Small Teams in Big Cities:

Inequality, City Size, and the Organization of Production[*]

Clara Santamaría[†]

December 4, 2018

Abstract

This paper studies the effect of spatial sorting on inequality through two channels: spatial differences in technology and the endogenous organization of production. First, I document a new fact on the spatial differences in the organization of production. The number of workers per manager is decreasing in city size, overall and within industries. I develop and quantify a model of a system of cities where workers with different skills organize in production teams. The model yields continuous wage distributions in cities of different sizes that resemble the data. I find that technology differs across cities in its productivity but also in its complexity, so there are no incentives for it to diffuse across cities. I then use the model to evaluate two local policies that are designed to address income inequality: a minimum wage and a housing subsidy. I find that a revenue-neutral housing subsidy is more effective than a minimum wage at reducing inequality, as measured by the variance of log wages.

1 Introduction

The organization of production differs systematically across cities of different sizes. Even within an industry, New York has on average 36% fewer workers per manager than cities with a quarter of New York's population, such as Detroit or Atlanta. Larger cities not only have fewer workers per manager but are also more productive and more unequal.[1] The goal of this paper is to understand how the sorting of individuals *across* cities and the organization of production *within* cities interact with technology to shape income inequality. To do this, I propose and quantify a spatial equilibrium model of team production that can reproduce key stylized facts on the organization of production and income inequality.

Understanding the forces that shape income inequality is an important and long-standing objective in economics. Following Katz and Murphy (1992), skill-biased technology has long been understood to be an important driver of income inequality.[2] More recently, sorting has received growing attention as an additional factor.[3] In particular, the well-documented differences in income distribution across cities suggest that spatial sorting may be a driving factor of inequality.[4] Spatial sorting affects overall income inequality through two channels: spatial differences in technology and the organization of production. First, if technology varies across space, the place where people live will determine the technology that they have access to. Technology affects people's marginal productivity and, as a result, overall inequality. Second, the place where people live also determines who lives close to them and with whom they can work. Living in New York not only gives people access to different technology — it also gives them access to a different set of potential coworkers. The characteristics of a person's coworkers affect their marginal productivity through the skill complementarities that arise when people work together. These complementarities, in turn, depend on the way production is organized in the

---

[1]The evidence on the productivity advantage associated with city size is reviewed by Rosenthal and Strange (2004). Moreover, Baum-Snow and Pavan (2013) document the correlation of city size and different measures of inequality, such as the variance of log wages and percentile ratios of the wage distribution.

[2]A well-developed body of literature studies skill-biased technical change as the main force behind the increase in income inequality in the US over the last few decades (Levy and Murnane (1992); Bound and Johnson (1992); Card and DiNardo (2002); Autor and Dorn (2013)).

[3]Notably, Combes et al. (2008) find sorting in skills to be an important factor to explain wage disparities for French workers.

[4]Differences in the income distribution across cities have been well established; see Rosenthal and Strange (2004); Glaeser et al. (2009); Baum-Snow and Pavan (2013).

city.

As mentioned above, the distribution of income in a city varies systematically with population size. Larger cities tend to have a higher average income, even after adjusting for housing costs (Rosenthal and Strange (2004)). Larger cities also have higher income inequality as measured by the variance of log wages. Higher income inequality appears not only in the upper tail, but dispersion is higher throughout the entire distribution. In particular, both the ratio of the 90th to the 50th percentile and the ratio of the 50th to the 10th percentile are increasing in city size (Baum-Snow and Pavan (2013)). Moreover, the dispersion in the income distribution is large enough so that the density both at the top and bottom tails of the real income distribution is higher in larger cities (Eeckhout et al. (2014)).

I start my analysis by providing novel evidence for how the organization of production changes with city size. In particular, I find that larger cities tend to have fewer workers per manager. In the spirit of Lucas (1978), I refer to the number of workers per manager as the span of control. This fact is robust to using alternative definitions of managers, such as using different occupational classifications or restricting attention to only heads of large companies. Moreover, the variation in the number of workers per manager across cities is not driven by the variation in industrial composition, but rather, is due to the span of control decreasing with city size within industries.[5,6] While this difference in span of control across cities within industries could reflect differences in production processes for the same good or service, it could also reflect differences in the type or quality of good or service being produced. I present evidence from the legal services and pharmaceutical industry to suggest that the latter possibility seems likely.

The basic Lucas setup with constant technology across cities cannot explain why larger cities have a smaller spans of control. In the Lucas framework, better managers benefit more from managing more workers. Since more skilled agents tend to reside in larger cities, we would expect larger cities to have larger spans of control — the opposite of what is seen in the data.[7]

---

[5]Large cities could have more managers per worker simply because they have more headquarter establishments. The within-industry analysis helps to address this possibility since headquarters are classified into a unique sector (NAICS 55). To investigate this further, I also control for the fraction of employment in the "Management of Firms" industry and find that the correlation between the fraction of managers and city size is still positive and significant.

[6]This finding complements the work of Duranton and Puga (2005), who document a positive correlation between the fraction of managers and city size within manufacturing. I find that this correlation holds within industries for a wide variety of sectors beyond manufacturing, and it is robust to various definitions of management.

[7]This logic assumes an elastic supply of workers. Otherwise, an abundance of high-skilled agents would increase competition for

In order to understand why larger cities have smaller spans of control, it is necessary to think about differences in technology across cities in a more nuanced way than simply a Hicks-neutral productivity shifter. This paper incorporates differences in the complexity of production and the need to communicate between managers and workers.

In order to study the variation in the organization of production across cities of different sizes, I develop a spatial equilibrium model of knowledge-based hierarchies, as in Garicano and Rossi-Hansberg (2006). In this framework, hiring workers with more skills allows a manager to hire more of them. This structure introduces complementarities between the skill of managers and the skill of workers. The resulting matching problem between managers and workers determines the optimal organization of production and depends critically on the skill distribution. Thus, the sorting of skills affects the mapping of skills into wages through the organization of production. The model consists of a set of cities that are heterogeneous in production technology, amenities, and housing stock, as well as a mass of agents who are heterogeneous in skill, labor supply, and preferences for cities. Agents have preferences for consumption, housing, and their location. They choose what city to live in and their occupation, that is, whether to become workers or managers.

I estimate the model through Simulated Method of Moments using data from the American Community Survey 5% IPUMS 2010–2014.[8] In the estimation, I recover the technology parameters for a representative large and small city in the US.[9] I find that cities use different technologies. However, I do not find the technology used in the large city to be a better technology than the one used in the small city. The large city uses a more *complex* technology that is more productive but addresses harder problems and requires more management time. This complex technology is better than the simple technology for some agents, but it is worse for others. In fact, if a social planner for the small city were given the opportunity of adopting the complex technology, the planner would not do so since it would decrease the average utility of

---

workers, drive wages, and make it less profitable to expand the span of control. Behrens et al. (2014) explore a Lucas setting in a system of cities and explore the interactions of sorting, agglomeration, and selection into entrepreneurship. Their setup predicts a constant fraction of entrepreneurs across cities. As the authors observe, this is consistent with the roughly constant fraction of self-employed in the data. However, it is not consistent with the increasing fraction of managers that this paper documents.

[8]In the estimation, I match three moments city by city: the median income of workers and managers and the number of workers per manager. In addition, I match two overall moments: the relative city size, and the overall variance of income.

[9]I also estimate the model for four particular cities in the US: New York, Chicago, Pittsburgh, and Springfield, MA. The results from this estimation, which are included in Section C.2 of the Appendix, are consistent with the results from the two city estimation.

the small city residents. This finding can rationalize technology differences across cities since the small city uses a simpler technology that is the preferred choice for its residents. Although my model has a single final good, one can also interpret the two cities as producing different types of this good that aggregate linearly, so that output is measuring the production of effective units of the good.

The equilibrium of the estimated model matches the entire distribution of income across cities well. In particular, it generates two features of the data that are not targeted in the estimation: first, larger cities are more unequal, and second, larger cities have fatter tails in the real income distribution.[10] The key for generating these stylized facts is that the slope of the wage schedule varies across cities. I refer to the slope of the wage schedule as the skill premium. Since the slope is not constant, the skill premium will vary across different skills. The main force that leads to more inequality in the large city is the higher skill premium that results from problems being both harder and more costly to communicate. This makes skills generally more useful in the large city.

According to this logic, it is surprising that there is a higher density at the very bottom of the real income and skill distribution in the large city. There are two main forces that generate this stylized fact in the model. First, the very lowest-skilled workers produce zero output on their own in either city, but they are more useful in the large city where they form a team with a more skilled manager. This happens because selection into management is tougher in the large city because high-skilled agents are more abundant there. Therefore the worst manager in the large city is better than the worst manager in the small city. Second, because there is a higher density of easy problems in the small city, workers with skills immediately above the lowest skill are more useful in the small city where they can solve more problems. Therefore, the lowest-skilled workers will be attracted to the large city and the middle-low skilled workers will be attracted to the small city.

I use the estimated technology parameters to perform policy counterfactuals in which both the sorting of individuals and the organization of production react endogenously to the policy change.

---

[10]In Eeckhout et al. (2014) the authors also present evidence of thick tails in the distributions of educational attainments as a proxy for skill.

I study the implementation of two policies in the large city designed to decrease inequality: a minimum wage and a revenue-neutral housing subsidy for low-income agents. These are two examples of some of the policies that are part of the debate on how to address income inequality in cities. I quantify how these policies affect agents throughout the skill distribution and find that the two policies have very different implications, as described below.

A minimum wage distorts the optimal organization of production by forcing the lowest-skilled managers into hiring higher-skilled workers and the lowest skilled workers into becoming self-employed. Moreover, the minimum wage causes low-skilled workers to relocate to the small city, driving down the wage of low-skilled workers in the small city. This, in turn, changes the organization of production in the small city by lowering the threshold skill at which workers prefer to become managers. As a result, the model shows that a minimum wage policy implemented in a large city can have spillover effects, raising inequality in smaller cities which are not targeted directly by the policy.

In contrast, the housing subsidy for low-income earners does not significantly distort the organization of production in the large city. Inequality in the large city decreases because the revenue-neutral subsidy redistributes income toward the lowest skilled. Additionally, the supply of low-skilled agents in the small city decreases since more low-skilled agents are attracted to the large city. This sorting effect drives up the wage of low-skilled workers in the small city, which toughens selection into management, increases team size, and lowers inequality.

Overall, both the minimum wage and the housing subsidy have small negative effects on the average utility of 1.6% and 0.6% respectively. This overall negative effect is expected since the original equilibrium allocation is efficient. More interesting, the minimum wage increases the variance of log utility by 44% while the affordable housing policy decreases the variance of log utility by 8%, indicating that the choice of policy matters for effectively reducing inequality.

My work contributes to several strands of the literature. First, I provide a framework to study sorting across cities and the organization of production within cities that matches well the entire distribution of income within cities and overall. Early work on the sorting of heterogeneous agents across cities by Abdel-Rahman and Wang (1997) featured two skill types in a core-periphery

6

model to study income disparities within and across regions. More recently, the sorting of heterogeneous agents has been studied to understand the positive correlation of income inequality with city size in Behrens et al. (2014), Davis and Dingel (2016), Davis and Dingel (2017), and Baum-Snow et al. (2017), and the fat tails in the real income distribution, which was documented in Eeckhout et al. (2014). This literature has proposed mechanisms through which sorting across ex-ante identical cities generates spatial differences in the income distribution. Instead, I measure technology differences across cities and allow for endogenous sorting and production organization. The measurement of technology by city provides new insights on the ways in which large cities differ from small cities. Eeckhout et al. (2014) document the presence of fat tails in the real income distribution of large cities. The authors propose a production function with extreme-skill complementarities that leads to the sorting of high- and low-skilled workers to large cities and middle-skilled workers to small cities. By modeling the endogenous organization of production, my framework provides a microfoundation for extreme-skill complementarities with a continuum of skills and generates sorting patterns where extreme skills sort disproportionately into larger cities.

Second, the literature on the organization of production pioneered by Lucas (1978) studies the way in which heterogeneous agents form teams and take on different roles in order to produce together. In a similar spirit, Calvo and Wellisz (1978) studies the role of managers monitoring workers. Early work on team production by Radner (1993) studied the role of teams to process information. More recently Garicano (2000) and Garicano and Rossi-Hansberg (2006) study the organization of knowledge in production hierarchies. In Garicano and Rossi-Hansberg (2015) the authors review the recent literature on the organization of knowledge. In this work, I build on these ideas by introducing spatial frictions so that production is organized within a city. I then endogenize the skill distribution within cities through sorting. By studying the organization of production in a spatial equilibrium framework, I can measure technology differences while taking into account endogenous differences in the skill distribution across cities.

Third, in this framework, agents with different skills face different incentives to locate in larger cities compared to smaller cities. This result speaks to the literature studying the pattern

of the sorting of skills across cities in the US. The sorting choices of different skills have been studied by Moretti (2012), who coined the term the "Great Divergence" to refer to the trend of increasingly differential sorting patterns for high-skilled relative to low-skilled agents. In recent work, Diamond (2016) estimates the welfare effects of this differential sorting. Additionally, Giannone (2018) quantifies the role of migration patterns of workers with different skills in response to a skill-biased technological change to explain regional wage dispersion in the US. In this paper, differential sorting patterns arise from differences in the complexity of technology across cities of different sizes.

Fourth, the finding that larger cities host smaller teams who use a more complex technology is related to the finding of Holmes and Stevens (2010) who find that small plants tend to produce specialty or customized goods. Similarly, Tian (2018) using data from Brazil, finds that firms in larger cities tend to have greater division of labor, and links it to the complexity of the products that these firms produce. It is also related to the work of Glaeser et al. (2010) who link smaller establishments with higher growth and an abundance of entrepreneurship.

Finally, the finding that larger cities have fewer workers per manager complements the work of Duranton and Puga (2005), and more recently Tian (2018). In their work, the authors document an increasing specialization of larger cities in management tasks, which they attribute to the geographical separation of headquarters from production plants. In contrast, this model does not allow for cross-city production teams. Instead, it focuses on the formation of local teams both inside headquarter establishments and production plants. To the extent that the industry classification is able to separate headquarter from production establishments, I do not find that location of headquarters can fully account for the smaller production teams found in larger cities. Moreover, by limiting production teams to be formed within a city, I am able to solve for a richer model of production hierarchies, where the span of control depends on the skill of the production workers. This assumption allows the model to generate smooth income distributions both within and across cities which resemble those observed in the data.

The rest of the paper proceeds as follows. Section 2 presents the stylized facts in the data, including new robust evidence on the negative correlation between the number of workers per

manager and city size. Section 3 introduces a spatial model of production organization. Section 4 details the estimation strategy for a version with two cities of different sizes. Section 5 analyzes the estimated technology differences in space. Section 6 interprets the model's implications for sorting on inequality. Section 7 evaluates the effectiveness of a housing subsidy or a minimum wage in reducing inequality. Finally, Section 8 provides concluding remarks.

2   Data and Empirical Regularities

In this section, I present a set of stylized facts regarding the cross-section of cities. In particular, I show that both the average income and the dispersion of log income are increasing in city population. These facts have already been documented in the literature and are suggestive of the importance of spatial sorting for the overall distribution of income.[11] I include them here for completeness since they will be used in the subsequent analysis. I then present evidence on differences in the organization of production across cities. Specifically, I show that the number of workers per manager is decreasing in city size. In the following section, I build a model that addresses both differences in the organization of production and the distribution of income across cities.

2.1   Data and Definitions

I use data from the American Community Survey (ACS) Public Use Microdata Sample (PUMS) 5-year sample covering the period from 2010 to 2014. The ACS samples 1 in every 40 addresses in the US every year. The 5-year sample is a combination of 1-year samples. I aggregate Public Use Microdata Areas (PUMAS) into Combined Statistical Areas (CSA) for the main analysis and use Metropolitan Statistical Areas (MSA) for robustness. The final sample contains 162 CSAs and 2.2 million observations.

---

[11]The empirical evidence on the correlation between average income and city size was reviewed by Rosenthal and Strange (2004). The correlation between city size and variance of income was documented by Baum-Snow et al. (2017), and the fat tails in the real income and skills were studied by Eeckhout et al. (2014).

CSA. Combined Statistical Areas are comprised of one or more adjacent Metropolitan and Micropolitan Statistical Areas (MSAs and μSAs) that have an employment exchange of at least 15% with the central county or counties of the parent MSA or μSA. MSAs and μ SAs are themselves comprised of adjacent counties that include a core urban area and outlying counties with an employment interchange of at least 25%.

Wages. The measure of wages includes pre-tax wages, salaries, commissions, cash bonuses, tips, and other money income received from an employer in the past 12 months. The top one-half percent of wages in each state are top-coded and assigned the average value of all top-coded wages in the state.[12] I include only individuals who worked for the last 12 months and who report having worked at least 35 weekly hours. In order to exclude outliers, I drop the bottom one-half percent of wages, which corresponds to annual wages below $6,250. I also exclude some sectors that are outside the scope of my analysis, such as the primary sectors (agriculture, fishing, forestry, and extraction) and the military. These sectors are not primarily urban. The average wage in the final sample is $60,756 with a standard deviation of $61,595.

I also use data from the Housing Survey of the ACS for the same period 2010-2014 in order to compute hedonic prices by city that allow me to construct real wages. I exclude individuals who live in group quarters, mobile homes, trailers, boats, tents, or farmhouses. Moreover, I only include individuals who are renting in order to avoid imputing rents on owned houses. The final sample from the Housing Survey contains 1.2 million observations of households who rent a building and pay on average $900/month with a standard deviation of $555/month for an apartment that on average has 2.8 rooms and was built about 5 years ago. I include the results from the hedonic regression in Section A.2 of the Appendix.

2.2   Income Distribution and City Size

In this section, I present two key stylized facts on the distribution of income across cities: first the positive correlation between city population and average income, and, second, the positive

---

[12]A list of the top codes for each state is included in Section A.1 of the Appendix.

correlation between city population and the variance of log income. Figure 2.1 presents both stylized facts. Panel (a) plots average income against city size and Panel (b) plots the variance of log income against city size where a city is defined as a Combined Statistical Area.[13]

Larger cities have a higher average income. A 100% increase in population is related to an increase in average annual income of almost $4,500 on average. The positive correlation between city population and the nominal income indicates higher productivity in larger cities to the extent that individuals are paid wages according to their productivity. However, these differences in nominal income may not translate into differences in purchasing power since living costs are higher in larger cities. In order to explore this further, I compute real income by dividing income by the price index implied by a Cobb-Douglas utility function with a housing share of 0.24 and where housing prices are the result of a hedonic regression.[14,15] Average real income is also positively correlated with city population, implying that at least part of the differences in nominal income translates into higher utility for the residents. A 100% increase in population is related to an increase in average real income of almost $3,000. The regression for real income on population is included in Section A.3 of the Appendix.

The positive correlation between city population and higher variance of log income indicates that large cities are more unequal. The higher inequality is due to higher inequality both at the top and at the bottom of the income distribution. For instance, both the ratio of the 90th to the 50th percentile and the ratio of the 50th percentile to the 10th percentile are increasing in city size. The regressions of percentile ratios on populations are included in Section A.3 of the Appendix.

These stylized facts summarize the way in which income distributions vary across cities. These differences in the income distribution come in part from differences in the composition of the people who live in different cities. For instance, if the most talented individuals live in the largest cities, we would expect the positive correlation between income and city size. But the distribution of income also differs because characteristics are compensated differently

---

[13]In Section A.3 of the Appendix, I present the stylized facts for Metropolitan Statistical Areas.

[14]Davis and Ortalo-Magné (2011) find that the share of expenditures on housing is remarkably constant across cities.

[15]The hedonic regression includes the number of rooms, the year the house was built, and the type of building. The results from these regressions are included in Section A.2 of the Appendix.

Figure 2.1: Income Distribution and City Size
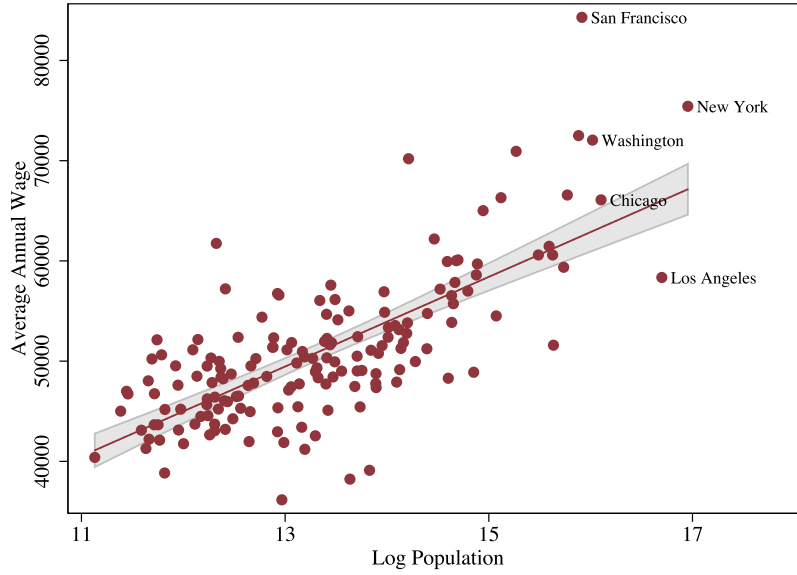
(a) Average Wage
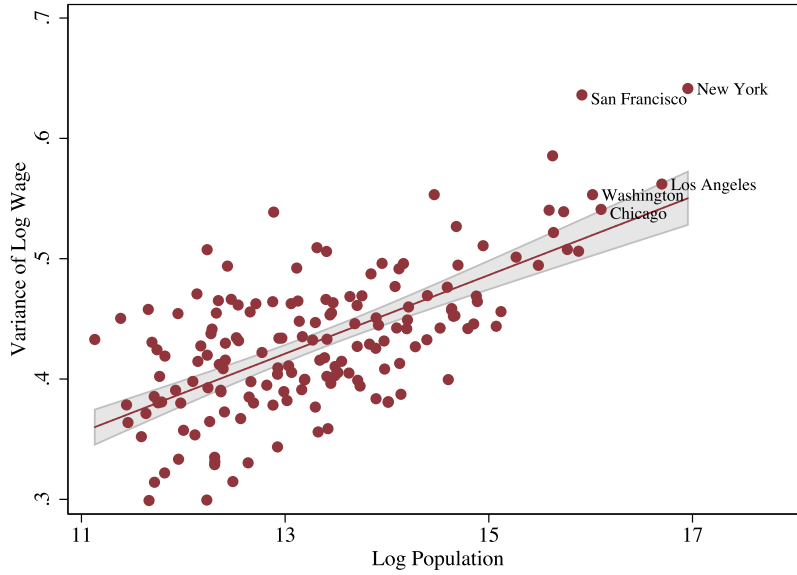
(b) Variance of Log Wage

Note: Panel (a) plots the average annual wage against logged population, and Panel (b) plots the variance of log wage against logged population. Each observation corresponds to a CSA. The source for the data is the American Community Survey, 5% IPUMS 2010–2014. The sample includes individuals who worked for the last 12 months for at least 35 hours per week. It excludes the military and primary sectors. The line represents a linear regression, and the grey region corresponds to the 95% confidence interval.

across cities. To illustrate this point, I first regress income on observable characteristics such as education, years of potential experience, gender, or race. Then I regress the average of residual income and the variance of log residual income on population and I find a positive correlation for both moments. A 100% increase in population is related to an increase in the average residual income of about $3,800. The results from these regressions are included in Section A.3 of the Appendix. The differences in compensation are potentially due to either differences in technology or labor complementarities that arise when people organize to produce together. In the next section, I present new evidence on the differences in the organization of production across cities.

2.3   Production Organization and City Size

In this section, I present new evidence on spatial differences in the organization of production. I focus on the number of workers per manager (span of control) as a key aspect of the organization of production. I first discuss the measurement of the span of control and then present the evidence on the negative correlation between the span of control and the population of a city.

