Technology, Spatial Sorting, and Job Polarization*

Jan Eeckhout†, Christoph Hedtrich‡, and Roberto Pinheiro§

October 2019

Preliminary version – Please do not cite

Abstract

Job polarization and wage inequality have increased substantially in the last four decades, being largely attributed to technological change. But there are two competing drivers: Skill Biased Technological Change (SBTC), leading to a rising college premium; and automation, leading to the replacement of routine occupations and the "hollowing out" of the income distribution. Using a novel data set on Information Technology (IT) adoption, we exploit geographical variation and the sorting patterns of differentially skilled workers to infer the main driver of job and wage polarization. We find strong evidence that there is more automation in big cities; big cities also have a disproportionate decrease in the share of routine cognitive jobs (clerical workers and low-level white collar workers). We propose an economic mechanism where the substitutability of routine workers by IT leads to higher IT adoption in large cities than in small cities. Wages and productivity are higher in large cities, whereas technology prices are constant across cities. This technology also generates thick tails in the skill distribution in large cities.

Keywords: Automation, Skill Distributions, City Sizes, Job and Wage Polarization.

JEL Codes: D21, J24, J31, R23.

*We would like to thank Jim Davis, Morris Davis, Tim Dunne, Bruce Fallick, Domenico Ferraro, Pawel Krolikowski, Jeff Lin, Stuart Rosenthal, Henry Siu, Chris Taber, Murat Tasci, the participants of the 2018 SOLE, 2018, 2018 UEA, 2018 Fall Midwest Macro, 2018 CLEF, 2018 Regional System, 2018 Summer Econometric Society, 2019 SED, UKentucky and Cleveland Fed Micro Workshop and participants of the HUDM Conference at USC Dornsife, 14th Nordic Summer Inst. in Labor Economics, among others for comments and suggestions. We would like to thank Albert Saiz and David Albouy for access to data and Ross Cohen-Kristiansen for research assistance. Any remaining errors are our responsibility. Jan Eeckhout acknowledges financial support from grant ECO2015-67655-P and ERC Advanced Grant 339186. The views in this article are those of the authors and do not necessarily reflect the views of the Federal Reserve System, the Federal Reserve Bank of Cleveland, or the Board of Governors.

†UPF and UCL jan.eeckhout@upf.edu.

‡Uppsala University, christoph.hedtrich@nek.uu.se

§Federal Reserve Bank of Cleveland, Roberto.Pinheiro@clev.frb.org.

1 Introduction

Job polarization and wage inequality have risen sharply since the early 1980s. In particular, the pay gap between the high and low educated, represented by the college premium, has gone up substantially, from 40% in 1980 to exceeding 97%.[1] Moreover, the college premium is at the highest level since 1915, the earliest year for which representative data are available.[2] The standard explanation first put forward by Katz and Murphy (1992) is skill-biased technological change (SBTC). New technologies make high skilled workers disproportionately more productive than low and middle skilled workers, thus leading to higher wages.

At the same time, technological change through automation has resulted in the "disappearing middle" of the income distribution. Automation has directly substituted capital for labor in routine tasks previously performed by moderately skilled workers. Hence, automation reduces job opportunities in the middle of the skill distribution, including clerical, administrative, production, and operative occupations. Jobs less affected by automation would demand either non-routine abstract tasks – requiring high levels of education and commanding high compensation – or non-routine manual tasks – which tend to be low-paying manual jobs. Consequently, we have a "hollowing out" of the income distribution

An open question is which of the technological forces – SBTC and automation – predominates in delivering job polarization and growing income inequality. To investigate these distinct drivers of job polarization and wage inequality, in this paper we exploit the geographical variation of technology adoption. The variation of technological change across locations informs us about the relative importance of technology on the college premium and on polarization.

We find that routine-task jobs are replaced by computers and software faster in large, expensive cities than in small, cheap cities. We show that living costs – in particular housing costs – play a key role. For example, let's consider two offices that are demanding for some standard accounting services that can be performed either by an accounting assistant or by an accounting software. One of these offices is located in New York city, the other in Akron, OH. In order to hire a new accounting assistant, the New York office must pay a wage that allows the new employee to live in an area close enough to the company's office in order to go to work every day. Since housing costs in the New York area are significantly higher than in Akron, OH, the New York-based firm must pay more to hire the same accounting assistant. In comparison, accounting software is the same price in both cities. Consequently, automation at a location-independent price is a more attractive substitute to the New York firm. In equilibrium, it is more likely that

[1]See Acemoglu and Autor (2011).

[2]According to Goldin and Katz (2009).

the New York firm will introduce the new software, while the Akron office hires an additional accounting assistant.

Our contribution is double. First, we use a novel data set collected by Aberdeen to analyze the role of investment in technology in local, geographically differentiated labor markets or CMSAs (Combined Metropolitan Statistical Areas). We have data for two measures: the total IT budget per worker and the expenditure on Enterprise Resource Planning (henceforth ERP) software. The combined use of IT budget per worker and exposure to ERP software gives us a diverse measure of technology adoption. On one hand, IT budget per worker is an accurate measure of investment in technology, being possibly used to either automate away routine tasks or complement non-routine cognitive tasks. Moreover, IT budget per worker is a continuous variable, and also has more detailed information and coverage across establishments. Instead, information on ERP software usage allows us to clearly identify the intensity of usage of automation technology.[3] Consequently, the introduction of ERP software reduces the need for clerical and low-level white collar workers. Moreover, in contrast with Personal Computers (PCs), which are general purpose technologies (Jovanovic and Rousseau (2005)), ERP software has as its main goal the replacement of clerical work.[4]

Our empirical results show that large and more expensive cities invest more in technology, measured either by the total IT budget per worker or by ERP software. At the same time, large expensive cities have also experienced the largest decrease in the fraction of routine cognitive workers in the population of employed workers.

Our second contribution is to propose a mechanism that can explain this correlation. We build an equilibrium model of heterogeneous workers' locations across cities that offers an economic mechanism to explain the empirical relation between investment in technology and the decline in routine tasks. In our model, housing prices play a key role in workers' city choices. Heterogeneously skilled citizens earn a living based on a competitive wage and choose housing in a competitive housing market. Under perfect mobility, their location choices make them indifferent between consumption-housing bundles, and therefore between different wage-housing price pairs across cities. Wages are generated by firms that compete for labor and that have access to a city-specific technology summarized by that city's total factor productivity (TFP).

[3]As pointed out by Bloom et al. (2014), ERP is the generic name for software systems that integrate several data sources and processes of an organization into a unified system. These applications are used to store, retrieve, and share information on any aspect of the sales and firm organizational processes in real time. This information includes not only standard metrics like production, deliveries, machine failures, orders and stocks, but also broader metrics on human resources and finance.

[4]Unfortunately, clear drawbacks of ERP measures are their coarseness – the only available information on ERP it is its type (no available information on type and number of licenses, for example), as well as the fact that we have information on software installation for 10% of the establishments in our sample.

This naturally gives rise to a price-theoretic measure of skills. Larger cities pay higher wages, and are more expensive to live in. Under worker mobility, revealed preference location choices imply that wages adjusted for housing prices are a measure of skills.

Within this framework, we introduce investment in technology capital. We start from the premise that that capital is produced globally and all cities are small open economies in the market for capital. Therefore, firms in all cities can rent any quantity of capital and take capital's rental rate as given.

In the presence of technological investment, we test the two competing hypotheses that have set out to analyze. On the one hand, the Skill Biased Technological Change (SBTC) hypothesis considers that capital and high-skill workers are complements, leading to a college wage premium. On the other hand, the automation hypothesis considers that mid-skill workers and capital are substitutes. While we believe that these hypothesis are not mutually exclusive, this simplification allows us to draw some stark comparisons in order to identify the driving forces behind the changes in the employment and wage distributions across cities.

We show that the automation hypothesis is able to match the empirical patterns that we find in the data particularly well. We observe an increasing substitution of routine cognitive jobs with ERP software and computers as the cost of investment of these technologies falls. Moreover, our model shows that the automation hypothesis is also able to deliver the thick tails distribution in the skill distribution, documented by Eeckhout et al. (2014). In contrast, in the same set-up, the SBTC hypothesis would deliver First Order Stochastic Dominance (FOSD) in the skill distribution. In this sense, while we do not discard the possibility of SBTC, our results point to the importance of including the automation hypothesis in order to match some key patterns presented by the empirical evidence.

Related Literature. Our paper is closest to Autor and Dorn (2013). They show that areas with a high concentration of workers performing routine tasks, there is a push towards automation. In this sense, we could imagine an initial large sunk cost of implementing automation – particularly true for routine manual workers – which would be more profitable the more workers the new machines would substitute. Our results point towards a different dynamics, that hinges on the differences of local prices. Through our results, even though clerical workers may be a somewhat smaller fraction of the labor force in New York City than in Akron OH, the fact that hiring a new accounting assistant is significantly more expensive in New York City makes it more attractive to New York-based firms to introduce the new software. Consequently, it is not necessarily the absolute fraction of the work force in routine tasks that induce automation, but the relative cost of introducing the new technology vs. routine task workers. Our results suit quite well the introduction of technologies that do not demand large initial sunk costs – such as the adoption of new softwares.

The notion that capital investment affects different skilled workers is of course not new. Krusell et al. (2000) were the first to argue that the college premium has risen so much because technological investment affects the high skilled more than the low skilled. The drop in the cost of such new technologies then further widens the gap between skilled and unskilled workers. Similarly, Beaudry et al. (2010) show that technology adoption – measured by PCs per worker – has occurred first in areas with relatively high supply of skill (or with low relative price of skill). They also show that these areas experienced the greatest increase in the return to education. Our analysis, while controlling for the relative supply of high skill workers in the MSA, highlights the importance of local prices in the sorting of workers and activities across space, which is mostly missing from Beaudry et al. (2010)'s analysis. Moreover, by allowing more than two types of workers, our framework is better suited to address the issues of job polarization and "disappearing middle" of the income distribution.

We are the first to document the effects of introducing new technologies while looking at technology investments that are not only tied to geographical locations, but also to a particular use. In this sense, we focus on software whose use is clearly related to the activities performed by routine cognitive workers, instead of general purpose technologies, such as PCs.[5]

In his 2019 Ely lecture, Autor (2019), like us, documents the variation of the disappearing middle across geographical locations. He also finds that this phenomenon is more pronounced in large cities. We go a step further, providing a mechanism to explain the economic phenomenon. Moreover, we use a direct measure of technology, namely the price of investment in technological capital. We have unique data on the use of technology at the establishment level. Acemoglu and Restrepo (2017) also analyze the role of technological change on the labor market, but they impute local level robot use based on national data. They posit that locations with lots of manufacturing have robots and have a decline in employment. Instead, we observe the adoption of new technologies at the establishment level.

Our paper is divided into six sections. Section 2 presents our model and theoretical results. Section 3 describes the data. Section 4 presents our empirical results. Section 5 estimates an extended version of the model that includes occupational choice and a housing supply sector. It also shows preliminary counterfactual experiments. Finally, section 6 concludes the paper. All proofs are presented in appendix section B.

[5]Autor and Dorn (2013) use the measure constructed by Doms and Lewis (2006) to pin down the number of PCs in 1990.

2 Model

Population. Consider an economy with heterogeneously skilled workers. Workers are indexed by a skill type i. For now, let the types be discrete: $i \in \mathcal{I} = \{1, ..., I\}$. Associated with this skill order is a level of productivity x_i. Denote the country-wide measure of skills of type i by M_i. Let there be J locations (cities) $j \in \mathcal{J} = \{1, ..., J\}$. The amount of land in a city is fixed and denoted by H_j. Land is a scarce resource.

Preferences. Citizens of skill type i who live in city j have preferences over consumption c_{ij}, and the amount of land (or housing) h_{ij}. The consumption good is a tradable numeraire good with price equal to one. The price per unit of land is denoted by p_j. We think of the expenditure on housing as the flow value that compensates for the depreciation, interest on capital, etc. In a competitive rental market, the flow payment will equal the rental price.[6] A worker has consumer preferences over the quantities of goods and housing c and h that are represented by: $u(c, h) = c^{1-\alpha}h^{\alpha}$, where $\alpha \in [0, 1]$. Workers are perfectly mobile, so they can relocate instantaneously and at no cost to another city. Because workers with the same skill are identical, in equilibrium each of them should obtain the same utility level wherever they choose to locate. Therefore for any two cities j, j' it must be the case that the respective consumption bundles satisfy $u(c_{ij}, h_{ij}) = u(c_{ij'}, h_{ij'})$, for all skill types $\forall i \in \{1, ..., I\}$.

Technology. Cities differ in their total factor productivity (TFP) which is denoted by A_j. For now, we assume that TFP is exogenous. We think of it as representing a city's productive amenities, infrastructure, historical industries, persistence of investments, etc.

In each city, there is a technology operated by a representative firm that has access to a city-specific TFP A_j. Output is produced by choosing the right mix of differently skilled workers i as well as the amount of capital k. While labor markets are local and workers must live in the city in which they are employed, capital markets are global and even large cities are small open economies in the capital markets. We also consider that firms rent capital that is owned by a zero measure of absentee capitalists. For each skill i, a firm in city j chooses a level of employment m_{ij} and produces output: $A_jF(m_{1j}, ..., m_{Ij}, k_j)$. Firms pay wages w_{ij} for workers of type i. It is important to note that wages depend on the city j because citizens freely locate between cities not based on the highest wage, but, given housing price differences, based on the highest utility. Like land and capital, firms are owned by absentee capitalists (or equivalently, all citizens own an equal share in the mutual fund that owns all the land and all the firms).

[6] We will abstract from the housing production technology; for example, we can assume that the entire housing stock is held by a zero measure of absentee landlords.

Finally, we consider that the rental price for capital is given by $r > 0$ which is determined in the global market and taken as given by firms in the different cities.

Market Clearing. In the country-wide market for skilled labor, markets for skills clear market by market, and for housing, there is market clearing within each city:

$$\sum_{j=1}^{J} C_j m_{ij} = M_i, \ \forall i \qquad \sum_{i=1}^{I} h_{ij} m_{ij} = H_j, \ \forall j. \tag{1}$$

where C_j denotes the number of cities with TFP A_j.

2.1 The Equilibrium Allocation

The Citizen's Problem. Within a given city j and given a wage schedule w_{ij}, a citizen chooses consumption bundles $\{c_{ij}, h_{ij}\}$ to maximize utility subject to the budget constraint (where the tradable consumption good is the numeraire, i.e. with price unity)

$$\begin{aligned} \max_{\{c_{ij},h_{ij}\}} u(c_{ij}, h_{ij}) &= c_{ij}^{1-\alpha} h_{ij}^{\alpha} \\ \text{s.t. } c_{ij} + p_j h_{ij} &\leqslant w_{ij} \end{aligned} \tag{2}$$

for all i, j. Solving for the competitive equilibrium allocation for this problem we obtain $c_{ij}^{\star} = (1-\alpha) w_{ij}$ and $h_{ij}^{\star} = \alpha \frac{w_{ij}}{p_j}$. Substituting the equilibrium values in the utility function, we can write the indirect utility for a type i as:

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{w_{ij}}{p_j^{\alpha}} \quad \Longrightarrow \quad w_{ij} = U_i p_j^{\alpha} \frac{1}{\alpha^{\alpha} (1-\alpha)^{1-\alpha}}, \tag{3}$$

where U_i is constant across cities from labor mobility. This allows us to link the wage distribution across different cities j, j'. Wages across cities relate as:

$$\frac{w_{ij}}{w_{ij'}} = \left(\frac{p_j}{p_{j'}}\right)^{\alpha}. \tag{4}$$

The Firm's Problem. All firms are price-takers and do not affect wages. Wages are determined simultaneously in each submarket i, j while capital rent is determined in the global market. Given the city production technology, a firm's problem is given by:

$$\max_{m_{ij}, \forall i} A_j F(m_{1j}, ..., m_{Ij}, k_j) - \sum_{i=1}^{I} w_{ij} m_{ij} - r k_j, \tag{5}$$

subject to the constraint that $m_{ij} \geqslant 0$ and $k \geqslant 0$. The first-order conditions are: $A_j F_{m_{ij}}(m_{ij}, k_j) = w_{ij}, \forall i$ and $A_j F_{k_j}(m_{ij}, k_j) = r$.[7]

Because there is no general solution for the equilibrium allocation in the presence of an unrestricted technology, we focus on variations of the Constant Elasticity of Substitution (CES) technology, where the elasticity is allowed to vary across skill types. As a benchmark therefore, we consider the following *separable* technology:

$$A_j F(m_{1j}, ..., m_{Ij}, k_j) = A_j \left(\sum_{i=1}^{I} m_{ij}^{\gamma_i} x_i + k_j x_k \right) \tag{6}$$

with $\gamma_i < 1$, $\forall i \in \{1, ..., I\}$. In this case the first-order conditions are $A_j \gamma_i m_{ij}^{\gamma_i - 1} x_i = w_{ij}, \forall i$ and $A_j \gamma_i k_j^{\gamma_i - 1} x_k = r$. Notice that if $\gamma_i \equiv \gamma$, $\forall i \in \{1, ..., I\}$ we have a CES production function.

In an on line appendix, we solve the allocation under *separable* technology as a special case of the more general technologies presented in the paper. Even without fully solving the system of equations for the equilibrium wages, observation of the first-order condition reveals that productivity between different skills i in a given city is governed by three components: (1) the productivity x_i of the skilled labor and how fast it increases in i; (2) the measure of skills m_{ij} employed (wages decrease in the measure employed from the concavity of the technology); and (3) the degree of concavity γ_i, indicating how fast congestion builds up in a particular skill. Without loss of generality, we assume that wages are monotonic in the order i.[8] This is consistent with our price-theoretic measure of skill.

We now proceed by introducing varying degrees of complementarity/substitutability between different skills and capital, starting from the *separable* technology. In this way, we are able to address different theories in terms of the impact of technology in either boosting the productivity of some types, as presented by the literature on Skill Bias Technological Change (henceforth SBTC) or replacing workers, as in the literature about automation. For tractability, let there be two cities, $j \in \{1, 2\}$ and three skill levels $i \in \{1, 2, 3\}$. We will also consider the degree of complementarity/substitutability by nesting a CES production function within the overall production function. Consequently, if we assume that there is a degree of complementarity

[7]In what follows, the non-negativity constraint on m_{ij} and k_j are dropped. This is justified whenever the technology satisfies the Inada condition that marginal product at zero tends to infinity whenever A_j is positive. This will be the case since we focus on variations of the CES technology.

[8]For a given order i, wages may not be monotonic as they depend on the relative supply of skills as well as on x_i. If they are not, we can relabel skills such that the order i corresponds to the order of wages. Alternatively, we can allow for the possibility that higher skilled workers can perform lower skilled jobs. Workers will drop job type until wages are non-decreasing. Then the distribution of workers is endogenous, and given this endogenous distribution, all our results go through. For clarity of the exposition, we will assume that the distribution of skills ensures that wages are monotonic.

between skill i and capital, while none between the remaining skills, then we consider that the technology can be written as $\left(m_{ij}^{\theta}x_i + k^{\theta}x_k\right)^{\frac{\gamma_i}{\theta}} + \sum_{l=-i} m_{lj}^{\gamma_j}x_l$. Notice that if $\gamma_i > \theta$, skill i and capital are gross complements, while if $\gamma_i < \theta$, capital and skill i are gross substitutes.

