 50/10 Ratio
Log(Total GDP)	0.0150**			-0.0160**		
	(0.00722)			(0.00618)		
Log(Private Ind. GDP)		0.0131*			-0.0158**	
		(0.00678)			(0.00651)	
Log(Population)			0.00745			-0.0130**
			(0.0121)			(0.00638)
Average Years of Edu.	0.903**	0.934**	1.075***	1.085***	1.095***	0.923***
	(0.349)	(0.350)	(0.397)	(0.190)	(0.195)	(0.179)
STD of Years of Edu.	0.266***	0.268***	0.287***	0.111***	0.115***	0.0877**
	(0.0496)	(0.0493)	(0.0561)	(0.0403)	(0.0425)	(0.0387)
Share of White Population	-0.0611	-0.0714	-0.111	-0.353***	-0.346***	-0.318***
	(0.0929)	(0.0913)	(0.0973)	(0.0709)	(0.0707)	(0.0575)
Constant	-1.224	-1.280	-1.582*	-1.582***	-1.602***	-1.196***
	(0.851)	(0.857)	(0.876)	(0.441)	(0.446)	(0.412)
Observations	254	254	264	254	254	264
R-squared	0.708	0.706	0.706	0.642	0.643	0.666

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Earnings Ratio by City Size, U.S. Data

Note: This table reports the results of estimating equation (1). The measures of inequality are various percentile ratios of total earnings within an MSA. For more details, see notes to Table 1.

VARIABLES	(1) 50/10 Ratio	(2) 50/10 Ratio	(3) 50/10 Ratio	(4) 50/10 Ratio	(5) 50/10 Ratio	(6) 50/10 Ratio	(7) 50/10 Ratio	(8) 50/10 Ratio
Population	0.0212***	0.00346	0.00907*	0.00184	-0.00599	-0.0112**	0.00252	-0.0150**
	(0.00562)	(0.00572)	(0.00487)	(0.00441)	(0.00632)	(0.00507)	(0.00454)	(0.00581)
Average Years of Edu.		0.344**			0.648***	0.591***		0.728***
		(0.133)			(0.181)	(0.125)		(0.161)
STD of Years of Edu.			0.0102		0.104**		-0.0477	0.0525
			(0.0325)		(0.0432)		(0.0331)	(0.0395)
Share of White Population				-0.252***		-0.341***	-0.292***	-0.318***
				(0.0640)		(0.0602)	(0.0662)	(0.0591)
Constant	0.553***	-0.0830	0.753***	0.772***	-0.562	-0.609**	0.669***	-0.815**
	(0.0732)	(0.314)	(0.0875)	(0.0536)	(0.370)	(0.298)	(0.0915)	(0.333)
Observations	264	264	264	264	264	264	264	264
R-squared	0.047	0.631	0.618	0.652	0.645	0.687	0.656	0.691
State Dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: City Size and Wage Inequality, Effects of Education and Racial Composition, Alternative Ordering

Note: This table reports the results of estimating Equation (1) by including all possible combinations of control variables. 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.

(a) Entrepreneur Wage Premium

VARIABLES	(1) Wage 1	(2) Wage 2	(3) Income 1	(4) Income 2	(5) Earning 1	(6) Earning 2
Entrepreneurs	0.183***	0.186***	0.282***	0.285***	0.261***	0.264***
	(0.0245)	(0.0237)	(0.0245)	(0.0236)	(0.0248)	(0.0239)
Entrepreneurs*Log(City Size)	0.0123***	0.0122***	0.00999***	0.00993***	0.0112***	0.0112***
	(0.00223)	(0.00218)	(0.00223)	(0.00218)	(0.00225)	(0.00220)
Age	0.0137***	0.0137***	0.0165***	0.0165***	0.0140***	0.0140***
	(8.63e-05)	(8.63e-05)	(8.60e-05)	(8.60e-05)	(8.82e-05)	(8.82e-05)
Years of Edu.	0.0674***	0.0674***	0.0778***	0.0778***	0.0736***	0.0736***
	(0.000635)	(0.000635)	(0.000750)	(0.000750)	(0.000707)	(0.000707)
Married	0.228***	0.228***	0.251***	0.251***	0.250***	0.250***
	(0.00188)	(0.00188)	(0.00190)	(0.00190)	(0.00190)	(0.00190)
Race == White	0.172***	0.172***	0.209***	0.209***	0.199***	0.199***
	(0.00289)	(0.00289)	(0.00296)	(0.00296)	(0.00287)	(0.00287)
Constant	4.576***	4.576***	8.255***	8.255***	8.320***	8.320***
	(0.0115)	(0.0115)	(0.0130)	(0.0130)	(0.0130)	(0.0130)
Observations	1,313,829	1,313,829	1,313,686	1,313,686	1,313,829	1,313,829
R-squared	0.371	0.371	0.422	0.422	0.390	0.390
MSA FE	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Occ FE	No	No	No	No	No	No

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) Entrepreneur Wage and City Size