2.3.1   Measuring the Span of Control

I define the span of control in a city as the number of workers per manager. This definition is based on the Lucas idea of management, where a manager is an expert whose talent gets leveraged by working with a team of workers. Managers will be identified in the data using self-reported occupations. The occupational classification system in the ACS includes "Management" as a category. However, this category corresponds to a more traditional view of management that has to do with particular tasks in a company such as hiring decisions or investment strategies. This is a narrower conception of management than that of an expert in a production team. In order to widen the definition of management, I follow Caliendo et al. (2015) and use the French occupational coding system, "Professions et Catégories Socioprofessionelles" (PCS). The ACS system groups occupations using similarity in tasks performed, while the French system incorporates information on the socioeconomic status of the job. As a result, the French

codings are naturally hierarchical and thus closer to the Lucas concept of a manager and will be my preferred definition. The PCS identifies five main categories: head of companies, upper management and professionals, intermediate professionals, white collar, and blue collar workers. The baseline definition I classify head of companies, upper management and professionals as "managers" and the remaining as "workers".

Although this grouping of PCS codes will be used as my preferred classification of managers, I will show that the stylized facts are robust to alternative definitions of managers. In particular, I will present the stylized facts using the management definition in the occupational classification used in the ACS. Moreover, I will use a more strict definition of manager that includes only heads of companies with more than 10 employees. This last definition will be based on the PCS classification that identifies the heads of companies and it is closer to the idea of a firm or an entrepreneur. I discuss the details of the French occupational classification as well as its correspondence with the classification used in the ACS in Section A.4.1 of the Appendix. In section A.4.2 of the Appendix, I look at changes in the organization of production in a more disaggregated way by using all of the PCS groupings and the main picture does not change.

2.3.2 Span of Control and City Size

In this section, I document the negative correlation between city size and span of control, that is, the number of workers per manager. Panel (a) in Figure 2 plots the span of control against the log population for Combined Statistical Areas. Note that I plot the level of the span of control for readability but the tables include regressions for the logged span of control so that the regression coefficients are less sensitive to level effects. The number of managers is identified using the PCS classification as described above. Larger cities tend to have lower spans of control. The difference goes from 2 workers per manager in some of the largest cities to 5 or 6 workers per manager is some of the smallest cities. An increase in the population of a city of 100% is related to 0.43 more workers per manager on average.

Some of this variation may be the result of variation in the industry composition. In particular, if larger cities specialize in industries with smaller spans of control then the span of control would

Figure 2.2: Decreasing Span of Control on City Size
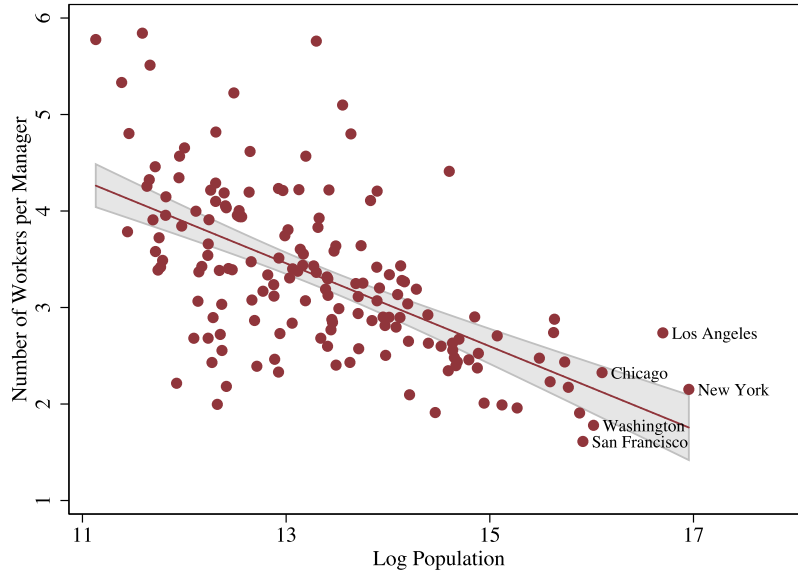
(a) Unconditional
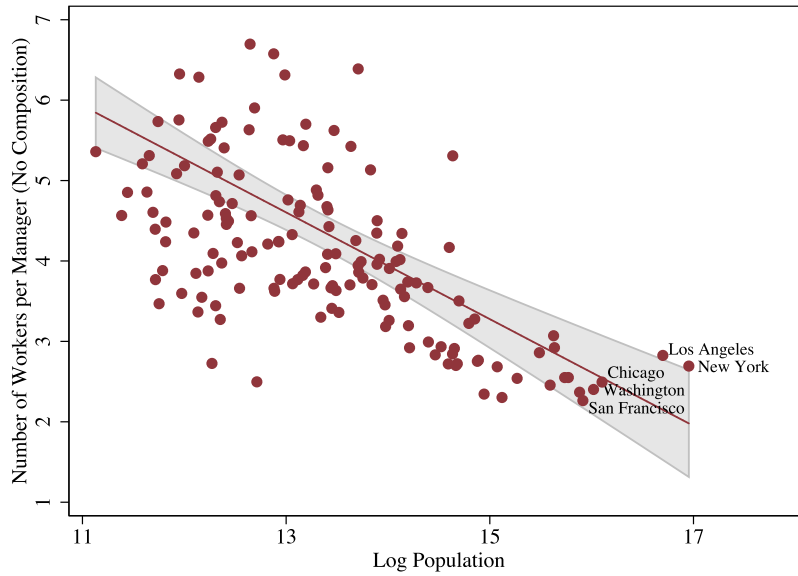
(b) Constant Industry Composition

Note: Panel (a) plots the number of workers per manager against logged population. Panel (b) plots the number of workers per manager, where the number of workers per manager is a weighted average of the number of workers per manager within an industry and the weights are given by the national industrial composition. Each observation corresponds to a CSA. The source for the data is the American Community Survey, 5% IPUMS 2010–2014. The sample includes individuals who worked for the last 12 months for at least 35 hours per week. It excludes the military and primary sectors. The line represents a linear regression and the grey region corresponds to the 95th confidence interval.

be decreasing in city size even if constant within an industry. To investigate the importance of industry composition in generating the negative correlation between the span of control and city size, I calculate the span of control in each city as a weighted average of the span of control in each industry, where the weights are fixed at the national industry composition. The resulting span of control is plotted in Panel (b) of Figure 2.2. The slope would be even steeper for a constant industry composition across cities. This implies that the span of control is decreasing within industries across city sizes and industry composition is changing in a way that makes the decrease in the span of control milder.

Table 2.1: Robustness to Management Definition

| | Log Span of Control | | |
	French PCS Occupations	ACS Occupations	Head of Companies
Log Population	-0.13**	-0.11**	-0.12**
Constant	2.95**	3.6**	7.03**
R^2	0.39	0.4	0.33
N	162	162	162

* p<0.05; ** p<0.01

Note: This table reports the results from regressing the logged span of control, or number of workers per manager, against logged population. Each observation corresponds to a CSA. It uses data from the American Community Survey, 5% IPUMS 2010–2014. The sample includes individuals who worked for the last 12 months for at least 35 hours per week. It excludes the military and primary sectors. The first column uses the French occupational classification system, Professions et Catégories Socioprofessionelles (PCS), to classify individuals into managers and workers. A more detailed overview of the PCS system is provided in Section A.4.1 of the Appendix. The second column uses the occupational classification system in the ACS. The third column uses the PCS and includes only heads of businesses with ten or more employees.

In order to test for robustness with respect to the definition of manager, Table 2.1 summarizes the results from regressing the log of the span of control on the log population for three definitions of a manager: 1) the PCS, or French classification system, 2) the OCC, or classification system used in the ACS, and, 3) an even stricter definition of manager that includes only heads of

Table 2.2: Robustness to Industry Composition

| | Log Span of Control | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Log Population	-0.13**	-0.08**	-0.12**	-0.1**	-0.12**	-0.09**
Constant	2.95**	2.49**	2.1*	6.54**	3.36**	3.17**
Industry Fixed Effect	No	No	Yes	Yes	Yes	Yes
Including "Manufacturing"	Yes	Yes	Yes	No	Yes	Yes
Including "Management of Firms"	Yes	Yes	Yes	Yes	No	Yes
Fraction of Employment in "Management of Firms"	-	-	-	-	-	-0.1**
R^2	0.37	0.01	0.77	0.8	0.76	0.77
N	162	9,773	9,773	7,240	9,745	9,611

* p<0.05; ** p<0.01

Note: This table reports the results from regressing the logged span of control, or number of workers per manager, against logged population. Observations in the first column correspond to a CSA. In Columns 2 to 6, an observation corresponds to an industry-CSA pair. It includes only industry-city pairs with more than 30 observations. Standard errors are clustered at the CSA. I use data from the American Community Survey, 5% IPUMS 2010–2014.

companies with 10 or more employees, classified using the PCS. I regress log on log in order to abstract from level effects by comparing elasticities. The main difference between those definitions is captured by the level effect, which, as expected, is higher for the OCC than the PCS and even higher for heads of companies. However, the elasticity of the span of control with respect to city size remains remarkably similar across these definitions. A 100% increase in population is related to between an 11% and a 13% increase in the span of control depending on the definition of a manager.

This result may simply reflect that the span of control differs among industries and industry composition may differ systematically with city size. In order to address this further, I include industry fixed effects in a regression of span of control against city size. The second column in

Table 2.2 reflects the results from this regression. Interestingly, the magnitude of the coefficient on city size increases when industry fixed effects are included from 0.08 to 0.12, meaning that the correlation between the span of control and city size is even stronger within industries.

This measure of the span of control could also be capturing the location of headquarters in larger cities. This possibility is partially addressed by including industry fixed effects since the administration of companies and headquarters have their own industry classification, which corresponds to NAICS 551114 "Corporate, subsidiary, and regional managing offices." Respondents are asked to report the main activity of the place where they work; if they work at the headquarters of a manufacturing company and manufacturing is not the main activity of the establishment, they should respond with "Corporate, subsidiary, and regional managing offices." However, it is possible that some of them misreport their industry. This misreporting is most concerning in the manufacturing sector since in this sector it is more common for production plants to be separated from the headquarters.

In order to further investigate this, Table 2.2 presents evidence of the robustness of the coefficient on log population to excluding manufacturing, excluding management of firms, and finally adding the fraction of employment in "Management of Firms" as a regressor. "Management of Firms" is the NAICS 55 sector that includes "Corporate, subsidiary, and regional managing offices." Unfortunately, the ACS does not identify more refined industries for this sector. The idea behind including the fraction of employment in Management of Firms is to pick up the correlation between population and the presence of headquarters in cities so that we can compare cities with varying population but a similar presence of headquarters. The coefficient only changes slightly; between -0.09 to -0.12. The result is robust to excluding manufacturing or management of firms as well as including the fraction of employment in the management.

2.4   A Closer Look at Two Industries

As shown in the previous section, even within the same industry, production is organized differently in larger cities. In order to make this fact more concrete, I look at a couple of examples

Figure 2.3: NAICS 3254: Manufacturing of Pharmaceuticals and Medicines

(a) Managers

(b) Workers
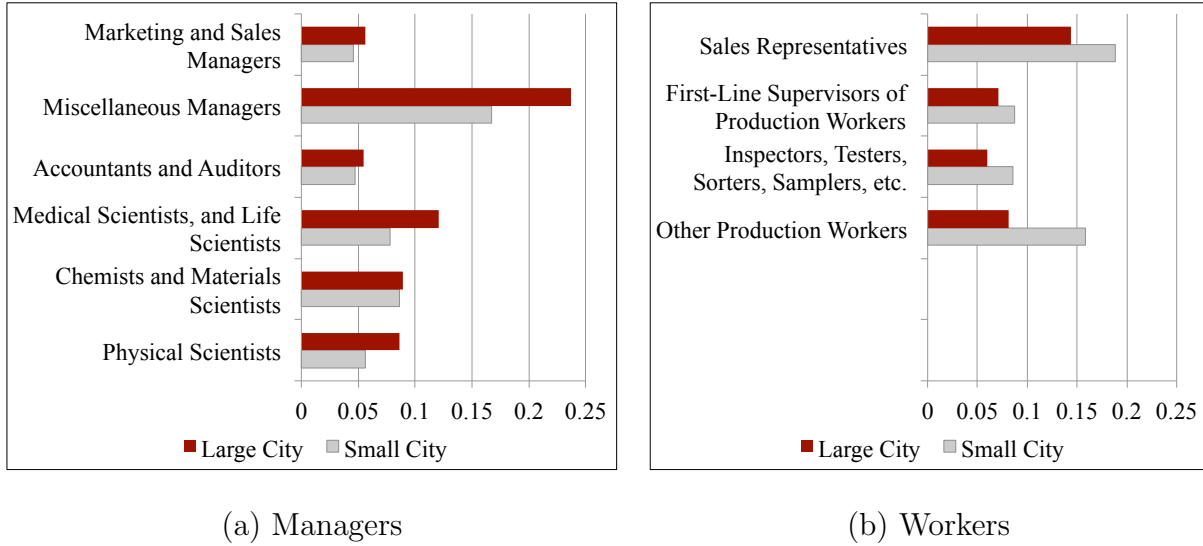
Note: This table compares the occupational composition of the "Manufacturing of Pharmaceuticals and Medicines" industry between large and small cities. Large cities include all CSAs with more than 2.5 million inhabitants, and small cities include those with less than 2.5 million inhabitants. It uses data from the American Community Survey, 5% IPUMS 2010–2014. This comparison includes only the 10 occupations with the highest share of employment in the industry at the national level. Panel (a) captures the composition for those occupations classified as "Managers" and Panel (b) captures the composition for those occupations classified as "Workers."

from two specific four-digit industries: NAICS 3254 — "Manufacturing of Pharmaceutical and Medicines"— and NAICS 5411 — "Legal Services."[16] I compare the fraction employment in each of the top 10 occupations for these two industries between cities of over 2.5 million inhabitants and in cities with fewer than 2.5 million.

The first example is illustrated in Figure 2.3 where I summarize the composition of the 10 most common occupations for "Manufacturing of Pharmaceutical and Medicines." In large cities, there is a larger fraction of not only managers but also of medical, life, chemical, and material scientists that perform the role a manager from the perspective of team leaders or problem solvers, while in smaller cities, there is a larger fraction of production workers and sales representatives. This difference in the occupational composition suggests that larger cities may be producing more innovative pharmaceuticals and medicines, and this requires the presence of a higher fraction of scientists, while smaller cities are producing more standard products that do not require as

---

[16]Examples for other industries can be found in Section A.4.4 of the Appendix.

Figure 2.4: NAICS 5411: Legal Services

(a) Managers

(b) Workers
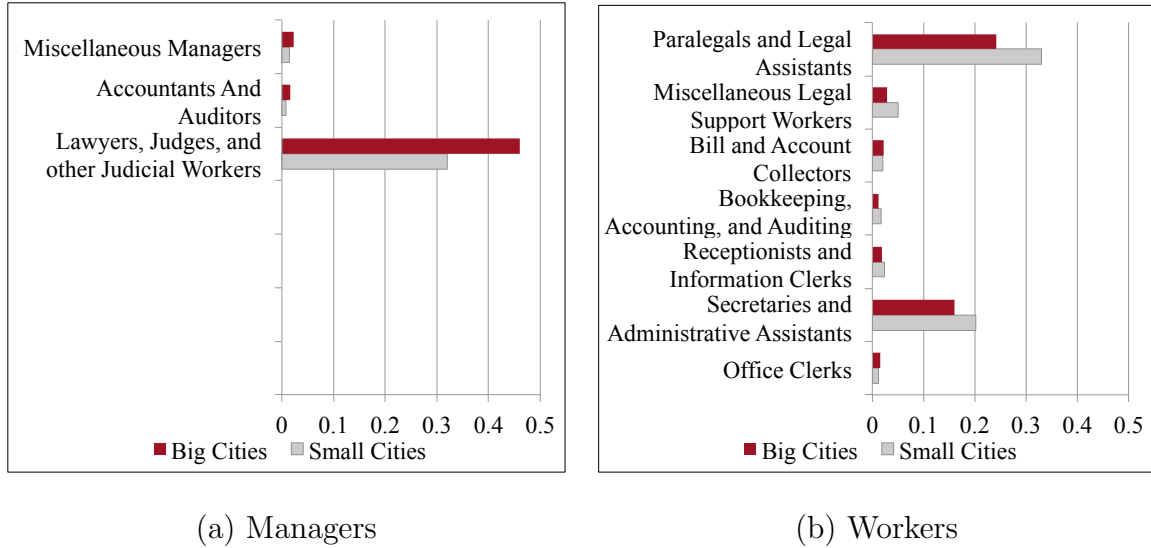
Note: This table compares the occupational composition of the "Legal Services" industry between large and small cities. Large cities include all CSAs with more than 2.5 million inhabitants and small cities include those with less than 2.5 million inhabitants. It uses data from the American Community Survey, 5% IPUMS 2010–2014. This comparison includes only the 10 occupations with the highest share of employment in the industry at the national level. Panel (a) captures the composition for those occupations classified as "Managers" and Panel (b) captures the composition for those occupations classified as "Workers."

many experts.

The second example presented in Figure 2.4 focuses on Legal Services and paints a similar picture.[17] In larger cities, we find a larger share of lawyers, while in the smaller cities, we find a larger fraction of paralegals and legal assistants. This difference suggests that in large cities, lawyers may deal with more complicated cases that require the work of many specialized lawyers, while in small cities, lawyers may handle more common, standard cases that can be handled mostly by paralegals and legal assistants supervised by only a few lawyers. For example, New York hosts some of the best law firms in the country, which specialize in a wide range of legal services from patent and copyright law to international corporate law.

Production is organized differently in larger cities even within the same industry. In particular, there is a larger fraction of managers in larger cities for a wide variety of industries. Differences in the organization of production might be reflecting the production of different quality levels of

---

[17]Garicano and Hubbard (2009) study the hierarchical organization of production in the legal services industry in the US and find that hierarchical production leads to a 30% increase in productivity.

the same product. It may be higher quality, more innovative, or more specialized commodities and services. Berkes and Gaetani (2017) find evidence of the concentration of knowledge-based activities in larger cities. It can also be a more complex set of tasks in the same production process. The production of these requires a higher fraction of managers or experts. The higher fraction of managers will result in differences in income inequality through two channels: first, if a higher fraction of managers is needed for production, managers will extract an income premium, and second, if only a few workers can work with each manager, that will generate higher inequality within workers. Better workers, as opposed to worse workers, will be able to match with better managers and be even more productive. In the next section, I present a model that speaks to these differences and links them to differences in the income distribution across cities.

3   A Spatial Model of Production Organization

The model embeds production in hierarchies as in Garicano and Rossi-Hansberg (2006) into a spatial equilibrium setting in the spirit of McFadden (1977).[18] There is a set of cities $C = \{1, ..., C\}$ indexed by c that differ in their production technology, amenities, and housing stock. The economy is populated by a unit mass of agents indexed by i that are heterogeneous in skill z, labor supply l and preferences for cities ϵ. Agents choose where to live and, conditional on their location, they choose whether to be a worker or a manager. In what follows, I describe in more detail the production technology as well as the choice of location and occupation.

3.1   Production

In this section, I describe how production takes place in a city taking as given the distribution of skills. The subscript c will indicate parameters that are specific to a city. In the following section, I describe the agent's location choice.

---

[18]There are abundant examples of spatial models in the same spirit. Some examples include Rosen (1979), Roback (1982), Eeckhout et al. (2014), Behrens et al. (2014), Diamond (2016), and Davis and Dingel (2016).

3.1.1 Hierarchical Team Production

Agents are heterogeneous in their skill z and their labor supply l. In a city, labor supply is distributed according to a lognormal distribution with mean 1 and variance $\left(exp\left(\sigma^2\right) - 1\right) exp\left(2 + \sigma^2\right)$, so that the variance of the log of labor supply is σ. Labor supply is independent of the distribution of skill and city preferences.[19] The unit of production is a team organized hierarchically in two layers. The top layer of a team is a single agent that I call a *manager* and the bottom layer consists of a mass of agents that I call *workers*.[20] Subscript w will denote the variables associated with the workers, and m will denote those associated with the manager. Individuals can also choose to produce on their own. However, in equilibrium, agents will optimally choose to produce as a part of a team and so I will abstract from this choice in what follows.

Production requires time and skill. Workers in a production team spend their working time encountering problems that need to be solved for production to happen. Skill is needed to solve the problems. For each unit of time that a worker spends working, they face a mass 1 of problems. The problems are heterogeneous in difficulty. The distribution of problem difficulties is given by $F_c\left(d\right) = d^{\alpha_c}$. I will refer to α_c as the *problem difficulty*. For each problem solved, a worker produces A_c units of output. I refer to A_c as the *productivity* of technology. A worker can only solve the fraction problems with difficulty lower than their skill. Therefore, a worker with skill z_w and labor supply l_w will solve a mass $F_c\left(z_w\right) l_w = z_w^{\alpha_c} l_w$ of problems, produce $A_c z_w^{\alpha_c} l_w$ units of output, and will be left with a mass $\left(1 - z_w^{\alpha_c}\right) l_w$ of unsolved problems. If the worker chooses to work on their own, their output will simply be $A_c z_w^{\alpha_c} l_w$ and the unsolved problems will be discarded. The worker is not able to assess the difficulty of the unsolved problems.

A manager of a team spends l_m units of time attempting to solve the unsolved problems that were encountered by the workers at the lower layer. A manager has to spend h_c units of time

[19]The variance of log working time, σ, is introduced in order to match residual income inequality. The distribution of labor supply will be independent from other characteristics of the agents within a city because of the assumption that the labor supply shock is realized after the choice of the city has been made.

[20]Although I abstract from the choice in the number of layers, I check for the restrictiveness of this assumption by allowing for the formation of a third layer in an extension of the model. Given the estimated parameters, a third layer is not optimal and therefore the two-layer assumption is not binding. For more details on how to endogenize the number of layers, refer to Garicano and Rossi-Hansberg (2006).

per unsolved problem to familiarize themselves with the problem. I refer to the time cost h_c as the *communication cost* since it controls how costly it is for workers to communicate unsolved problems to managers. A manager that supplies l_m units of time will be able to address l_m/h_c unsolved problems.

If a manager hires workers of skill z_w, these workers will solve a fraction $z_w^{\alpha_c}$ of the problems they encounter and pass on to the manager the remaining fraction $(1 - z_w^\alpha)$. Therefore, given their labor supply, l_m, the manager will be able to hire $L_w(z_w)$ units of the time of workers of skill z_w, where

$$L(z_w)\left(1 - z_w^\alpha\right) h_c = l_m. \tag{3.1}$$

The manager will then solve the problems of difficulty lower than his skill z_m so that a fraction $z_m^{\alpha_c}$ of the mass $L_w(z_w)$ of problems encountered by the workers in the lower level of the hierarchy are solved either by the manager or the workers. As a result, the output per unit of a manager's time of a hierarchy formed by workers of skill z_w and a manager of skill z_m is given by $Y(z_m, z_w) = \frac{A_c z_m^{\alpha_c}}{h_c\left(1 - z_w^{\alpha_c}\right)}$. Note that the output function $Y(z_m, z_w)$ is supermodular in the skill of the manager and the skill of the workers. Therefore the matching function will be increasing in equilibrium, so better managers will hire better workers. The proof for positive assortative matching is included in Section B.1 of the Appendix.

The production technology is thus fully characterized by three parameters: the productivity A_c, the problem difficulty α_c, and the communication cost h_c. In the next section, the matching problem is decentralized by letting managers hire workers and pay out wages.