Definition 1 *Consider the following technologies:*

I. ***Automation.*** *Capital and middle skill workers are substitutes.*

$$A_j F(m_{1j}, m_{2j}, m_{3j}, k) = A_j \left\{ m_{1j}^{\gamma_1}x_1 + \left(m_{2j}^{\theta}x_2 + k_j^{\theta}x_k\right)^{\frac{\gamma_2}{\theta}} + m_{3j}^{\gamma_3}x_3 \right\} \quad \textit{where} \quad \gamma_2 < \theta \tag{7}$$

II. ***Skill-Bias Technological Change.*** *Capital and high skill workers are complements.*

$$A_j F(m_{1j}, m_{2j}, m_{3j}, k) = A_j \left\{ m_{1j}^{\gamma_1}x_1 + \left(m_{3j}^{\theta}x_3 + k_j^{\theta}x_k\right)^{\frac{\gamma_3}{\theta}} + m_{2j}^{\gamma_2}x_2 \right\} \quad \textit{where} \quad \gamma_3 > \theta \tag{8}$$

2.1.1 Automation

We first derive the equilibrium conditions for case I, Automation. The first-order conditions (henceforth FOCs) are for each j and all skill types i and capital, respectively:

$$\begin{array}{ll} (m_{1j}): & A_j\gamma_1 m_{1j}^{\gamma_1-1}x_1 = w_{1j}, \quad \forall j \in J; \\ (m_{2j}): & A_j\frac{\gamma_2}{\theta}\left(m_{2j}^{\theta}x_2 + k_j^{\theta}x_k\right)^{\frac{\gamma}{\theta}-1}\theta m_{2j}^{\theta-1}x_2 = w_{2j}, \quad \forall j \in J; \\ (k_j): & A_j\frac{\gamma_2}{\theta}\left(m_{2j}^{\theta}x_2 + k_j^{\theta}x_k\right)^{\frac{\gamma_2}{\theta}-1}\theta k_j^{\theta-1}x_k = r, \quad \forall j \in J; \\ (m_{3j}): & A_j\gamma_3 m_{3j}^{\gamma_3-1}x_3 = w_{3j}, \quad \forall j \in J; \end{array} \tag{9}$$

Using labor mobility, we can write the wage ratio in terms of the house price ratio for all i, $\frac{w_{i2}}{w_{i1}} = \left(\frac{p_2}{p_1}\right)^{\alpha}$ and equate the first-order condition in both cities for a given skill. If we then compare the results for low- and high skill workers and use both the utility equalization condition, due to labor mobility, and the housing market clearing conditions for cities 1 and 2 we have:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1-1}} M_1}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1-1}}\right\}} \quad \text{and} \quad m_{31} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_3-1}} M_3}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_3-1}}\right\}} \tag{10}$$

and likewise for city 2. Finally, using the FOCs for skill 2 and capital, jointly with utility equalization and labor market condition for skill 2 in city 1, we have:

$$m_{21}=\frac{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}}\frac{k_1}{k_2}}{\left[1+\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}}\frac{k_1}{k_2}\right]}M_2 \quad \text{and} \quad k_2=\frac{M_2x_2^{\frac{1}{\theta}}-\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}} \tag{11}$$

and likewise for city 2.

So far we have consumer optimization for consumption and housing, the location choice by the worker, and firm optimization given wages. The next step is to allow for market clearing in the housing market given land prices. The system is static and solved simultaneously, which is reported in the appendix section A. In what follows, we assume $H_j = H$ for all cities j. Below, we will discuss the implications where this simplifying assumption has bite.

The Main Theoretical Results. First we establish the relationship between TFP and house prices. When cities have the same amount of land, we can establish the following result.

Proposition 1 (Automation, TFP, and Housing Prices) *Assume* $\gamma_2 < \theta$. $A_i > A_j \Rightarrow p_i > p_j, \forall j \in \{1, 2\}$

Consequently, the city with the highest TFP is also the one with the highest housing prices. We establish this result for cities with an identical supply of land. Clearly, the supply of land is important in our model since in a city with an extremely small geographical area, labor demand would drive up housing prices all else equal. This may therefore make it more expensive to live in even if the productivity is lower. Because in our empirical application we consider large metropolitan areas (NY city for example includes large parts of New Jersey and Connecticut), we believe that this assumption does not lead to much loss of generality.[9]

We now focus on the demand for capital and TFP. As proposition 2 shows, the city with higher TFP also demands more capital. The intuition is straightforward. In cities with higher TFP, housing prices are higher and workers must be compensated in order to afford living in a more expensive place. Furthermore, since firms with higher TFP hire more of all skill levels, the decreasing marginal returns are also more strong, pushing towards the increase in the use of capital in order to replace middle skills in this case. Hence, high-TFP cities demand more capital.

[9] In fact, the equal supply of housing condition is only sufficient for the proof, but not necessary. However, our model does not address the important issue of within-city geographical heterogeneity, as analyzed for example in Lucas and Rossi-Hansberg (2002). In our application, all heterogeneity is absorbed in the pricing index by means of the hedonic regression.

Proposition 2 (Automation, TFP and capital demand) *Assume* $\gamma_2 < \theta$. $A_i > A_j \Rightarrow k_i > k_j$.

Then, in theorem 1 we show that the city with the high TFP is also larger. In fact, we are able to show that, in equilibrium, the high-TFP city has more workers at all skill levels.

Theorem 1 (Automation and City Size) *Assume* $\gamma_2 < \theta$ *and* $A_1 > A_2$. *We have that* $S_1 > S_2$.

Finally, theorem 2 shows that, in the case in which $\gamma_i \equiv \gamma$ for all skills and $\gamma < \theta$, high-TFP city has proportionately more of high and low skill workers than low-TFP cities. This is true even though high-TFP cities have more of all types. Consequently, the high-TFP city is more unequal in terms of its skill distribution.

Theorem 2 (Automation and Spatial Sorting) *Assume* $\gamma_i \equiv \gamma,\ \forall i \in \{1, 2, 3\}$ *and* $\gamma < \theta$. *If* $A_1 > A_2$ *we have that city 1 has thick tails in the skill distribution.*

In appendix section C, we simulate the automation model to get a better understanding of the model's mechanisms. We focus on parameter changes related to the observed evolution of computer power and prices over the last twenty years. We also take into account changes in the employed labor force's skill distribution. Our counterfactual exercises show that, while the changes in the skill distribution may be responsible for the bulk of the change in the overall shape of the distributions between 1995 and 2015, changes in technology's cost and productivity are the leading factors explaining why big cities are increasingly more unequal when compared to smaller ones. Finally, the exercise highlights that automation by itself is unlikely to explain the increase in compensation growth observed by high-skill workers. In other words, SBTC is likely needed in order to boost high-skill workers income growth.

2.1.2 Skill Biased Technological Change

We now consider the case of *Skill-Bias Technological Change* (henceforth SBTC) in which capital and high-skill workers are complements. In this case, the FOCs for each city j, skill type i, and capital, respectively are:

$$\begin{aligned}
&(m_{1j}) : A_j \gamma_1 m_{1j}^{\gamma_1 - 1} x_1 = w_{1j} \\
&(m_{2j}) : A_j \gamma_2 m_{2j}^{\gamma_2 - 1} x_2 = w_{2j} \\
&(m_{3j}) : A_j \gamma_3 \left(m_{3j}^{\theta} x_3 + k_j^{\theta} x_k \right)^{\frac{\gamma_3}{\theta} - 1} m_{3j}^{\theta - 1} x_3 = w_{3j} \\
&(k_j) : A_j \gamma_3 \left(m_{3j}^{\theta} x_3 + k_j^{\theta} x_k \right)^{\frac{\gamma_3}{\theta} - 1} k_j^{\theta - 1} x_k = r
\end{aligned} \qquad (12)$$

Using labor mobility, we can write the wage ratio in terms of the house price ratio for all i, $\frac{w_{i2}}{w_{i1}} = \left(\frac{p_2}{p_1}\right)^{\alpha}$ and equate the first-order condition in both cities for a given skill. If we then compare the results for low- and middle-skill workers and use both the utility equalization condition, due to labor mobility, and the housing market clearing conditions for cities 1 and 2 we have:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1 - 1}} M_1}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1 - 1}}\right\}} \quad \text{and} \quad m_{21} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_2 - 1}} M_2}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_2 - 1}}\right\}} \tag{13}$$

and likewise for city 2. Finally, using the FOCs for skill 3 and capital, jointly with utility equalization and labor market condition for skill 2 in city 1, we have:

$$m_{31} = \frac{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}} \frac{k_1}{k_2}}{\left[1 + \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}} \frac{k_1}{k_2}\right]} M_3 \quad \text{and} \quad k_2 = \frac{M_3 x_3^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1 \gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3 - \theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1 \gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3 - \theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{14}$$

and likewise for city 2.

So far we have consumer optimization for consumption and housing, the location choice by the worker, and firm optimization given wages. The next step is to allow for market clearing in the housing market given land prices. The system is static and solved simultaneously, which is reported in appendix section A. In what follows, we assume $H_j = H$ for all cities j. Below, we will discuss the implications where this simplifying assumption has bite.

The Main Theoretical Results. First we establish the relationship between TFP and house prices. When cities have the same amount of land, we can establish the following result.

Proposition 3 (SBTC, TFP, and Housing Prices) *Assume* $\gamma_3 > \theta$. $A_i > A_j \Rightarrow p_i > p_j$, $\forall j \in \{1, 2\}$

Consequently, the city with the highest TFP is also the one with the highest housing prices. We establish this result for cities with an identical supply of land. Clearly, the supply of land is important in our model since in a city with an extremely small geographical area, labor demand would drive up housing prices all else equal. This may therefore make it more expensive to live in even if the productivity is lower. Because in our empirical application we consider large metropolitan areas (NY city for example includes large parts of New Jersey and Connecticut), we believe that this assumption does not lead to much loss of generality.

We now focus on the demand for capital and TFP. As proposition 4 shows, the city with higher TFP also demands more capital.

Proposition 4 (SBTC, TFP, and capital demand) *Assume* $\gamma_3 > \theta$. $A_i > A_j \Rightarrow k_i > k_j$.

Corollary 1 shows that the high TFP city also attracts more high-skill workers.

Corollary 1 (SBTC and demand for high skill) *Assume* $\gamma_3 > \theta$. $A_i > A_j \Rightarrow m_{3i} > m_{3j}$.

Finally, theorem 3 shows that in the case in which $\gamma_i \equiv \gamma$ for all skills and $\gamma > \theta$, high-TFP city attracts proportionately more skilled workers. In particular, we show that the skill distribution in the high-TFP city stochastically dominates in first order the skill distribution in the low-TFP city.

Theorem 3 *Assume* $\gamma_i \equiv \gamma$, $\forall i \in \{1, 2, 3\}$ *and* $\gamma > \theta$. *If* $A_1 > A_2$, *we have that city 1's skill distribution F.O.S.D. city 2's skill distribution.*

Differently from the case of Automation, **SBTC does not imply that the high-TFP city is larger**. In appendix section D, we present two examples that illustrate that results can go either way, i.e., depending on the parameters we may have the high-TFP city to be either larger or smaller than the low-TFP city.

3 Data Sources and Measurement

Data on Workers

Our main data source is the Census Public Use Microdata. We use the 5% Samples for 1980, 1990, and 2000 and for 2013-2015 we combine the American Community Survey yearly files. From these files, we construct labor force and price information at the Metropolitan Statistical Area (MSA) level. The definition of a MSA we use is the 2000 Combined Metropolitan Statistical Areas (CMSA) by the Census for all MSAs that are part of an CMSA or otherwise the MSA itself. For simplicity, we will refer to this definition as MSA from now on. We follow the same procedure as Baum-Snow and Pavan (2013) in order to match the Census Public Use Microdata Area (PUMA) of each Census sample to the 2000 Census Metropolitan Area definitions. The Census data restricts us to consider only MSAs which are sufficiently large, as they are otherwise not identifiable due to the minimal size of a PUMA. For each year we then construct information on the labor force in each MSA and the local price level. We focus our attention to full-time, full-year workers aged 25-54. In order to obtain an estimate of the price level at the MSA level, we consider a simple price index including both consumption goods – which sell at the same price across different locations – and housing, which is priced differently in each MSA. Based on a hedonic regression using rental data and building characteristics, we calculate the difference in housing values across cities. In large parts of our empirical analysis we focus on the occupational composition of MSAs. To do so, we aggregate the census occupations into broad groups based on their task content as in Cortes et al. (2014). Table 1 shows the classification into groups by task components and the corresponding titles of occupation groups in the Census 2010 Occupation Classification system[10].

Table 2 presents sample averages and standard deviations in the subsample of MSAs for which we have data in all years in the Census and information on technology adoption. We present descriptive statistics for the main variables used in the analysis: occupation shares, employment levels, and our MSA rent index.

Table 1: Occupation Groups by Tasks

Tasks	Census Occupations
Non-routine Cognitive	Management
	Business and financial operations
	Computer, Engineering and Science
	Education, Legal, Community Service, Arts and Media Occupations
	Healthcare Practitioners and Technical Occupations
Non-routine Manual	Service Occupations
Routine Cognitive	Sales and Related
	Office and Administrative Support
Routine Manual	Construction and Extraction
	Installation, Maintenance and Repair
	Production
	Transportation and Material Moving

Data on technology

Our technology data comes from the Ci Technology Database, produced by the Aberdeen Group (formerly known as Harte-Hanks). The data has detailed hardware and software information for over 200,000 sites in 2015[11], including not only installed capacity but also expected future expenses in technology. Their data also includes detailed geographical location for the

[10] See https://www.census.gov/people/io/files/2010_OccCodeswithCrosswalkfrom2002-2011nov04.xls for the detailed list of Census 2010 Occupations and Cortes et al. (2014) for the mapping to previous Census Occupation Classifications

[11] In fact, the overall sample is significantly larger than 200,000, but we are restricting the sample to the plants and sites to which we have detailed software information.

Table 2: Descriptive Statistics

	1980 mean (st. dev.)	2015 mean (st. dev.)
MSA's Occupation Shares		
Non-Routine Cognitive	34.6% (3.95)	45.3% (5.46)
Non-Routine Manual	9.9% (2.43)	14.8% (2.38)
Routine Cognitive	29.8% (2.12)	22.9% (1.96)
Routine Manual	25.3% (4.71)	16.7% (3.08)
MSA's Rent and Size		
log rent index	0.01 (0.13)	0.01 (0.23)
Employment in 000s	861.61 (1049.25)	1535.77 (1678.15)
No. of MSAs	261	261

Note: Averages and standard deviations are weighted by MSA employment. Subsample of MSAs for which we have complete data in all years.

interviewed sites, as well as aggregation to the firm level. Finally, they also collect some basic information about the sites, such as detailed industry code, number of employees, and total revenue. We have available information for the years 1990, 1996, and 2000-2015. Our current analysis focuses on the information from 2015 not only due to a larger sample size, but also due to more detailed information on IT budget and software installation.

We consider several measures of investment in technology. Initially, we consider a broad measure of investment in technology: the total IT budget per worker. While this measure may overstate the investment in technology made to either boost the productivity or replace a given set of workers, it has several advantages. First, this measure is available for all the establishments in our sample. Second, the portion of our database that includes IT budget information covers a significant fraction of the employed labor force as well as establishments, once compared to other standard databases.[12] In particular, table 3 shows that, compared to the National Establishment Time-Series (NETS), our sample covers on average 53% of the MSA's employed

[12] National Establishment Time-Series (NETS) and the County Business Pattern (CBP), for example.

labor force. Moreover, while table 3 shows that our sample covers on average only 13% of the MSA's establishments, table 4 shows that this is mostly due to a low coverage of establishments with 1 to 4 employees. In fact, the coverage is on average above 60% for establishments with 20 employees or more. In terms of industry coverage, while our sample is more heavily concentrated in manufacturing, all but two sectors have average coverage in the MSA above 30% (see table 5).[13] Third, it is an easily interpretable continuous variable, i.e., it does not suffer from potential biases or judgment calls in the variable construction. Fourth, IT budget per worker is highly correlated to several different categories of investment in technology. In particular, in 2015, in our sample of more than 170,000 establishments, the correlation between IT budget per worker and hardware budget per employee, software budget per employee, and PC budget per employee is always above 0.95. Consequently, overall IT budget per worker gives us a good summary statistic for the variation in technology adoption observed across both establishments and MSAs.

Alternatively, we may focus on measures that target the degree of complementarity or substitutability between a group of occupations and technology. In particular, we focus on the adoption of Enterprise Resource Planning software (ERP) in order to measure the establishments intent in automate routine cognitive tasks. As pointed out by Bloom et al. (2014), ERP software systems integrate several data sources and processes of an organization into a unified system, reducing the need for clerical and low-level white collar workers. We consider ERPs that help managing the following areas: Accounting, Human Resources, Customer & Sales Force, Collaborative and Integration, Supply Chain Management, as well as bundle software like the ones produced by SAP, which are usually called Enterprise Applications.

There are benefits and drawbacks in using ERP measures. The main benefit is that ERP is a clear measure of a establishment's intent in automating. In this sense, ERP softwares are quite distinct from aggregate measures such as IT budget and other general purpose technologies, such as the adoption of personal computers. The key drawbacks are twofold. First, there is a significant reduction in establishment coverage. As shown in table 3, our information on ERP adoption covers on average only 16% of workers and 1% of establishments in the MSA, compared to NETS. Moreover, even after controlling for establishment size, MSA average coverage is above 30% only for establishments that have 250 employees or more. Second, we need to focus on coarser measures of technology adoption. Our leading measure of ERP adoption is the fraction of establishments in the MSAs that adopted ERP softwares. This measure, while being easy to calculate and robust to outliers, does not capture the intensive margin of ERP adoption. For example, consider two establishments, A and B, that adopt ERP softwares at different degrees.

[13] Our results are also robust to sub-samples focused on private establishments. Consequently, the inclusion of state-run or governmental departments in our sample do not drive our results.

Establishment A adopts a relatively simple accounting software that may replace the work of a few accounting assistants. Differently, establishment B adopts an integrated ERP software system that allows it to automate several processes within the firm – sales, HR, inventory, accounting, etc. Both establishments would be classified as "adopters" and contribute the same for our leading measure. Consequently, our leading measure will be biased towards finding no effect.

Due to the significant drawbacks of the ERP measure, we focus our analysis on the IT budget per worker in section 4. However, we present the results for ERP measures in appendix section E. While results are understandably weaker for ERP – due to smaller sample size – they are still qualitatively similar to the ones presented in section 4.

Finally, in terms of geographical coverage and summary statistics, figure 1a shows the geographical dispersion of IT budget per worker across the country in 2015. First of all, corroborating the results presented in table 3, notice that the geographical coverage is quite good, with only very few MSAs missing. In fact, the missing MSAs are due to the matching procedure of the Census PUMA to the 2000 Census Metropolitan Area definitions as described by Baum-Snow and Pavan (2013).

Table 6 presents the summary statistics for IT budget per worker across MSAs. First of all, notice that there is a difference in the definition of the unit of count between the first row and rows 2-4 in table 6. In the first row, we calculate the MSA's IT budget per worker by dividing the sum of the total IT budget of all establishments in the MSA by the sum of these establishments labor force. In this sense, we obtain an average IT budget per worker that puts more weight on larger establishments. Differently, for the summary statistics presented in rows 2-4, we first calculate the IT budget per worker for each establishment and then look at the average, median, and standard deviation of IT budget per worker across establishments within a given MSA. Consequently, rows 2-4 have an establishment as the unit of measure, reducing the weight of larger establishments in the overall count. In this sense, rows 2-4 allows us to evaluate within- and between-MSA IT budget per worker dispersion across establishments. While our analysis focuses on the definition of MSA's IT budget per worker presented in table 6's row 1, rows 2-4 show that there is significant within-MSA variation of IT budget per worker across establishments. Moreover, our empirical results are robust to the different ways to calculate the IT budget per worker presented in table 6. As we can see in row 1 of table 6, there is significant variation in IT budget per worker across MSAs.