VARIABLES	(1) Wage 1	(2) Wage 2	(3) Income 1	(4) Income 2	(5) Earning 1	(6) Earning 2
Entrepreneurs*Log(City Size)	0.0213***	0.0210***	0.0196***	0.0193***	0.0205***	0.0203***
	(0.00236)	(0.00231)	(0.00235)	(0.00230)	(0.00240)	(0.00235)
Age	0.0128***	0.0128***	0.0153***	0.0153***	0.0129***	0.0129***
	(8.22e-05)	(8.22e-05)	(8.08e-05)	(8.08e-05)	(8.25e-05)	(8.25e-05)
Years of Edu.	0.0467***	0.0467***	0.0544***	0.0544***	0.0508***	0.0508***
	(0.000462)	(0.000462)	(0.000546)	(0.000546)	(0.000514)	(0.000514)
Married	0.198***	0.198***	0.217***	0.217***	0.215***	0.215***
	(0.00186)	(0.00186)	(0.00186)	(0.00186)	(0.00186)	(0.00186)
Race == White	0.136***	0.136***	0.167***	0.167***	0.158***	0.158***
	(0.00272)	(0.00272)	(0.00276)	(0.00276)	(0.00269)	(0.00269)
Constant	5.534***	5.539***	9.385***	9.391***	9.406***	9.411***
	(0.0317)	(0.0309)	(0.0327)	(0.0319)	(0.0328)	(0.0320)
Observations	1,313,829	1,313,829	1,313,686	1,313,686	1,313,829	1,313,829
R-squared	0.408	0.408	0.466	0.466	0.433	0.433
MSA FE	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Occ FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: U.S. Entrepreneurial Wage Premium and City Size.

Note: Results of estimating Equation (3) with the benchmark definition of entrepreneurs. For the details of definition, refer to Table 10. Data source: IPUMS-USA, 2000.

(a) Narrow Definition

VARIABLES	(1) Wage 1	(2) Wage 2	(3) Income 1	(4) Income 2	(5) Earning 1	(6) Earning 2
Entrepreneurs	0.392***	0.392***	0.622***	0.620***	0.490***	0.490***
	(0.0543)	(0.0527)	(0.0576)	(0.0558)	(0.0559)	(0.0542)
Entrepreneurs*Log(City Size)	0.0142***	0.0143***	0.00603	0.00628	0.0131***	0.0132***
	(0.00481)	(0.00473)	(0.00514)	(0.00504)	(0.00497)	(0.00488)
Age	0.0139***	0.0139***	0.0167***	0.0167***	0.0143***	0.0143***
	(8.92e-05)	(8.92e-05)	(8.80e-05)	(8.80e-05)	(9.13e-05)	(9.13e-05)
Years of Edu.	0.0721***	0.0721***	0.0834***	0.0834***	0.0793***	0.0793***
	(0.000633)	(0.000633)	(0.000738)	(0.000738)	(0.000701)	(0.000701)
Married	0.236***	0.236***	0.262***	0.262***	0.260***	0.260***
	(0.00187)	(0.00187)	(0.00189)	(0.00189)	(0.00189)	(0.00189)
Race == White	0.182***	0.182***	0.220***	0.220***	0.211***	0.211***
	(0.00285)	(0.00285)	(0.00292)	(0.00292)	(0.00283)	(0.00283)
Constant	4.520***	4.520***	8.188***	8.188***	8.251***	8.251***
	(0.0116)	(0.0116)	(0.0131)	(0.0131)	(0.0134)	(0.0134)
Observations	1,313,829	1,313,829	1,313,686	1,313,686	1,313,829	1,313,829
R-squared	0.364	0.364	0.413	0.413	0.379	0.379
MSA FE	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Occ FE	No	No	No	No	No	No

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) Broad Definition

VARIABLES	(1) Wage 1	(2) Wage 2	(3) Income 1	(4) Income 2	(5) Earning 1	(6) Earning 2
Entrepreneurs	0.0739***	0.0768***	0.286***	0.287***	0.155***	0.158***
	(0.0273)	(0.0263)	(0.0242)	(0.0233)	(0.0275)	(0.0265)
Entrepreneurs*Log(City Size)	0.0127***	0.0126***	0.00651***	0.00649***	0.0117***	0.0116***
	(0.00251)	(0.00245)	(0.00223)	(0.00219)	(0.00252)	(0.00247)
Age	0.0135***	0.0135***	0.0159***	0.0159***	0.0137***	0.0137***
	(8.76e-05)	(8.76e-05)	(8.51e-05)	(8.51e-05)	(8.89e-05)	(8.89e-05)
Years of Edu.	0.0687***	0.0687***	0.0769***	0.0769***	0.0744***	0.0744***
	(0.000623)	(0.000623)	(0.000730)	(0.000730)	(0.000694)	(0.000694)
Married	0.228***	0.228***	0.247***	0.247***	0.249***	0.249***
	(0.00193)	(0.00193)	(0.00193)	(0.00193)	(0.00195)	(0.00195)
Race == White	0.174***	0.174***	0.206***	0.206***	0.200***	0.200***
	(0.00290)	(0.00290)	(0.00294)	(0.00294)	(0.00287)	(0.00287)
Constant	4.551***	4.551***	8.253***	8.253***	8.298***	8.298***
	(0.0118)	(0.0118)	(0.0134)	(0.0134)	(0.0139)	(0.0139)
Observations	1,313,829	1,313,829	1,313,686	1,313,686	1,313,829	1,313,829
R-squared	0.365	0.365	0.424	0.424	0.384	0.384
MSA FE	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Occ FE	No	No	No	No	No	No

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: U.S. Entrepreneurial Wage Premium and City Size, Alternative Definitions.