3.1.2 Manager's problem

Consider the problem of an agent with skill z_m and labor supply l_m who chooses to be a manager. The manager takes as given the wage per unit of time for each worker's skill, $w_c(z)$ and chooses the skill of the workers they want to hire z_w. The earnings of a manager are given by the residual team output after paying out wages. Therefore, a manager of skill z_m that supplies

l_m units of time solves the following optimization problem,[21]

$$max_{z_w} \left(A_c z_m^{\alpha_c} - w_c (z_w) \right) L_w (z_w), \tag{3.2}$$

subject to

$$L_w (z_w) \left(1 - z_w^{\alpha_c} \right) = \frac{l_m}{h_c}. \tag{3.3}$$

Let the matching function $m_c (z_w)$ be defined as the skill of the manager that finds it optimal to hire workers of skill z_w. The solution to the manager's problem will be given by the inverse of the matching function. By plugging in the time constraint, we can get an implicit expression for the matching function, namely,

$$m_c^{-1} (z_m) = argmax_{z_w} \frac{\left(A_c z_m^{\alpha_c} - w_c (z_w) \right) l_m}{h_c \left(1 - z_w^{\alpha_c} \right)} = argmax_{z_w} \frac{A_c z_m^{\alpha_c} - w_c (z_w)}{1 - z_w^{\alpha_c}}. \tag{3.4}$$

Importantly, notice that the matching function does not depend on the time supplied by the manager, which simplifies the labor market clearing condition. Now we can write the manager's optimal payoff per unit of time as a function solely of the manager's skill,

$$R_c (z_m) = \frac{A_c z_m^{\alpha_c} - w \left(m_c^{-1} (z_m) \right)}{h_c \left(1 - m_c^{-1} (z_m)^{\alpha_c} \right)}. \tag{3.5}$$

Finally, the income earned by a manager with skill z_m that supplies working time l_m is given by $R_c (z_m) l_m$ and the income earned by a worker of skill z_w that supplies l_w units of time is given by $w_c (z_w) l_w$.

3.2   Agents

In this section, I describe the agents in the model, their preferences, and the choice of city given an income function per unit of time $I_c (z)$ in each city that is the result of the production problem described in the previous section.

---

[21]Note that I have assumed that a manager with ability z_m hires workers of homogenous skill z_w. This assumption is without loss of generality as proven in Antràs et al. (2006).

3.2.1 Preferences

There is a mass 1 of agents with skill $z \in (0, 1)$ distributed uniformly in the population.[22] Let us denote the overall distribution of skill by $Q(z)$. Agents' supply of working time l is distributed as a log-normal with variance $\left(exp\left(\sigma^2\right) - 1\right) exp\left(2 + \sigma^2\right)$ and mean of 1 and is independent from their skill and city preferences. The supply of working time is realized after a city has been chosen. This shock can be interpreted either as a shock to the available time of the agent or as a shock to the preferred supply of working time. Both interpretations are equivalent for the purpose of the model. Importantly, the shock is realized after the city choice is made, so it will not affect the relative inequality of cities and will be independent from the skill distribution within cities.

Agents derive utility from consumption of the numeraire, x, housing, h, city amenities that are enjoyed by all the residents, a_c, and an idiosyncratic amenity that is independently distributed across both agents and cities, ϵ_c^i.[23]

Conditional on choosing city c, agents solve the following optimization problem given their skill z and labor supply l,

$$max_{x,h} U^i(x, h, c) = max_{x,h} \left\{ h^\gamma x^{1-\gamma} + a_c + \epsilon_c^i \right\}. \tag{3.6}$$

$$subject\ to:\ I_c(z)\, l \geq p_c h + x, \tag{3.7}$$

where p_c denotes the housing price, $I_c(z)$ refers to the income per unit of time in city c as a function of skill, and ϵ_c^i is an idiosyncratic amenity that is independently distributed as an Extreme Value Type I distribution with variance β. Namely, $H(\epsilon) = exp\left\{-exp\left\{-(\epsilon)/\beta\right\}\right\}$. The

---

[22]Notice that the choice of distribution for the skill distribution cannot be separated from the distribution of problem difficulty, and therefore the choice of a uniform distribution of skill is a normalization. In order to illustrate this, let $\bar{z}$ be the skill that solves y problems $\bar{z} = F^{-1}\left(\frac{y}{A}\right)$. Then we must look for the mass of people with skill at least $\bar{z}$, that is $G\left(F^{-1}\left(\frac{y}{A}\right)\right)$. The answer, therefore, depends on $G \circ F^{-1}$. If we allow for enough flexibility on both distributions, F can always undo the effect of G. Therefore, I normalize G to a uniform distribution to emphasize that skill is a mere ranking and that the way it translates into productivity and payoffs will depend on the production technology.

[23]The common amenity a_c can be interpreted as the mean of the idiosyncratic amenity ϵ_c^i. City amenities enter additively in the utility of agents. This assumption allows for a simple expression for the probability of allocating to a particular city. Alternatively, I could rewrite amenities multiplicatively and assume a Fréchet distribution for the idiosyncratic amenities.

variance of idiosyncratic amenities β controls the strength of mobility frictions. The variance is inversely related to mobility. Intuitively, when the variance of the idiosyncratic amenity is higher, the two cities are worse substitutes and differences in incomes across cities result in lower mobility.

3.2.2   Occupational choice

Agents choose whether to become managers or workers in order to maximize income and since the time endowment does not affect this choice, income per unit of time $I_c(z) = max\{w_c(z), R_c(z)\}$. Let the equilibrium occupational choice be characterized by a threshold skill z_c^* such that agents with lower skill than the threshold will become workers, and agents with higher skill will become managers. It turns out this assumption is without loss of generality as shown in Antràs et al. (2006). The required condition required is that at the threshold, the derivative of the wage function must be lower than the derivative of the managerial rent,

$$w_c'\left(z_c^*\right) < R_c'\left(z_c^*\right). \tag{3.8}$$

Intuitively, this extra condition guarantees that at the threshold skill, where $w_c\left(z_c^*\right) = R_c\left(z_c^*\right)$, a small decrease in skill will decrease wages by less than managerial income, so the agent will prefer to become a worker, while a small increase in skill will increase managerial income more than the wage so that the agent will prefer to be a manager. This condition also guarantees that workers do not want to be self-employed. For the set of parameters for which this condition holds, the threshold equilibrium exists and it is always unique. The proof of the existence and uniqueness of a threshold equilibrium in this setting is presented in Garicano and Rossi-Hansberg (2006). During the estimation, I check that the estimated parameters fall in this region.

3.2.3 Sorting into locations

Agents sort into locations in order to maximize their expected indirect utility. The expected utility from living in a particular city c for an agent of skill z is given by:

$$E[V_c(z)] = \Gamma \frac{I_c(z)}{p_c^{\gamma}} + a_c + \varepsilon_c = \tilde{V}_c(z) + \varepsilon_c,$$ (3.9)

where $\Gamma = \gamma^{\gamma}(1-\gamma)^{\gamma}$ and γ is the Cobb-Douglas weight on housing. The density of skills in a city $g_c(z)$ in equilibrium, with CDF $G_c(.)$, is obtained by multiplying the overall density of skills $q(z)$ by the fraction of those skills that live in each city $\pi(c,z)$. Under the Extreme Value Type I distributional assumption, the density of agents of skill z that optimally choose to live in city c is given by,[24]

$$g_c(z) = \pi(c,z)q(z) = \frac{exp\left\{\tilde{V}_c(z)/\beta\right\}}{\left(\sum_{j=1}^{C} exp\left\{\tilde{V}_j(z)/\beta\right\}\right)}q(z).$$ (3.10)

Recall that the global distribution of skills was normalized to a uniform $[0,1]$. So for all z, $\sum_{c \in C} g_c(z) = 1$.

3.3 Housing Market

Each city has a fixed supply of land H_c, which is owned by absentee landlords. The expenditure on housing is therefore not part of the income received by agents. This is a simplifying assumption that allows me to abstract from the choice of acquiring land in a city. Cobb-Douglas preferences imply that agents spend a fixed fraction γ of their income on housing. The housing price clears the market,

$$\int_0^1 \gamma \frac{I_c(z)}{p_c}g_c(z)\,dz = H_c.$$ (3.11)

---

[24] For the derivation of this result and a discussion of the multinomial logit limitations, the reader can refer to McFadden (1977).

3.4   Labor Market

The equilibrium matching function clears the labor market in each city. For each skill z_w the amount of time supplied by workers of skill lower or equal than z_w must be equal to the amount of time demanded by the managers who will hire those workers. Namely,

$$\int_0^{z_w} \int_0^\infty l g_c(s) \, g_c(l) \, dl \, ds = \int_{z_c^*}^{m_c(z_w)} \int_0^\infty \frac{l g_c(s)}{h_c \left(1 - m_c^{-1}(s)^{\alpha_c}\right)} g_c(l) \, dl \, ds, \ \forall z_w \in \left[0, z_c^*\right]. \tag{3.12}$$

Because the supply of labor is independent from the distribution of skills within cities, it integrates to 1 in the labor market clearing condition, which can be written as

$$\int_0^{z_w} g_c(s) \, ds = \int_{z_c^*}^{m_c(z_w)} \frac{g_c(s)}{h_c \left(1 - m_c^{-1}(s)^{\alpha_c}\right)} ds, \ \forall z_w \in \left[0, z_c^*\right]. \tag{3.13}$$

Differentiating both sides with respect to z_w results in the following differential equation involving the matching function,

$$g_c(m_c(z_w)) \, m_c'(z_w) = g_c(z_w) \, h_c \left(1 - z_w^\alpha\right). \tag{3.14}$$

The initial and terminal conditions for this differential equation follow from market clearing and positive assortative matching. The lowest-skilled worker is hired by the lowest-skilled manager, $m_c(0) = z_c^*$ and the highest-skilled worker is hired by the highest-skilled manager $m_c(z_c^*) = 1$. The solution to that differential equation, together with the initial condition, $m_c(0) = z_c^*$ results in an expression for the matching function,

$$m_c(z) = G_c^{-1} \left(G_c(z_c^*) + \int_0^z g_c(s) \, h_c(1 - s^{\alpha_c}) \, ds\right). \tag{3.15}$$

The terminal condition for the matching function $m_c(z_c^*) = 1$ provides an expression for the threshold skill that depends exclusively on parameters and the skill distribution,

$$G_c(1) = G_c(z_c^*) + \int_0^{z_c^*} g_c(s) h_c(1 - s^{\alpha_c}) ds. \tag{3.16}$$

Notice that the matching function depends solely on the skill distribution, the communication cost, and the problem difficulty so that shifts in the Hicks neutral productivity A_c will only affect the matching through the skill distribution. More importantly, the matching function does not depend on the wage schedule. Otherwise, solving for the equilibrium would require finding a solution to a system of differential equations on both the matching function and the wage function.

3.5 Equilibrium

In this section, I describe the equilibrium for this economy. I start by defining a threshold equilibrium and then I will proceed to characterize the equilibrium.

Definition 1. Spatial threshold equilibrium

An equilibrium for this economy is a set of thresholds $\{z_c^*\}_{c=1}^C$, matching functions $\{m_c(z)\}_{c=1}^C$, wages $\{w_c(z)\}_{c=1}^C$, housing prices $\{p_c\}_{c=1}^C$, and density functions $\{g_c(z)\}_{c=1}^C$ such that:

1. Managers optimally choose the skill of the workers they want to hire $m_c^{-1}(z)$ taking the wage function $w_c(z)$ as given (Eq. 3.4).

2. The threshold skill summarizes the optimal occupational choice. Agents of skill $z < z_c^*$ optimally choose to become workers and those of skill $z > z_c^*$ optimally choose to become managers, and $w_c(z_c^*) = R_c(z_c^*)$.

3. The skill distribution is the result of optimal city choice. The density functions $g_c(z)$ are the result of the optimal choice of city (Eq. 3.10).

4. Housing market: housing prices p_c^h are such that all local housing markets clear (Eq. 3.11).

5. Labor market: wage functions $w_c(z)$ are such that all local labor markets clear (Eq. 3.13).

3.5.1 Characterization of Wage Function

The wage schedule is such that the matching function solves the manager's problem. The first order condition from the manager's problem results in the following differential equation for the wage function, conditional on the matching function,

$$
w_c(z) \frac{\alpha_c z^{\alpha_c - 1}}{1 - z^{\alpha_c}} - w_c'(z) = A_c \frac{\alpha_c z^{\alpha_c - 1} m(z)^{\alpha_c}}{1 - z^{\alpha_c}}. \tag{3.17}
$$

The solution for this differential equation together with the initial condition on the wage function $w_c(0)$ deliver the following expression for the wage function,

$$
w_c(z) = (1 - z^{\alpha_c}) \left(w_c(0) + \int_0^z \frac{\alpha_c s^{\alpha_c - 1}}{(1 - s^{\alpha_c})^2} A_c m_c(s)^{\alpha_c} \, ds \right), \tag{3.18}
$$

where the initial condition $w_c(0)$ needs to be found. Agents at the threshold skill must be indifferent between being a manager or a worker, $w_c(z_c^*) = R_c(z_c^*)$. This provides the condition that determines the initial wage. Namely,

$$
w_c(0) = \left(h_c \left(1 - z_c^{*\alpha_c} \right) + 1 \right)^{-1} \left(A_c z_c^{*\alpha_c} - h_c \left(1 - z_c^{*\alpha_c} \right) \int_0^{z_c^*} \frac{\alpha_c s^{\alpha_c - 1}}{(1 - s^{\alpha_c})^2} A_c m_c(s)^{\alpha_c} \, ds \right), \tag{3.19}
$$

where $R_c(z)$ denotes the managerial income given the optimal choice of workers' skill.

3.5.2 Existence of Equilibrium

Recall that the density of skills and cities is the result of the optimal choice of individuals and is given by the following expression:

$$
g_c(z) = \frac{exp \left\{ \left(\Gamma \frac{I_c(z; g_c(.))}{p_c^\gamma} + a_c \right) / \beta \right\}}{\left(\sum_{j=1}^C exp \left\{ \left(\Gamma \frac{I_j(z; g_j(.))}{p_j^\gamma} + a_j \right) / \beta \right\} \right)}. \tag{3.20}
$$

Notice that I have made explicit the fact that the income function depends on the whole skill distribution through the matching function and the wage schedule. This equilibrium condition defines a mapping of skill distributions on skill distributions $\hat{g}(z) = \langle g_1(z), ..., g_c(z), ..., g_C(z) \rangle$. Proposition 1 formally states the existence of an equilibrium as the existence of a fixed point for this mapping.

Proposition 1. *Existence of Equilibrium*

Let $\overrightarrow{g}(z) = \langle g_1(z), ..., g_C(z) \rangle$ be a vector valued function mapping $[0, 1] \rightarrow \mathbb{R}^C$, such that $\sum_{n=1}^{C} g_n(z) = 1$ and $g_n(z) > 0$. Let T be the mapping defined by:

$$T\left(\overrightarrow{g}(.)\right) = \left\langle \frac{exp\left\{\frac{1}{\beta}(I_1(z; g_1(.))) + a_1)\right\}}{\sum_{j=1}^{C} exp\left\{\frac{1}{\beta}(I_j(z; g_j(.)) + a_j)\right\}}, ..., \frac{exp\left\{\frac{1}{\beta}(I_C(z; g_C(.))) + a_1)\right\}}{\sum_{j=1}^{C} exp\left\{\frac{1}{\beta}(I_j(z; g_j(.)) + a_j)\right\}} \right\rangle \quad .$$

There exists a vector valued function, $\overrightarrow{g}(z)^*$, such that it is a fixed point of the mapping T, $T\left(\overrightarrow{g}(z)^*\right) = \overrightarrow{g}(z)^*$.

The proof of Proposition 1 is an application of Schauder's fixed point and is included in Section B.2 of the Appendix. Unfortunately, there is no proof of uniqueness, so when taking the model to the data I will try different starting points for the distribution of skills and check that the solution converges to the same distribution.

4 Quantification

I quantify the model to match two representative cities: a large city that includes all the CSAs with a population larger than 2.5 million inhabitants, and a small city that includes CSAs with a population between 100,000 and 2.5 million inhabitants. For reference, Denver is the city closest to the threshold of 2.5 million inhabitants. Two parameters are taken from the literature. The first of these parameters is the Cobb-Douglas weight on housing, γ, which is set to 0.24 to match the average expenditure on housing in the US. This expenditure share is found to be remarkably

constant across cities in Eeckhout et al. (2014). The second is the variance of the Extreme Value Type I distribution, β, which controls the mobility. This value captures the elasticity of city size with respect to income. In order to estimate this elasticity directly from the data, it is crucial to correctly instrument for income since it is correlated with many confounding factors affecting city size. I will take this elasticity from the estimation in Diamond (2016). The author estimates two elasticities in this paper, one for college educated and the other for non-college educated. I will use an elasticity of 0.3, which is roughly in between the two elasticities. Table 3 summarizes the parameters borrowed from the literature.

The rest of the parameters are estimated in three stages. In a first stage, I estimate the housing prices, p_c from hedonic regressions and use them directly in the model to retrieve the housing stock, H_c. In a second stage, given the housing prices, and the parameters borrowed from the literature, I set up a Simulated Method of Moments to jointly estimate amenities a_c and the technology parameters (A_c, h_c, α_c). In a third stage, I obtain the variance of the log labor supply shock, σ, to match the residual overall variance of log wages. In what follows, I explain each of the stages in more detail.

Housing prices are the result of a hedonic housing price regression estimated using the housing section of the American Community Survey, Public Use Microdata Sample for the years 2010–2014. The hedonic regression includes the number of rooms, the year the unit was built, and the type of apartment, in order to control for differences in the characteristics of the housing stock. The estimated housing prices are 1.2 for the large city and 0.9 for the small city, and the corresponding housing stocks are 0.96 for the large city and 0.08 for the small city. The complete results from these regressions are presented in section A.2 of the Appendix. Recall that even though there is a unit mass of agents in the economy, they do not consume a unit of housing, but they can choose the optimal level of housing, given housing prices. Because the housing stock does not have natural units, I normalize the housing prices so that the average price is 1. As expected, the large city is bigger in terms of housing stock and also has a more expensive price per unit of housing.

In the second stage, I estimate amenities and production technology in an exactly identified

Table 4.1: Calibrated Parameters

	Description	Value	Target
γ	Cobb-Douglas Housing Weight	0.24	Fraction of Housing Expenditure Davis and Ortalo-Magné (2011)
H_c^s	Housing Stock	[0.9642, 0.0775]	Hedonic Price Regressions (p_c = [1.2080, 0.9111])
β	Mobility	0.3	Income Elasticity of City Size Diamond (2016)

Note: This table reports the parameters in the model that are quantified before the Simulated Method of Moments.

Simulated Method of Moments. There are four parameters to be estimated — three from the production function for each city, specifically, the communication cost (h), the problem difficulty (α), and the productivity (A) — and a fourth from the amenity for the large city (a) (since the amenity for the small city is normalized to zero). Although there is not a one-to-one relationship between the parameters and the moments: 1) changing the communication cost controls the span of control, 2) the problem difficulty controls the difference between the median income of managers and workers, 3) the productivity controls the level of both the median income of managers and workers, 4) the amenity level governs the relative city size, and 5) the variance of log time supply affects the overall variance of income in the economy. The estimated parameters are reported in Table 4.2. The fit of the estimation is reported in Section C.1 of the Appendix.

Finally, in a third stage, I calculate the variance of the log supplied working time (σ) that matched the residual variance of log wages. It is possible to calculate this variance after the Simulated Method of Moments since it does not interact with the targeted moments. The independence of the distribution of the labor supply and the distribution of skill follows from the assumption that the labor supply shock is realized after the choice of city. Otherwise, the choice of city will depend on the labor supply in a manner that interacts with skill. Given the independence of the distribution of labor supply and skill, the total variance of income is the

sum of the variance of log income per unit of time and variance of log labor supply, which is

$$Var\left(log\left(Observed\ Income\right)\right) = Var\left[E\left(log\left(I_c\left(z\right)l\right)\right)\right] + E\left[Var\left(log\left(I_c\left(z\right)l\right)\right)\right] \tag{4.1}$$

$$= Var\left[E\left(log\left(I_c\left(z\right)\right)\right)\right] + E\left[Var\left(log\left(I_c\left(z\right)\right)\right)\right] + \sigma. \tag{4.2}$$

Therefore, the variance of log labor supply, σ, is the difference between the observed variance of log income and the model-generated variance of log income. The variance of the log labor supply shock is 40% of the total variance of log income, implying that the model can explain the remaining 60%. For reference, the classical Mincer equation can explain around a third of total variance, Mincer (1975).

Table 4.2: Estimated Parameters

		Large City	Small City
A	Productivity	8.4139	6.7370
h	Communication cost	0.7367	0.6776
α	Problem Difficulty	0.8725	0.5011
a	Amenity	1.0471	0
σ	Variance of Log Working Time	0.337	

Note: This table contains the parameters that result from the Simulated Method of Moments, plus the variance of log working time. The variance of log working time is calculated as a residual after the estimation procedure.

The production function parameters that come out of the estimation are consistent with the idea that production in larger cities is organized differently in part because larger cities use a technology that embeds harder problems that are harder to communicate. Larger cities may be producing higher quality, more innovative, or more customized products that require more management time as well as higher skill. The large city is characterized by a higher communication cost and a higher problem difficulty that is compensated by a higher productivity. I will refer to the production technology in the large city as the complex technology since the higher productivity comes at the cost of more skill and management intensity as compared to

the small city's technology, which I will refer to as the simple technology.

Interestingly, the complex technology used by the large city does not dominate the simple technology. The complex technology is better than the simple one in terms of productivity, but it is worse both in terms of the higher communication cost and in terms of the higher problem difficulty. This can potentially explain the persistent technology differences across cities of different sizes. I explore below whether agents in the small city would benefit from adopting the complex technology in the following section.

Table 4.3: Non-Targeted Moments

	Large City		Small City		Ratio Large to Small	
	Model	Data	Model	Data	Model	Data
Average Income	$51,536	$66,971	$44,504	$55,606	1.16	1.2
Variance of Log Income	0.5798	0.5758	0.4678	0.4744	1.24	1.21
90th to 50th Percentile Ratio	$26,241	$26,650	$24,341	$24,788	1.08	1.07
50th to 10th Percentile Ratio	$27,338	$25,236	$23,831	$22,548	1.15	1.12

Note: This table includes the fit of the model with respect to moments that were not targeted in the estimation.

The estimated model is able to replicate a series of patterns in the data that were not targeted in the estimation. In Table 4.3 I report the average income and the variance of log income in both the large and the small city, as well as the 90th to 50th and the 50th to 10th percentile ratios. The average income in the model is a bit lower than in the data because of the long upper tail in the income distribution in the data that is not matched in the model. However, the ratio of average income of the large city to the small city is close to the one in the data. Recall that the observed income from the American Community Survey is top-coded at the top one-half percent of incomes in each state and replaced by the average income of all top-coded incomes in the state. Average incomes will be particularly sensitive to this top coding and for that reason, I chose to target median incomes instead.