In our estimates at the metropolitan area level, we follow the literature in their data adjustments. In particular, we control for the distribution of establishment sizes across cities following

Table 3: Coverage Ci Aberdeen relative to NETS

	Mean	**S.D.**	**p10**	**p25**	**p50**	**p75**	**p90**	**N**
IT Budget Sample								
Fraction Emp. in Ci	53%	9%	44%	50%	55%	58%	61%	272
Fraction Est. in Ci	13%	3%	9%	11%	13%	15%	15%	272
Fraction Sales in Ci	54%	9%	45%	51%	55%	59%	63%	272
ERP Sample								
Fraction Emp. in Ci	16%	5%	10%	13%	15%	18%	21%	272
Fraction Est. in Ci	1%	0%	1%	1%	1%	1%	1%	272
Fraction Sales in Ci	17%	6%	10%	14%	17%	20%	24%	272

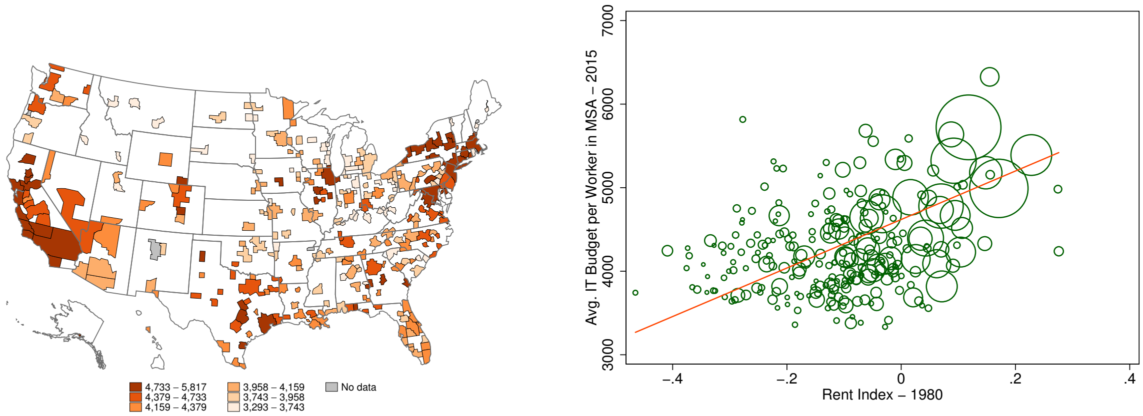
Table 4: Coverage Ci Aberdeen relative to NETS by Establishment Size

	Mean	**S.D.**	**p10**	**p25**	**p50**	**p75**	**p90**	**N**
IT Budget Sample								
1 to 4 Employees	3%	1%	2%	3%	3%	4%	4%	272
5 to 9 Employees	27%	4%	22%	25%	27%	29%	31%	272
10 to 19 Employees	56%	7%	50%	53%	57%	59%	61%	272
20 to 49 Employees	61%	7%	57%	59%	62%	65%	67%	272
50 to 99 Employees	68%	8%	62%	65%	68%	72%	74%	272
100 to 249 Employees	69%	9%	62%	66%	70%	73%	76%	272
250 to 499 Employees	78%	12%	67%	72%	77%	83%	90%	272
500 to 999 Employees	84%	27%	67%	75%	82%	90%	100%	272
1,000 or more Employees	84%	23%	58%	73%	83%	100%	110%	270

Table 5: Ci Coverage relative to NETS: Employment by Industry

	Mean	**S.D.**	**p10**	**p25**	**p50**	**p75**	**p90**	**N**
IT Budget Sample								
Manufacturing	70%	12%	59%	65%	72%	78%	82%	272
Construction	46%	8%	36%	41%	46%	51%	55%	272
Information	66%	13%	51%	60%	67%	74%	81%	272
Finance	47%	10%	37%	42%	47%	53%	59%	272
Professional & Bus Services	35%	10%	24%	30%	35%	41%	47%	272
Education and Health	68%	10%	60%	65%	70%	73%	76%	272
Leisure and Hospitality	21%	8%	13%	16%	20%	24%	29%	272
Public Adm	71%	11%	57%	68%	73%	77%	82%	272
Trade, Transp., and Util.	33%	7%	25%	29%	33%	37%	41%	272
Mining	55%	24%	15%	43%	60%	72%	81%	271
Other Services	28%	7%	20%	24%	28%	31%	36%	272

a similar approach to Doms and Lewis (2006), in which we net the technology adoption variable



(a) Avg. IT Budget per worker

(b) Avg. IT per worker vs local price level

Figure 1: Geographical distribution of IT across CMSAs – 2015

Table 6: Descriptive statistics of technology adoption across MSAs – 2015

	Mean	Median	S.D.	Min	Max	N
IT Budget						
MSA's IT Budget/Emp.	4,919	4,381	2,436	2,710	33,905	272
Avg. IT Budget/Emp. by site	4,238	4,159	515	3,293	5,817	272
Median IT Budget/Emp. by site	2,888	2,860	342	2,062	3,750	272
St. Dev. IT Budget/Emp. by site	8,865	4,917	11,453	3,123	97,557	272

from the effects of industry and establishment size. Let $\gamma_{i,c,t}$ be the technology for establishment i in city c and time t. We estimate, using OLS, the following model:

$$\gamma_{i,c,t} = \sum_t \left[\beta_{I,t} Ind_{i,t} \times Size_{i,t} + \beta_{C,t} City_{i,t} + \beta_{Y,t} Year_{i,t}\right] + \varepsilon_{i,t} \tag{15}$$

where *Ind*, *Size*, and *City* are vectors of dummy variables of industry (3 digit SIC) of the establishment, size of the establishment (8 employment size classes, following CBP[14]). In this case, $\beta_{C,t}$ is the key measure, capturing the differences in technology use across cities, after controlling for over 950 industry/size interactions.

Following Doms and Lewis (2006), we also drop observations where the Aberdeen coverage is particularly slim, such as retail, farming, and mining. We also exclude establishments in the IT producing sector.

[14]Doms and Lewis (2006) are not clear about which categories they are. However, since they weight their regression based on the CBP and limit their sample to establishments with 5 employees or more, the class sizes are likely: 5 to 9 employees,10 to 19 employees, 20 to 49 employees, 50 to 99 employees, 100 to 249 employees, 250 to 499 employees, 500 to 999 employees, and more than 1000 employees.

Data on Metropolitan Areas' Characteristics

In order to control for metropolitan area characteristics, we gather information on housing supply elasticity, natural amenities, and industry composition in the MSA.

In terms of housing Supply Elasticity, we consider 3 possible measures. Our key measure is based on Saiz (2010)'s housing supply elasticity. This measure takes into account both land use restrictions as well as geographical restrictions in building in different areas. In robustness checks presented in an online appendix, we consider two additional measures. The Wharton Residential Land Use Regulation Index (WRLURI), based on work by Gyourko et al. (2008), which takes into account building regulations. Finally, we consider Ganong and Shoag (2017)'s Land regulation index, which is based on the number of state supreme and appellate court cases containing the phrase "land use" over time. Each one of this measures has its pros and cons. The benefit of the Ganong and Shoag (2017)'s measure is its time-series nature. In particular, we can control for proxies for land-use restrictions in 1980, which is exogenous to changes that happened in the period analyzed. On the other hand, this is a coarse measure, being available only at the state level. Differently, the Saiz (2010)'s housing elasticity supply is a much richer control, including detailed information on both regulation and geographical constraints at the PMSA level. Unfortunately, this measure is only available for 2007.

We consider two sources of natural amenities. Our main measures of natural amenities come from the United States Department of Agriculture (USDA). In particular, we focus on the following measures: Mean Temperature for January (1941-1970); Mean Temperature for July (1941-1970); Mean hours of Sunlight for January (1941-1970); ln(% of Water Area); Mean relative humidity for July (1941-1970). In an online appendix, we measure of natural amenities from the National Climatic Data Center (2008), computed by Albouy (2012). In particular, we focus on the following measures: Heating and cooling degree days (annual); average sunshine as a percentage of possible; average slope of the land in the metropolitan area; average distance to the closest coastline.[15]

In order to control for the metropolitan areas industry composition, we follow Beaudry et al. (2010) and include controls that reflect a city's employment mix across 12 industry groups in 1980. In particular, we control for the share of employment in industry categories that correspond roughly to one-digit SICs (public sector is the excluded category): Agriculture and Mining; Construction; Non-durable Manufacturing; Durable Manufacturing; Transportation and Utilities; Wholesale; Retails; Finance, Insurance, and Real Estate; Business and Repair Services;

[15] In unreported regressions, we also controlled for Albouy (2012)'s Adjusted Quality of Life Index. Similarly, we control for USDA's Natural Amenities Scale. However, neither of these aggregate indexes is statistically significant, differently from the direct measures of natural amenities.

Other low-skill Services; Entertainment; Professional Services. To calculate this share, we gather information on employment across industry sectors within MSAs using the 1980 County Business Pattern (CBP).

4 Empirical Evidence

In this section, we describe our evidence regarding the adoption of automation technology and the occupational composition of cities. We focus on the two main predictions of the theory: (1) locations with higher housing costs should implement automation technology at higher rates and (2) locations with higher housing costs should also see decreasing shares of their workforce being employed in middle-skill occupations, whose tasks are being replaced by automation technology. As discussed in section 3, we focus on IT budget per worker as our key variable on technology investment. In appendix section E, we present the results using Enterprise Resource Planning (ERP) software adoption by the establishment as the technology adoption indicator. Results are qualitatively similar in both cases.

Table 7 shows the results for MSA-level linear regression models of the log of the average IT budget per worker, adjusted for plant employment interacted with three-digit SIC industry, following Beaudry et al. (2010) and Doms and Lewis (2006). In this table, we present initial evidence that supports our theoretical results, even after taking into account alternative explanations put forward by previous work. In particular, the importance of the presence of a local supply of skilled workers (Doms and Lewis (2006) and Bessen (2002)) as well as the presence of a large share of automatable routine jobs in the area, jointly with a decline in the costs of automation and consumer preferences over varieties and services (Autor and Dorn (2013)). Our goal is to show that the mechanisms presented in our model are still key drivers for technology investment, even after controlling for alternative hypothesis.

Table 7 shows that MSA's rental price index in 1980 helps to explain the variation in IT budget per worker across MSAs, even after controlling for the presence of natural amenities, housing supply elasticity, and industry composition.[16] In specification (1), a one standard deviation increase in local price index (an increase of 13% in the local price index) is associated with an increase of $90.00 in the MSA's average IT budget per worker. This magnitude corresponds to an increase of 3% in the average IT budget per worker. Specification (2) finds no statistically

[16] As pointed out by Beaudry et al. (2010), in this case the industry mix controls are on top of the detailed industry adjustment already preformed on the dependent variable (three-digit SIC × establishment size). The industry mix controls therefore capture any additional indirect or "spillover" effects of industry mix in the IT regressions.

significant correlation between the area's ratio of college equivalents to non-college equivalents and the average IT budget per worker in 2015. Similarly, specification (3) finds no statistically significant correlation between the MSA's share of routine-cognitive jobs in 1980 and the average IT budget per worker in 2015. Specification (4) includes all controls presented in specifications (1)-(3) together, with only a marginal decline in the magnitude of the impact of local rent prices. Finally, specification (5) controls for the MSA's average degree of offshorability of the local jobs in 1980 – using the task offshorability index presented by Autor and Dorn (2013). We find that the impact of local housing prices declines somewhat in magnitude. A one standard deviation increase in local price index (an increase of 13% in the local price index) is associated with an increase of $75.09 in the MSA's average IT budget per worker. This magnitude corresponds to an increase of 2.53% in the average IT budget per worker.

Table 7: IT budget per worker – 2015

	(1)	(2)	(3)	(4)	(5)
	log(IT)	log(IT)	log(IT)	log(IT)	log(IT)
MSA log rent index 1980	0.2446***			0.2392**	0.2099**
	(0.0936)			(0.0982)	(0.0989)
MSA routine cognitive share 1980		-0.6043		-0.6401	-0.9040**
		(0.4438)		(0.4184)	(0.4217)
MSA's log $\left(\frac{S}{U}\right)$ in 1980			0.0295	0.0072	-0.0048
			(0.0354)	(0.0349)	(0.0332)
MSA Offshorability 1980					0.1956
					(0.1261)
Housing supply elasticity	-0.0032	-0.0050	-0.0088	-0.0001	-0.0000
	(0.0065)	(0.0066)	(0.0067)	(0.0064)	(0.0065)
USDA's Amenities	Yes	Yes	Yes	Yes	Yes
MSA's Industry Mix Controls	Yes	Yes	Yes	Yes	Yes
F statistic	20.82	19.50	17.73	22.01	20.28
Adj. R^2	0.544	0.527	0.523	0.546	0.550
MSAs	222	222	222	222	222

Standard errors in parentheses. The dependent variable in all columns is the logarithm of the average IT budget per employee in the metro area, adjusted for plant employment interacted with three digit SIC industry. Each observation (a MSA) is weighted by its employment in 2015. Stars represent: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

However, results presented in table 7 may suffer from selection on unobservables. In particular, the types of firms that select themselves into more expensive MSAs may be significantly different from the ones that locate in less expensive places, biasing our results. In order to control for this effect, in table 8 we run establishment-level linear regression models of the establishment's IT budget per employee on MSA and establishment level variables. In particular, we include firm- and industry-fixed effects. As a result, our results on local price level highlight the within-firm variation across establishments in different locations.[17] Results presented in table 7, where we restrict our sample to establishments with at least 50 employees and we cluster our standard errors at the MSA level. These results highlight the importance of local prices on the establishment's IT budget per employee, even after controlling for firm and industry fixed effects. In fact, from specification (1), we observe that a one standard deviation increase in local price index (an increase of 13.4% in the local price index) is associated with an increase in the establishment's average IT budget per worker of about $59.40. This magnitude corresponds to an increase of 2.12% in the average IT budget per worker. While this effect seems small, we must keep in mind that we are already controlling for firm- and industry-fixed effects, as well as establishment's size and revenue and MSA's natural amenities and industry mix. Moreover, notice that the coefficient of local prices index on IT budget per worker does not vary significantly across the different specifications presented in table 8. Finally, the coefficients of the share of routine-cognitive workers in 1980, MSA's average degree of offshorability of the local jobs in 1980, and MSA's ratio of college equivalent workers are all statistically insignificant.

We now turn to the second prediction of the theory: High cost locations should feature a decline in the share of workers, whose tasks can be automated after the introduction of new technology. We use 1980 as the pre-technology period and compare to the occupational composition in 2015. Our focus on such a long span of time is motivated by the fact that we compare steady state predictions of the model and ignore short-term dynamics.

Table 9 presents the results of linear regressions of the change in the routine-cognitive share of MSAs between 1980 and 2015. Specification (1) indicates that a one standard deviation increase in local price index (an increase of 13.6% in the local price index) is associated with a 0.9 percentage point larger drop in the routine-cognitive share over 1980-2015. Thus, the most expensive places have about a 5.8 percentage point larger drop in the routine-cognitive share relative to the cheapest locations. This is a significant difference compared to the average routine-cognitive share of 23% in 2015. Specification (2) highlights the impact of the 1980's share of routine-cognitive workers. A one standard deviation increase in the 1980's share of routine-cognitive workers (an increase of 2.7 percentage points in the local share of routine-cognitive jobs) is associated with a 2 percentage point larger drop subsequently. Specification (3) shows that a one standard deviation increase in the share of "college equivalent" workers relative to non-"college equivalent" workers (representing an 11 percentage-point increase in this

[17] In our sample, 62.2% of firms are single establishment (36,845 firms). From the multi-establishment firms (22,409 firms), 54.5% (12,207 firms) have all their establishments in the same MSA (11,788 firms), while the remaining 45.5% (10,202 firms) have establishments distributed across MSAs with significant differences in local prices.

Table 8: IT Investment by Establishment - Firm and Industry FE

	(1) log(IT)	(2) log(IT)	(3) log(IT)	(4) log(IT)	(5) log(IT)
MSA log rent index 1980	0.167***	0.194***	0.156***	0.167***	0.186***
	(0.050)	(0.054)	(0.050)	(0.050)	(0.054)
MSA's log $\left(\frac{S}{U}\right)$ in 1980		-0.025			-0.033*
		(0.019)			(0.019)
MSA Offshorability 1980			0.056		0.079
			(0.071)		(0.079)
MSA routine cognitive share 1980				0.001	0.001
				(0.002)	(0.003)
log(Site's Size)	-0.064***	-0.064***	-0.064***	-0.064***	-0.064***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
log(Site's Revenue)	2.400***	2.400***	2.400***	2.400***	2.401***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Housing Elasticity	-0.001	-0.000	-0.001	-0.002	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Firm FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
MSA Controls	Yes	Yes	Yes	Yes	Yes
F statistic	1,219.12	1,161.07	1,246.97	1,224.84	1,183.50
Adj. R^2	0.7822	0.7823	0.7823	0.7823	0.7823
No. of Sites	187,642	187,642	187,642	187,642	187,642
No. of Firms	59,254	59,254	59,254	59,254	59,254

Standard errors in parentheses. The dependent variable in all columns is the logarithm of the average IT budget per employee in the establishment. Each observation (an establishment) is weighted by the probability weight from a match between the Aberdeen data and the 2015 County Business Pattern. Stars represent: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

share) is correlated with a 1.5 percentage point larger drop in the routine-cognitive share over 1980-2015. Specification (4), we combine all three regressors plus controls in a multivariate regression. All three variables are strongly related to the decline in the routine-cognitive share of workers, even after accounting for their covariation. However, the partial effect of each is smaller. The effect of a one standard deviation higher house price drops to 0.5 percentage point. Similarly, the effects of a one standard deviation higher initial routine share and ratio of "college equivalent" workers drop to 1.8 and 0.9 percentage points, respectively. Finally, in specification (5) we observe magnitudes and statistical significance to drop after we control for the average degree of offshorability of the jobs in the MSA. The effect of a higher local price index drops to about half of the observed effect in specification (1). Similarly, the effects of the share of routine-cognitive workers in 1980 and the ratio of "college equivalent workers drop by 30% and

Table 9: Change in routine-cognitive share, 1980-2015

	(1)	(2)	(3)	(4)	(5)
	Δrout-cog	Δrout-cog	Δrout-cog	Δrout-cog	Δrout-cog
MSA log rent index 1980	-0.0779***			-0.0434**	-0.0348*
	(0.0260)			(0.0196)	(0.0177)
1980's Share Routine-Cognitive 1980		-0.7691***		-0.6863***	-0.6091***
		(0.0802)		(0.0700)	(0.0788)
1980's log $\left(\frac{S}{U}\right)$			-0.0475***	-0.0303***	-0.0268***
			(0.0079)	(0.0069)	(0.0066)
1980's Offshorability Index					-0.0572**
					(0.0290)
Housing supply elasticity	-0.0086***	-0.0030*	-0.0061***	-0.0037**	-0.0037**
	(0.0019)	(0.0016)	(0.0018)	(0.0015)	(0.0015)
USDA's Amenities	Yes	Yes	Yes	Yes	Yes
MSA's Industry Mix Controls	Yes	Yes	Yes	Yes	Yes
F statistic	21.32	39.20	35.47	53.76	50.92
Adj. R^2	0.675	0.768	0.722	0.814	0.819
MSAs	222	222	222	222	222

Standard errors in parentheses. Each observation (an MSA) is weighted by its employment in 2015. Stars represent: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

50%, respectively. While our measure of offshorability only highlights the occupation's potential exposure to offshoring, it is not unlikely that both offshoring and automation have happened concomitantly during the 1980-2015 period. Overall, our results confirm the prediction that expensive locations have seen a larger decline in their share of routine-cognitive workers.

In appendix section F, we present evidence of increase in spatial dispersion based on measures of concentration of skills across regions. These measures allow us to test if we have observed an increase in the spatial dispersion of skills across MSAs in the last 30 years, while abstracting from issues of long-run trends in the composition of labor force. Consequently, we are able to focus on the correlation between the spatial dispersion of skills and MSAs characteristics – in particular size and cost of housing. We consider three simple measures: The location quotient that compares the skill distribution in the MSA against the overall skill distribution in the economy, the Ellison and Glaeser (1997) index of industry concentration, and an adjusted version of this index proposed by Oyer and Schaefer (2016).[18] These measures corroborate that there was an increase in the concentration of routine cognitive and routine manual jobs in small cities. Moreover, cognitive occupations have seen a (small) increase in concentration over time. Finally, while both small and large cities have seen a reduction in concentration over time, the

[18]The latter two indexes attempt to measure concentration by comparing it against a distribution that would be obtained by chance (the "dartboard approach").

reduction has been on average larger at large cities, i.e. large cities became relatively more unequal over time.

5 Estimation

In order to complement the descriptive evidence in the previous sections and to perform quantitative counterfactuals, we estimate an extended version of the model which we estimated by indirect inference (Gourieroux et al., 1993).

The extended model embeds a more realistic housing market by introducing Stone-Geary preferences and a finite supply elasticity of housing. Furthermore, the production function allows for generic substitution patterns between capital and labor across occupations. Finally, individuals are heterogeneous in their skill, which can differ across occupations, and in their preferences for locations. Workers choose location and occupation jointly, thus we allow locations biased towards a certain type of job to attract more workers skilled in that particular job.