Note: Results of estimating Equation (3) with two alternative definitions of entrepreneurial. Narrow definition restricts entrepreneurs to CEOs, and the broad definition extends entrepreneurs to include the self-employed. For the details of definition, refer to Table 10 and the main text. Data source: IPUMS-USA, 2000.

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%

Table 10: Definition of Entrepreneurs, U.S.

Note: This table reports the definition of entrepreneurs used in the estimation of Equation (3). 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.

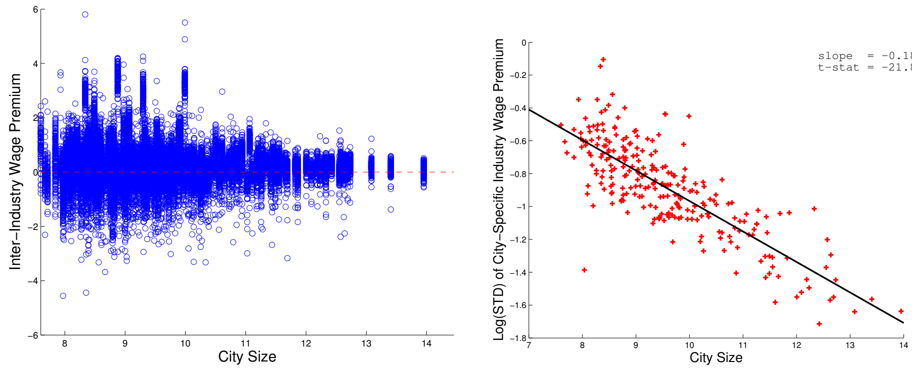
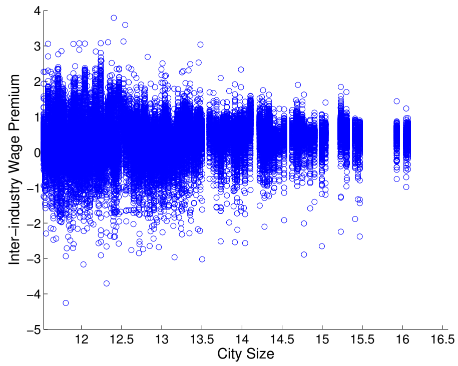
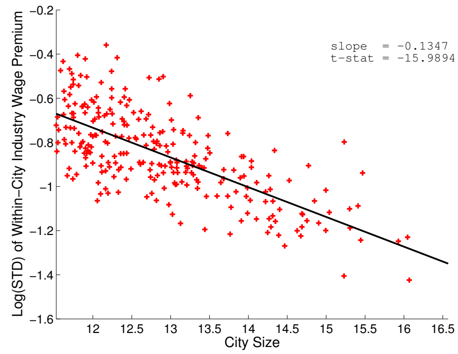


Figure 10: Inter-Industry Wage Premium by City, Robustness Checks

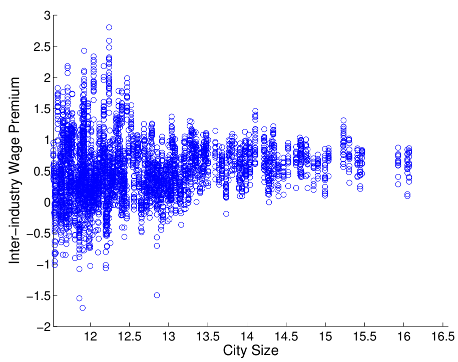
Notes: Panel (a) plots the industry wage premium estimated in each city against the city size measured in the log of private GDP. Panel (b) plots the log of standard deviation of industry wage premium within each city against the same measure of city size. The slope and the t-statistics reported in Panel (b) are based on a simple linear regression between the two variables. Data source: IPUMS-USA 2000 and Population Census 2000.



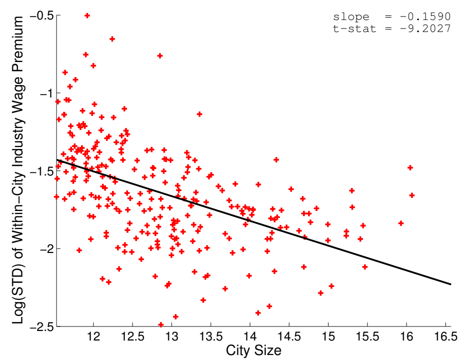
(a) Full Sample, Unweighted Average



(b) Full Sample, Unweighted Average



(c) Restricted Sample, Unweighted Average



(d) Restricted Sample, Unweighted Average

Figure 11: Monte-Carlo Simulation, Robustness Checks

Notes: Robustness checks of the Monte-Carlo Simulation with unweighted average. Data source: IPUMS-USA 2000 and Population Census 2000.