The model matches well the ratio of the variance of log income in the large city to the

Figure 4.1: Density of Log Income and Real Log Income

(b) Log Income Density



(b) Log Real Income Density
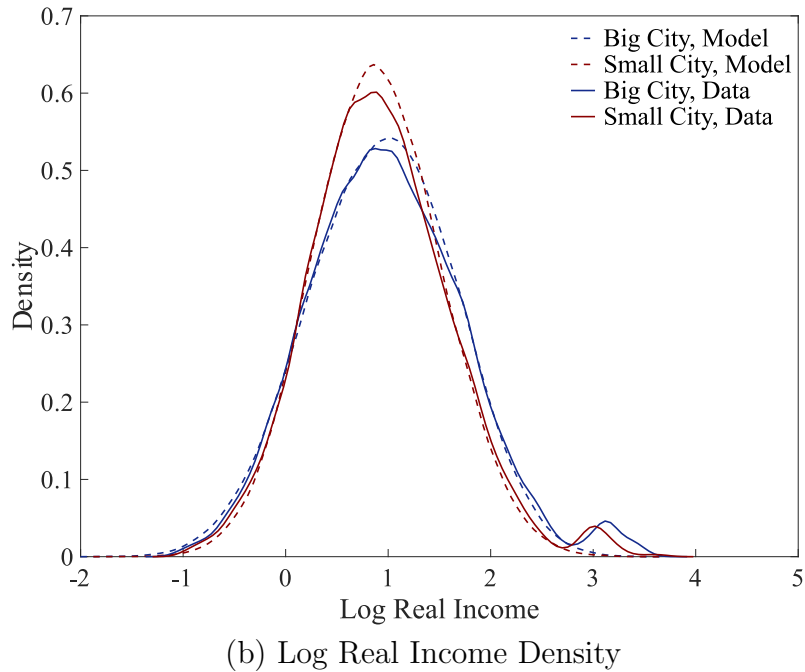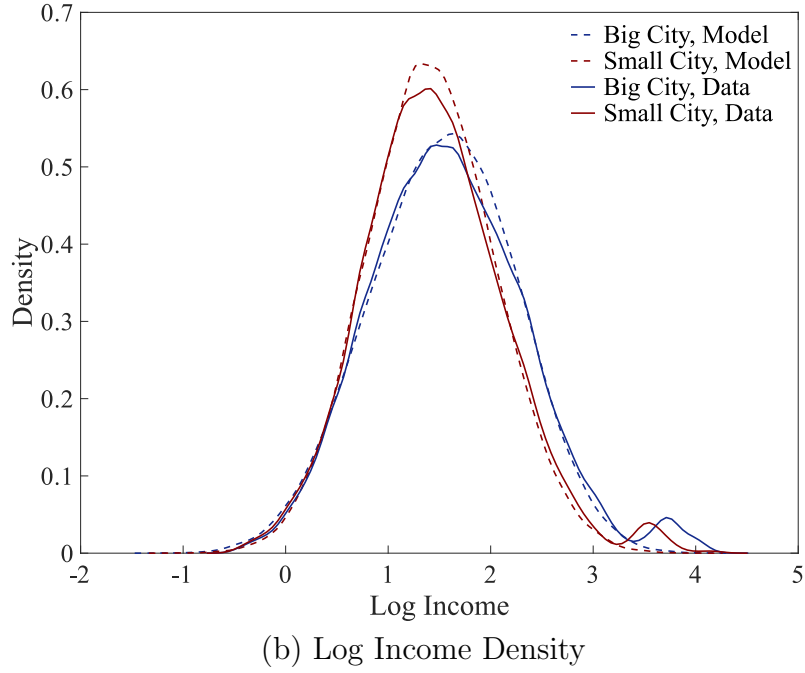
Note: This figure plots the density of log income and real log income in the data and in the estimated model using a normal kernel smoothing with bandwidth 0.1. The real income is calculated by dividing income by the price index implied by the model, i.e, the housing price raised to the Cobb-Douglas weight on housing p^γ. The large city is formed by CSAs with a population larger than 2.5 million inhabitants, and the small city is formed by all the CSAs with a population smaller than 2.5 million in habitants.

variance of log income in the small city. As shown in Table 4.3, larger cities have a higher income inequality compared to small cities. The model also does well in terms of the 90th to 50th and the 50th to 10th percentile ratios. The model does generate slightly higher 90th to 50th percentile ratios and slightly lower 50-10 percentile ratios compared to the data, but the ratio of large to small for both ratios is remarkably close to the data.

Next, I compare the real income distribution to the one in the data. The model-based real income is calculated by dividing income by the price index. Namely, $I_c^{Real}(z) = \frac{I_c(z)}{p_c^{\gamma}}$. As usual, the price index is given by p^{γ} as a result of Cobb Douglass preferences (given the price of the consumption good is normalized to 1). Both the nominal and real income distributions are plotted in Figure 4.1. The model does well in matching the fact that the income distribution in the large city is shifted to the right and is more dispersed than the income distribution in the small city. Moreover, the model is able to generate a fat tail at the bottom of the income distribution, which was documented to be a robust feature of the data in Eeckhout et al. (2014). Not surprisingly, the model does not generate the bump at the top of the income distribution which is the result of the top-coding of income in the ACS. Recall that income above the 95.5th percentile in each state is top-coded and replaced by the mean income of all top-coded values in the state.

Finally, the bottom fat tail is not only present in the real income distribution but also in the distribution of skills. This is consistent with the fat tail in large cities in the distribution of measures of skill, such as educational attainment, documented by Eeckhout et al. (2014). Figure 4.2 plots the fraction of each skill type that locates in the large city. Recall that the overall skill distribution is uniform, so skills can be interpreted as percentiles of the population. For example, a skill of 0.1 corresponds to a productivity level such that 10% of the population is below that level. This figure captures the fact that a higher fraction of the very low skill types sorts to the large city as compared to the immediately higher skilled workers. The kinks in this fraction are due to kinks in the income functions at the threshold skills. For reference, the large city attracts 93% of all the agents in the economy but relatively more of the lowest and highest skilled.

Figure 4.2: Fraction Located in the Large City
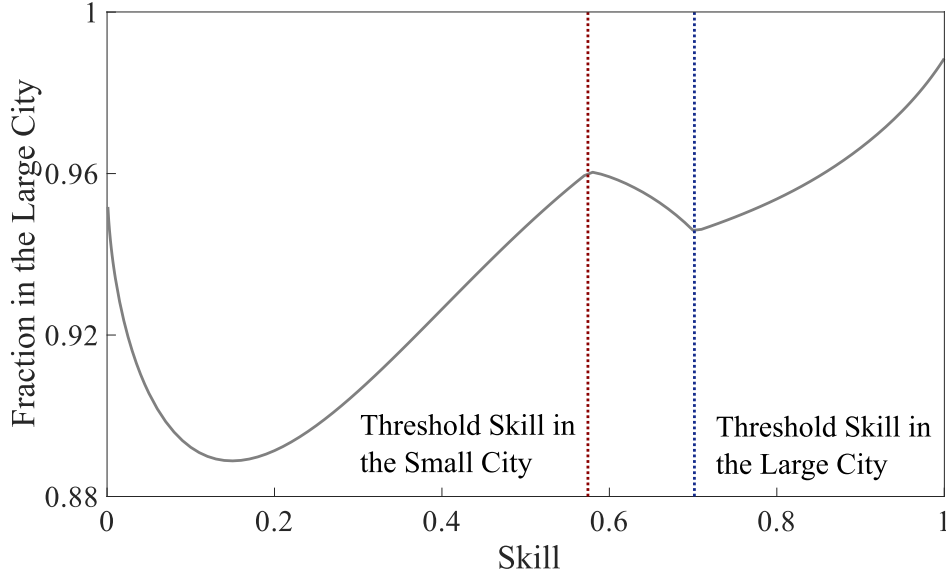
Note: This figure plots the fraction of each skill type that chooses to locate to the large city in the estimated model. The dotted lines represent the threshold skill in each city such that agents of skill lower than the threshold optimally choose to become workers and those with skill higher than the threshold optimally choose to become managers.

In order to gain intuition for why the large city is more unequal and has fatter tails in the real income distribution, it is helpful to first look at the income schedule, which is plotted in Figure 4.4. The income schedule is steeper in the large city for most skills, which generates higher inequality in the large city. The difference in the slope of the income functions in the large and small city determines the slope of the skill distribution. The fraction of skills in the large city is decreasing when the slope of the income function in the small city is steeper than in the large city. This is because the small city becomes increasingly attractive when the income function is steeper. The income schedule in the small city is steeper for the lowest-skilled workers compared to the larger city. The higher skill premium at the bottom implies that the small city becomes more attractive as we move up the skill distribution for the lowest skills, and so the large city attracts more of the lowest-skilled workers compared to the middle-skilled workers. This explains that the fraction of agents that locate in the large city is decreasing for the lowest skills. At some point, the income function in the large city becomes steeper than in the small city, and the fraction of skills in the large city becomes increasing again. The middle downward

Figure 4.3: Income Schedule
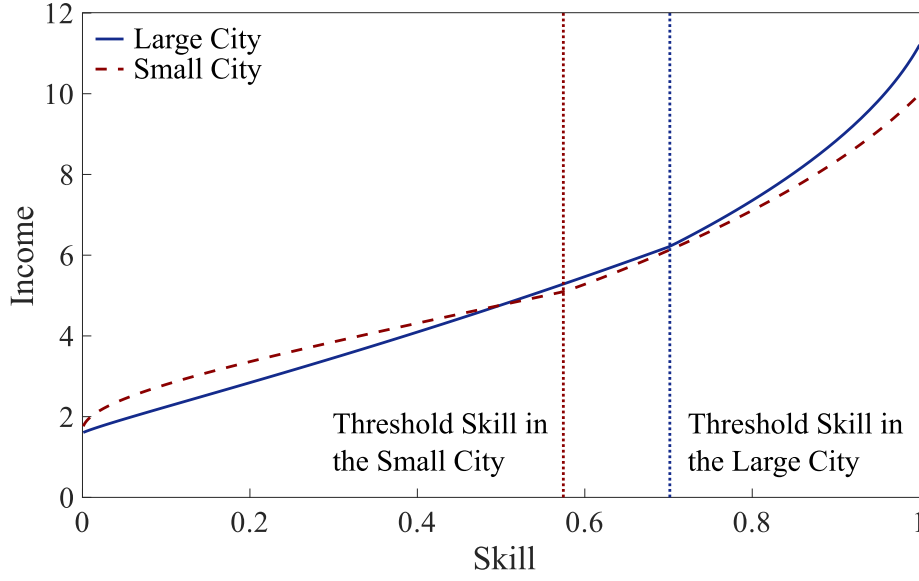
Note: This figure plots the income function of the large city and the small city in equilibrium. The vertical lines correspond to the threshold skill in the large and the small city so that agents with skills below the threshold are workers and those above are managers.

sloping interval corresponds to skills that choose to be managers in the small city but workers in the large city. For these skills, the income function in the small city is steeper than in the large city.

In order to understand the shape of the income schedule, it is useful to think about the productivity of the hierarchy. In what follows, I refer to the workers plus manager as a hierarchy and to the workers as a team. The total output of a hierarchy formed by a unit of time from a manager of skill z_m and workers of skill z_w is given by the product of two terms. First, the output per unit of worker's time, $A_c z_m^\alpha$; which depends on the skill of the manager, to which I will refer as the *manger's productivity*. Second, the units of workers time $\frac{1}{h_c(1-z_w^\alpha)}$, which I will refer to as *team size*. The productivity of the hierarchy is given by the total output divided by the sum of the time from the workers and the manager. Therefore, the hierarchy productivity is increasing both in the manger's productivity and in the team size. In equilibrium, higher-skilled workers match with higher-skilled managers so that the productivity of a hierarchy formed by higher-skilled agents is higher both through the manager's productivity and the team size.

Figure 4.4: Matching
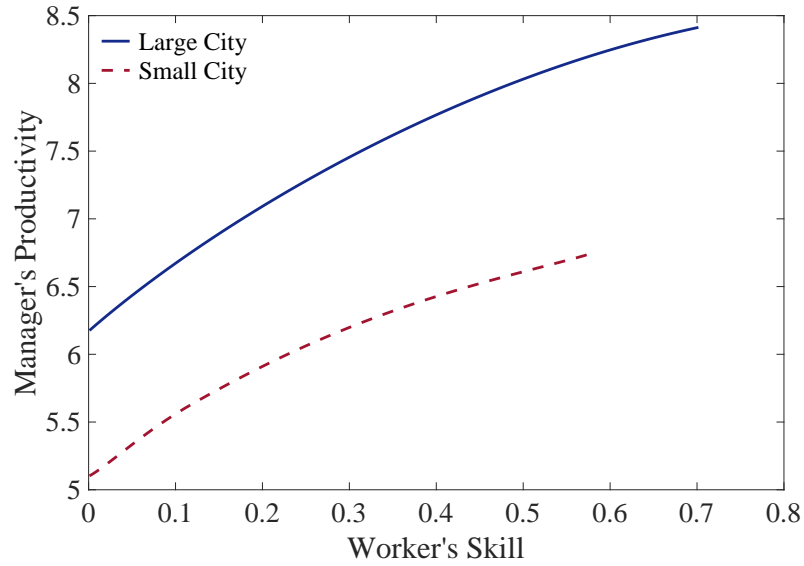
(a) Matching to Manager's Productivity
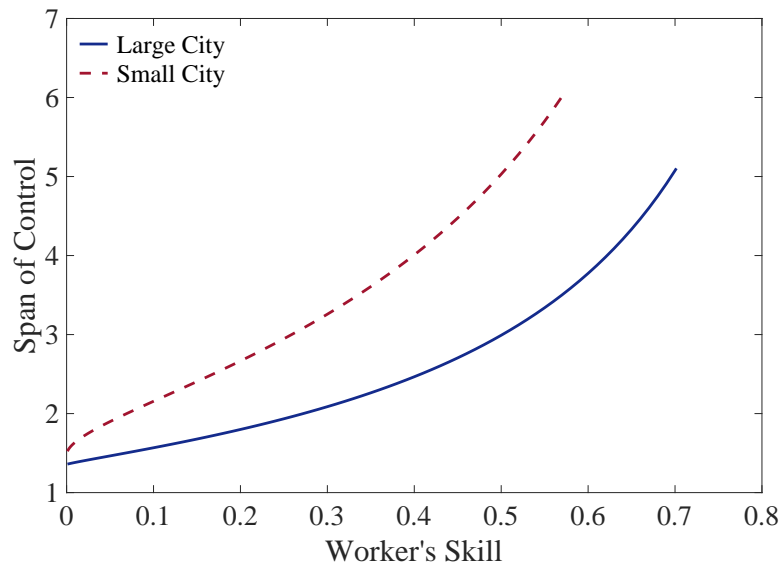
(b) Team Size

Note: This figure plots the matching to manager's productivity in panel (a) and the size of the team, or number of workers in panel (b) for the estimated equilibrium.

In order to see how these two effects vary differently with the skill of the worker in the large and the small city, Figure 4.4 plots the matching of worker's skill to manager's skill, to manager's productivity and to team size in each city. The advantage of the simple technology used in the small city comes from larger teams, while the advantage from the complex technology comes from higher manager's productivity. The team size effect is especially strong for the lowest skills because an increase in team size has a larger effect on the hierarchy's productivity for a smaller hierarchy than for a large one. The steep increase in team size translates in the steeper income function in the small city for the lower-skilled workers. The manger's productivity effect dominates for the rest of the skills, generating more inequality in the large city.

5   Spatial Differences in Technology

In the previous section, I estimated production technologies in both representative cities. Differences in technology may come from long-term differences in characteristics across cities such as access to natural resources, geographical location, infrastructure, or institutions. New York is not only different because of the people who live there, but also because of infrastructure, such as access to international airports, or institutions, such as the stock exchange. These persistent characteristics may give New York a comparative advantage using a more complex technology to solve harder and more costly to communicate problems. To the extent that the technology differences measured in the estimation are tied to exogenous or long-term differences across cities, taking technology differences as exogenous is a reasonable assumption when evaluating policy changes.

In this section, I investigate to what extent the differences in technology that were estimated are desirable. Since technology is three dimensional, the ranking of technologies will be more subtle than with the usual one-dimensional technology. In particular, the simple technology of the small city is not dominated in every dimension by the complex technology of the large city. Even though it has lower productivity, it is also characterized by easier problems that are less costly to communicate, which makes the technology a better fit for some low-skilled individuals.

For this exercise, I consider two available technologies: the complex technology used by the large city with high productivity (high A) but hard problems (high α) that are costly to communicate (high h), and the simple technology with low productivity (low A) but easy problems (low α) that are cheap to communicate (low h).[25]

I define the social welfare function in a city as the utility per capita. This social welfare function deviates slightly from aggregate utility as in the classical utilitarian function at the city level. If aggregate utility is used as the social welfare in a city, then social welfare would increase by adding new people to the city even if the average utility went down.[26] I will first determine whether the technological differences are desirable for the small city and the large city, and then I will determine whether they are desirable from an aggregate perspective.

I consider two criteria to determine whether the technological differences are desirable for a city. According to the first criterion, the welfare function of city c is defined as the utility per capita for residents of city c, regardless of whether the increase in utility is due to composition changes in the residents of the city. According to the second criterion, the welfare function for city c is calculated as the utility per capita of the current residents of the city c, regardless of whether these residents move out of the city as a result of changing technology.[27] When reporting the results, I refer to the first criterion as "City Welfare" and the second criterion as "Residents Welfare." Table 5.1 reports the percentage change in welfare that would result first from the large city adopting the simple technology and then from the small city adopting the complex technology. I report both welfare criteria and the change for the large and the small city.

On the one hand, the large city only loses slightly, about 2%, from adopting the small city's technology in terms of utility. The reason is precisely that even though the simple technology

---

[25]I also a quantify the production technology for four cities in the US: New York, Chicago, Pittsburgh, and Springfield, MA. The results are included in Section C.2 of the Appendix. In the four-cities estimation, I find a similar pattern where productivity is monotonically increasing in city size. However, the lower-productivity technologies have either easier problems or lower cost communication so that the technologies of the smaller cities like Pittsburgh or Springfield are not dominated by the technology of New York or Chicago.

[26]The average utility has been used as the standard measure of social welfare in the optimal growth literature following Samuelson (1975).

[27]In order to calculate the change in "Residents Welfare," I simulate the economy with each individual being characterized by a draw of idiosyncratic preferences for the large city from an Extreme Value Type I distribution with variance β. I calculate the optimal city choice for everyone in the initial equilibrium and the optimal city in equilibrium after the policy is implemented. The change in the "Residents Welfare" in the large city is then calculated as the change in average utility for those agents who chose the large city in the initial equilibrium regardless of whether they choose the large or the small city after the policy implementation. The change in "City Welfare" for the large city is the change of average welfare between those who chose the large city before the policy and those who chose the large city after the policy. I draw a new working-time shock for each individual following the new choice of a city after the policy implementation.

Table 5.1: Welfare Effect from Changes in Technology

	Large City Adopting Simple Tech.		Small City Adopting Complex Tech.	
	Large City	Small City	Large City	Small City
City Welfare	-2.12%	16.82%	-1.24%	20.65%
Residents Welfare	-1.33%	4.13%	0.08%	-2.75%

Note: This table contains the welfare effect of a change in technology. The first row calculates welfare as the change in average utility of the people who were living in each city before and after the technological change. In particular, it includes changes in the average utility coming from changes in the composition of who decides to live in each city. The second row calculates welfare as the change in average utility of those who were living in each city before the technological change. In particular, it includes the utility of those who move out of the city but not of those who move into the city. The first two columns correspond to the large city adopting the simple technology while the small city maintains the simple technology and the next two columns correspond to the small city adopting the complex technology while the large city maintains the complex technology.

is worse for the top skills, it is better for the bottom skills. Because the large city is so much larger than the small city, its skill distribution is close to uniform and the utility loss for the high-skilled agents is compensated by the utility gain for the low-skilled agents so that there is only a small aggregate effect. The small city would benefit from a 17% increase in welfare if the large city adopted the simple technology. This large effect is mostly for the small city attracting more of the highest-skilled agents. In terms of the small city residents, they would experience a 4% increase in per capita utility due to the increase in demand for low-skilled workers generated by the incoming high skill.

On the other hand, the small city experiences a large effect from adopting the large city's technology. The average utility would increase by 20% comparing the original residents to the new residents. A large part of this effect is due to the small city's becoming a more attractive place for high-skilled agents. The increase in the abundance of high-skill agents increases income and utility through a composition effect. However, this increase is naturally at the expense of the large city, which loses about 1% in average utility.

However, if the criterion is to maximize the utility of the original residents, adopting the complex technology would decrease average utility by 2.75%. Therefore, in terms of the welfare

Table 5.2: Percentage Change in Average Utility and Total Output

	Both Cities Using Simple Tech.	Both Cities Using Complex Tech.	Large City Using Simple Tech. & Small City Using Complex Tech.
Average Utility	-1.07%	-0.05%	-0.8%
Total Output	-1.99%	-0.27%	-1.34%

Note: This table contains the effect of a change in technology for the aggregate economy. The first row contains the effect on average utility and the second row contains the effect on aggregate output.

function of the residents of the small city, it is not desirable to switch technologies. In that sense, the fact that the small city uses a simpler technology is perpetuated by the behavior of the small city residents, a least conditional on the initial equilibrium.

In order to understand the optimal choice of technology for the economy as a whole, I consider the problem of a social planner who chooses the technology of each city, allowing for the optimal allocation of agents across cities. In Table 5.2, I plot the percentage change in the aggregate output and aggregate utility from changing technologies first so that both cities use the complex technology, second so that both cities use the simple technology, and third so that the large city uses the simple technology and the small city uses the complex technology. Interestingly, I find that the current allocation of technologies is optimal both in terms of utility and output. It is only slightly better than both cities using the complex technology.

This exercise illustrates that the idea that small cities have worse technology is an oversimplification. In reality, the diffusion of the complex technology to the small cities is not necessarily a good thing, which is consistent with the fact that we do not see it happening. This result speaks to the literature on place-based policies as in Kline and Moretti (2014) and Chetty et al. (2014), in which the authors study the circumstances under which placed-based policies are preferred to people-based policies. This framework does not feature any of the market imperfections that can potentially justify placed-based policies. However, since the transfer of technology is not modeled, the framework allows for the small city to have a worse technology than the large city, in which case it would be beneficial for the small city to adopt the big city's technology. The

Table 6.1: Percentage Change in the Variance of Log Income

	Large City	Small City	Ratio of Large to Small	Overall
Complex Tech. in Both Cities	1.62%	19.07%	-14.52%	2.49%
Simple Tech. in Both Cities	-26.54%	-11.45%	-16.93%	-25.81%
Uniform Skills in Both Cities	0.98%	-9.31%	11.29%	0.19%

Note: This table contains the change in the variance of log income that results from three exercises. The columns respectively contain the change in the variance of log income for the large city, the small city, the ratio of the variance in the large city to the variance in the small, and the variance in the overall economy. The first and second rows correspond to a change in the production technology allowing for agents to reallocate across cities and occupations (manager and worker) optimally. The first row considers a change in technology so that both cities use the complex technology. The second row considers a change in technology so that both cities use the simple technology. The third row contains the change in the variance of log income that would result from reallocating agents across cities so that both cities display a uniform distribution of skills, maintaining the relative population size of the cities, and allowing agents to re-optimize the choice of occupation (manager and worker).

fact that this is not the case is a result that comes out of the estimation.

6   Income Inequality: Technology, and Sorting

Differences in the income distribution across cities are only partially due to differences in technology. Spatial sorting is also crucial in determining the income distribution both through the organization of production and its effect on the income schedule. In this section, I will discuss the effect of both technology and spatial sorting on overall income inequality and income inequality across cities.

To better understand these forces, I look first at the effect of equalizing technology across cities. I then look at the effect of imposing a uniform sorting that results in the distribution of skills being uniform in both cities while maintaining the relative size of the large and the small city. Table 5.2 summarizes the results from these exercises.

Eliminating differences in technology reduces the ratio of income inequality between large cities and small cities by about 15% if both cities adopt the complex technology and about 17%

if both cities adopt the small cities technology. For comparison, a 19% decrease in the ratio would equate income inequality across cities.

Equalizing the technology also has an important effect on overall inequality. Diffusing the complex technology to both cities increases overall inequality by 2.5% while diffusing the simple technology to both cities decreases inequality by about 26%. The increase in inequality from the small city adopting the large city technology is yet another argument for why it may not be beneficial to incentivize the diffusion of the complex technology everywhere. Moreover, this increase in inequality as a result of the diffusion of the complex technology is consistent with the idea in Caicedo et al. (2017) that the increasing income inequality in the US over the last several decades may be coming from increasing complexity of the technology.

A uniform spatial sorting has the opposite effect on the ratio of income inequality across cities. It increases differences in income inequality, making the large city more unequal and the small city less unequal. Moreover, a uniform spatial sorting would decrease overall income inequality. Spatial sorting affects the income distribution across cities by dampening the effects of differences in technology and bringing the income functions closer together. This happens as the higher-skilled agents move to the large city where the skill premium is higher, bringing the skill premium down and closer to the one in the small city.