In the following, we shortly introduce the model extensions and then discuss identification of the main model parameters. The identification arguments motivate the moment selection for the estimation protocol.

5.1 Extended Model Setup

We extend the model to capture the key features of housing, labor, and capital allocations in the data.

Cities $j \in \mathcal{J}$ are characterized by their production opportunities, housing supply, and amenities. Each city produces a single final output that is a combination of different occupations i. Each occupation produces output by combining labor in efficiency units m_{ij} with capital k_{ij}. The production function F has a nested CES structure given by

$$F(\mathbf{m}_j, \mathbf{k}_j, \mathbf{A}_j) = A_j \left\{ \sum_i \left[m_{ij}^{\gamma_i} A_{l,ij} + k_{ij}^{\gamma_i} A_{k,i} \right]^{\frac{\lambda}{\gamma_i}} \right\}^{\frac{1}{\lambda}}, \tag{16}$$

where m_{ij} are efficiency units of labor in occupation i in city j. A_j is general TFP. $A_{l,ij}$ is labor-enhancing productivity in occupation i in city j and capital enhancing productivity is $A_{k,i}$. Factor markets are competitive, thus both labor and capital are paid according to their marginal product. Each efficiency unit of labor costs $\tilde{w}_{i,j}$ and capital price is r. The output price is normalized to 1. Firms maximize profits and input as well as output markets are competitive, i.e. firms are price takers. Capital supply is fully elastic at price r, which is taken as given. The firms maximize profits, which implies input demand follows

$$\tilde{w}_{i,j} = A_j \left\{ \sum_i \left[m_{ij}^{\gamma_i} A_{l,ij} + k_{ij}^{\gamma_i} A_{k,i}\right]^{\frac{\lambda}{\gamma_i}} \right\}^{\frac{1}{\lambda}-1} \left[m_{ij}^{\gamma_i} A_{l,ij} + k_{ij}^{\gamma_i} A_{k,i}\right]^{\frac{\lambda}{\gamma_i}-1} m_{ij}^{\gamma_i - 1} A_{l,ij} \tag{17}$$

$$r = A_j \left\{ \sum_i \left[m_{ij}^{\gamma_i} A_{l,ij} + k_{ij}^{\gamma_i} A_{k,i}\right]^{\frac{\lambda}{\gamma_i}} \right\}^{\frac{1}{\lambda}-1} \left[m_{ij}^{\gamma_i} A_{l,ij} + k_{ij}^{\gamma_i} A_{k,i}\right]^{\frac{\lambda}{\gamma_i}-1} k_{ij}^{\gamma_i - 1} A_{k,i}. \tag{18}$$

The total supply of labor is measured in efficiency units, i.e. $m_{i,j} = \int s_i dP(s_i|j,i)$, where $P(s_i|j,i)$ denotes the distribution of skills of workers working in occupation i in city j. Be aware that $P(s_i|j,i)$ integrates to the number of workers in location j working in occupation i, not 1. Labor supply at the city level is determined endogenously and will be described in the following..

Workers are heterogeneous in their skills and preferences for locations. There is a given economy wide distribution of skills and preferences for locations. Workers consume the final good and housing, where housing must be consumed in the same city as the workplace. Preferences over consumption and housing follow

$$u(c,h) = c^{1-\alpha}(h - \underline{h})^{\alpha}, \tag{19}$$

where $\underline{h}$ represents a minimal housing requirement each worker consumes. As before, workers maximize utility subject to their budget constraint

$$c + p_j h \leqslant w_j^*, \tag{20}$$

where income w_j^* follows from workers' optimal occupation choice, which will be described next. Each worker is endowed with a set of skills for each occupation, summarized by the vector $\mathbf{s} = [s_1, \ldots, s_I]$. The skill vector represents how many efficiency units of labor a worker could deliver in each occupation. Income in an occupation follows $w_{i,j}(\mathbf{s}) = s_i \tilde{w}_{i,j}$. The economy-wide distribution of skills is given by $G(\mathbf{s})$. The indirect utility of a location for a worker with a given set of skills $\mathbf{s}$ is

$$V_j(\mathbf{s})\varepsilon_j = t_j \max_i a_i \frac{(w_{i,j}(\mathbf{s}) - p_j \underline{h})}{p_j^{\alpha}} \varepsilon_j, \tag{21}$$

where amenity a_i represents a common taste for a type of job. A worker chooses the occupation optimally, taking into account real income and the amenity value of the job. The optimal occupation choice can be expressed as a cutoff rule of the maximal skill in other occupations i' such that the worker would choose occupation i. The highest skill in occupation i' for which a worker still chooses occupation i is

$$\bar{s}_{i',j}(s_i) = \frac{a_i \tilde{w}_{i,j} s_i - p_j \underline{h}(a_i - a_{i'})}{a_{i'} \tilde{w}_{i',j}}. \tag{22}$$

That is for $s_{i'} \leqslant \bar{s}_{i',j}(s_i)$ a worker with skills $s_i, s_{i'}$ choose occupation i. An occupation is chosen if that occupation fulfills

$$\textit{Choose Occupation } i, \textit{ if } s_{i'} \leqslant \bar{s}_{i',j}(s_i) \quad \forall i' = 1, ..I; i' \neq i. \tag{23}$$

The general amenity t_j of location j is enjoyed by all workers and ε_j represents idiosyncratic tastes for different locations. The distribution of idiosyncratic tastes is i.i.d. across locations and individuals, following a Fréchet distribution with scale parameter τ. The location parameter is normalized to 1, as t_j determines the level of amenity values in each location. The share of workers choosing a location j, then follows

$$P(j|\mathbf{s}) = \frac{V_j(\mathbf{s})^\tau}{\sum_j V_j(\mathbf{s})^\tau}. \tag{24}$$

The skill distribution in each location is $P(\mathbf{s}|j) = \frac{P(j|\mathbf{s})G(\mathbf{s})}{P(j)}$ where $P(j) = \int \ldots \int P(j|\mathbf{s}) dG(\mathbf{s})$. Given the skill distribution in each location, we can directly construct the skill distribution within each occupation – denoted by $P(s_i|j, i)$ – by using the occupation decision rule (23).

The Housing Market is competitive. Housing supply follows the price-quantity schedule

$$p(H) = \phi_j H^{\epsilon_p}. \tag{25}$$

In equilibrium, housing supply H adjusts such that the housing amount demanded by workers equals the amount supplied. The revenue from the housing market is consumed by absentee landowners. Housing demand in a city is given by

$$H_j^D = \int (1-\alpha)\underline{h} + \alpha \frac{w_j^*(\mathbf{s})}{p_j} dP(\mathbf{s}|j). \tag{26}$$

Furthermore, note that a worker's share of spending on housing is linear in income and prices

$$\frac{hp}{w} = \alpha + (1-\alpha)\underline{h}\frac{p}{w} \tag{27}$$

Here w is shorthand notation for a worker's total income.

5.2 Identification

We will discuss identification of model parameters in general, the arguments will motivate the subsequent moment selection. To discuss identification, consider observing the wage distribution in each city and occupation, the housing price of each location, the spending share on housing and the value of capital used in each city and occupation.

1. *Skill Distribution:*

   For a general discussion of identification of skill distributions in Roy models see French and Taber (2011). We use a parametric approach and assume the skill distribution $G(\mathbf{s})$ has marginal distributions following $Beta(1, \beta_i)$ in each skill dimension over a given interval. Furthermore, skills are uncorrelated $G(\mathbf{s}) = \prod G_i(s_i)$. Note that we normalize the location of the skill distribution to the interval $[\underline{s}, \overline{s}]$, because the level of productivity is already determined in the production function. Given the parametric restriction, we can identify β_i from the dispersion of wages in each occupation. The model implies a distribution of skills by occupation in each city $P(s_i|j, i)$. The distribution of skills in a location is $P(\mathbf{s}|j) = \frac{P(j|\mathbf{s})G(\mathbf{s})}{P(j)}$. Thus given prices, it depends one to one on the economy-wide skill distribution. $P(\mathbf{s}|j)$ then directly translates into the skill distribution within chosen occupations $P(s_i|j, i) = \int \mathbf{1}(s_i > \max(\{\bar{s}_{i',j}\}_{i' \in I, i' \neq i}) dP(\mathbf{s}|j)$. Thus using properties of $P(s_i|j, i)$ we can infer properties about $G(\mathbf{s}) = \prod G_i(s_i)$. Here we use the standard deviation of wages to infer the right tail parameter of the Beta distributions.

2. *Productivity Parameters:*

   We normalize the productivity parameter of labor in occupation 1 $A_{l,1j} = 1$. Out of the productivity parameters we have to normalize one, as they all enter multiplicative. Thus, consider all productivity parameters as relative to occupation 1. From optimal labor demand (17) it follows that the level of wages in a city depends one to one on TFP A_j, given relative factor inputs. Thus A_j can be identified from average wages. Furthermore, optimal labor demand $m_{i,j}$ is locally monotone in its own productivity $A_{l,ij}$ given input prices and other factor demands. Thus $A_{l,ij}$ is identified, but note that we already use average wages to identify TFP. Therefore, we only have $J(I-1)$ wage equations left to identify parameters $A_{l,ij}$. This also means we cannot identify the elasticity of substitution between occupations, governed by λ, from just the cross-section. Capital productivity can be identified analogously. Given prices and capital usage, we can infer capital's level of productivity. Lastly, we identify the elasticity of substitution between capital and labor in an occupation from relative demand across locations with different relative prices. We assume capital prices and capital productivity are identical across locations. It follows that

$$\frac{r}{\tilde{w}_{ij}} = \frac{k_{ij}^{\gamma_i - 1} A_{k,i}}{m_{ij}^{\gamma_i - 1} A_{l,ij}}.$$

Thus γ_i is identified from capital demand relative to labor demand across locations with different labor costs.

3. *Utility Function Parameters:*
   The utility function parameters α and $\underline{h}$ are identified from the relation of spending shares on housing with house prices and wages:

$$\frac{hp}{w} = \alpha + (1-\alpha)\underline{h}\frac{p}{w}.$$

4. *Location Amenities:*
   A location's amenity t_j can be identified from the workers' optimal location choice problem (24), as each location's value is monotone and increasing in its amenity value (21). Thus, given a location's wages and housing prices, the share of workers that choose the location is directly determined by its amenity value relative to other locations ($t_1 = 1$). The Fréchet scale parameter τ is not identified and normalized to 2.

5. *Occupation Amenities a_i:*
   An occupation's amenity value a_i can be determined from worker's optimal occupation choice, since the share of workers choosing an occupation is increasing in the occupation's amenity value, as we see in equation (23).

6. *Housing Supply:*
   A cities housing supply schedule is determined by ϕ_j and the elasticity of house prices to quantity ϵ_p. The elasticity is normalized following the estimates by Saiz (2010). The local house price shifter ϕ_j can be identified from local house prices and total housing demand, which equals supply in equilibrium.

5.3 Moments and Parameter Estimates

We estimate the model by fitting a set of data moments. We calculate the model solution with 5 cities. We use the American Community Survey and Aberdeen Ci Technology Database, as in the previous sections, to construct data moments. We construct city level measures of employment and IT usage from the data. The IT capital variables by occupation are constructed as the product of average IT budget per employee and the share of employees exposed to softwares that are related to their occupations. See table 31 for the list of softwares and assignment. The moments we use are all constructed for the year 2015, where we pool years 2014-2016 from the American Community Survey and 2014-2015 for the Aberdeen Data to improve coverage. The moments in the data are calculated within three bins, where we categorize cities by total employment and split them into bins of approximately equal employment shares.

Table 10 shows moments that we target at the city level. Average wages are substantially higher in large cities compared to small cities. We also fit the share of workers employed in the different occupation categories. Routine manual jobs are particularly abundant in small cities, while routine cognitive jobs are represented in approximately equal shares across cities. The non-routine cognitive occupations on the other hand show a strong bias towards large cities. In large cities, not only wages are higher, but also house prices. Similarly, average spending on IT is substantially lower in small cities compared to large cities. Furthermore, the difference in IT spending is larger (in logs) than in terms of log size, which means that per employee spending is also rising with city size. This is in line with the previously shown empirical evidence.

Table 10: City Moments 2015 and Model Fit

	Smallest Cities		Small Cities		Middle Cities		Large Cities		Largest Cities	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
Average log wage	2.5	2.6	2.6	2.7	2.7	2.7	2.8	2.8	2.8	2.9
	(0.011)		(0.024)		(0.037)		0.061		0.066	
Share employed in										
routine manual	0.28	0.25	0.23	0.22	0.2	0.21	0.2	0.21	0.21	0.21
	(0.0063)		(0.013)		(0.0083)		(0.016)		(0.023)	
routine cognitive	0.22	0.22	0.23	0.23	0.23	0.22	0.21	0.2	0.22	0.21
	(0.0022)		(0.0027)		(0.0065)		(0.0093)		(0.0081)	
non-routine cognitive	0.38	0.4	0.43	0.42	0.46	0.44	0.49	0.46	0.46	0.44
	(0.0062)		(0.013)		(0.013)		(0.023)		(0.03)	
Log rent (index)	-0.88	-0.89	-0.76	-0.74	-0.65	-0.65	-0.51	-0.5	-0.33	-0.33
	(0.011)		(0.033)		(0.061)		(0.12)		(0.098)	
Log size relative to largest cities	-4.2	-4.3	-2.6	-2.6	-1.6	-1.6	-0.72	-0.7		
	(0.26)		(0.26)		(0.25)		(0.21)		()	
Log IT spending difference to largest cities										
routine cognitive	-4.6	-4.6	-2.8	-2.8	-1.7	-1.7	-0.76	-0.74		
	(0.31)		(0.31)		(0.29)		(0.26)			
non-routine cognitive	-4.5	-4.6	-2.7	-2.8	-1.7	-1.7	-0.79	-0.73		
	(0.31)		(0.3)		(0.29)		(0.25)			

Note: Data Moments calculated from American Community survey public use sample 2015-2017, accessed through IPUMS. IT capital data from Aberdeen Ci Technology database. Unit of observation is a CMSA as defined in section 3. Standard Errors in parentheses calculated by Bootstrap.

In table 11 we present the moments we target only at the occupation level. We include average log wages and the standard deviation of log wages by occupation. Furthermore, we target overall estimated spending on IT by occupation. Average wages vary as expected across occupation categories. The standard deviation of wages within occupation groups is substantially higher in cognitive occupations than in manual occupations. IT spending we estimate form the Aberdeen Ci Technology database and find similar spending amounts in both routine and non-routine cognitive occupations. IT spending in manual occupations is assumed to be zero.

Table 11: Occupation Moments 2015 and Model Fit

	Data	Model
Average log wage		
routine manual	2.5	2.5
	(8.2e-6)	
routine cognitive	2.6	2.5
	(1.4e-5)	
non-routine cognitive	3.1	3.0
	(9.8e-6)	
Standard Deviation of log wages		
non-routine manual	0.56	0.63
	(1.4e-5)	
routine manual	0.57	0.62
	(1.0e-5)	
routine cognitive	0.66	0.65
	(1.7e-5)	
non-routine cognitive	0.68	0.64
	(1.2e-5)	
IT Spending		
routine cognitive	0.9	0.89
	0.14	
non-routine cognitive	1.0	1.0
	0.15	

Note: Standard Errors in parentheses. Data Moments calculated from American Community survey public use sample 2015-2017, accessed through IPUMS. IT capital data from Aberdeen Ci Technology database. IT spending by occupation corresponds to the product of IT spending and the share of employees exposed to software specific to each occupation. See table 31 for the list of softwares.

The estimated model parameters at the city level are presented in table 12 and 13.

Table 12: City TFP A, amenities t and housing price shifter ϕ

	Smallest Cities	Small Cities	Middle Cities	Larger Cities	Largest Cities
TFP A_j	52.0	56.0	61.0	62.0	63.0
	(9.2)	(13.0)	(14.0)	(14.0)	(14.0)
Amenities t_j		1.0	1.0	0.97	0.91
		(0.017)	(0.018)	(0.018)	(0.021)
House Price Shifter ϕ_j	3.6	1.8	1.2	0.98	0.86
	(0.025)	(0.012)	(0.0096)	(0.0088)	(0.011)

Note: Standard Errors in parentheses.

Table 13: Labor productivity A_l

Occupation	Smallest Cities	Small Cities	Middle Cities	Large Cities	Largest Cities
RM	1.9	1.2	0.91	0.97	1.0
	(0.12)	(0.093)	(0.078)	(0.12)	(0.12)
RC	1.4	1.5	1.4	1.3	1.4
	(0.36)	(0.35)	(0.36)	(0.35)	(0.37)
NRC	1.4	1.4	1.5	1.5	1.5
	(0.31)	(0.36)	(0.37)	(0.38)	(0.38)

Note: Standard Errors in parentheses.

The capital productivity parameters $A_{k,i}$ and γ_i, which governs the substitutability between capital and labor in each occupation, are presented in table 14. Capital is relatively more substitutable with labor in routine-cognitive occupations, compared to non-routine cognitive occupations. This is in line with suggestions by the previous literature, and in line with the motivating empirical evidence we presented. This suggests also, that potentially both the automation of routine-cognitive occupations and a skill bias in technological change are import in explaining employment trends.

Table 14: Capital productivity A_k and γ

Occupation	Productivity A_k	γ
RC	0.77	0.29
	(0.25)	(0.065)
NRC	0.43	0.12
	(0.089)	(0.042)

Note: Standard Errors in parentheses.

Table 15: Occupation amenity a and Skill Distribution

Occupation	Amenity a_i	Skill Distribution parameter $B(1, \beta_i)$
NRM		15.0
		(0.046)
RM	1.6	9.4
	(0.023)	(0.019)
RC	1.7	15.0
	(0.025)	(0.045)
NRC	1.3	6.1
	(0.022)	(0.014)

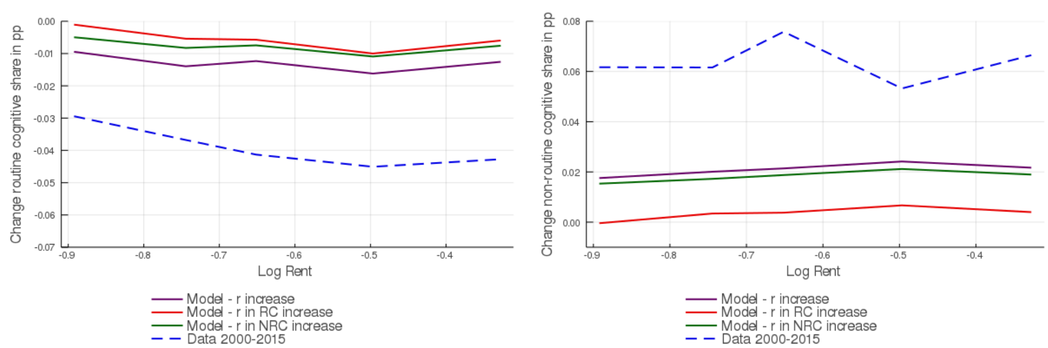
Note: Standard Errors in parentheses. Skill distribution $G(\mathbf{s})$ is independent across dimensions, thus $G(\mathbf{s})$ is the product of the CDF's of the estimated Beta distributions.

5.4 Results – Spatial Sorting

We consider an experiment where the quality adjusted price of IT capital increases from its 2016 level to approximately its value in 2000. The IT capital price is measured by the Software Price Index of the BLS. IT prices from 2000 to 2015 fell by just over 50%. With this experiment, we evaluate to what extent the fall in prices of IT can explain the change in sorting of jobs to cities over the last 2 decades in the United States.

Using the estimated parameters for 2015 we show how the model reacts to an increase in capital prices and compare the changes in the model to the evolution in the data from 2000-2015. We focus here on the allocation of jobs to cities, particularly the share of workers in cognitive occupations.

In figure 2a we show the model's change of employment shares π_i with respect to IT prices. And compare it to the change in the data between 2000 and 2015. We see that routine-cognitive jobs sort away from expensive locations over time, as shown by the dashed line. The model matches well the cross-sectional differences in the decline of routine-cognitive employment. However, the overall level of the decline is only about a third of that in the data. Comparing a fall in IT prices for capital used only in routine cognitive occupations with the overall fall, we find that the falling prices for IT in both occupations matter substantially for routine cognitive employment. Overall results suggest, that the spread of IT technology has substantially contributed to the fall in routine-cognitive employment.



(a) Change in routine cognitive employment share by city in response to a fall in IT prices by 60%.

(b) Change in non-routine cognitive employment share by city in response to a fall in IT prices by 60%.

The evolution of non-routine cognitive employment, shown in figure 2b, is almost a mirror image to that of routine-cognitive employment. However, the concentration of the rise in non-routine cognitive employment was not so much concentrated in only the largest cities, but also in mid-sized cities. The model response in employment in non-routine cognitive occupations is about one third of the changes in the data. Thus, based on the estimated parameters the declining prices of IT capital can explain a substantial share of the rising employment shares in non-routine cognitive employment. In terms of the sorting across cities, the model features larger rises in non-routine cognitive employment in more expensive, larger cities. Most of this effect is driven by the fall in prices for capital in non-routine cognitive occupations.

6 Conclusion

In this paper, we show that the substitution of routine jobs and tasks with machines, computers, and software has not happened evenly in space. In fact, the relative benefit of replacing middle-skill workers that perform routine tasks by computers and software depend on the cost of hiring a worker in this particular location. Consequently, living costs – in particular housing costs – play a key role. Our empirical results show that the share of routine-abstract jobs has gone down proportionately more in expensive and large cities. Moreover, these areas also have seen a larger investment in technologies directly associated with the tasks previously exercised by routine-abstract workers. In order to rationalize the observed empirical patterns, we propose an equilibrium model of location choice by heterogeneously skilled workers where each location is a small open economy in the market for computers and software. We show that if computers are substitutes to middle-skill workers – commonly known as the automation hypothesis – we have that in equilibrium large and expensive cities will invest more in automation, as they are more likely to substitute middle-skill workers with computers. Intuitively, in large and expensive cities, the relative benefit of substituting computers for routine cognitive workers is higher than in cheaper and smaller places, since computers have the same price everywhere, while workers must reside locally, having to be compensated for the high local housing prices.

A Theory - Preliminary Steps - Automation

A.1 Automation

Closing the Model

The final steps to close the model involve simplifying the model such that we have a system with only two equations and two unknowns (k_1 and $\frac{p_1}{p_2}$. Based on the calculations presented in the paper for k_2, k_1 and their respective FOCs, we obtain:

$$F_j(m_{1j}, m_{2j}, m_{3j}, k_j) = A_j \left[m_{1j}^{\gamma_1} x_1 + \left(m_{2j}^{\theta} x_2 + k_j^{\theta} x_k \right)^{\frac{\gamma_2}{\theta}} + m_{3j}^{\gamma_3} x_3 \right] \tag{28}$$

FOCs:

$$\begin{aligned}
&(m_{1j}) : A_j \gamma_1 m_{1j}^{\gamma_1 - 1} x_1 = w_{1j} \\
&(m_{2j}) : A_j \gamma_2 \left(m_{2j}^{\theta} x_2 + k_j^{\theta} x_k \right)^{\frac{\gamma_2}{\theta} - 1} m_{2j}^{\theta - 1} x_2 = w_{2j} \\
&(m_{3j}) : A_j \gamma_3 m_{3j}^{\gamma_3 - 1} x_3 = w_{3j} \\
&(k_j) : A_j \gamma_2 \left(m_{2j}^{\theta} x_2 + k_j^{\theta} x_k \right)^{\frac{\gamma_2}{\theta} - 1} k_j^{\theta - 1} x_k = r
\end{aligned}$$

Since from utility equalization, we have:

$$\frac{w_{ij}}{w_{ij'}} = \left(\frac{p_j}{p_{j'}} \right)^{\alpha}, \quad \forall i \in \{1, 2, 3\} \text{ and } \forall j \in \{1, 2\} \tag{29}$$

From (m_{11}), (m_{12}), and feasibility condition for skill 1, we have:

$$m_{11} = \frac{\left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_1 - 1}} M_1}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_1 - 1}}} \tag{30}$$

Similarly, for skill 3:

$$m_{31} = \frac{\left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_3 - 1}} M_3}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_3 - 1}}} \tag{31}$$

From (m_{21}), (k_1), (m_{22}), (k_2), labor market clearing, and the utility equalization condition, we have:

$$\left(\frac{m_{21}}{m_{22}} \right) = \left(\frac{p_1}{p_2} \right)^{\frac{\alpha}{\theta - 1}} \frac{k_1}{k_2} \tag{32}$$

Now let's go back to the expression for (k_1). Manipulating it, we have that:

$$m_{21} = \left\{ \frac{1}{x_2} \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1 - \gamma_2)}{\gamma_2 - \theta}} - x_k \right] \right\}^{\frac{1}{\theta}} k_1 \tag{33}$$

Similarly, for (k_2), we have:

$$m_{22} = \left\{ \frac{1}{x_2} \left[\left(\frac{r}{A_2 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_2^{\frac{\theta(1 - \gamma_2)}{\gamma_2 - \theta}} - x_k \right] \right\}^{\frac{1}{\theta}} k_2 \tag{34}$$

Dividing (33) by (34)and substituting (32), we have:

$$\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta-1}} = \left\{ \frac{\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]}{\left[\left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]} \right\} \tag{35}$$

Manipulating and simplifying it, we have:

$$k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} = \left(\frac{A_2}{A_1}\right)^{\frac{\theta}{\gamma_2-\theta}} \left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} + \left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\theta-\gamma_2}} \left[1 - \left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}}\right] x_k$$

Now, we also can use the fact that $m_{21} + m_{22} = M_2$. Then, we have that:

$$M_2 x_2^{\frac{1}{\theta}} = \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1 + \left[\left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_2 \tag{36}$$

Substituting (35) and manipulating, we have:

$$k_2 = \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{37}$$

Substituting (37) into (36) and manipulating, we have:

$$\begin{aligned} &\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} = \\ &= \left(\frac{A_2}{A_1}\right)^{\frac{\theta}{\gamma_2-\theta}} \left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} + \left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\theta-\gamma_2}} \left[1 - \left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}}\right] x_k \end{aligned} \tag{38}$$

which implicitly pins down k_1 as a function of $\frac{p_1}{p_2}$.

Finally, in order to pin down the equilibrium, we need to work with the housing market equilibrium conditions. Looking at the ratio of the housing market clearing conditions, we have:

$$\frac{w_{11}m_{11} + w_{21}m_{21} + w_{31}m_{31}}{w_{12}m_{12} + w_{22}m_{22} + w_{32}m_{32}} = \frac{p_1}{p_2}$$

Now substituting wages and labor demands and rearranging it, we have:

$$\begin{array}{c}\left\{\begin{array}{c}\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}m_{21}^{\theta}x_2-\\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{22}^{\theta}x_2+k_2^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}m_{22}^{\theta}x_2\end{array}\right\}=\\ \left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\ +\left(\frac{M_3}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3}x_3\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_3}{\gamma_3-1}}\right]\end{array}\right\}\end{array}\tag{39}$$

Then, from the ratio of (m_{21}) and (m_{22}), we have:

$$\left(m_{22}^{\theta}x_2+k_2^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}=\left(\frac{p_2}{p_1}\right)^{\alpha}\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}\times\left(\frac{m_{21}}{m_{22}}\right)^{\theta-1}\times\left(\frac{A_1}{A_2}\right)\tag{40}$$

Substituting (40) into (39) and rearranging, we have:

$$\begin{array}{c}\left\{\left[1-\left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_2-m_{21}}{m_{21}}\right]\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}m_{21}^{\theta}x_2\right\}=\\ \left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\ +\left(\frac{M_3}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3}x_3\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_3}{\gamma_3-1}}\right]\end{array}\right\}\end{array}\tag{41}$$

But then, from equation (33), we have that:

$$m_{21}^{\theta}x_2=\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\theta)}{\gamma_2-\theta}}-k_1^{\theta}x_k\tag{42}$$

Similarly, from (k_1), we have:

$$\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}=\left(\frac{r}{A_1\gamma_2x_k}\right)k_1^{1-\theta}\tag{43}$$

Then, from (42) and (43), we have:

$$\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}m_{21}^{\theta}x_2=\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\gamma_2}{\gamma_2-\theta}}k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2-\theta}}-\frac{r}{A_1\gamma_2}k_1\tag{44}$$

Substituting equation (37) into (32) and manipulating, we have:

$$\frac{M_2 - m_{21}}{m_{21}} = \frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}}{k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \tag{45}$$

Consequently:

$$\left[1 - \left(\frac{p_1}{p_2} \right)^{1-\alpha} \frac{M_2 - m_{21}}{m_{21}} \right] = \frac{\left\{ \begin{array}{c} \left(1 + \left(\frac{p_1}{p_2} \right)^{1-\alpha} \right) k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} \\ - \left(\frac{p_1}{p_2} \right)^{1-\alpha} M_2 x_2^{\frac{1}{\theta}} \end{array} \right\}}{k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \tag{46}$$

Then, from equations (44) and (46), we have that:

$$\begin{gathered} \left[1 - \left(\frac{p_1}{p_2} \right)^{1-\alpha} \frac{M_2 - m_{21}}{m_{21}} \right] \left(m_{21}^{\theta} x_2 + k_1^{\theta} x_k \right)^{\frac{\gamma_2 - \theta}{\theta}} m_{21}^{\theta} x_2 = \\ \frac{\left\{ \begin{array}{c} \left(1 + \left(\frac{p_1}{p_2} \right)^{1-\alpha} \right) k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} \\ - \left(\frac{p_1}{p_2} \right)^{1-\alpha} M_2 x_2^{\frac{1}{\theta}} \end{array} \right\}}{k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \times \left\{ \left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\gamma_2}{\gamma_2 - \theta}} k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2 - \theta}} - \frac{r}{A_1 \gamma_2} k_1 \right\} \end{gathered} \tag{47}$$

Notice that the LHS of equation (47) is the same of the one of equation (41). Substituting it back, we have:

$$\begin{gathered} \frac{\left\{ \begin{array}{c} \left(1 + \left(\frac{p_1}{p_2} \right)^{1-\alpha} \right) k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} \\ - \left(\frac{p_1}{p_2} \right)^{1-\alpha} M_2 x_2^{\frac{1}{\theta}} \end{array} \right\}}{k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \times \left\{ \left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\gamma_2}{\gamma_2 - \theta}} k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2 - \theta}} - \frac{r}{A_1 \gamma_2} k_1 \right\} = \\ \left\{ \begin{array}{c} \left(\frac{M_1}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_1 - 1}}} \right)^{\gamma_1} x_1 \left[\frac{A_2}{A_1} \frac{p_1}{p_2} - \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_1}{\gamma_1 - 1}} \right] \\ + \left(\frac{M_3}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_3 - 1}}} \right)^{\gamma_3} x_3 \left[\frac{A_2}{A_1} \frac{p_1}{p_2} - \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_3}{\gamma_3 - 1}} \right] \end{array} \right\} \end{gathered} \tag{48}$$

Finally, notice that equations (48) and (38) generate a system with two equations and two unknowns (k_1 and $\frac{p_1}{p_2}$):

$$\frac{\left\{ \left(1+\left(\frac{p_1}{p_2}\right)^{1-\alpha}\right) k_1 \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} - \left(\frac{p_1}{p_2}\right)^{1-\alpha} M_2 x_2^{\frac{1}{\theta}} \right\}}{k_1 \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} \times \left\{ \left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\gamma_2}{\gamma_2-\theta}} k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2-\theta}} - \frac{r}{A_1\gamma_2}k_1 \right\} = \left\{ \begin{array}{l} \left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1 \left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right] \\ + \left(\frac{M_3}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3} x_3 \left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_3}{\gamma_3-1}}\right] \end{array} \right\} \tag{F.1}$$

$$\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} = \tag{F.2}$$

$$= \left(\frac{A_2}{A_1}\right)^{\frac{\theta}{\gamma_2-\theta}} \left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} + \left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\theta-\gamma_2}} \left[1-\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}}\right] x_k$$

Preliminary Results

In this subsection, we present some preliminary results that will help us to show the main results presented in the paper.

Lemma A.1: *The distribution of skills across cities is identical if and only if* $\frac{m_{i1}}{m_{i2}} = constant, \forall i \in \{1,2,3\}$.

Proof: ($\Rightarrow$) Consider that the distribution across cities is constant, then $pdf_{i1} = pdf_{i2}, \forall i \in \{1,2,3\}$, i.e.:

$$\frac{m_{i1}}{m_{11}+m_{21}+m_{31}} = \frac{m_{i2}}{m_{12}+m_{22}+m_{32}} \tag{49}$$

But that means that $\frac{m_{i1}}{m_{i2}} = \eta = \frac{S_1}{S_2} = \frac{m_{11}+m_{21}+m_{31}}{m_{12}+m_{22}+m_{32}}$. The other direction is trivial. ■

Lemma A.2: *Assume* $\gamma_2 < \theta$. $p_1 = p_2$ *if and only if* $A_1 = A_2$.

Proof: Towards a contradiction, let's assume that $A_1 = A_2$ and $p_1 > p_2$. From the RHS of

$(F.1)$, we have:

$$\left\{\begin{array}{c} \left(\frac{M_1}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1\left[\frac{p_1}{p_2}-\left(\frac{p_1}{p_2}\right)^{\frac{\gamma_1\alpha}{\gamma_1-1}}\right] \\ +\left(\frac{M_3}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3} x_3\left[\frac{p_1}{p_2}-\left(\frac{p_1}{p_2}\right)^{\frac{\gamma_3\alpha}{\gamma_3-1}}\right] \end{array}\right\}>0$$

Since $p_1 > p_2$, $\gamma_1 < 1$, and $\gamma_3 < 1$. Therefore, the LHS of $(F.1)$ must also be positive in order for the equality to be satisfied. Then, from equation (44), we have:

$$\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}} m_{21}^{\theta}x_2=\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\gamma_2}{\gamma_2-\theta}} k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2-\theta}}-\frac{r}{A_1\gamma_2}k_1$$

So the second term on the LHS of $(F.1)$ must be positive. Moreover, from (43), we have that:

$$k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}=m_{21}x_2^{\frac{1}{\theta}}>0$$

Consequently, in order to satisfy $(F.1)$, we must have:

$$\frac{M_2x_2^{\frac{1}{\theta}}-k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}{\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}<k_1\left(\frac{p_1}{p_2}\right)^{\alpha-1}$$

Dividing both sides by $\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}$, we have:

$$\frac{M_2x_2^{\frac{1}{\theta}}-k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}<k_1\left(\frac{p_2}{p_1}\right)^{\left(1+\frac{\alpha\theta}{1-\theta}\right)} \tag{50}$$

Now, from $(F.2)$, we have that, due to $p_1 > p_2$ and $\gamma_2 < \theta$:

$$\frac{M_2x_2^{\frac{1}{\theta}}-k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}>\left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta}\times\frac{\theta-\gamma_2}{\theta(1-\gamma_2)}}k_1 \tag{51}$$

Then, notice that:

$$1 + \frac{\alpha\theta}{1-\theta} - \frac{\alpha\theta}{1-\theta} \times \frac{\theta - \gamma_2}{\theta(1-\gamma_2)} = 1 + \frac{\alpha\theta}{1-\theta}\left[1 - \frac{\theta-\gamma_2}{\theta(1-\gamma_2)}\right] = 1 + \frac{\alpha\theta}{1-\theta}\left[\frac{\gamma_2(1-\theta)}{\theta(1-\gamma_2)}\right] > 0 \quad (52)$$

Therefore the exponent at $\frac{p_2}{p_1}$ is higher at the RHS of (50). Since $\frac{p_2}{p_1} \in (0,1)$, we have that:

$$k_1 \left(\frac{p_2}{p_1}\right)^{\left(1+\frac{\alpha\theta}{1-\theta}\right)} < \left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta} \times \frac{\theta-\gamma_2}{\theta(1-\gamma_2)}} k_1$$

consequently, equations (50) and (51) give us a contradiction.

Now, again towards a contradiction, let's assume $p_2 > p_1$. In this case, from the RHS of $(F.1)$, we have:

$$\left\{ \begin{array}{l} \left(\frac{M_1}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1 \left[\frac{p_1}{p_2} - \left(\frac{p_1}{p_2}\right)^{\frac{\gamma_1\alpha}{\gamma_1-1}}\right] \\ + \left(\frac{M_3}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3} x_3 \left[\frac{p_1}{p_2} - \left(\frac{p_1}{p_2}\right)^{\frac{\gamma_3\alpha}{\gamma_3-1}}\right] \end{array} \right\} < 0$$

Since $p_1 < p_2$, $\gamma_1 < 1$, and $\gamma_3 < 1$. Therefore, the LHS of $(F.1)$ must also be negative. Since we already showed that the second term in the LHS and the denominator of the first term in the LHS must be positive, this requirement of a negative LHS implies, after dividing both sides by $\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}$:

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} > k_1 \left(\frac{p_2}{p_1}\right)^{\left(1+\frac{\alpha\theta}{1-\theta}\right)} \quad (53)$$

Then, from $(F.2)$, since $p_1 < p_2$, the last term on the RHS is positive. Consequently, once $\gamma_2 < \theta$, we have:

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} < \left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta} \times \frac{\theta-\gamma_2}{\theta(1-\gamma_2)}} k_1 \quad (54)$$

Since:

$$1 + \frac{\alpha\theta}{1-\theta} - \frac{\alpha\theta}{1-\theta} \times \frac{\theta - \gamma_2}{\theta(1-\gamma_2)} = 1 + \frac{\alpha\theta}{1-\theta}\left[1 - \frac{\theta-\gamma_2}{\theta(1-\gamma_2)}\right] = 1 + \frac{\alpha\theta}{1-\theta}\left[\frac{\gamma_2(1-\theta)}{\theta(1-\gamma_2)}\right] > 0$$

and $p_2 > p_1$, we have that:

$$k_1 \left(\frac{p_2}{p_1}\right)^{\left(1+\frac{\alpha\theta}{1-\theta}\right)} > \left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta}\times\frac{\theta-\gamma_2}{\theta(1-\gamma_2)}} k_1$$

Consequently, equations (53) and (54) give us a contradiction. Therefore, we have that $p_1 = p_2 \Leftrightarrow A_1 = A_2$. ■

A.2 SBTC

The final steps to close the model involve simplifying the model such that we have a system with only two equations and two unknowns (k_1 and $\frac{p_1}{p_2}$. Based on the calculations presented in the paper for k_2, k_1 and their respective FOCs, we obtain:

$$F_j(m_{1j}, m_{2j}, m_{3j}, k_j) = A_j \left[m_{1j}^{\gamma_1} x_1 + \left(m_{3j}^{\theta} x_3 + k_j^{\theta} x_k\right)^{\frac{\gamma_3}{\theta}} + m_{2j}^{\gamma_2} x_2\right] \tag{55}$$

FOCs:

$$\begin{aligned}
&(m_{1j}) : A_j \gamma_1 m_{1j}^{\gamma_1 - 1} x_1 = w_{1j} \\
&(m_{2j}) : A_j \gamma_2 m_{2j}^{\gamma_2 - 1} x_2 = w_{2j} \\
&(m_{3j}) : A_j \gamma_3 \left(m_{3j}^{\theta} x_3 + k_j^{\theta} x_k\right)^{\frac{\gamma_3}{\theta} - 1} m_{3j}^{\theta - 1} x_3 = w_{3j} \\
&(k_j) : A_j \gamma_3 \left(m_{3j}^{\theta} x_3 + k_j^{\theta} x_k\right)^{\frac{\gamma_3}{\theta} - 1} k_j^{\theta - 1} x_k = r
\end{aligned}$$

Since from utility equalization, we have:

$$\frac{w_{ij}}{w_{ij'}} = \left(\frac{p_j}{p_{j'}}\right)^{\alpha}, \quad \forall i \in \{1, 2, 3\} \text{ and } \forall j \in \{1, 2\} \tag{56}$$

From (m_{11}), (m_{12}), and feasibility condition for skill 1, we have:

$$m_{11} = \frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1 - 1}} M_1}{1 + \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1 - 1}}} \tag{57}$$

Similarly, for skill 2:

$$m_{21} = \frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2 - 1}} M_2}{1 + \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2 - 1}}} \tag{58}$$

From (m_{31}), (k_1), (m_{32}), (k_2), labor market clearing, and the utility equalization condition, we have:

$$\left(\frac{m_{31}}{m_{32}}\right) = \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}} \frac{k_1}{k_2} \tag{59}$$

Now let's go back to the expression for (k_1). Manipulating it, we have that:

$$m_{31} = \left\{\frac{1}{x_3}\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]\right\}^{\frac{1}{\theta}} k_1 \tag{60}$$

Similarly, for (k_2), we have:

$$m_{32} = \left\{\frac{1}{x_3}\left[\left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]\right\}^{\frac{1}{\theta}} k_2 \tag{61}$$

Dividing (33) by (61)and substituting (59), we have:

$$\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta-1}} = \left\{\frac{\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]}{\left[\left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]}\right\} \tag{62}$$

Manipulating and simplifying it, we have:

$$k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} = \left(\frac{A_2}{A_1}\right)^{\frac{\theta}{\gamma_3-\theta}} \left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} + \left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\theta-\gamma_3}} \left[1-\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}}\right] x_k$$

Now, we also can use the fact that $m_{31} + m_{32} = M_3$. Then, we have that:

$$M_3 x_3^{\frac{1}{\theta}} = \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1 + \left[\left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_2 \tag{63}$$

Substituting (62) and manipulating, we have:

$$k_2 = \frac{M_3 x_3^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{64}$$

Substituting (64) into (63) and manipulating, we have:

$$\left\{ \frac{M_3 x_3^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1 \gamma_3 x_k} \right)^{\frac{\theta}{\gamma_3 - \theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1 \gamma_3 x_k} \right)^{\frac{\theta}{\gamma_3 - \theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} = \tag{65}$$
$$= \left(\frac{A_2}{A_1} \right)^{\frac{\theta}{\gamma_3 - \theta}} \left(\frac{p_1}{p_2} \right)^{\frac{\alpha\theta}{1-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} + \left(\frac{r}{A_2 \gamma_3 x_k} \right)^{\frac{\theta}{\theta - \gamma_3}} \left[1 - \left(\frac{p_1}{p_2} \right)^{\frac{\alpha\theta}{1-\theta}} \right] x_k$$

which implicitly pins down k_1 as a function of $\frac{p_1}{p_2}$.