The adjustment of the skill premium occurs through changes in the organization of production. To illustrate this, consider the effect for the small city of allowing for spatial sorting. Figure 6.1 plots the income schedule for the estimated skill distribution and the uniform skill distribution. The small city starts with a simple technology that is not very productive but which faces many easy problems that are cheap to communication. Management is only profitable for the very high-skilled agents and the few managers hire large teams of workers. For the lowest-skilled agents, the skill premium is very high because of easy problems and communication allow teams to get larger quickly as skill increases even by just a little bit. If we then allow agents to reallocate, the small city will attract the low skilled and in particular the middle-low skilled, while the high-skilled agents will like to move to the large city where the technology is complex but more productive. As the supply of high-skilled managers decreases, the wage for the workers

Figure 6.1: Income Schedule
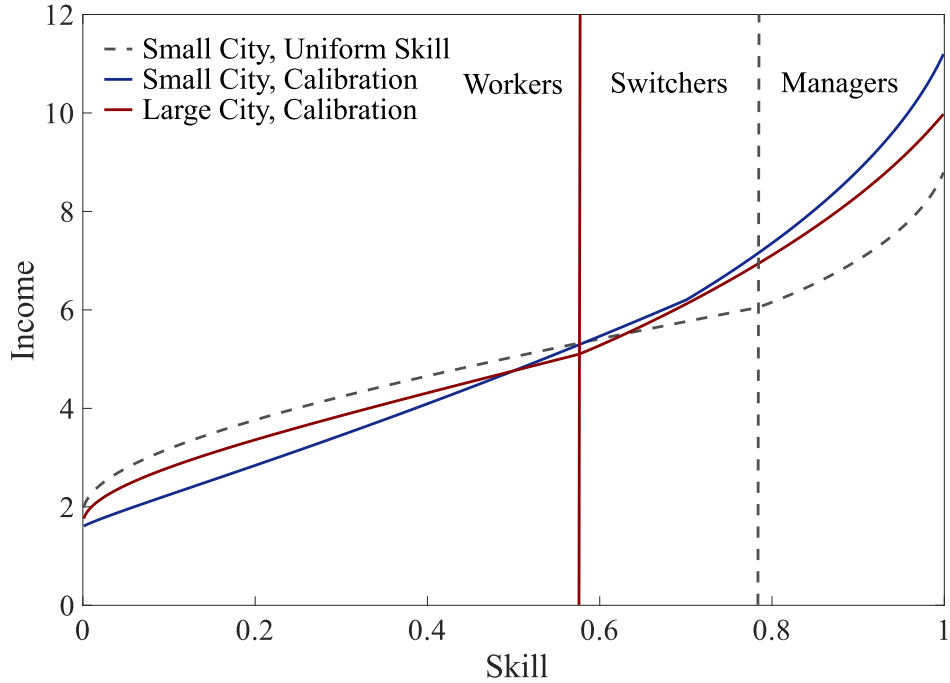
Note: This figure plots the change in the income function from a change in the sorting of agents so that both cities display a uniform distribution of skills but maintain their relative size. The solid lines correspond to the income functions in equilibrium, while the dashed line corresponds to the income function under symmetric sorting (uniform distribution in both cities). The vertical lines correspond to the threshold skill in the small city so that agents with skill below the threshold optimally choose to become workers and those with skill above the threshold optimally choose to become managers.

will decrease and the profitability of management will increase, and the better workers will switch into management. As a result of these adjustments, the income schedule will become closer to the one in the large city.

This effect has interesting consequences for policy evaluation. If we observe the income function across cities, we may conclude that differences across cities are not very large, and we may be tempted to implement policies that attract talent to the smaller cities hoping to take advantage of possible human capital externalities. However, a uniform sorting will result in larger differences in the income distribution than those in equilibrium. Not taking this effect into account will result in a miscalculation of the consequences of policies that distorts sorting.

7 Policy counterfactuals

In this section, I look at the effect of city-level policies taking into account the endogenous response of sorting and production organization. I take the differences in technology as exogenous. While it is possible that part of the difference in technology is due to agglomeration economies, to the extent that the policy exercises do not have a large effect on the size of the city, agglomeration forces should remain reasonably invariant to these policies.

7.1 Minimum Wage

In recent years, there has been a rise in the number of cities implementing minimum wage increases beyond the state or federal level in an attempt to address the growing levels of inequality. In August 2014, the U.S. Conference of Mayors "Cities of Opportunity Task Force" issued a list of strategies to fight income inequality. The strategies included efforts to increase the minimum wage. The National Employment Law Project report "City Minimum Wage Laws: Recent Trends and Economic Evidence" includes a list of all the local minimum wage ordinances passed since 2003. Table 7.1 reports the most recent minimum wage ordinances for some of the major cities in the US. In brackets, I include the year by which the minimum wage targets will be implemented. For comparison, Table 7.1 also includes the state minimum wage, which is in all cases above the federal minimum wage of $7.25 an hour.

In this section, I evaluate the impact of imposing a minimum wage that is 50% higher than the current lowest wage in the large city. I relegate the analysis of imposing a minimum wage in the small city to Section D.1 of the Appendix. The lowest income in the large city implied by the model is $16,084 per year which is equivalent to $8.36 per hour if we assume a 40-hour week. This is slightly above the federal minimum wage of $7.25 per hour but lower than many state levels. I will impose a minimum wage of $12.56 per hour. This is comparable to the increase implemented by the city of Seattle, which increased its minimum wage from $9.47 per hour to $13 per hour between 2015 and 2016. In a recent paper, Jardim et al. (2017) study the effect of

Table 7.1: City Minimum Wage Ordinances

	Year Passed	Minimum Wage Target	State Minimum Wage
Washington, DC	2013	$15 (by 2020)	$12.50
Chicago, IL	2014	$13 (by 2018)	$8.25
San Francisco, CA	2014	$15 (by 2018)	$10.50
Seattle, WA	2014	$15 (by 2018-21)	$11
Los Angeles, CA	2015	$15 (by 2020)	$10.50

Sources: National Employment Law Project, City Minimum Wage Laws, 2016; National Conference on State Legislatures, 2017

the minimum wage increase in Seattle. I will use their estimated effects on hours and wages as a benchmark. Another study, by the University of California at Berkeley's Institute for Research on Labor and Employment, found no employment effects for the restaurant industry.

In order to implement the minimum wage, I impose that managers are not able to pay wages per unit of time below the minimum wage, and I relax the labor market clearing condition. The new equilibrium is now characterized by two threshold skills: a lower bound threshold $\underline{z}$ such that agents below that skill are unemployed and the management threshold z^* such that agents above that threshold become managers. Agents who are not hired by any managers have access to a self-employment production technology so their earnings will be positive and increasing in skill. I assume that agents know about the minimum wage regulation and its impact on the income schedule, so they choose the optimal city taking this into account. As a result, many of the low-skill agents will prefer to live in the small city, but those with a high enough idiosyncratic preference will remain in the large city and use the self-employment technology.

I start by considering the effect of this minimum wage on the decrease in employment and the effect on average wages and compare it to the magnitudes found in the Jardim et al. (2017) study on Seattle. The authors focus on the effect on low-skill workers, which they identify as workers earning below $19 an hour. They find that for these workers employment fell by 9% and

Table 7.2: Welfare Effect of a Minimum Wage

	Large City	Small City	Overall
	All Residents		
Average Utility	0.21%	-15.91%	-1.45%
Var Log Utility	25.62%	102.31%	44.02%
	Original Residents		
Average Utility	-1.43%	-1.84%	-1.45%
Variance Log Utility	48.09%	22.91%	44.02%
	Employed Agents		
Average Utility	5.2%	-15.91%	2.9%
Variance Log Utility	-25.45%	102.31%	7.08%

Note: This table contains the effect on average utility and variance of log utility that result from the large city implementing a minimum wage 50% higher than the current lowest wage of the large city. The columns respectively contain the effect for the large city, the small city, and the overall economy. The first two rows contain the effect from comparing the utility of agents who were living in the corresponding city in each column before and after the minimum wage. The next two rows contain the effect on the utility of agents who were living in the city corresponding to the column before the minimum wage policy. Notice that this consideration does not affect the calculation for the overall economy. The final two rows compare the effect on utility for agents who were living in the city corresponding to the column before and after, and who are either a manager or a worker hired by a manager. Notice that this consideration does not change the calculation for the small city, so that the effect on the first two rows and the last two rows for the small city are unchanged.

wages increased by 3% resulting in a total decrease in payroll of \$125 per month. I will follow the same specification as in the Jardim et al. (2017) study and consider the effect on workers making below \$36,480 annually, which is equivalent to \$19 an hour for a 40-hour week and full year. I do not take into account the self-employed in this calculation. I find that the average wage increases by 15% and employment decreases by 33% so that an average worker loses \$506 per month assuming that the self-employed earn zero dollars. The average cost is within the same order of magnitude. However, recall that the change in the minimum wage studied here is larger than the one experienced in Seattle. These numbers are simply for comparison, and they do not pretend to be a quantification of the effect in Seattle.

Next, I quantify the impact of this policy on welfare, measured by average utility, and on inequality measured by the variance of log utility. In order to better understand the consequences of implementing an increase in the minimum wage, I consider the effect of the policy for the original residents, comparing the new residents to the old residents, and considering only employed agents. Table 7.2 summarizes the results from the three scenarios for the large city, the small city, and the overall effect.

I find that a minimum wage does slightly increase average utility by 0.2% in the large city, where it is implemented. This is, however, due to a composition effect as low-skilled agents who cannot get hired in the large city move out. Indeed the large city loses 2.34% of its population as a result of the policy and low skill agents move out at larger rates. Restricting attention to the original residents of the large city, I find that the minimum wage decreases average utility by 1.43%.

The effect on the small city is large and negative. If we compare residents before and after, they experience a 16% drop in average utility. In part, this effect is due to the inflow of low-skill agents. The original residents experience a smaller decrease of almost 2% in average utility and an increase in the variance of 23%. The minimum wage hurts the small city both in terms of average utility but also by increasing inequality.

Overall, there is a small decrease in the average utility of 1.45% but a large increase in the variance of 44%. However, if we only consider employed agents, I find an increase in average

utility of 3% and an increase in the variance of log utility of 7%.

The negative effect on overall utility comes mostly from low-skill agents losing their job. Looking only at those who remain employed, I do find a positive impact with a 5% higher average utility and 25% lower inequality. It is important to keep in mind that the equilibrium in this model is efficient and has none of the usual mechanisms that will generate positive effects from an increase in the minimum wage, such as monopsony power (Bhaskar and To (1999)), job search (Flinn (2006)), endogenous productivity(Rebitzer and Taylor (1995)), or subsistence income (Dessing (2002)). Therefore, it is not surprising that the minimum wage has an overall negative impact. The richness of the model in the matching of heterogeneous workers is particularly well suited to understanding the effect of the policy on inequality and identifying the winners and losers of the policy throughout the income distribution.

Implementing a minimum wage has differential effects throughout the income distribution due to changes in the organization of production. Figure 7.1 plots the percentage change in utility by skill in the large city and small city. Let us first consider the effect in the large city. The lowest-skilled agents who were earning below the minimum wage cannot find a manager to pay the high wages and are forced to produce on their own or move to the small city. These are the main losers from the policy. As a result of the lowest-skilled workers moving out of the labor market, two things happen: first, the demand for the remaining workers goes up, driving wages up for most workers, and second, the lowest-skilled managers switch into becoming workers. The higher wages imply increases in utility for workers and decreases in utility for most managers. Finally, since the threshold for being a manager increases, the very top managers who hire the best workers can now hire better workers than before, so they also experience an increase in utility.

Next, let us consider the effect of this policy on the small city. The small city now receives a large mass of low-skill workers who moved out of the large city because they could not find a manager to hire them. The inflow of low-skill workers drives wages down in order for the labor market to clear. This decrease in wages translates into lower utility for most workers. As a result of the lower wages, the highest-skilled workers are now able to hire workers and become

Figure 7.1: Change in the Utility of the Original Residents
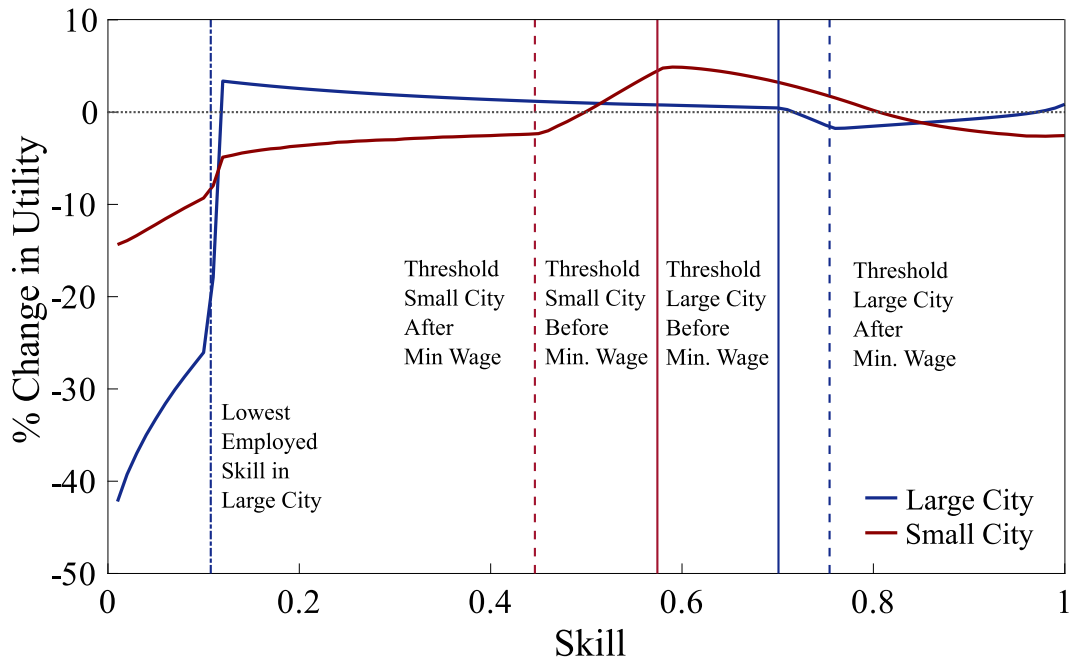
Note: This figure contains the effect on utility that result from the large city implementing a minimum wage 50% higher than the current lowest wage of the large city. There is a distribution of utilities within a skill type that reflect the distribution in idiosyncratic amenities and working-time supply. The figure plots the percentage change in average utility for skill bins, where each bin corresponds to 1% of the population in the economy. That is, the first bin includes skills from 0 to 0.01, the second bin includes skills from 0.01 to 0.02, etc. The vertical line on the left corresponds to the lowest skilled that is hired by a manager in the large city. Agents with skills below this threshold work on their own using the production technology without a manager. The following vertical lines correspond to the threshold skill in the large and small city before and after the minimum wage, such that agents below the threshold skill optimally choose to become workers and those with skill above the threshold optimally choose to become managers.

managers. These top workers that switch into management, along with most managers that benefit from the decreasing wages, are the main winners of the policy. Finally, due to the lower threshold to become a manager, the top manager who used to hire the top workers now has to hire worse workers than before, so these top managers will suffer losses in utility.

To sum up, the minimum wage decreases overall utility mostly from forcing the lowest-skilled agents into unemployment. On top of that, it has a negative impact on the managers that used to hire those low-skilled workers. However, it increases the average utility within employed agents, in part due to a composition effect. Remarkably, the minimum wage policy fails to decrease overall inequality even within people who do not lose their jobs due to the large increase in inequality in the small city that outweighs the decrease in inequality within employed workers in the large city. The supply of low-skilled workers in the small city depresses wages, thus hurting workers and benefiting middle managers. Interestingly, top managers also lose in the small city due to the change in the organization of production that moves the workers they used to hire into management and forces them to hire worse workers than before.

Table 7.3: Subsidized Housing Summary Statistics

	US	New York City
Annual rent	$3, 984	5,496
Household Income	$13,726	$19,306
% Household income spent on rent	29%	28%
% Below 50% of Median	94%	90%
% Below 30% of Median	73%	74%

Source: HUD, A Picture of Subsidized Housing 2016

7.2 Affordable Housing

The US government devotes about $40 billion each year to means-tested housing programs Collison et al. (2016). Access to affordable housing is a pressing concern in some of the larger cities where housing costs keep rising. In New York City, more than 735,000 people currently reside in public housing or receive subsidized assistance through the New York City Housing Authority.[28] For context, Table 7.3 presents some statistics on the characteristics of individuals receiving subsidized housing in the US and in New York City as New York City is not only the largest city in the US but also has the largest housing authority.

Table 7.4: Welfare Effect of Housing Subsidy

	Large City	Small City	Overall
	Current Residents		
Average Utility	-0.72%	2.08%	-0.59%
Var Log Utility	-7.24%	-12.5%	-8.09%
	Original Residents		
Average Utility	-0.61%	-0.18%	-0.59%
Variance Log Utility	-8.55%	-6.24%	-8.09%

Note: This table contains the effect on average utility and variance of log utility that result from the large city implementing a 50% housing subsidy for agents earning below half of the median income in the large city. The columns respectively contain the effect for the large city, the small city, and the overall economy. The first two rows contain the effect from comparing the utility of agents who were living in the corresponding city in each column before and after the minimum wage. The next two rows contain the effect on the utility of agents who were living in the city corresponding to the column before the minimum wage policy. Notice that this consideration does not affect the calculation for the overall economy.

In this section, I quantify the impact of a housing subsidy that resembles Section 8 Voucher programs. Typically, housing vouchers are granted to individuals earning below half of the

---

[28]Statistics from the New York City Housing Authority Website: www1.nyc.gov/site/nycha/about

median income. The voucher covers housing costs that exceed 30% of the recipients' income, that is, recipients are guaranteed to spend no more than 30% of their income on housing. In a comprehensive review of recent literature on housing policy in the US, Olsen (2003) includes details on Section 8 and other low-income housing programs. Given the assumption of Cobb-Douglas preferences for housing and consumption goods, the share of housing expenditure is constant and given by the Cobb-Douglas share on housing γ. This share γ is calibrated to match average overall expenditure of 24% following the empirical finding in Davis and Ortalo-Magné (2011). As a result, everyone spends 24% of their income on housing. In reality, the expenditure share varies with income, Eeckhout et al. (2014) find the expenditure share of housing varies from 35% for low-income agents to 22% for high-income agents.

In order to quantify the effect of the subsidy, I need to know the reduction in housing costs that result from receiving a voucher. This number is not observed in the data. There are also no current estimates for this number in the literature. One of the most exhaustive studies is Olsen (2003), in which he compares the effect from eight previous studies on different programs using data from the 1970s and 1980s. The estimates for the percentage increase in housing consumption range widely from 22% to 82% for public housing programs, from 26% to 58% for subsidized projects, and from 16% to 63% for voucher programs. Given the wide range of estimates, I choose a roughly reasonable subsidy of 50% for housing costs.

The affordable housing implemented will consist of a 50% subsidy to the cost of housing for people earning below half of the median income in the large city. In order to make the policy revenue-neutral, a lump sum tax is collected equally from everyone in the city. Table 7.4 summarizes the effect of this subsidy. The subsidy impacts both the large city, where the policy is implemented, and the small city, through sorting decisions of agents. The average utility of the original residents of the large city goes slightly down by 0.61%; however, it does have a positive effect on inequality by decreasing the variance of log utility by 8.55%. The effects for the original residents in the small city are similar but smaller, with a decrease in the average utility of 0.18% and a decrease of the variance of log utility of 6.24%. Overall inequality measured by the variance of log utility goes down by 8%, while the average utility also decreases slightly, by

0.59%. It is also worth noting that the average utility in the small city comparing old versus new residents will go up by 2% due to a composition effect caused by low-skill agents moving into the large city attracted by the housing subsidy.

There are naturally winners and losers from this policy. In Figure 7.2, I plot the percentage change in utility for each skill level of the original residents in both the large and the small city. As expected, agents who earned below half of the median income in the large city are the big winners of this policy; in the large city, utility at the lowest skills increases around 8%. More interesting are the spillover effects that also drive utility up in the small city for the bottom skills by around 3%. This increase in the small city is the result of a stronger incentive for low-skill individuals to move to the large city. This incentive decreases the number of low-skill workers in the small city, pushing wages up. Interestingly, in the small city, the losers of the policy are middle-skilled agents who were the lowest-skilled managers hiring the lowest skilled workers. Now, these managers have to pay their workers higher wages due to the scarcity of low-skill workers who are now more likely to move to the large city.

The two policies considered here, minimum wage and housing subsidy, have opposite effects on the small city. While the minimum wage increases the supply of low-skill workers in the small city, hurting the bottom skill residents of the small city, affordable housing has the opposite effect. Affordable housing in the large city decreases the supply of low-skill workers in the small city and drives their income up. As a result, inequality in the small city increases with the minimum wage and decreases with affordable housing.

Summing up, both the minimum wage and the housing subsidy policies have small negative overall effects on the average utility of 1.45% and 0.59% respectively. The negative effect is not surprising since there are no inefficiencies in the decentralized equilibrium. However, the minimum wage increases the variance of log utility by 44%, while the housing subsidy decreases the variance of log utility by 8%. This leads us to conclude that housing subsidies are more effective at reducing inequality than minimum wage policies.

Figure 7.2: Change in the Utility of the Original Residents
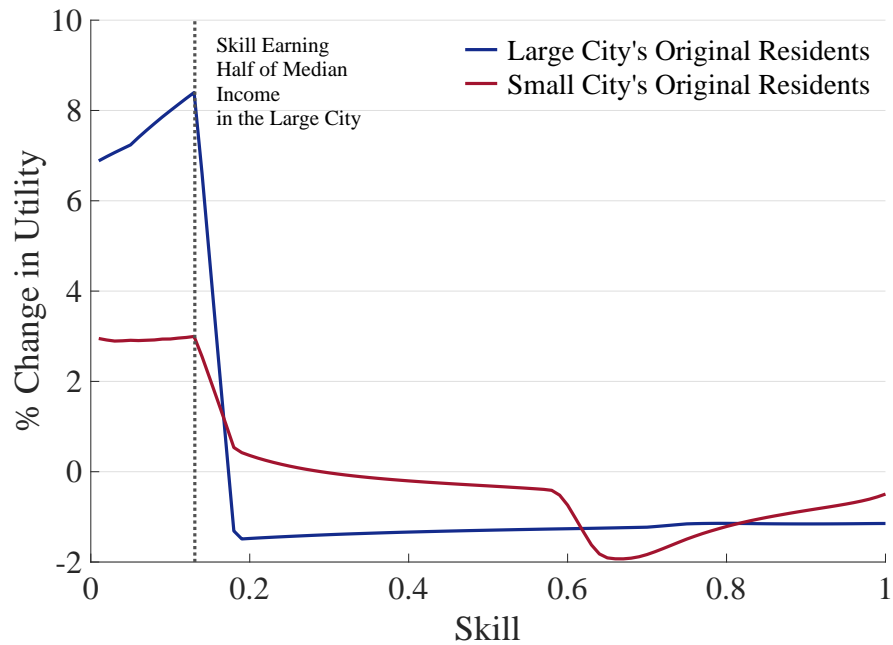
Note: This figure contains the effect on utility that result from the large city implementing a 50% housing subsidy for agents earning below half of the median income in the large city. There is a distribution of utilities within a skill type that reflects the distribution in idiosyncratic amenities and working-time supply. The figure plots the percentage change in average utility for skill bins, where each bin corresponds to 1% of the population in the economy. That is, the first bin includes skills from 0 to 0.01, the second bin includes skills from 0.01 to 0.02, etc. The vertical line on the left corresponds to the skill of the agents that earn half of the median income in the large city. Agents with skill below this level will receive a housing subsidy in the large city.

8 Conclusion

In recent years there has been rising concern about extreme income inequality and its consequences. This concern is leading to a series of policies implemented at the city level designed to address inequality. The most common policies involve affordable housing and increases in the minimum wage. Understanding the way in which spatial sorting interacts with technological differences across cities is key in order to evaluate the effect of these city policies as these policies typically distort the incentives to sort across cities.

The spatial sorting with heterogeneous agents literature has focused on explaining differences across ex-ante identical locations. While this question is extremely interesting and helps us understand the emergence of heterogeneous cities, it is not the best set-up to evaluate the effect of policies in the short and medium run. Moreover, by allowing for exogenous differences in technology, I am able to match the empirical differences across cities even while allowing for sorting to have an equalizing effect that follows from the basic economic intuition of supply and demand.