Finally, in order to pin down the equilibrium, we need to work with the housing market equilibrium conditions. Looking at the ratio of the housing market clearing conditions, we have:

$$\frac{w_{11} m_{11} + w_{21} m_{21} + w_{31} m_{31}}{w_{12} m_{12} + w_{22} m_{22} + w_{32} m_{32}} = \frac{p_1}{p_2}$$

Now substituting wages and labor demands and rearranging it, we have:

$$\left\{ \begin{array}{c} \left(m_{31}^{\theta} x_3 + k_1^{\theta} x_k \right)^{\frac{\gamma_3 - \theta}{\theta}} m_{31}^{\theta} x_3 - \\ - \frac{A_2}{A_1} \frac{p_1}{p_2} \left(m_{32}^{\theta} x_3 + k_2^{\theta} x_k \right)^{\frac{\gamma_3 - \theta}{\theta}} m_{32}^{\theta} x_3 \end{array} \right\} = \tag{66}$$
$$\left\{ \begin{array}{c} \left(\frac{M_1}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_1 - 1}}} \right)^{\gamma_1} x_1 \left[\frac{A_2}{A_1} \frac{p_1}{p_2} - \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_1}{\gamma_1 - 1}} \right] \\ + \left(\frac{M_2}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_2 - 1}}} \right)^{\gamma_2} x_2 \left[\frac{A_2}{A_1} \frac{p_1}{p_2} - \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_2}{\gamma_2 - 1}} \right] \end{array} \right\}$$

Then, from the ratio of (m_{31}) and (m_{32}), we have:

$$\left(m_{32}^{\theta} x_3 + k_2^{\theta} x_k \right)^{\frac{\gamma_3 - \theta}{\theta}} = \left(\frac{p_2}{p_1} \right)^{\alpha} \left(m_{31}^{\theta} x_3 + k_1^{\theta} x_k \right)^{\frac{\gamma_3 - \theta}{\theta}} \times \left(\frac{m_{31}}{m_{32}} \right)^{\theta - 1} \times \left(\frac{A_1}{A_2} \right) \tag{67}$$

Substituting (67) into (66) and rearranging, we have:

$$\left\{ \left[1 - \left(\frac{p_1}{p_2} \right)^{1-\alpha} \frac{M_3 - m_{31}}{m_{31}} \right] \left(m_{31}^{\theta} x_3 + k_1^{\theta} x_k \right)^{\frac{\gamma_3 - \theta}{\theta}} m_{31}^{\theta} x_3 \right\} = \tag{68}$$
$$\left\{ \begin{array}{c} \left(\frac{M_1}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_1 - 1}}} \right)^{\gamma_1} x_1 \left[\frac{A_2}{A_1} \frac{p_1}{p_2} - \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_1}{\gamma_1 - 1}} \right] \\ + \left(\frac{M_2}{1 + \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_2 - 1}}} \right)^{\gamma_2} x_2 \left[\frac{A_2}{A_1} \frac{p_1}{p_2} - \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_2}{\gamma_2 - 1}} \right] \end{array} \right\}$$

But then, from equation (60), we have that:

$$m_{31}^{\theta} x_3 = \left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\theta)}{\gamma_3-\theta}} - k_1^{\theta} x_k \tag{69}$$

Similarly, from (k_1), we have:

$$\left(m_{31}^{\theta} x_3 + k_1^{\theta} x_k\right)^{\frac{\gamma_3-\theta}{\theta}} = \left(\frac{r}{A_1\gamma_3 x_k}\right) k_1^{1-\theta} \tag{70}$$

Then, from (69) and (70), we have:

$$\left(m_{31}^{\theta} x_3 + k_1^{\theta} x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta} x_3 = \left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\gamma_3}{\gamma_3-\theta}} k_1^{\frac{\gamma_3(1-\theta)}{\gamma_3-\theta}} - \frac{r}{A_1\gamma_3} k_1 \tag{71}$$

Substituting equation (64) into (59) and manipulating, we have:

$$\frac{M_3 - m_{31}}{m_{31}} = \frac{M_3 x_3^{\frac{1}{\theta}} - k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{72}$$

Consequently:

$$\left[1 - \left(\frac{p_1}{p_2}\right)^{1-\alpha} \frac{M_3 - m_{31}}{m_{31}}\right] = \frac{\left\{\begin{array}{c} \left(1 + \left(\frac{p_1}{p_2}\right)^{1-\alpha}\right) k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} \\ - \left(\frac{p_1}{p_2}\right)^{1-\alpha} M_3 x_3^{\frac{1}{\theta}} \end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{73}$$

Then, from equations (71) and (73), we have that:

$$\begin{gathered} \left[1 - \left(\frac{p_1}{p_2}\right)^{1-\alpha} \frac{M_3 - m_{31}}{m_{31}}\right] \left(m_{31}^{\theta} x_3 + k_1^{\theta} x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta} x_3 = \\ \frac{\left\{\begin{array}{c} \left(1 + \left(\frac{p_1}{p_2}\right)^{1-\alpha}\right) k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} \\ - \left(\frac{p_1}{p_2}\right)^{1-\alpha} M_3 x_3^{\frac{1}{\theta}} \end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \times \left\{\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\gamma_3}{\gamma_3-\theta}} k_1^{\frac{\gamma_3(1-\theta)}{\gamma_3-\theta}} - \frac{r}{A_1\gamma_3} k_1\right\} \end{gathered} \tag{74}$$

Notice that the LHS of equation (74) is the same of the one of equation (68). Substituting it back, we have:

$$\frac{\left\{\begin{array}{c}\left(1+\left(\frac{p_1}{p_2}\right)^{1-\alpha}\right)k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}}k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}-x_k\right]^{\frac{1}{\theta}}\\ -\left(\frac{p_1}{p_2}\right)^{1-\alpha}M_3 x_3^{\frac{1}{\theta}}\end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}}k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}-x_k\right]^{\frac{1}{\theta}}}\times\left\{\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\gamma_3}{\gamma_3-\theta}}k_1^{\frac{\gamma_3(1-\theta)}{\gamma_3-\theta}}-\frac{r}{A_1\gamma_3}k_1\right\}=$$
$$\left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\ +\left(\frac{M_2}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2}x_2\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_2}{\gamma_2-1}}\right]\end{array}\right\} \tag{75}$$

Finally, notice that equations (75) and (65) generate a system with two equations and two unknowns (k_1 and $\frac{p_1}{p_2}$):

$$\left\{\begin{array}{ll}\frac{\left\{\begin{array}{c}\left(1+\left(\frac{p_1}{p_2}\right)^{1-\alpha}\right)k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}}k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}-x_k\right]^{\frac{1}{\theta}}\\ -\left(\frac{p_1}{p_2}\right)^{1-\alpha}M_3 x_3^{\frac{1}{\theta}}\end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}}k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}-x_k\right]^{\frac{1}{\theta}}}\times\left\{\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\gamma_3}{\gamma_3-\theta}}k_1^{\frac{\gamma_3(1-\theta)}{\gamma_3-\theta}}-\frac{r}{A_1\gamma_3}k_1\right\}= & \\ \left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\ +\left(\frac{M_2}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2}x_2\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_2}{\gamma_2-1}}\right]\end{array}\right\} & (F.1)\\ \left\{\frac{M_3 x_3^{\frac{1}{\theta}}-\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}}k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}-x_k\right]^{\frac{1}{\theta}}k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}}k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}-x_k\right]^{\frac{1}{\theta}}}\right\}^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}= & (F.2)\\ =\left(\frac{A_2}{A_1}\right)^{\frac{\theta}{\gamma_3-\theta}}\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}}k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}+\left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\theta-\gamma_3}}\left[1-\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}}\right]x_k & \end{array}\right.$$

Preliminary Results

In this subsection, we present some preliminary results that will help us to show the main results presented in the paper.

Lemma A.3: *The distribution of skills across cities is identical if and only if* $\frac{m_{i1}}{m_{i2}} = constant, \forall i \in \{1,2,3\}$.

Proof: ($\Rightarrow$) Consider that the distribution across cities is constant, then $pdf_{i1} = pdf_{i2}, \forall i \in \{1,2,3\}$, i.e.:

$$\frac{m_{i1}}{m_{11}+m_{21}+m_{31}} = \frac{m_{i2}}{m_{12}+m_{22}+m_{32}} \tag{76}$$

But that means that $\frac{m_{i1}}{m_{i2}} = \eta = \frac{S_1}{S_2} = \frac{m_{11}+m_{21}+m_{31}}{m_{12}+m_{22}+m_{32}}$. The other direction is trivial. ■

Lemma A.4: *Assume* $\gamma_3 > \theta$. $p_1 = p_2$ *if and only if* $A_1 = A_2$.

Proof: Towards a contradiction, let's assume that $A_1 = A_2$ and $p_1 > p_2$. Consequently, $\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta-1}} < 1$. From (62), we have $k_1 < k_2$. But then, from equation (59), we obtain $m_{31} < m_{32}$. Finally, from the RHS of (39), we have:

$$\left\{\begin{array}{c} \left(\frac{M_1}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1 \left[\frac{p_1}{p_2} - \left(\frac{p_1}{p_2}\right)^{\frac{\gamma_1\alpha}{\gamma_1-1}}\right] \\ + \left(\frac{M_2}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2} x_2 \left[\frac{p_1}{p_2} - \left(\frac{p_1}{p_2}\right)^{\frac{\gamma_2\alpha}{\gamma_2-1}}\right] \end{array}\right\} > 0$$

Since $p_1 > p_2$, $\gamma_1 < 1$, and $\gamma_2 < 1$. However, given the results we obtained from (62) and (59), the LHS of (66) gives us:

$$\left\{\begin{array}{c} \left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 - \\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3 \end{array}\right\} < 0$$

which is a contradiction.

Similarly, again towards a contradiction, let's consider $A_1 = A_2$ and $p_1 < p_2$. Then $\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta-1}} > 1$. Again from (62), we have $k_1 > k_2$. Similarly, from (59), we obtain $m_{31} > m_{32}$. But then, from (66), we have that:

$$\left\{\begin{array}{c} \left(\frac{M_1}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1 \left[\frac{p_1}{p_2} - \left(\frac{p_1}{p_2}\right)^{\frac{\gamma_1\alpha}{\gamma_1-1}}\right] \\ + \left(\frac{M_2}{1+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2} x_2 \left[\frac{p_1}{p_2} - \left(\frac{p_1}{p_2}\right)^{\frac{\gamma_2\alpha}{\gamma_2-1}}\right] \end{array}\right\} < 0$$

given $p_1 < p_2$. Then RHS(66) < 0. While

$$\left\{ \begin{array}{c} \left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 - \\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3 \end{array} \right\} > 0.$$

which again gives you a contradiction. Therefore, we have that $p_1 = p_2$. Consequently, we have that $A_1 = A_2 \Rightarrow p_1 = p_2$.

Now, let's show that $p_1 = p_2 \Rightarrow A_1 = A_2$. Assume $p_1 = p_2$. Then, from (59), we have:

$$\frac{m_{31}}{m_{32}} = \frac{k_1}{k_2} \tag{77}$$

From (62) , we have

$$\frac{k_1}{k_2} = \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{78}$$

Combining (77) and (78), we have:

$$\frac{m_{31}}{m_{32}} = \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{79}$$

But then, from LHS(66), substituting (77) and (79) given $p_1 = p_2$, we have:

$$\left\{ \begin{array}{c} \left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 - \\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3 \end{array} \right\} = \left[\left(\frac{A_1}{A_2}\right)^{\frac{\gamma_3}{1-\gamma_3}} - \frac{A_2}{A_1}\right]\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3 \tag{80}$$

while the RHS(66) gives us:

$$\left\{ \begin{array}{c} \left(\frac{M_1}{1+\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_1}}}\right)^{\gamma_1} x_1 \left[\frac{A_2}{A_1} - \left(\frac{A_1}{A_2}\right)^{\frac{\gamma_1}{1-\gamma_1}}\right] \\ + \left(\frac{M_2}{1+\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_2}}}\right)^{\gamma_2} x_2 \left[\frac{A_2}{A_1} - \left(\frac{A_1}{A_2}\right)^{\frac{\gamma_2}{1-\gamma_2}}\right] \end{array} \right\} \tag{81}$$

Then, consider the case in which $A_1 > A_2$. From (80), we have that LHS(66)> 0, while (81) gives us RHS(66)< 0. Similarly, if $A_1 < A_2$, (80) gives us LHS(66)< 0 while (81) gives us RHS(66)> 0. Consequently, (66) is only satisfied if $A_1 = A_2$, concluding our proof.■

B Proofs

Proof of Proposition 1

Proof. Towards a contradiction, assume that $A_2 > A_1$ and $p_1 > p_2$. Then, the RHS of $(F.1)$ is positive. Consequently, in order to satisfy $(F.1)$, $(F.1)$'s LHS must also be positive. Following the same argument presented in the proof of Lemma A.2, we have that inequality (50) must hold. Then, from $(F.2)$ we have that, given that $p_1 > p_2$, the last term in $(F.2)$'s RHS – $\left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\theta-\gamma_2}}\left[1-\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{1-\theta}}\right]x_k$ – is negative. We also know that since $A_2 > A_1$ and $\gamma_2 < \theta$, $\left(\frac{A_2}{A_1}\right)^{\frac{\theta}{\gamma_2-\theta}} < 1$. Therefore, $(F.2)$ gives us:

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} > \left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta}\times\frac{\theta-\gamma_2}{\theta(1-\gamma_2)}} k_1 \tag{82}$$

Given (52) we have that, once $\frac{p_2}{p_1} \in (0,1)$:

$$k_1\left(\frac{p_2}{p_1}\right)^{\left(1+\frac{\alpha\theta}{1-\theta}\right)} < \left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta}\times\frac{\theta-\gamma}{\theta(1-\gamma)}} k_1$$

Consequently, (50) and (82) give us a contradiction. Following the same procedure we can easily show that $A_1 > A_2$ and $p_2 > p_1$ give us the same contradiction. Since lemma A.3 shows that price equality is only achieved through TFP equality, this concludes our proof. ■

Proof of Proposition 2

Proof. Without loss of generality, assume $A_1 > A_2$, Then, based on proposition 2, we have that $p_1 > p_2$. Then, from equation (35), we have:

$$\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta-1}} = \left[\frac{\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}{\left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}\right] \tag{83}$$

Then, since $\theta < 1$, we have $\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta-1}} < 1$. Consequently:

$$\left[\frac{\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}{\left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}\right] < 1 \tag{84}$$

Rearranging it:

$$\left(\frac{k_1}{k_2}\right)^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} < \left(\frac{A_1}{A_2}\right)^{\frac{\theta}{\gamma_2-\theta}} \tag{85}$$

Since $\gamma_2 < \theta$, this implies that $\left(\frac{k_1}{k_2}\right)^{\frac{\theta(1-\gamma_2)}{\theta-\gamma_2}} > \left(\frac{A_1}{A_2}\right)^{\frac{\theta}{\theta-\gamma_2}}$. Since $A_1 > A_2$, we must have that $\frac{k_1}{k_2} > \frac{A_1}{A_2} \Rightarrow k_1 > k_2$. ■

Before we prove Theorem 1, let's prove some preliminary results that will be important for the theorems' proofs.

Lemma 1 *If $A_1 > A_2$ we must have that $\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2} > 1$.*

Proof. From proposition 1 we have that $A_1 > A_2 \Rightarrow p_1 > p_2$. Now, let's focus on $(F.1)$'s RHS. This term is positive or negative depending on the following term:

$$\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2}\right]^{\frac{\gamma_i}{1-\gamma_i}}, \ \forall i \in \{1,3\} \tag{86}$$

Now, towards a contradiction, let's assume that $A_1 > A_2$ and $\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2} < 1$. Consequently, the second term in expression (86) is less than one. Similarly, $\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2} < 1 \Rightarrow \frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha} > 1$. Since $\alpha < 1$ and $\frac{p_1}{p_2} > 1$, this gives us that

$$\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2}\right]^{\frac{\gamma_i}{1-\gamma_i}} > 0, \ \forall i \in \{1,3\}$$

and the $(F.1)$'s RHS is positive. Then, $(F.1)$'s LHS must also be positive. Following the same argument presented in the proof of lemma A.2, we have that inequality (50) must hold.