This paper contributes to the literature by measuring these differences in technology across cities and separating the effect of technology and sorting. I find that technology is the main driver of inequality differences across cities, while sorting works to dampen those differences. Interestingly, I find that these technology differences may be optimal since the small city uses a simpler technology that is not Pareto-dominated by the large city technology. Recovering the structural parameters of production, I am able to evaluate the effect of policies with endogenous reallocation of individuals across space as well as the endogenous reorganization of production. I find that a subsidy to housing is more effective at reducing inequality than a minimum wage. Understanding the initial source of technological differences across cities is left for future work.

This paper treats technology as exogenous and measures two technologies, one for the large city and one for the small city. In reality, there are many different technologies of varying complexity being used in both cities. This paper measures some composite of those underlying technologies. It is outside the scope of this paper to endogenize the technology differences.

However, the fact that the skill distribution in the small city is better suited to use the simple technology and the skill distribution in the large city is better suited to use the complex technology suggests a way to endogenize technology differences in future work. If new technologies are being invented and they could choose the optimal city, the complex technologies would select into the skill-abundant city. Furthermore, if large cities attract higher-skilled agents disproportionately, we would also expect the complex technologies to select into the large city.

References

Abdel-Rahman, Hesham M., and Ping Wang. 1997. "Social Welfare and Income Inequality in a System of Cities." *Journal of Urban Economics*, 41(3): 462–483.

Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg. 2006. "Offshoring in a Knowledge Economy." *Quarterly Journal of Economics*, 121(1): 31–77.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review*, 103(5): 1553–1597.

Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan. 2017. "Why Has Urban Inequality Increased?" *American Economic Journal: Applied Economics (Forthcoming)*.

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

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive cities: Sorting, selection, and agglomeration." *Journal of Political Economy*, 122(3): 507–553.

Berkes, Enrico, and Ruben Gaetani. 2017. "Income Segregation and the Rise of the Knowledge Economy." *Unpublished manuscript.*

Bhaskar, Venkataraman, and Ted To. 1999. "Minimum Wages for Ronald McDonald Monopsonies: A Theory of Monopsonistic Competition." *The Economic Journal*, 109(455): 190–203.

Bound, John, and George Johnson. 1992. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *American Economic Review*, 82(3): 371–92.

Caicedo, Santiago, Robert E. Lucas Jr., and Esteban Rossi-Hansberg. 2017. "Learning, Career Paths, and the Distribution of Wages." *Unpublished manuscript.*

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The Anatomy of French Production Hierarchies." *Journal of Political Economy*, 123(4): 809–852.

Calvo, Guillermo A., and Stanislaw Wellisz. 1978. "Supervision, Loss of Control, and the Optimum Size of the Firm." *Journal of Political Economy*, 86(5): 943–952.

Card, David, and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics*, 20(4): 733–783.

Chetty, Rak, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Income Mobility in the United States." *The Quarterly Journal of Economics*, 129(4): 1553–1623.

Collison, Robert, Ingrid Gould Ellen, and Jens Ludwig. 2016. "Low-Income Housing Policy." *Economics of Means-Tested Transfer Programs in the United States*, 2.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial Wage Disparities: Sorting Matters!." *Journal of Urban Economics*, 63(2): 723–742.

Davis, Donald, and Jonathan Dingel. 2016. "A Spatial Knowledge Economy." *Unpublished manuscript.*

Davis, Donald, and Jonathan Dingel. 2017. "The Comparative Advantage of Cities." *Unpublished manuscript.*

Davis, Morris A., and François Ortalo-Magné. 2011. "Household Expenditures, Wages, Rents." *Review of Economic Dynamics*, 14(2): 248–261.

Dessing, Maryke. 2002. "Labor Supply, the Family and Poverty: the S-shaped Labor Supply Curve." *Journal of Economic Behavior and Organization*, 49(4): 433–458.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *The American Economic Review.*

Duranton, Gilles, and Diego Puga. 2005. "From Sectoral to Functional Urban Specialisation." *Journal of Urban Economics*, 57 343–370.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014. "Spatial Sorting." *Journal of Political Economy*, 122(3): 554–620.

Flinn, Christopher. 2006. "Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates." *Econometrica*, 74(4): 1013–1062.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, 108(5): 874–904.

Garicano, Luis, and Thomas Hubbard. 2009. "Earnings Inequality and Coordination Costs: Evidence From U.S. Law Firms." *NBER Working Papers 1471.*

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *The Quarterly Journal of Economics*, 121(4): 1383–1435.

Garicano, Luis, and Esteban Rossi-Hansberg. 2015. "Knowledge-Based Hierarchies: Using Organizations to Understand the Economy." *Annual Review of Economics*, 123(4): 809–852.

Giannone, Elisa. 2018. "Skill-Biased Technical Change and Regional Convergence." *Unpublished manuscript.*

Glaeser, Edward L., William R. Kerr, and Giacomo A.M. Ponzetto. 2010. "Clusters of entrepreneurship." *Journal of Urban Economics*, 67(1): 150–168.

Glaeser, Edward L., Matt Resseger, and Kristina Tobio. 2009. "Inequality in Cities." *Journal of Regional Science*, 49(4): 617–646.

Holmes, Thomas J., and John J. Stevens. 2010. "An Alternative Theory of the Plant Size Distribution with an Application to Trade." NBER Working Papers 15957, National Bureau of Economic Research, Inc.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma van Inwegen andJacob Vigdor, and Hilary Wething. 2017. "Minimum Wage Increases, Wages, and Low-Wage Employment: Evidence from Seattle." *NBER Working Paper*(23532): .

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, 107(1): 35–79.

Kline, Patrick, and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics*, 6(1): 629–662.

Levy, Frank, and Richard J. Murnane. 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trades and Proposed Explanations." *Journal of Economic Literature*, 30(3): 1333–81.

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics*, 9(2): 508–523.

McFadden, Daniel. 1977. "Modelling the Choice of Residential Location." *Cowles Foundation Discussion Papers*(477): .

Mincer, Jacob. 1975. *Education, Experience, and the Distribution of Earnings and Employment: An Overview.* NBER Chapters: National Bureau of Economic Research, Inc, 71-94.

Moretti, Enrico. 2012. *The New Geography of Jobs.*: Houghton Mifflin Harcourt.

Olsen, Edgar O. 2003. *Means-Tested Transfer Programs in the United States.* Chap. Housing Programs for Low-Income Households(6): : The University of Chicago Press.

Radner, Roy. 1993. "The organization of decentralized information processing." *Econometrica*, 61(5): 1109–1146.

Rebitzer, James, and Lowell J. Taylor. 1995. "The Consequences of Minimum Wage Laws Some New Theoretical Ideas." *Journal of Public Economics*, 56(2): 245–255.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy*, 90(6): 1257–1278.

Rosen, Sherwin. 1979. "Wage-based Indexes of Urban Quality of Life." *Current Issues in Urban Economics.*

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

Samuelson, Paul A. 1975. "The Optimum Growth Rate for Population." *International Economic Review*, 16(3): 531–538.

Spanos, Grigorios. 2018. "Firm Organization and Productivity Across Locations." *Unpublished manuscript.*

Tian, Lin. 2018. "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil." *Unpublished manuscript.*

A Data Appendix

In this section, I expand on some details of the American Community Survey data, and present the results from the hedonic price regressions. I also expend on the main stylized facts on the distribution of income and perform some additional robustness checks.

A.1 American Community Survey: Top-coded income levels

Income reported in the American Community Survey is top-coded at the 95.5 percentile level for each state in order to preserve confidentiality. The top-coded incomes are replaced by the average of the top-coded incomes within the State. Table A.1 reports the level of top-coding for incomes in each state.

A.2 Hedonic Price Regression

In this subsection, I present the results from hedonic price regressions where rent paid is regressed on a constant, the number of rooms, the number of units in a building, the year the building was built in and an indicator for geography. I use two types of indicators, the CSA in order to report stylized facts on the distribution of real income across cities, and an indicator for "large city" and "small city" in order to estimate the model in a two city setting. The representative large city includes all CSAs with a population larger than 2.5 million inhabitants, while the small city includes all CSAs with a population below 2.5 million.

The regression equations for each geographic indicator are:

$$lrent_i = \beta_0 + \beta_1 NumRooms_i + \beta_2 NumUnits_i + \beta_3 YearBuilt_i + \beta_4 CSA_i + \epsilon_i, \qquad \text{(A.1)}$$

$$lrent_i = \beta_0 + \beta_1 NumRooms_i + \beta_2 NumUnits_i + \beta_3 YearBuilt_i + \beta_4 SizeCategory + \epsilon_i. \qquad \text{(A.2)}$$

Table A.1: Income Top Codes, ACS 5% PUMS 2010–2014

State	Income Top code	State	Income Top Code
Alabama	$311,000	Montana	$267,000
Alaska	$366,000	Nebraska	$320,000
Arizona	$309,000	Nevada	$315,000
Arkansas	$325,000	New Hampshire	$431,000
California	$455,000	New Jersey	$539,000
Colorado	$419,000	New Mexico	$281,000
Connecticut	$642,000	New York	$587,000
Delaware	$343,000	North Carolina	$379,000
District of Columbia	$582,000	North Dakota	$450,000
Florida	$393,000	Ohio	$360,000
Georgia	$392,000	Oklahoma	$327,000
Hawaii	$325,000	Oregon	$344,000
Idaho	$313,000	Pennsylvania	$410,000
Illinois	$456,000	Rhode Island	$384,000
Indiana	$340,000	South Carolina	$323,000
Iowa	$359,000	South Dakota	$338,000
Kansas	$392,000	Tennessee	$385,000
Kentucky	$305,000	Texas	$410,000
Louisiana	$314,000	Utah	$341,000
Maine	$330,000	Vermont	$330,000
Maryland	$467,000	Virginia	$426,000
Massachusetts	$538,000	Washington	$413,000
Michigan	$331,000	West Virginia	$237,000
Minnesota	$459,000	Wisconsin	$366,000
Mississippi	$291,000	Wyoming	$354,000
Missouri	$344,000		

Table A.2 reports the results from these regressions. For conciseness, I report only the coefficient on the number of rooms and number of units in a building. The coefficients on the year built and the CSA are as expected. Older buildings are associated with lower rent. For example, with respect to 2005, buildings built before 1939 are associated with 0.29 log points lower rent and buildings built in 2013 are associated with 0.19 log points higher rent. The intuition on the coefficients on CSAs is more subtle but it does follow common knowledge on expensive and inexpensive cities. For example, relative to Albany, living in San Francisco is related to 0.55 log points higher rent, while living in Pittsburgh is related to 0.25 log points lower rent.

A.3   Distribution of Income and City Size

In this subsection, I include additional details and robustness to the main stylized facts presented in the main text. First, I begin by presenting the two main stylized facts at the Metropolitan Statistical Area level in order to compare them to the Combined Statistical Areas level used for the main analysis. Panel (a) in Figure A.1 plots the average annual wage against the log population for all MSAs in the United States and Panel (b) plots the variance of log wage on log population. Larger cities tend to have a higher average income and a higher variance of log income. Notice that the outliers are often part of a larger Combined Statistical Area. Bridgeport and Trenton belong to New York's CSA, Boulder belongs to Denver's CSA, and San Jose belongs to San Francisco's CSA.

Larger cities are also typically more expensive places to live. The higher nominal income, although an indication of a higher productivity, may not reflect a higher utility from living in larger cities. In order to explore this further, Figure A.2 plots the average real income against log population. The real income is calculated by dividing income by the price index implied by a model with Cobb-Douglas preferences on consumption and housing where consumption is the numeraire. The price index is therefore given by the price of housing to the power of the Cobb-Douglas weight on housing expenditure. The housing prices are the result of the hedonic regressions presented in Section A.2 of the Appendix. Larger cities not only have a higher average

Table A.2: Hedonic Price Regression

	Log Rent (CSA)	Log Rent (2 Cities)
0 Rooms	-0.51**	-0.55**
1 Room	-0.46**	-0.53**
2 Rooms	-0.25**	-0.33**
3 Rooms	-0.14**	-0.19**
4 Rooms	Omitted	Omitted
5 Rooms	0.06**	0.08**
6 Rooms	0.09**	0.06**
7 Rooms	0.05*	0.01
8 Rooms	-0.11**	-0.12**
9 or More Rooms	-0.2**	-0.19**
1-family house, detached	0.05**	-0.12**
1-family house, attached	Omitted	Omitted
2-family house, attached	-0.07**	-0.16**
3-4 family building	-0.06**	-0.12**
5-9 family building	-0.04**	-0.1**
10-19 family building	0.01**	-0.02**
20-49 family building	0.003	0.03**
50+ family building	0.02**	0.09**
Small City Dummy	-	-0.2**
Constant	7.05**	7.21**
R^2	0.24	0.08
N	1,221,418	1,221,418

*$p<0.05$, **$p<0.01$*

Note: This table uses data from the ACS IPUMS 5% 2010–2014 Housing Survey. It includes households that pay rent on their apartments and does not include those living in group quarters, mobile homes, boats, tents, or farm houses. Some regressors are omitted for space considerations.

Figure A.1: Distribution of Income at the MSA level
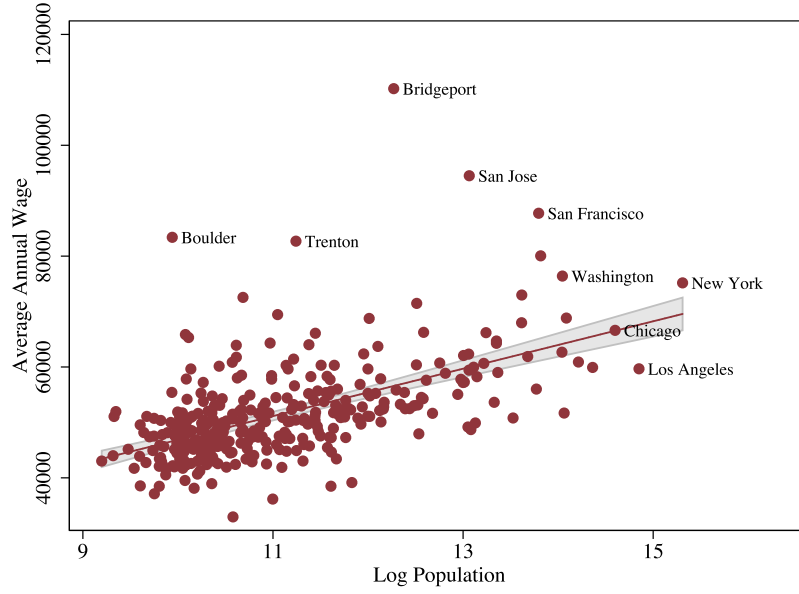
(a) Average Income
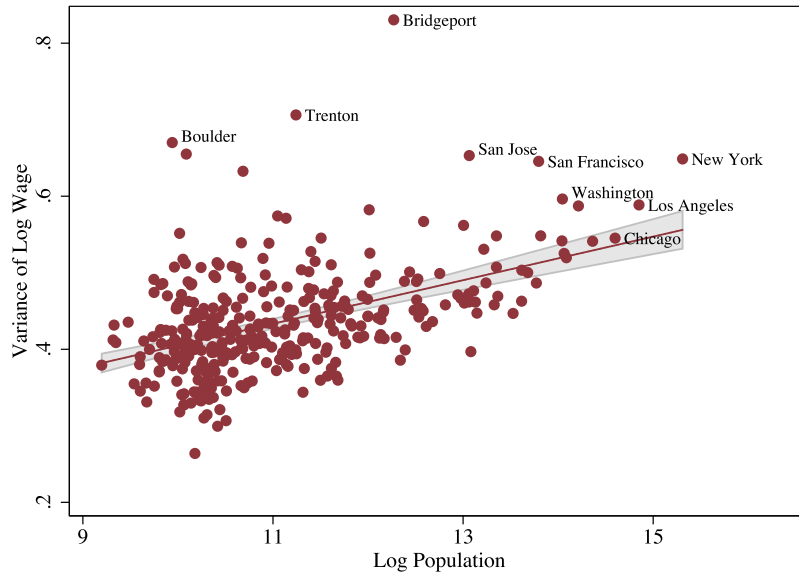
(b) Variance of Log Income

Note: This figure plots the average annual wage and the variance of log annual wage against logged population for Metropolitan Statistical Areas. The data source used is the American Community Survey 5% IPUMS for 2010–2014. The line corresponds to a linear regression and the shaded area corresponds to the 95th confidence interval.
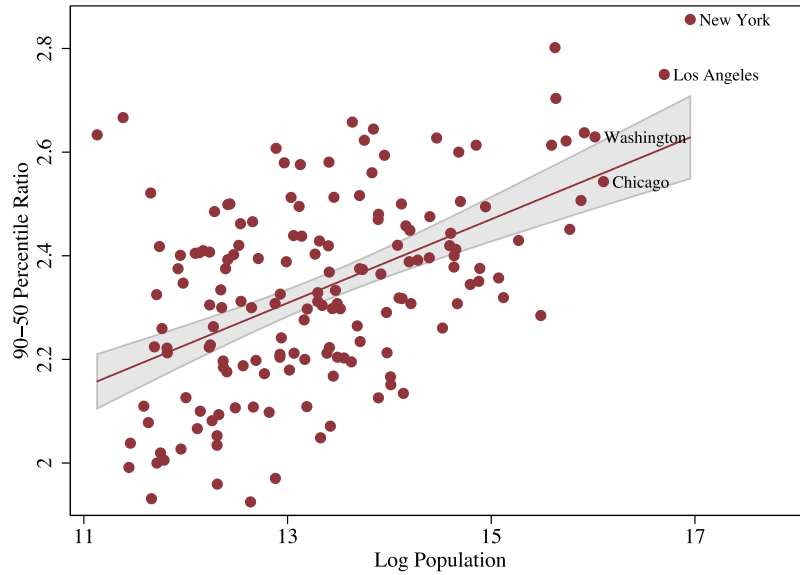
Figure A.2: Model-based adjustment for housing prices
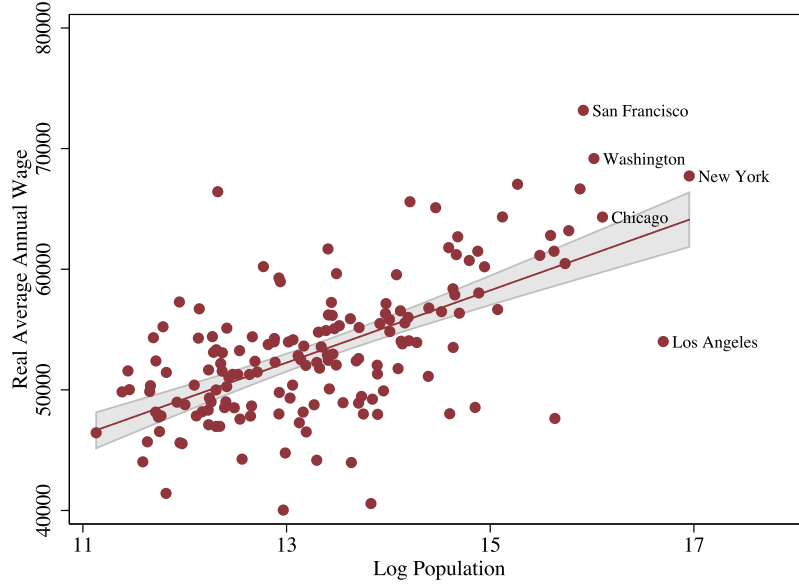
Note: This figure plots the average real annual wage against logged population for Combined Statistical Areas. The data source used is the American Community Survey 5% IPUMS for 2010–2014. The real wage is computed by dividing wages by the price index implied by the model. Given Cobb-Douglas utility and consumption as the numeraire, the price index is simply the housing price raised to the Cobb-Douglas housing share. The housing prices are the result of hedonic regressions reported in Section A.2 of the Appendix. The line corresponds to a linear regression and the shaded area corresponds to the 95th confidence interval.

nominal income but also a higher average real income. Notice that dividing income by the price index does not affect the variance of log income and so the regression will be unchanged.

The variance of log income is one possible measure of inequality. In order to further explore whether income inequality is higher for difference percentiles, Figure A.3 plots the ratio of the 90th to 50th percentiles in Panel (a) and the ratio of the 50th to 10th percentiles in Panel (b). I find that both ratios are increasing in city size. Larger cities are more unequal both at the top and at the bottom of the income distribution.

Part of the differences in the income distribution across cities may come from differences in the composition of who lives in larger cities versus smaller cities. If more productive agents sort into the larger cities then these will have higher average incomes. In order to explore to what extent the composition in observable characteristics can explain the differences in income distribution across cities, I first run the following regressions of income and log income on years

Figure A.3: Percentile Ratios

(a) 90th to 50th Percentiles Ratio
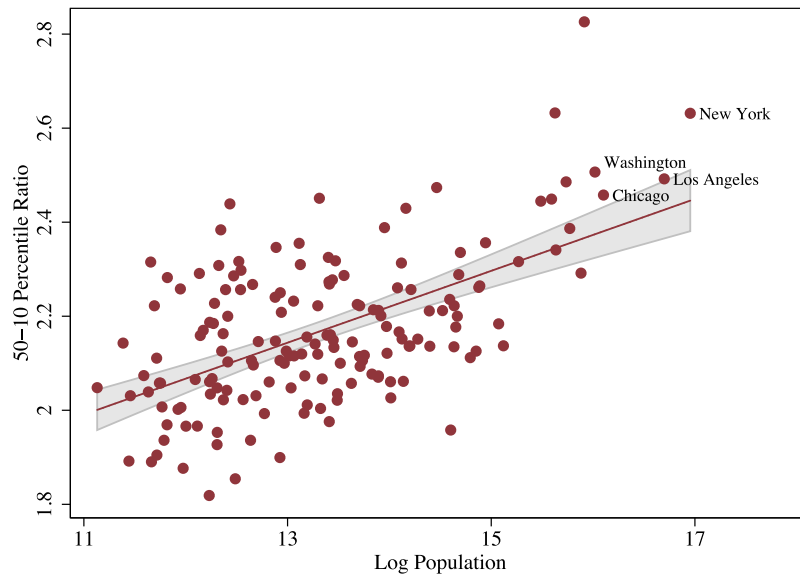
(b) 90th to 10th Percentiles Ratio

Note: This figure plots the ratio of the 90th to 50th percentiles and the ratio of the 50th to 10th percentiles against logged population for Combined Statistical Areas. The data source used is the American Community Survey 5% IPUMS for 2010–2014. The line corresponds to a linear regression and the shaded area corresponds to the 95th confidence interval.

Table A.3: Mincer Regressions

	Income	Log Income
Years of Schooling	9,262**	0.13**
Years of Potential Experience	926**	0.01**
Female Dummy	-22,317**	-0.3**
White Race Dummy	7,245**	0.12**
Constant	-82,369**	8.67**
R^2	0.19	0.29
N	2,211,219	2,211,219

*p<0.05, **p<0.01*

Note: This table contains the results of regressing income and log income on years of schooling, years of potential experience, a female dummy, and a white race dummy using data from the ACS 5% IPUMS for 2010–2014. The formulation for this regression follows Mincer (1975). Years of schooling are assigned as follows: one year of schooling per grade for grades 1 to 12; 12 years of schooling for grade 12 with no diploma, high school diploma, GED, or alternative credential; 14 years of schooling for some college with no degree and an Associate's degree; 16 years of schooling for a Bachelor's degree; 17 years of schooling for a Professional degree, and 18 years of schooling for Master's and Doctorate degrees. Years of potential experience are assigned by subtracting years of schooling plus six from age.

of schooling, potential experience, gender, and race, following Mincer (1975):

$$Income_i = \beta_0 + \beta_1 Schooling_i + \beta_2 Experience_i + \beta_3 Female_i + \beta_4 White_i + \epsilon_i \tag{A.3}$$

$$LogIncome_i = \beta_0 + \beta_1 Schooling_i + \beta_2 Experience_i + \beta_3 Female_i + \beta_4 White_i + \epsilon_i. \tag{A.4}$$

Results from these regression are included in Table A.3.

I then use the residuals from regression A.3 to calculate the average residual income by city and the residuals from regression A.4 to calculate the variance of log residual income by city. Figure A.4 plots the average of the residual income and the variance of the log residual income against population size. Observable characteristics can only explain part of the pattern but both the average residual income and the variance of the log residual income are still increasing in

Table A.4: Average Income and City Size

	Avg. Income	Avg. Real Income	Avg. Residual Income
Log Population	4,472**	2,997**	3,764**
R^2	0.52	0.38	0.57
N	162	162	162

* p<0.05; ** p<0.01

Note: this table contains the results from regressing average income, average real income and average residual income on logged population using data from the American Community Survey 5% IPUMS 2010-14. Each observation is a Combined Statistical Area. he real wage is computed by dividing wages by the price index implied by the model. Given Cobb-Douglas utility and consumption as the numeraire, the price index is simply the housing price raised to the Cobb-Douglas housing share. The housing prices are the result of hedonic regressions reported in Section A.2 of the Appendix. The residual income is the residual from regressing income on years of schooling, potential experience, gender and race. The line represents the linear fit and the shaded area is the 95 percent confidence interval.

city size.

As a summary, Table A.4 includes the result from regressing average income, average real income, and variance or residual income against log population. The three coefficients are significantly positive and the R^2 from the regressions are high. Interestingly, the R^2 is highest for the residual income meaning that city size has more explanatory power for the residual income.

A.4   Span of Control

A.4.1   Definition of Managers: PCS

There are six major groups in the PCS: (1) agricultural workers, (2) self-employed and business owners, (3) upper managers and high professionals, (4) intermediate professions: education, health, administrator, sellers, technicians, (5) white-collar workers, and (6) blue-collar workers. I exclude agricultural and self-employed workers and classify business owners, upper managers, and high professionals (classes 2 and 3) as "managers" and intermediate professions, white and

Figure A.4: Distribution of Residual Income
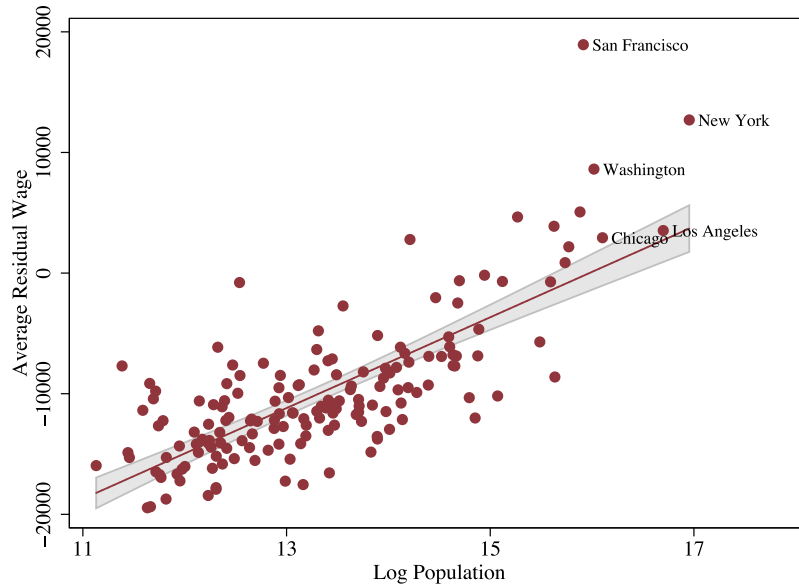
(a) Average Residual Income
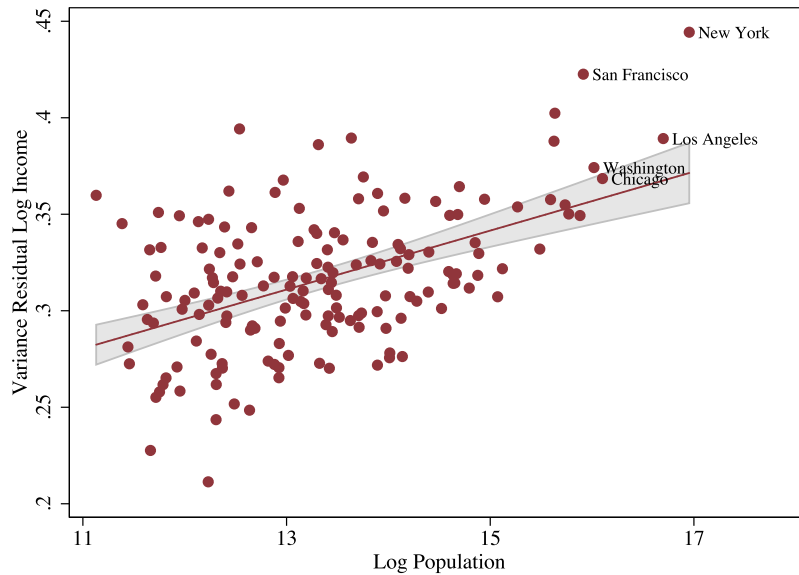
(b) Variance of Log Residual Income

Note: This figure plots the average residual income and the variance of log residual income for Combined Statistical Areas using data from the American Community Survey 5% IPUMS 2010–2014. The residual income is the residual from regressing income on years of schooling, potential experience, gender, and race. The line represents the linear fit and the shaded area is the 95% confidence interval.

blue-collar workers (classes 4, 5, and 6) as "workers." The crosswalk between the PCS and ACS occupational classifications is not one-to-one so each occupation will be assigned a probability of being a manager given its ACS occupation. The result is that on top of the ACS manager occupations, all get a probability of one for being a manager, but there are additional occupations with a positive probability of corresponding to a manager. For example, 41% of most science and engineering occupations will be classified as mangers. Other occupations are entirely classified as managers even if they would not fall under the classic definition of manager, such as directors and producers.

Table A.5: Production Organization with Multiple Layers

| | Fraction of Employees that are | | | |
	Blue & White Collar Workers	Layer 1 Managers	Layer 2 Managers	Layer 3 Managers
Log Population	-2.22**	-0.06	2.07**	0.38**
Constant	77.21**	27.44**	-8.48**	-0.60
R^2	0.34	0.00	0.38	0.35
N	162	162	162	162

*p<0.05; ** p<0.01*

Note: This table reports the results from regressing share of employment in each layer of management on to logged population for Combined Statistical Areas. The data source used is the American Community Survey 5% IPUMS for 2010–2014.

A.4.2   Multiple Layers using the PCS

In this section, I relax the simplification of describing the organization of production with two layers. I use the PCS codes to divide occupations into four layers: production workers (including both blue and white collar workers), intermediate professionals (layer 1 of management), professionals and upper management (layer 2 of management), and business owners (layer 3 of

management). Figure plots how the share of employment in each of this layers changes with city population. I find that doubling city size is associated with an average decrease of 2.2 percentage points in the share of production workers, no significant change in the share of intermediate professionals, a 2 percentage points increase in the share of upper management and professionals, and a 0.4 percentage point increase in the share of business owners.

The span of control increases in every layer except at the top layer since the number of business owners increases by less than upper management and professionals. The lower increase in business owners could be due to the fact that in small cities there may be more firms with only one level of management, and that would be the business owner. However, as I show in section 2 the main finding of decreasing spans of control with city size still holds when classifying only heads of companies as managers. Since I do not have firm identifiers I am not able to speak to the number of firms with a given number of layers. However, in recent work on France, Spanos (2018) finds that the distribution of the number of layers changes slightly with city size. Doubling population density is associated with 1.5% fewer firms with one layer of management, and 1.2% more firms with two layers of management.

A.4.3   Robustness

In this subsection, I present some additional details to the evidence on the organization of production and city size. I start by checking the robustness of the findings with respect to the definition of city. Figure A.5 plots the number of workers per manager on the log of population for Metropolitan Statistical Areas. The span of control is decreasing with city size for MSAs. Notice that the outliers are often part of a larger Combined Statistical Area. Bridgeport and Trenton belong to New York's CSA, Boulder belongs to Denver's CSA and San Jose belongs to San Francisco's CSA, and interestingly, they were also outliers in the regressions of average income and variance of log income.

Next, I check the robustness of the empirical regularity with respect to the definition of management. Figure A.6 plots the number of workers per manager for two alternative definitions of manager. In panel (a), the definition of manager is the head of a company with at least 10

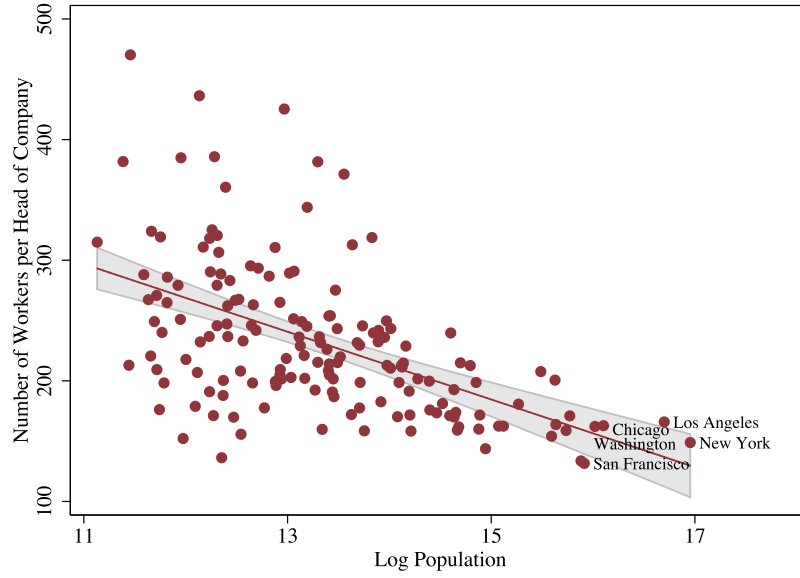## Figure A.5: Span of Control at the MSA Level
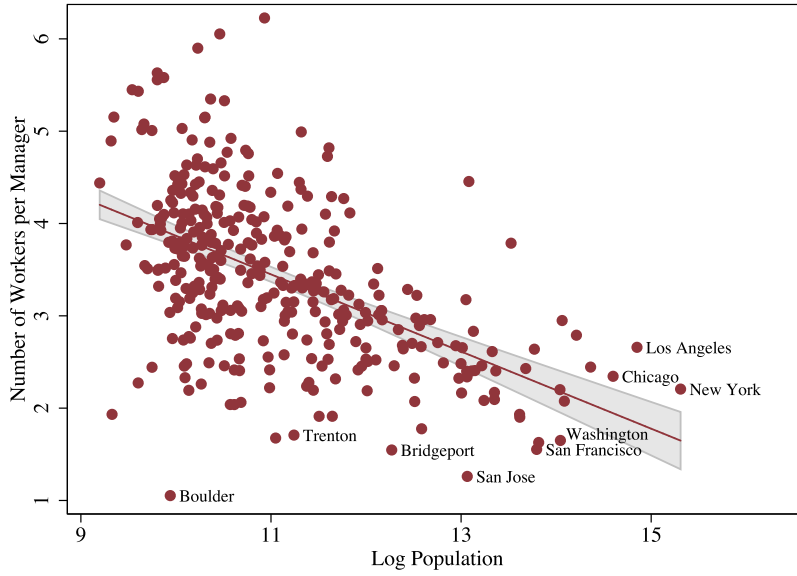
Note: This figure plots the number of workers per manager against logged population for Metropolitan Statistical Areas. The data source used is the American Community Survey 5% IPUMS for 2010–2014. The line corresponds to a linear regression and the shaded area corresponds to the 95th confidence interval.

employees according to the French occupational classification (PCS). In panel (b), a manager is defined using the ACS classification for managers. The span of control is decreasing in city size for both of these definitions although the level of span of control varies greatly with the definitions. It is between 100 and 500 workers per manager when using head of company, and between 5 and 15 when using the ACS classification.

Manufacturing is a particular sector where it is common for production plants to be located outside of cities while headquarters are located within cities. In order to explore whether the patter is coming from this effect, Figure A.7 plots the number of workers per manager against log population excluding the manufacturing sector. The span of control is still decreasing in city size.

City size is correlated with average income, variance of log income, and span of control. In order to investigate whether the span of control is correlated with the moments of the income distribution beyond what is captured by population, I first regress average income, variance of log income, and span of control on population in three separate regressions. Then in Figure A.8,

## Figure A.6: Alternative Management Definitions

(a) Head of Company
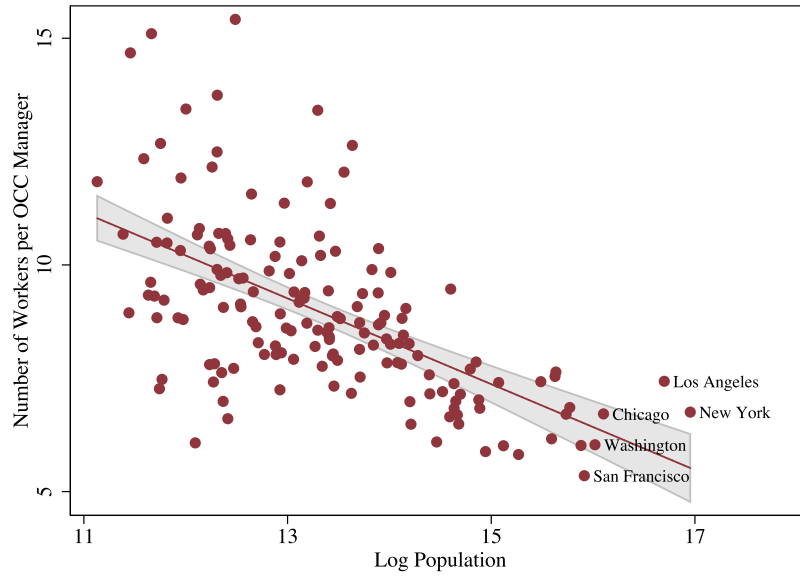
(b) ACS Occupational Classification

Note: This figure plots the number of workers per manager against logged population for Combined Statistical Areas. The data source used is the American Community Survey 5% IPUMS for 2010–2014. The line corresponds to a linear regression and the shaded area corresponds to the 95th confidence interval. Panel (a) plots the number of workers per head of company of more than 10 employees classified using the French occupational classification system (PCS). Panel (b) plots the number of workers per manager, where agents are classified as managers using the ACS occupational classification system.

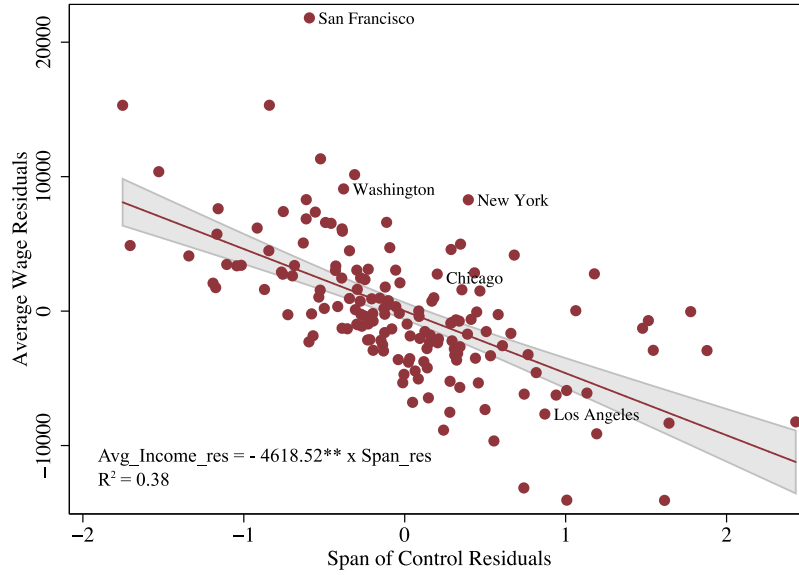Figure A.7: Span of Control, Excluding Manufacturing
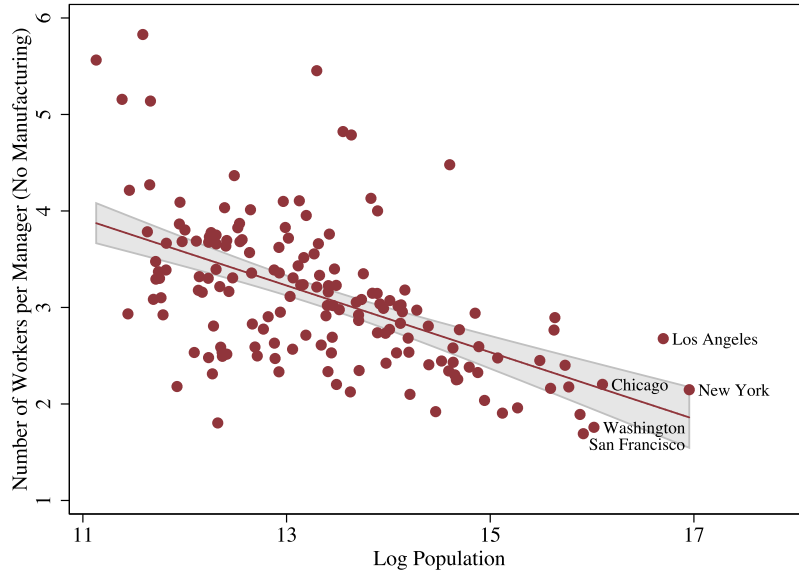
Note: This figure plots the number of workers per manager against logged population for Combined Statistical Areas excluding the Manufacturing sector. The data used is the American Community Survey 5% IPUMS for 2010–2014. The line corresponds to a linear regression and the shaded area corresponds to the 95th confidence interval.

I regress the residual average income and the residual variance of log income on the residual span of control. Cities with smaller spans of control than predicted by their size tend to have a higher average income and a higher variance of log income than predicted by their size.
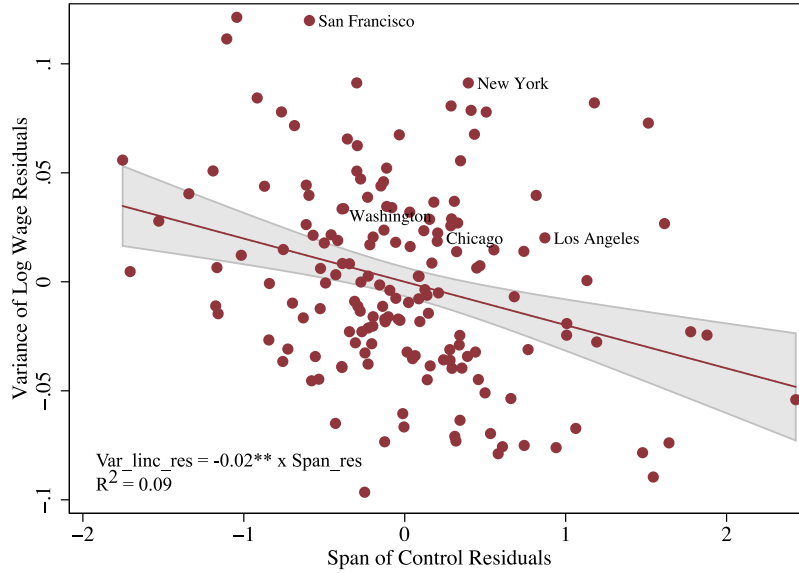
A.4.4 Additional Industry Examples

In this section, I present a couple of additional examples of occupational composition of industries across large and small cities. Figure A.9 includes the composition for the ten largest occupations of the "Insurance Carriers" industry. I compare a representative large city formed by all CSAs with population larger than 2.5 million inhabitants and a small city with population below 2.5 million. The large city has a larger fraction of financial managers and miscellaneous managers and a larger fraction of accountants and auditors, while the smaller city has a larger fraction of insurance claims and policy processing clerks as well as claims adjusters, appraisers, examiners, and investigators. This difference in the occupational composition could be because in

79

Figure A.8: Partialling Out Population

Avg_Income_res = - 4618.52** x Span_res
R² = 0.38

(a) Residual Average Wage


Var_linc_res = -0.02** x Span_res
R² = 0.09

(b) Residual Variance of Log Wage

Note: This figure plots the residual average wage and the residual variance of log wage against the residual span of control (number of workers per manager) where the residuals are the result of three independent regressions of average wage, variance of log wage, and span of control on population. The data source used is the American Community Survey 5% IPUMS for 2010–2014. The line corresponds to a linear regression and the shaded area corresponds to the 95th confidence interval.

large cities insurance carriers handle more complex cases, or they may carry out a more complex set of tasks such as the design of insurance contracts, which require a higher fraction of managers per worker. Meanwhile, in the smaller city, they perform simpler tasks such as processing and investigating insurance claims, which require a few fraction of managers per worker.
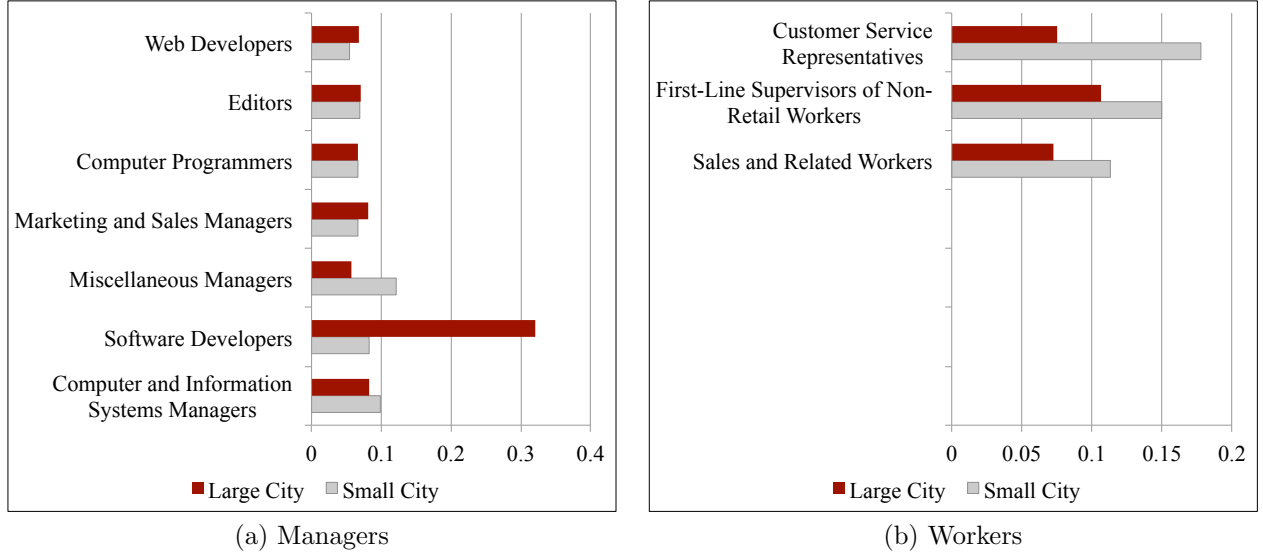
Figure A.9: Insurance Carriers

(a) Managers

(b) Workers
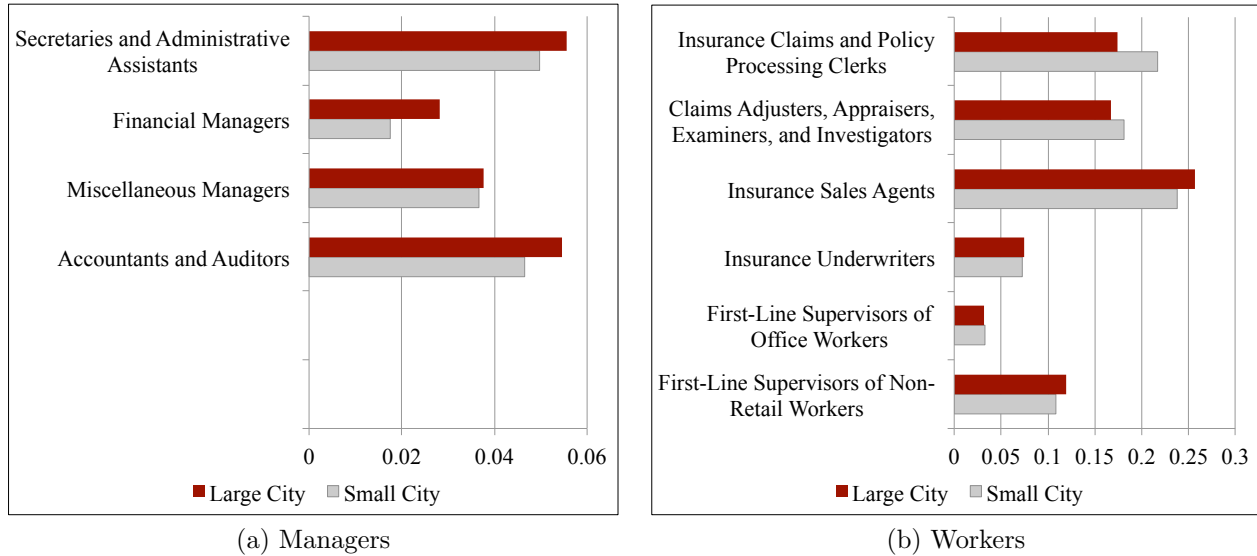
Note: This table compares the occupational composition of the "Insurance Carriers" industry between large and small cities. Large cities include all CSAs with more than 2.5 million inhabitants and small cities include those with less than 2.5 million inhabitants. It uses data from the American Community Survey, 5% IPUMS 2010-14. This comparison includes only the 10 occupations with the highest share of employment in the industry at the national level. Panel (a) captures the composition for those occupations classified as "Managers," and Panel (b) captures the composition for those occupations classified as "Workers."