Similarly, from $p_1 > p_2$, we have that the last term on $(F.2)$'s RHS is negative. Therefore, since $\gamma_2 < \theta$, we have:

$$\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \right\} > \left(\frac{A_2}{A_1} \right)^{\frac{1}{1-\gamma_2}} \left(\frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta} \times \frac{\gamma_2 - \theta}{(1-\gamma_2)}} k_1 \tag{87}$$

Then, we have that:

$$\frac{\text{RHS}(50)}{\text{RHS}(87)} = \left(\frac{p_2}{p_1} \right)^{1 + \frac{\alpha\theta}{1-\theta} \left[1 - \frac{\gamma_2 - \theta}{\theta(1-\gamma_2)} \right]} \left(\frac{A_1}{A_2} \right)^{\frac{1}{1-\gamma_2}} \tag{88}$$

Notice that $1 - \frac{\gamma_2 - \theta}{\theta(1-\gamma_2)} = \frac{\gamma_2(1-\theta)}{\theta(1-\gamma_2)}$. Consequently:

$$\frac{\text{RHS}(50)}{\text{RHS}(87)} = \left(\frac{p_2}{p_1} \right)^{1 + \frac{\gamma_2 \alpha}{(1-\gamma_2)}} \left(\frac{A_1}{A_2} \right)^{\frac{1}{1-\gamma_2}} = \left\{ \left(\frac{p_2}{p_1} \right)^{1-\gamma_2(1-\alpha)} \frac{A_1}{A_2} \right\}^{\frac{1}{1-\gamma_2}} \tag{89}$$

But then, notice that $1 - \gamma_2(1-\alpha) - \alpha = (1-\alpha)(1-\gamma_2) > 0$. Therefore, $1 - \gamma_2(1-\alpha) > \alpha$. Since $p_2 < p_1$, we have that:

$$\left(\frac{p_2}{p_1} \right)^{1-\gamma_2(1-\alpha)} \frac{A_1}{A_2} < \left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} < 1 \tag{90}$$

where the last inequality comes from our assumption for the contradiction. Then, since $\frac{1}{1-\gamma_2} > 0$, we have $\frac{\text{RHS}(50)}{\text{RHS}(87)} < 1$. But then inequalities (50) and (87) cannot both be satisfied and we have a contradiction. ■

Corollary 2 *If $A_1 > A_2$ we must have $m_{11} > m_{12}$ and $m_{31} > m_{32}$.*

Proof. From the expression for m_{11}, we have:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \right]^{\frac{1}{\gamma_1 - 1}} M_1}{\left\{ 1 + \left[\left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \right]^{\frac{1}{\gamma_1 - 1}} \right\}} = \frac{\left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma_1}} M_1}{\left\{ 1 + \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma_1}} \right\}} \tag{91}$$

Since from lemma 1 we have $\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} > 1$, we must have that $\frac{\left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma_1}} M_1}{\left\{ 1 + \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma_1}} \right\}} > \frac{M_1}{2}$. Consequently $m_{11} > m_{12}$. The identical argument shows that $m_{31} > m_{32}$. ■

Proof of Theorem 1

Proof. We already know that $m_{11} > m_{12}$ and $m_{31} > m_{32}$. So, the only way in which we may have $S_2 > S_1$ is that $m_{22} > m_{21}$. Therefore, towards a contradiction, assume that $m_{22} > m_{21}$. From (45):

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}}{\left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} > k_1 \tag{92}$$

Then, back to $(F.2)$, we have:

$$\begin{aligned} &\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} = \\ &= \left(\frac{A_2}{A_1} \right)^{\frac{\theta}{\gamma_2 - \theta}} \left(\frac{p_1}{p_2} \right)^{\frac{\alpha\theta}{1-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} + \left(\frac{r}{A_2 \gamma_2 x_k} \right)^{\frac{\theta}{\theta - \gamma_2}} \left[1 - \left(\frac{p_1}{p_2} \right)^{\frac{\alpha\theta}{1-\theta}} \right] x_k \end{aligned} \tag{93}$$

Since $A_1 > A_2$ we know from previous results that $p_1 > p_2$. Consequently, the last term in $(F.2)$'s RHS is negative and we have:

$$\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_1}{\left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} > \left(\frac{A_2}{A_1} \right)^{\frac{\theta}{\gamma_2 - \theta}} \left(\frac{p_1}{p_2} \right)^{\frac{\alpha\theta}{1-\theta} \times \left[1 + \frac{1-\gamma_2}{\gamma_2 - \theta} \right]} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} \tag{94}$$

Now, from (92) we have that, since $\gamma_2 < \theta$:

$$\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}}{\left[\left(\frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} < k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} \tag{95}$$

Now, substituting (95) into (94), we have:

$$k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} > \left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[\left(\frac{r}{A_1\gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}}}{\left[\left(\frac{r}{A_1\gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} > \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{\theta}{\theta-\gamma_2}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} \quad (96)$$

From lemma 2 and the fact that $\theta > \gamma_2$, we have that $\left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{\theta}{\theta-\gamma_2}} > 1$. Consequently, we found a contradiction. Therefore, we must have $m_{21} > m_{22}$ and $S_1 > S_2$. ■

Before presenting the proof for theorem 2, let's consider a final intermediary result:

Claim 1 *Assume $\gamma_2 < \theta$. If $A_1 > A_2$ we must have* $\frac{m_{21}}{m_{22}} < \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma_2}}$

Proof. From lemma 1, we have that if $A_1 > A_2$ we must have $\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} > 1$. Then, from $(F.2)$, since $p_1 > p_2$, we must have:

$$\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}} k_1}{k_1 \left[\left(\frac{r}{A_1\gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} < \left\{ \frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right\}^{\frac{\theta}{\gamma_2-\theta}}$$

From (45) and $\gamma_2 < \theta$, we have $\frac{m_{21}}{m_{22}} < \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma_2}}$, concluding the proof. ■

Proof of Theorem 2:

Proof. Assume that $\gamma_i \equiv \gamma, \forall i \in \{1,2,3\}$ and $\gamma < \theta$. Assume that $A_1 > A_2$ as well. From theorem 1 and claim 1 we have $S_1 < \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma_2}} S_2$. Then, notice that $pdf_{1i} = \frac{m_1 i}{S_i}$. Therefore $\frac{pdf_{11}}{pdf_{12}} = \frac{m_{11}}{m_{12}} \times \frac{S_2}{S_1}$. Since $\frac{m_{11}}{m_{12}} = \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma}}$ and $\frac{S_2}{S_1} > \frac{1}{\left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma}}}$, we have that:

$$\frac{pdf_{11}}{pdf_{12}} > \left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma}} \times \frac{1}{\left[\left(\frac{p_2}{p_1} \right)^{\alpha} \frac{A_1}{A_2} \right]^{\frac{1}{1-\gamma}}} \quad (97)$$

Consequently $pdf_{11} > pdf_{12}$. The same calculation gives us $pdf_{31} > pdf_{32}$. Since density functions must add to one, we must also have $pdf_{21} < pdf_{22}$ ■

Proof of Proposition 3

Proof. Towards a contradiction, assume that $A_1 > A_2$ and $p_2 > p_1$. Then, from (62), after some manipulations and using $\gamma_3 > \theta$, and $\frac{p_2}{p_1} > 1$ we have:

$$\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} > \left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}}$$

i.e.:

$$\frac{k_1}{k_2} > \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{98}$$

From equation (59), we have:

$$\frac{m_{31}}{m_{32}} > \frac{k_1}{k_2} \Rightarrow \frac{m_{31}}{m_{32}} > \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{99}$$

Then, from LHS(66), substituting (98) and (99), we have:

$$\left\{\begin{array}{c} \left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 - \\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3 \end{array}\right\} > \left[\left(\frac{A_1}{A_2}\right)^{\frac{\gamma_3}{1-\gamma_3}} - \frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)\right]\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3 > 0 \tag{100}$$

While from RHS(66), we have that:

$$\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_i}{\gamma_i-1}}\right] < 0, \forall \gamma_i < 1$$

Consequently RHS(39)< 0, which gives us a contradiction. Since we showed in lemma A.2 that $p_1 = p_2$ only happens if $A_1 = A_2$, we must have that $A_1 > A_2 \Rightarrow p_1 > p_2$. Following the same procedure we can easily show that $A_2 > A_1 \Rightarrow p_2 > p_1$. ■

Proof of Proposition 4

Proof. Without loss of generality, assume that $A_1 > A_2$. From proposition 3 we have that $A_1 > A_2 \Rightarrow p_1 > p_2$. From (64) and (F.2), given that $p_1 > p_2$, we have – after some manipulations:

$$\frac{k_1}{k_2} > \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \left(\frac{p_2}{p_1}\right)^{\frac{\alpha(\gamma_3-\theta)}{(1-\gamma_3)(1-\theta)}}$$

While from (59), we have that:

$$\frac{m_{31}}{m_{32}} > \left(\frac{p_2}{p_1}\right)^{\frac{\alpha}{1-\theta}} \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \left(\frac{p_2}{p_1}\right)^{\frac{\alpha(\gamma_3-\theta)}{(1-\gamma_3)(1-\theta)}}$$

Simplifying it:

$$\frac{m_{31}}{m_{32}} > \left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right]^{\frac{1}{1-\gamma_3}} \tag{101}$$

Let's consider two cases:

Case 1: $\left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right] \geqslant 1$ – In this case, equation (101) already implies that $m_{31} \geqslant m_{32}$. From (59) and $\theta < 1$, we have that:

$$\frac{k_1}{k_2} \geqslant \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} > 1 \tag{102}$$

Consequently, $k_1 > k_2$, concluding this part of the proof.

Case 2: $\left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right] < 1$ – In this case, from RHS(66), we have that:

$$\left\{\begin{array}{l} \left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right]^{\frac{\gamma_1}{1-\gamma_1}}\right] \\ + \left(\frac{M_2}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2} x_2\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right]^{\frac{\gamma_2}{1-\gamma_2}}\right] \end{array}\right\}$$

Given $\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha} < 1$, notice that:

$$\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha} < 1 \Rightarrow \frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha} > 1 \Rightarrow \frac{A_2}{A_1}\frac{p_1}{p_2} > 1$$

Consequently, $\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right]^{\frac{\gamma_i}{1-\gamma_i}}\right] > 0$ for $i \in \{1,2\}$ and RHS(66)> 0.

But then, from (68), given that $\left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 > 0$, we would need to have:

$$1 - \left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_3 - m_{31}}{m_{31}} > 0$$

Rearranging it:

$$\frac{m_{31}}{m_{32}} > \left(\frac{p_1}{p_2}\right)^{1-\alpha} > 1 \tag{103}$$

From (103) and (59), we have:

$$\left(\frac{p_2}{p_1}\right)^{\frac{\alpha}{1-\theta}}\frac{k_1}{k_2} > \left(\frac{p_1}{p_2}\right)^{1-\alpha} \Rightarrow \frac{k_1}{k_2} > \left(\frac{p_1}{p_2}\right)^{1+\frac{\alpha\theta}{1-\theta}} \tag{104}$$

Consequently, (104) implies that $k_1 > k_2$, concluding our proof. ■

Proof of Corollary 1

Proof. Proof of proposition 4 already showed this result for all cases but $\left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right] = 1$. In this case, notice that:

$$\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha} = 1 \Rightarrow \frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha} = 1$$

Since $\alpha < 1$ and $p_1 > p_2$, we have that $\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right) > 1$. Again, we can show that the RHS(39)> 0. Following the same steps presented in the proof of proposition 4, we can conclude that $m_{3i} > m_{3j}$. ■

Proof of Theorem 3

Proof. Towards a contradiction, assume that $pdf_{31} \leqslant pdf_{32}$. In this case, we must have:

$$\frac{m_{31}}{m_{11} + m_{21} + m_{31}} \leqslant \frac{m_{32}}{m_{12} + m_{22} + m_{32}}$$

Rearranging and simplifying it, we obtain:

$$m_{31}m_{12} - m_{32}m_{11} + m_{31}m_{22} - m_{32}m_{21} \leqslant 0 \tag{105}$$

From equations (57) and (58) and labor market clearing conditions, we have:

$$m_{11} = \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}} m_{12} \quad \text{and} \quad m_{21} = \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}} m_{22} \tag{106}$$

As a result, we have:

$$m_{31}m_{12} - m_{32}m_{11} = m_{32}m_{12}\left\{\frac{m_{31}}{m_{32}} - \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}}\right\} > 0 \tag{107}$$

and

$$m_{31}m_{22} - m_{32}m_{21} = m_{32}m_{22}\left\{\frac{m_{31}}{m_{32}} - \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}}\right\} > 0 \tag{108}$$

where the inequalities come from $\frac{m_{31}}{m_{32}} > \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}}$ as shown in equation (101). Consequently, equations (105), (107), and (108) jointly show a contradiction. As a result, $pdf_{31} > pdf_{32}$.

Similarly, towards a contradiction, consider that $pdf_{21} \geqslant pdf_{22}$. In this case, we must have:

$$\frac{m_{21}}{m_{11} + m_{21} + m_{31}} \geqslant \frac{m_{22}}{m_{12} + m_{22} + m_{32}}$$

Rearranging and simplifying it, we obtain:

$$m_{12}m_{21} - m_{22}m_{11} + m_{32}m_{21} - m_{31}m_{22} \leqslant 0 \tag{109}$$

From (106), after some manipulations, we have:

$$m_{12}m_{21} - m_{22}m_{11} = 0 \tag{110}$$

and

$$m_{32}m_{21} - m_{31}m_{22} = m_{32}m_{22}\left\{\left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}} - \frac{m_{31}}{m_{32}}\right\} < 0 \tag{111}$$

where the inequalities come from $\frac{m_{31}}{m_{32}} > \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}}$ as shown in equation (101). Consequently, equations (109), (110), and (111) jointly show a contradiction. As a result, $pdf_{21} < pdf_{22}$.

Finally, towards a contradiction, assume that $pdf_{11} \geqslant pdf_{12}$. In this case, we must have:

$$\frac{m_{11}}{m_{11} + m_{21} + m_{31}} \geqslant \frac{m_{12}}{m_{12} + m_{22} + m_{32}}$$

Rearranging and simplifying it, we obtain:

$$m_{11}m_{22} - m_{12}m_{21} + m_{32}m_{11} - m_{31}m_{12} \leqslant 0 \tag{112}$$

In equation (110), we already showed that $m_{11}m_{22} - m_{12}m_{21} = 0$. Then, from (106) and (101), we have:

$$m_{32}m_{11} - m_{31}m_{12} = m_{32}m_{12}\left\{\left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}} - \frac{m_{31}}{m_{32}}\right\} < 0 \tag{113}$$

Consequently, equations (112), (110), and (113) jointly show a contradiction. As a result, $pdf_{11} < pdf_{12}$, concluding our proof that pdf_1 F.O.S.D. pdf_2. ■

C Numerical Example

In this section, we simulate the model in order to get a better understanding of the model's mechanisms and how changes in the parameters may affect the two regions' labor markets. We focus on two parameter changes that are related to the observed evolution of computing power prices over the last twenty years. First, the price of PCs and software went down significantly over this time period. Second, personal computers became significantly more powerful, being able to do operations that needed servers or computer networks previously. While this distinction seems subtle at first sight, it is an important difference for the model. Reductions in price, while increasing the benefit of renting more capital, do nothing to counteract the decreasing marginal contribution of capital. Differently, increases in computer power per machine, by increasing x_k, avoids the decreasing forces of marginal productivity. Moreover, we also believe it is an important distinction in reality. Increasing computer power through the use of servers or connected networks, while possible, demands a lot of coordination and knowledge by its users. These additional user costs reduce the widespread implementation of internal networks and local servers. Moreover, while prices for information technology have gone down, the wide decline in the price indexes for technology, presented in figures (a) and (c) in figure 3 are mostly due to the increase in the processing power which is factored in by the Bureau of Labor Statistics (BLS). Furthermore, even though there is some evidence that the gross investment in personal computers and peripherals has stalled in the latter period, once we control for processing power, the investment in computers has continued to go up, as we present in figures (b) and (d) in figure 3. Consequently, it is important to take into account a potential difference between quality and quantity when we are dealing with changes due to technological progress over time.

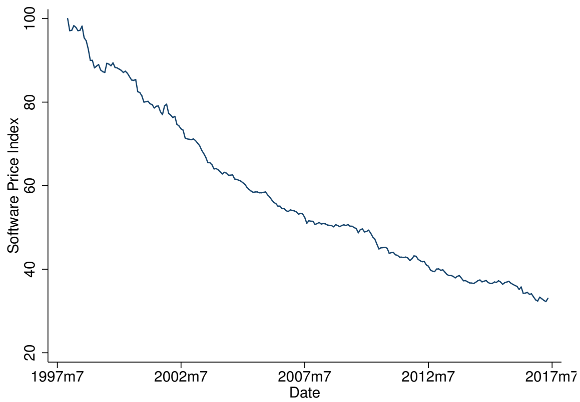
C.1 Benchmark parametrization

In this section, we show a simple numerical example that illustrates the results of the model. In order to be able to interpret the results more properly, we use results found in the previous literature and data in order to calibrate our parameters. We start using parameter values described by Eeckhout et al. (2014)'s table 2 in order to pin down the values for city TFP and workers' labor productivity. We consider the case that $\gamma_i \equiv \gamma, \forall i \in \{1, 2, 3\}$ and use Eeckhout et al. (2014)'s table 2 to set γ as well. Moreover, we follow Davis and Ortalo-Magné (2011) and set $\alpha = 0.24$. Finally, we must specify values for both θ and the housing stock. We will keep these values as given at $H_i = 62,559,000, \forall i \in \{1, 2\}$ which is close to the BEA's estimate for half of the total housing units for the United States in 2005Q2, and $\theta = 0.85$. We present these parameters in table 16. We assume that these parameters are fixed over time in our numerical exercise.

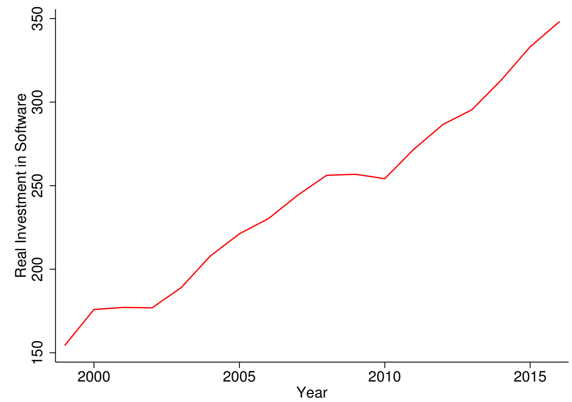
Table 16: Maintained Parameters – from Eeckhout et al. (2014)

γ	θ	A_1	A_2	y_1	y_2	y_3	H_i	α
0.8	0.85	19,118	9,065	0.3189	1	1.4733	62,559,000	0.24

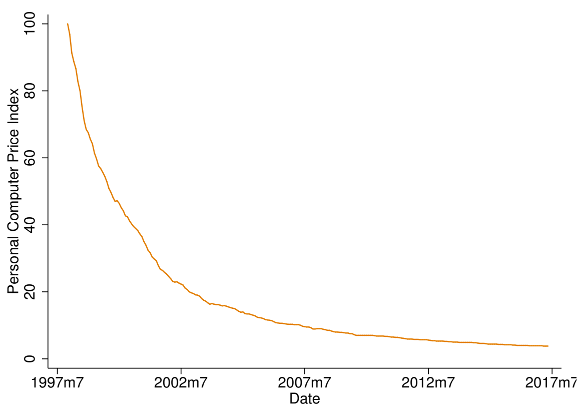
We then consider two periods in time: 1995 and 2015. We consider changes in the size and



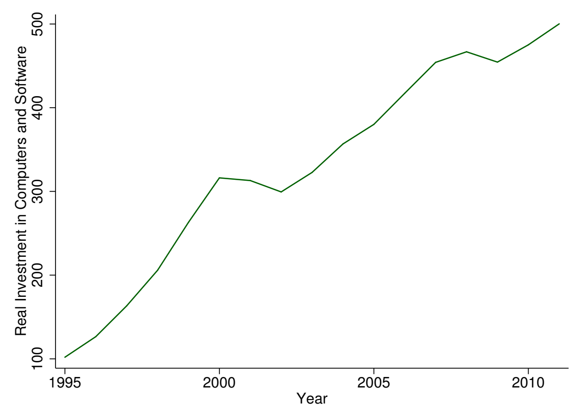
(a) Software's Price Index: 1997–2017
Source: Bureau of Labor Statistics



(b) Real Investment in Software: 1999–2017
Source: Bureau of Economic Analysis



(c) PC's Price Index: 1997–2017
Source: Bureau of Labor Statistics



(d) Real Investment in PCs: 1995–2011
Source: Bureau of Economic Analysis

Figure 3: Price Index and Real Investment in Technology

composition of the population – measured by the size of the labor force and the distribution across occupations. We follow the distribution of the population across routine and non-routine manual and cognitive occupations for the years 1989 and 2014 as presented by Cortes et al. (2016). We combine routine cognitive and manual occupations to form the middle-skill measure, while we consider non-routine cognitive occupations as high skill and non-routine manual as low skill. Finally, we disregard the unemployed. Similarly, we consider changes in the technology. We pin down x_k by normalizing it at 1 in 1995 and using the estimates for multi-factor productivity (MFP) growth for softwares as presented by Byrne et al. (2017)'s table 3B in order to pin down x_k in 2015. Similarly, in order to consider the changes in the price for technology, we normalize $r = 700$ in 1995 – close to the value that Eeckhout et al. (2014) implied for a middle-skill worker in the small city – and use Byrne and Corrado (2017)'s estimate of price decrease in the cost of ICT investments (Table 4 – software), in order to pin down the value for r in 2015. The calibrated values are presented in table 17.