The next example I present here is from the "Internet Publishing and Broadcast, and Web Search Portals" industry. Figure A.10 contains the occupational composition for the ten most common occupations across large and small cities for this industry. Larger cities have a much larger fraction of software developers than smaller cities. In comparison, smaller cities have a higher fraction of customer service and sales workers. This suggests that the larger cities specialize in more complex tasks like the development of new software, which requires a higher fraction of managers, in this case software and web developers per worker. The smaller cities specialize in simpler tasks like customer support for the clients using well-tested products, and these tasks require fewer managers per worker.

Figure A.10: Internet Publishing and Broadcast, and Web Search Portals

(a) Managers

(b) Workers

Note: This table compares the occupational composition of the "Internet Broadcast, and Web Search Portals" industry between large and small cities. Large cities include all CSAs with more than 2.5 million inhabitants and small cities include those with less than 2.5 million inhabitants. It uses data from the American Community Survey, 5% IPUMS 2010–2014. This comparison includes only the 10 occupations with the highest share of employment in the industry at the national level. Panel (a) captures the composition for those occupations classified as "Managers," and Panel (b) captures the composition for those occupations classified as "Workers."

B   Model Appendix

B.1   Positive Assortative Matching

Let $Y(z_m, z_w)$ denote the rents per unit of time of a manager of skill z_m that hires workers of skill z_w. We know that $Y(m(z), z) = R(m(z))$ if $m(.)$ is the equilibrium matching function.

In equilibrium, since managers choose the skill of the workers optimally: $\frac{\partial Y(m(z),z)}{\partial z} = 0$

Totally differentiating this equation:

$$\frac{\partial z_m}{\partial z_w} = -\frac{\partial^2 Y(z_m, z_w)/\partial^2 z_w}{\partial^2 Y(z_m, z_w)/\partial z_m \partial z_w}. \tag{B.1}$$

The numerator has to be negative because since managers are maximizing rents in equilibrium.

To show that the denominator is positive, notice that

$$Y(z_m, z_w) = \frac{A_c z_m^{\alpha_c}}{h_c \left(1 - z_w^{\alpha_c}\right)}, \tag{B.2}$$

$$\frac{\partial Y(z_m, z_w)}{\partial z_m} = \frac{\alpha_c A_c z_m^{\alpha_c - 1}}{h_c \left(1 - z_w^{\alpha_c}\right)}, \tag{B.3}$$

$$\frac{\partial Y(z_m, z_w)}{\partial z_m \partial z_w} = \frac{\alpha_c^2 A_c z_m^{\alpha_c - 1} z_w^{\alpha_c - 1}}{h_c \left(1 - z_w^{\alpha_c}\right)^2} > 0. \tag{B.4}$$

Hence,

$$\frac{\partial z_m}{\partial z_w} > 0.$$

Since the argument is valid for all workers, an equilibrium matching function $m'(z) > 0$ for all workers of skill z.

B.2 Schauder's fixed point theorem

Let V be the Hausdorff topological vector space of $\overrightarrow{g}(z) = \langle g_1(z), ..., g_C(z) \rangle$ bounded continuous vector valued functions mapping $[0, 1] \to \mathbb{R}^C$, with the uniform norm and the topology defined by all open sets. Note that since $\sum_{n=1}^{C} g_n(z) = 1$ and $g_n(z) > 0$, $g_n(z) \in [0, 1]$. Let K be the nonempty convex subset of V consisting of all functions mapping $[0, 1] \to [0, 1]^C$. Let T be the mapping defined by

$$T\left(\overrightarrow{g}(.)\right) = \left\langle \frac{exp\left\{\frac{1}{\beta}(Y_1(z; g_1(.))\} + a_1)\right\}}{\sum_{j=1}^{C} exp\left\{\frac{1}{\beta}(Y_j(z; g_j(.)) + a_j)\right\}}, ..., \frac{exp\left\{\frac{1}{\beta}(Y_C(z; g_C(.))\} + a_1)\right\}}{\sum_{j=1}^{C} exp\left\{\frac{1}{\beta}(Y_j(z; g_j(.)) + a_j)\right\}} \right\rangle.$$

T is a continuous mapping and the image of this mapping is contained in K. By Schauder's fixed point theorem: T has a fixed point.

B.3 Production with Three-Layer Teams

In this section, I present an extension of the model that allows for the formation of teams with three layers. A team with three layers is formed by workers, with subscript w, middle managers, with subscript m, and one top manager, with subscript t. Middle managers will be agents that both learn about unsolved problems and communicate unsolved problems to another agent. Top managers will be agents that learn about unsolved problems but do not communicate unsolved problems. Finally, a worker is an agent who encounters production problems and communicates unsolved problems.

A worker of skill z_w living in city c and with time supply l_w, encounters a mass 1 of problems per unit of time they spend working. The worker then solves the mass of problems with difficulty lower than their skill, $F_c(z_w) = z_w^{\alpha_c}$ where F_c is the distribution of problem difficulty in city c. For each mass 1 of problems solved, the worker produces A_c units of output. Therefore, the total output produced by this worker is $A_c z_w^{\alpha_c} l_w$. A mass of workers of skill z_w that supply L_w units of working time will generate $A_c z_w^{\alpha_c} L_w$ units of output and a mass $L_w \left(1 - z_w^{\alpha_c}\right)$ of unsolved problems. The workers cannot assess the difficulty of the unsolved problems but they can communicate them to a middle manager.

A middle manager of skill z_m and working time l_m has to spend h_c units of time per unsolved problem that they learn from the workers in the lower layer. Let $L_{m \to w}$ denote the amount of time the middle manager hires from workers. The time constraint faced by a middle manager who hires workers of skill z_w is given by:

$$L_{m \to w} \left(1 - F_c(z_w)\right) h_c \leq l_m. \tag{B.5}$$

Let us denote the maximum amount of workers' time that a middle manager can hire from workers of skill z_w per unit of time by $\bar{L}_{m \to w}(z_w)$. This function is given by the binding time constraint of the middle manager:

$$\overline{L}_{m\to w}(z_w) = \frac{1}{(1 - F_c(z_w))\, h_c}. \tag{B.6}$$

Therefore, the total output produced by a team formed by a mass of middle managers of skill z_m that supply L_m units of time and a mass of workers of skill z_w that supply $\overline{L}_{m\to w}(z_w)\, L_m$ units of time is $A_c F_c(z_m)\, \overline{L}_{m\to w}(z_w)\, L_m$. The team also generates a mass $(1 - F_c(z_m))\, \overline{L}_{m\to w}(z_w)\, L_m$ of unsolved problems. The middle managers are not able to assess the difficulty of the unsolved problems, but they can communicate them to a top manager.

A top manager of skill z_t and working time l_t has to spend h_c units of time per unsolved problem that they learn from the middle managers in the lower layer. Therefore, the time constraint for a top manager who hires a mass $L_{t\to m}$ of middle managers of skill z_m who in turn hire a mass $L_{t\to m}\overline{L}_{m\to w}(z_w)$ workers of skill z_w is given by:

$$(1 - F_c(z_m))\, \overline{L}_{m\to w}(z_w)\, L_{t\to m} h_c \leq l_t. \tag{B.7}$$

Let us denote the maximum amount of middle managers' time that a top manager can hire from middle managers of skill z_m per unit of time by $\overline{L}_{t\to m}(z_w, z_m)$. This function is given by the binding time constraint of the middle manager:

$$\overline{L}_{t\to m}(z_w, z_m) = \frac{1 - F_c(z_w)}{1 - F_c(z_m)}. \tag{B.8}$$

Therefore, the total output produced by a team formed by a top manager of skill z_t and working time l_t who hires $\overline{L}_{t\to m}(z_w, z_m)$ units of time from middle managers of skill z_m who in turn hire $\overline{L}_{m\to w}(z_w)$ units of time from workers of skill z_w is $A_c F_c(z_t)\, \overline{L}_{t\to m}(z_w, z_m)\, \overline{L}_{m\to w}(z_w)\, l_t$.

In order to decentralize the equilibrium, I assume that the top manager keeps the total output from the production team and pays out wages. For notational simplicity, wages are paid per unit of a worker's time. The top manager pays wages to the middle manager $w_c^m(z_m)$ per unit of worker time that the middle manager hires. Then, the middle manager pays wages to workers $w_c^w(z_w)$ per unit of time.

B.3.1 Optimization Problem of Top Managers

A top manager of skill z_t and working time supply l_t chooses the skill of the middle managers to hire, z_m in order to maximize profits.

$$R_t(z_t) l_t = max_{z_m} \Pi_t(z_t, z_m) = max_{z_m} \left(\frac{A_c F_c(z_t) - w_c^m(z_m)}{h_c(1 - F_c(z_m))} \right) l_t. \tag{B.9}$$

The solution to this problem will define a matching function between middle managers and top managers, $m_{mt}(z_m)$:

$$m_{mt}^{-1}(z_t) = argmax_{z_m} \Pi_t(z_t, z_m) = max_{z_m} \left(\frac{A_c F_c(z_t) - w_c^m(z_m)}{h_c(1 - F_c(z_m))} \right) l_t. \tag{B.10}$$

B.3.2 Optimization Problem of Middle Managers

A middle manager of skill z_m and working time supply l_m chooses the skill of the workers to hire, z_w in order to maximize profits.

$$R_m(z_m) l_m = max_{z_w} \Pi_m(z_m, z_w) = max_{z_w} \left(\frac{w_c^m(z_m) - w_c^w(z_w)}{h_c(1 - F_c(z_w))} \right) l_m. \tag{B.11}$$

The solution to this problem will define a matching function between workers and middle managers $m_{wm}(z_w)$:

$$m_{wm}^{-1}(z_m) = argmax_{z_w} \Pi_m(z_m, z_w) = max_{z_m} \left(\frac{w_c^m(z_m) - w_c^w(z_w)}{h_c(1 - F_c(z_w))} \right) l_m. \tag{B.12}$$

B.3.3 Labor Market Clearing

In order for the labor market to clear, two conditions need to hold. First, the amount of time supplied by workers with skill lower than z_w needs to equal the amount of time demanded of workers with skill lower than z_w by middle managers. Namely,

$$\int_0^{z_w} g(s)\,ds = \int_{z_{wm}^*}^{m_{wm}(z_w)} \frac{g(s)}{h\left(1 - F\left(m_{wm}^{-1}(s)\right)\right)}\,ds. \tag{B.13}$$

Differentiating both sides with respect to z_w, we obtain a differential equation for the matching function of workers to middle managers:

$$g(z)\,h(1 - F(z)) = g(m_{wm}(z))\,m_{wm}'(z). \tag{B.14}$$

Integrating both sides from 0 to z_w we obtain the solution to the differential equation:

$$m_{wm}(z) = G^{-1}\left(G\left(z_{wm}^*\right) + h\int_0^z g(s)(1 - F(s))\,ds\right), \tag{B.15}$$

with initial conditions given by: $m_{wm}(0) = z_{wm}^* \; m_{wm}\left(z_{wm}^*\right) = z_{mt}^*$.

Second, the amount of time supplied by middle managers with skill lower than z_m has to be equal to the amount of time demanded of middle managers with skill lower than z_m by top managers.

$$\int_{z_{wm}^*}^z g(s)\,ds = \int_{z_{mt}^*}^{m_{mt}(z)} \frac{1 - F\left(m_{wm}^{-1}\left(m_{mt}^{-1}(z)\right)\right)}{1 - F\left(m_{mt}^{-1}(z)\right)} g(s)\,ds. \tag{B.16}$$

Differentiating both sides with respect to z_m, we obtain a differential equation for the matching function of middle managers to top managers:

$$g(z)\frac{1 - F(z)}{1 - F\left(m_{wm}^{-1}(z)\right)} = g(m_{mt}(z))\,m_{mt}'(z),$$

with initial conditions given by: $m_{mt}\left(z_{wm}^*\right) = z_{mt}^*$ and $m_{mt}\left(z_{mt}^*\right) = 1$.

Finally, integrating both sides from z_{wm}^* to z_m, we obtain an expression for the matching function of middle managers to top managers:

$$m_{mt}(z) = G^{-1}\left(G\left(z^*_{mt}\right) + \int_{z^*_{wm}}^{z} g(s)\frac{(1-F(s))}{1-F\left(m^{-1}_{wm}(s)\right)}ds\right).$$

The two initial conditions provide a system of two equations and two unknowns on the two thresholds: z^*_{wm} and z^*_{mt}:

$$z^*_{mt} = G^{-1}\left(G\left(z^*_{wm}\right) + h\int_0^{z^*_{wm}} g(s)(1-F(s))ds\right) \tag{B.17}$$

$$1 = G^{-1}\left(G\left(z^*_{mt}\right) + \int_{z^*_{wm}}^{z^*_{mt}} g(s)\frac{(1-F(s))}{1-F\left(m^{-1}_{wm}(s)\right)}ds\right). \tag{B.18}$$

B.3.4   Wage Functions

The first order condition for the optimization problem of the top manager in equation (27) with respect to the skill of the manager results in the following differential equation for the wage function:

$$w'_m(z_m) + w_m(z_m)\frac{F'(z_m)}{1-F(z_m)} = AF(z_t)\frac{F'(z_m)}{1-F(z_m)}. \tag{B.19}$$

The solution for this differential equation is a wage function for middle managers:

$$w_m(z_m) = (1-F(z_m))\left(\frac{w_m\left(z^*_{wm}\right)}{1-F(z^*_{wm})} + \int_{z^*_{wm}}^{z_m} AF(m_{mt}(s))\frac{F'(s)}{(1-F(s))^2}ds\right), \tag{B.20}$$

where $w_m\left(z^*_{wm}\right)$ is the initial condition and will de derived below from the indifference condition at the threshold.

The first order condition for the optimization problem of the middle manager in equation (29) with respect to the skill of the manager results in the following differential equation for the wage function:

$$w'_w(z_w) + w_w(z_w) \frac{F'(z_w)}{1 - F(z_w)} = w_m(z_m) \frac{F'(z_w)}{1 - F(z_w)}. \tag{B.21}$$

The solution for this differential equation is a wage function for workers:

$$w_w(z_w) = (1 - F(z_w)) \left(w_w(0) + \int_0^{z_w} w_m(m_w(s)) \frac{F'(s)}{(1 - F(s))^2} ds \right), \tag{B.22}$$

where $w_w(0)$ is the initial condition and will be derived below from the indifference condition at the threshold.

The indifference condition at the threshold skill z^*_{mt} between being the worst top manager or the best intermediate manager results in an equation for the initial wage of the middle managers:

$$\frac{w_m(z^*_{mt}) - w_w(z^*_{wm})}{h(1 - F(z_{wm}))} = \frac{AF(z^*_{mt}) - w_m(z^*_{wm})}{h(1 - F(z_{wm}))}, \tag{B.23}$$

$$w_m(z^*_{wm}) = (1 - F(z^*_{wm})) \left(\frac{AF(z^*_{mt})}{(1 - F(z^*_{mt}))} - \int_{z^*_{wm}}^{z^*_{mt}} AF(m_{mt}(s)) \frac{F'(s)}{(1 - F(s))^2} ds \right). \tag{B.24}$$

Finally, the indifference condition at the threshold skill z^*_{wm} between being the worst middle manager or the best worker results in an equation for the initial wage of the middle managers:

$$w_w(z^*_{wm}) = \frac{w_m(z^*_{wm}) - w_w(0)}{h} \tag{B.25}$$

$$w_w(0) = \left(\frac{1 + h(1 - F(z^*_{wm}))}{h(1 - F(z^*_{wm}))} \right) \left(\frac{w_m(z^*_{wm})}{h(1 - F(z^*_{wm}))} - \int_0^{z^*_{wm}} w_m(m_w(s)) \frac{F'(s)}{(1 - F(s))^2} ds \right). \tag{B.26}$$

C   Quantification Appendix

C.1 Baseline Estimation

Table C.1 contains the target moments and the models simulated in the model.

Table C.1: Target Moments

	Large City		Small City	
	Model	Data	Model	Data
Median Income of Managers	$80,590	$80,674	$66,398	$66,455
Median Income of Workers	$38,208	$38,295	$35,536	$35,586
Span	2.2514	2.2525	2.8162	2.8166
City Size	0.9327	0.9326		
	Overall Moments			
Variance of Log Income	0.5725		0.5725	

C.2 Four Cities Quantification

In this section, I quantify the technology for four cities in the US: New York, Chicago, Pittsburgh, and Springfield, MA. I select these cities because they are distributed throughout the size distribution of cities and they approximately trace out the linear regression for average income, variance of log income, and span of control, so they can be taken as representative cities for their size. Table 19 reports the estimated technology parameters for these four cities and Table 20 reports the match to the moments in the data.

The main takeaway from this quantification is that for these four cities there is a pattern similar to the one found with two cities. The largest city has the most productive technology and the productivity is monotonically increasing in city size. However, the less productive tech-

Table C.2: Estimated Parameters

		New York, NY	Chicago, IL	Pittsburgh, PA	Springfield, MA
A	Productivity	9.0256	8.0808	7.0730	4.6055
h	Communication cost	0.7293	0.7336	0.8562	0.6206
α	Problem Difficulty	0.9306	0.7295	0.4927	0.4858
a	Amenity	2.8149	2.3933	1.8451	0

Table C.3: Target Moments

	New York, NY		Chicago, IL		Pittsburgh, PA		Springfield, MA	
	Model	Data	Model	Data	Model	Data	Model	Data
Median Income of Managers	$86,972	$87,043	$77,756	$77,827	$64,531	$64,554	$48,208	$48,234
Median Income of Workers	$40,198	$40,337	$39,264	$39,389	$37,383	$37,515	$30,287	$30,354
Span	2.2040	2.2084	2.4036	2.3867	2.5280	2.5254	3.5542	3.5548
City Size	0.6476	0.6412	0.2806	0.2781	0.0705	0.0693	-	-

nologies are not dominated by the more productive ones because they either have easier problems or lower communication cost. In fact, none of these technologies is completely dominated. The communication cost is not monotonic and it is particularly high for Pittsburgh; however, it is compensated by particularly easy problems.

D   Counterfactuals Appendix

D.1   Minimum Wage in the Small City

In this section, I repeat the analysis on imposing a minimum wage for the small city. The lowest earned income in the small city is $17,619 which is equivalent to $9.17 an hour and I will quantify the impact of imposing a $13.76 an hour.

I start by analyzing the effect on the low skilled, identified by those agents earning below $19 an hour following the strategy of Jardim et al. (2017) on the minimum wage in Seattle. The authors find that for these low-income workers, employment fell by 9% and wages increased by 3% resulting in a total decrease in payroll of $125 per month. I find that employment for the low skill workers decreases by 16.36% and their average wage increases by 6.8 percent. On average they lose $263 per month. The numbers are about twice as high as the effects found for Seattle.

Next, I quantify the impact of this policy on welfare, measured by average utility, and on inequality measured by variance of log utility. In order to better understand the consequences of implementing an increase in the minimum wage, I consider the effect of the policy for the original residents, comparing the new residents to the old residents, and considering only employed agents. Table D.1 summarizes the results from the three scenarios for the large city, the small city, and the overall effect.

Overall, the minimum wage has a small negative effect on average utility of 0.04% and a small effect on the variance of log utility, of 1.03%. The effects for the large city are even smaller. The effect of this policy on the original residents of the city is slightly negative in terms of average

Table D.1: Effect of a 50% Increase in the Minimum Wage of the Small City

	Large City	Small City	Overall
	All Residents		
Average Utility	-0.25%	3.1%	-0.04%
Var Log Utility	4.4%	-6.43%	2.28%
	Original Residents		
Average Utility	-0.02%	-0.38%	-0.04%
Variance Log Utility	0.23%	26.2%	2.28%
	Employed Agents		
Average Utility	-0.25%	4.03%	0.02%
Variance Log Utility	4.4%	-20.52%	1.03%

Figure D.1: Change in the Utility of the Original Residents
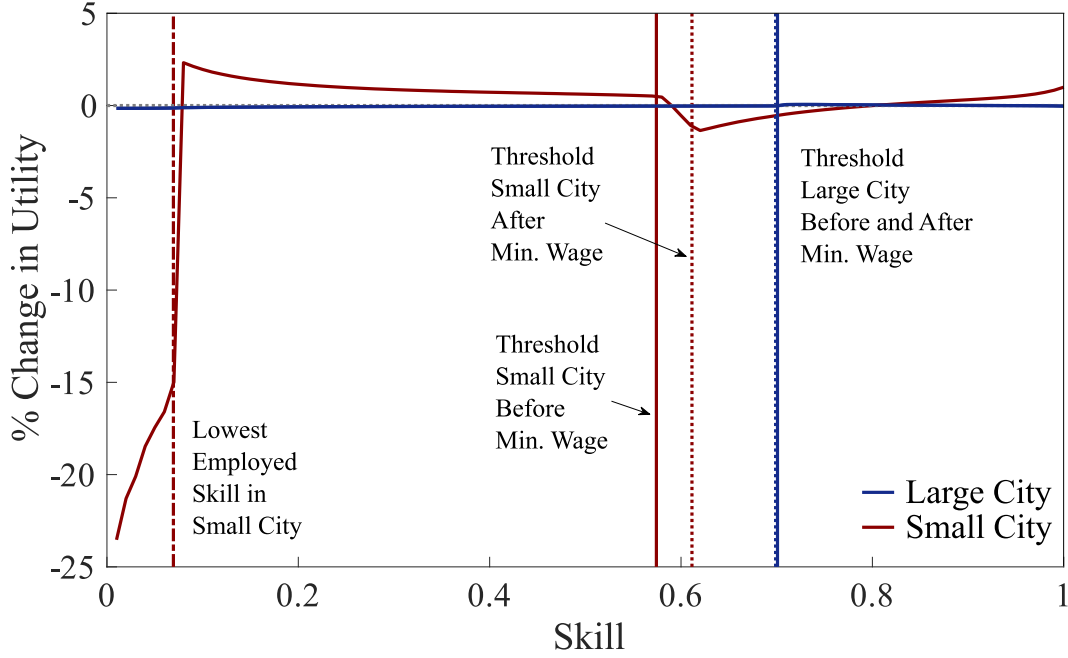
utility, with a 0.38% decrease. Moreover, the inequality among the original residents of the small city increases by 26.2%. The increase in inequality comes from the losses for the low-skill workers that become self-employed. If we only look at agents that stay in the labor market, they experience a 4% increase in average utility and a 20.52% reduction in the variance of log utility.

Finally, I quantify the differential effect of the policy across skill levels. Figure D.1 plots the percentage change in average utility for the original residents of each skill type both in the small city and in the large city. On the one hand, the effects in the large city are negligible because the large city represents such a large fraction of total population, the mobility effect from the policy almost does not affect the overall distribution of skills in the large city. This effect highlights that the effect the minimum wage in the large city had on the small city was coming from the impact on the skill distribution of the small city. On the other hand, the effects on the small city are similar to the ones in the large city although smaller in magnitude. The main losers from the policy are the lowest-skilled agents and the middle managers, while the big winners are the workers above the minimum wage and the very top managers who can now hire better workers.