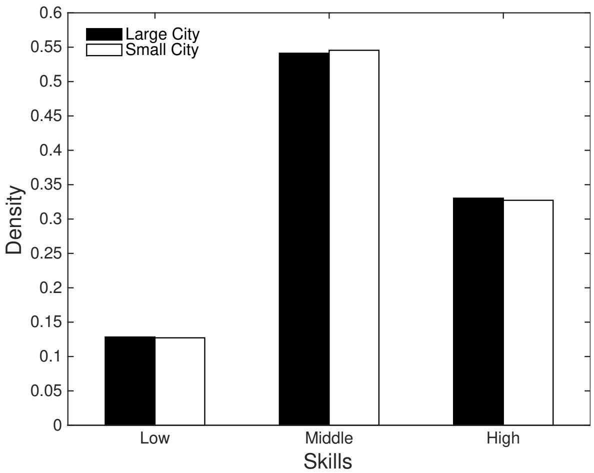
Table 17: Adjusted Parameters - Experiments

	y_k	r	M_1	M_2	M_3
1995	1	700	15,836,150	66,973,717	40,745,094
2015	1.333	635.58	26,640,565	67,576,067	61,078,368

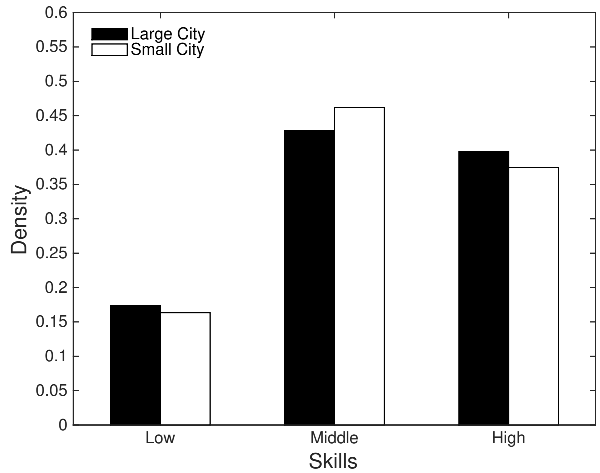
Results are presented in figure 4 and table 18. As we can see from figures 4(a) and 4(b) and table 18's panel B, between 1995 and 2015, city 1 not only became even bigger than city 2, but it also became more unequal – the proportion of mid-skilled workers went down significantly more in city 1 than in city 2. While this result is in line with the overall increase in inequality that we observed over time, jointly showing a geographical component, it does not clearly indicates the underlying reason for this increase in inequality. From our parameters in table 17, we have that many things changed between 1995 and 2015. First, not only the population has grown, but the distribution of skills across the overall population has developed fatter tails. Second, technology became cheaper as well as more productive. In order to disentangle these effects, we consider two counterfactuals. In the first counterfactual, we keep the overall population size and skill distribution at its 1995 levels and only allow technology to become cheaper and more productive, presented in figure 4(c) and in table 18 **Pop. Fixed** lines. In the second counterfactual, we keep technology at its 1995 levels of cost and productivity, while allowing population and skill distribution to adjust to its 2015 levels, presented in 4(d) and in table 18 **Tech. Fixed** lines. As we can see from the results, while changes in population may be responsible for the bulk of the change in the overall shape of the distributions between 1995 and 2015, the changes in technology cost and productivity are the leading factors behind the big cities becoming increasingly more unequal compared to smaller ones.

D Skill Biased Technological Change and City Size - Numerical Examples

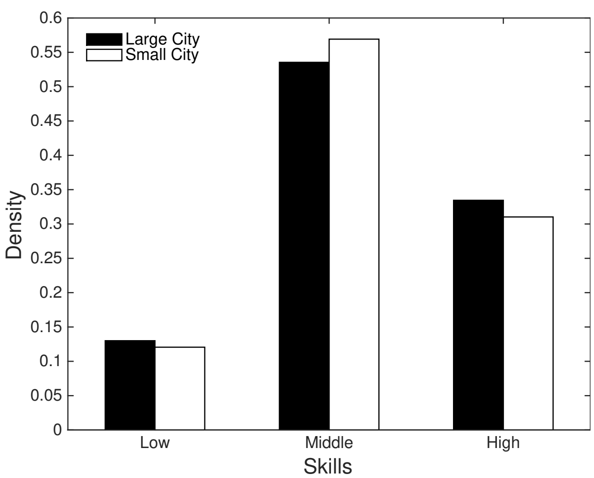
Differently from the case of Automation, **SBTC does not imply that the high-TFP city is larger**. In this section, we present two examples that illustrate that results can go either way.



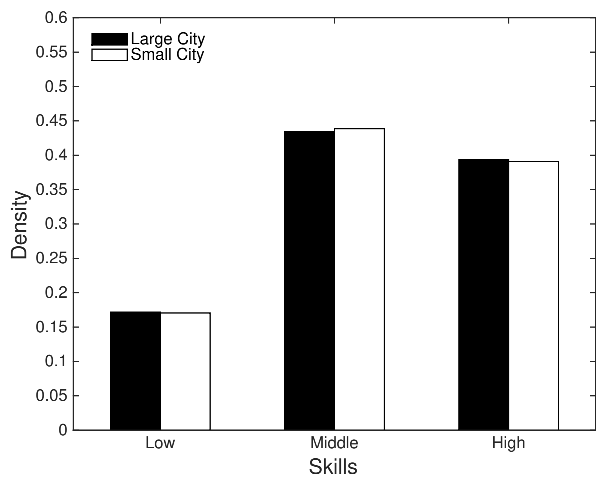
(a) Skill Distribution – 1995



(b) Skill Distribution – 2015



(c) Skill Distribution – Population Constant



(d) Distribution – Technology Constant

Figure 4: Skill Distribution across cities – 1995 vs. 2015

D.1 High-TFP city is smaller (the "Boulder" case)

In this case, we consider the parameter values presented in table 19. Equilibrium prices and quantities are presented in table 20. As we can see, the high-TFP city, while paying higher wages, investing more in capital, having higher housing prices, and having more high-skill workers, it is still smaller than the low-TFP city. In particular, the high-TFP city has fewer low- and mid-skill workers than the low TFP city. Finally, as expected, the skill distribution in the High-TFP city skill dominates in first order the skill distribution in the Low-TFP city, as we see in figure 5.

Table 18: Numerical Exercise Results

Panel A: Prices and Wages

	p_1	p_2	w_{11}	w_{12}	w_{21}	w_{22}	w_{31}	w_{32}
1995	188.38	28.193	184.73	117.10	432.80	274.36	706.45	447.82
2015	224.91	34.466	166.30	106.02	422.28	269.21	650.81	414.91
Pop. Fixed	185.38	28.572	184.48	117.77	422.85	269.95	705.52	450.4
Tech. Fixed	227.91	34.084	166.48	105.52	432.05	273.83	651.53	412.94

Panel B: City Size and Skill Distribution

	S_1	f_{11}	f_{21}	f_{31}	S_2	f_{12}	f_{22}	f_{32}
1995	99,936,000	12.84%	54.12%	33.04%	23,620,000	12.72%	54.55%	32.73%
2015	125,058,000	21.71%	42.86%	39.79%	30,237,100	16.33%	46.21%	37.45%
Pop. Fixed	99,342,000	12.93%	53.54%	33.45%	24,213,500	12.06%	56.92%	31.02%
Tech. Fixed	125,633,000	21.60%	43.43%	39.39%	29,664,100	17.05%	43.85%	39.09%

Table 19: Parameters – "Boulder" case

x_1	x_2	x_3	A_1	A_2	H	x_k	r
0.3189	1	1.4733	19118	19000	62559000	1.333	635.58
γ_1	γ_2	γ_3	θ	α	M_1	M_2	M_3
0.8	0.8	0.82	0.5	0.24	15836150	66973717	40745094

Table 20: Equilibrium Outcomes – "Boulder" case

	m_{1j}	m_{2j}	m_{3j}	S_j	p_j
City 1	7683472.87	32494687	21148855.3	61327015.1	496.69
City 2	8152677.13	34479030	19596238.7	62227945.9	460.71

	w_{1j}	w_{2j}	w_{3j}	k_j
City 1	204.68	481.03	5308.34	1207650344
City 2	201.02	472.42	5213.4	1020740138

D.1.1 High-TFP city is larger (the "NYC" case)

In this case, we consider the parameter values presented in table 21. In order to make a simple comparison, the parameters are the same of the "Boulder" case, apart from a higher A_1. Equilibrium prices and quantities are presented in table 22. Notice that the high-TFP city is larger. However, we still have fewer low- and mid-skill workers. As before, all other results follow through, including the F.O.S.D. of the skill distribution of the high-TFP city, as seen in figure 6.

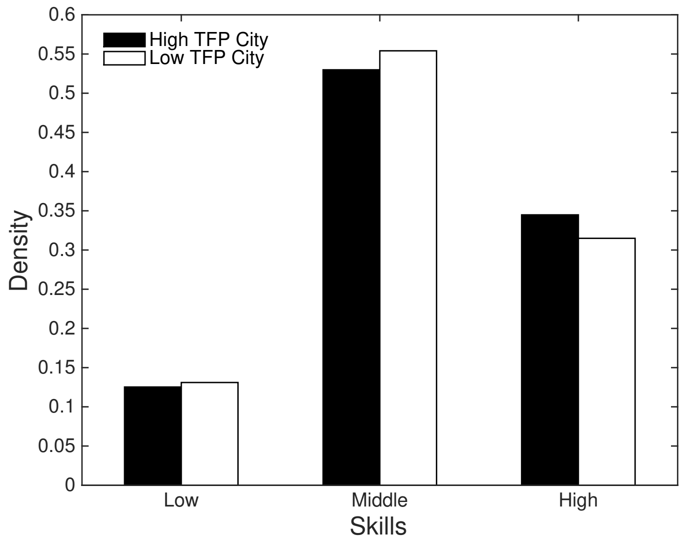


Figure 5: Skill Distribution: High vs. Low TFP cities – "Boulder" case

Table 21: Parameters – "NYC" case

x_1	x_2	x_3	A_1	A_2	H	x_k	r
0.3189	1	1.4733	21118	19000	62559000	1.333	635.58
γ_1	γ_2	γ_3	θ	α	M_1	M_2	M_3
0.8	0.8	0.82	0.5	0.24	15836150	66973717	40745094

Table 22: Equilibrium outcomes – "NYC" case

	m_{1j}	m_{2j}	m_{3j}	S_j	p_j
City 1	7841477.23	33162913.76	24434572.45	65438963.44	663.10
City 2	7994672.77	33810803.24	16310521.55	58115997.56	420.08

	w_{1j}	w_{2j}	w_{3j}	k_j
City 1	225.17	529.19	6283.30	1954872597.36
City 2	201.81	474.28	5631.30	995018365.67

E Empirical Evidence - Alternative technology measures

E.1 Enterprise Resource Planning (ERP) software

In this section, we discuss the coverage of our sample that includes information on ERP adoption, as well as the empirical evidence on the relationship between ERP adoption and local rental price index as well as 1980's share of routine-cognitive jobs in the local labor force.

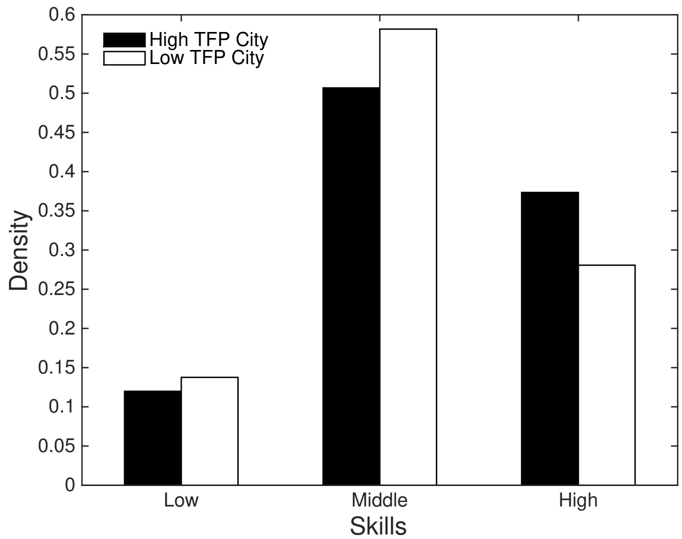


Figure 6: Skill Distribution: High vs. Low TFP cities – "NYC" case

E.1.1 Data Coverage

As discussed in section 3 and presented in table 3, our ERP sample is limited. Our information on ERP adoption covers on average only 16% of workers and 1% of establishments in the MSA, compared to NETS. Moreover, as presented in table 23, even after controlling for establishment size, MSA average coverage is above 30% only for establishments that have 250 employees or more. Finally, table 24 shows that employment coverage is below 30% in all industry sectors.

Table 23: Coverage Ci Aberdeen relative to NETS by Establishment Size

	Mean	**S.D.**	**p10**	**p25**	**p50**	**p75**	**p90**	**N**
ERP Sample								
1 to 4 Employees	0.1%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	272
5 to 9 Employees	0.4%	0.2%	0.0%	0.3%	0.4%	0.5%	0.7%	272
10 to 19 Employees	1.8%	0.6%	1.0%	1.5%	1.8%	2.1%	2.5%	272
20 to 49 Employees	6.2%	1.6%	4.0%	5.3%	6.2%	7.1%	8.1%	272
50 to 99 Employees	14.3%	3.6%	10.0%	12.1%	14.1%	16.4%	19.2%	272
100 to 249 Employees	26.2%	6.0%	20.0%	22.0%	26.4%	29.9%	33.8%	272
250 to 499 Employees	31.9%	11.7%	20.0%	25.0%	30.0%	36.8%	45.2%	272
500 to 999 Employees	41.1%	18.4%	22.0%	30.4%	38.1%	50.0%	61.9%	272
1,000 or more Employees	43.4%	22.1%	20.0%	30.0%	40.0%	53.9%	68.3%	270

Nonetheless, table 25 shows that there is a lot of dispersion in the ERP shares across MSAs even in 2015, when we should expect already a more widespread use of technology. As we can see, we have at least some information on 272 MSAs across the country. Moreover, we can see

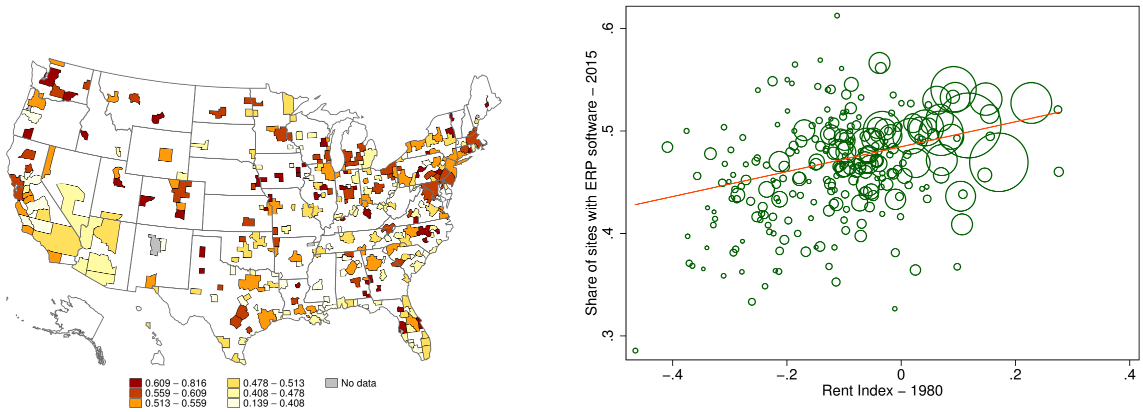
Table 24: Ci Coverage relative to NETS: Employment by Industry

	Mean	S.D.	p10	p25	p50	p75	p90	N
ERP Sample								
Manufacturing	28%	12%	12%	20%	28%	35%	42%	272
Construction	7%	5%	2%	4%	6%	9%	12%	272
Information	20%	11%	8%	13%	19%	26%	34%	272
Finance	8%	7%	2%	4%	7%	11%	16%	272
Professional & Bus Services	8%	5%	2%	4%	7%	11%	14%	272
Education and Health	24%	8%	15%	19%	24%	28%	33%	272
Leisure and Hospitality	6%	6%	2%	3%	5%	7%	11%	272
Public Adm	18%	8%	9%	12%	17%	21%	27%	272
Trade, Transp., and Util.	7%	4%	3%	4%	6%	9%	12%	272
Mining	9%	16%	0%	0%	0%	11%	30%	271
Other Services	5%	4%	1%	3%	5%	7%	9%	272

that, while on average about 47% of the establishments have at least some form of ERP, there is substantial variation across the country. Some MSAs have a fraction as low as 29%, while others have more than 61% of establishments with some form of ERP. Even more, as we show in figure 7b, the degree of adoption seems closely tied to the size as well as cost of living in the MSA, proxied by the rental index. Finally, figure 7a shows the geographical dispersion of ERP concentration across the country in 2015. First of all geographical coverage is quite good, with only very few MSAs completely missing. In fact, the missing MSAs are due to the matching procedure of the Census PUMA to the 2000 Census Metropolitan Area definitions as described by Baum-Snow and Pavan (2013).

Table 25: Descriptive statistics of technology adoption across MSAs – 2015

	Mean	Median	S.D.	Min	Max	N
ERP Share						
Share of Workers in Est. w/ ERP	51.45%	52.09%	11.88%	9.76%	86.94%	272
Share of Establishments w/ ERP	46.34%	46.64%	5.09%	28.57%	61.25%	272
No. of ERPs						
Avg. No. of ERPs per Est.	0.77	0.78	0.11	0.41	1.17	272
Median No. of ERPs per Est.	0.24	0	0.42	0	1	272
St. Dev. Of No. ERP per Est.	1.05	1.06	0.11	0.73	1.36	272



(a) Share of workers exposed to ERP

(b) ERP share vs local price level

Figure 7: Geographical distribution of ERP across MSAs – 2015

E.1.2 Empirical Evidence

Table 26 presents the same specifications as presented in table 7, replacing IT budget per worker with the fraction of establishments in the MSA with at least one ERP software. As we can observe, results for local price indexes are similar to the ones observed in table 7, even though results for the alternative theories are somewhat stronger. Results controlling for firm and industry fixed effects at the establishment level are presented in table 27. As expected, due to a significant decrease in sample size, results are weaker and lose statistical significance in some cases. However, the overall pattern is still the same as the one presented in table 8, i.e., establishments in more expensive MSAs are more likely to adopt ERP softwares.

F Measures of Concentration

We now calculate measures of concentration of skills across regions. These measures allow us to test if we have observed an increase in the spatial dispersion of skills across MSAs in the last 30 years. Moreover, these measures abstract from issues of long-run trends in the composition of labor force. Consequently, we are able to focus on the correlation between the spatial dispersion of skills and MSAs characteristics – in particular size and cost of housing. We consider three simple measures: The location quotient that compares the skill distribution in the MSA against the overall skill distribution in the economy, the Ellison and Glaeser (1997) index of industry concentration, and an adjusted version of this index proposed by Oyer and Schaefer (2016). The latter two indexes attempt to measure concentration by comparing it against a distribution that would be obtained by chance (the "dartboard approach").

Table 26: Fraction of Establishments with ERP software in CMSA – 2015

	(1) ERP	(2) ERP	(3) ERP	(4) ERP	(5) ERP
MSA log rent index 1980	0.2483***			0.1966***	0.1662**
	(0.0634)			(0.0655)	(0.0657)
MSA routine cognitive share 1980		0.9091***		0.7735**	0.4998
		(0.3259)		(0.3123)	(0.3501)
MSA's log $\left(\frac{S}{U}\right)$ in 1980			0.0859***	0.0463**	0.0338
			(0.0239)	(0.0226)	(0.0239)
MSA Offshorability 1980					0.2028*
					(0.1034)
Housing supply elasticity	-0.0050	-0.0149**	-0.0118**	-0.0111*	-0.0110*
	(0.0057)	(0.0058)	(0.0054)	(0.0059)	(0.0057)
USDA's Amenities	Yes	Yes	Yes	Yes	Yes
CMSA's Industry Mix Controls	Yes	Yes	Yes	Yes	Yes
F statistic	3.48	2.92	3.67	4.28	5.29
Adj. R^2	0.183	0.161	0.179	0.228	0.242
CMSAs	222	222	222	222	222

Standard errors in parentheses. The dependent variable in all columns is the fraction of establishments with ERPs in the metro area, adjusted for plant employment interacted with three digit SIC industry. Each observation (a CMSA) is weighted by its employment in 2015. Stars represent: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 27: ERP Presence by Establishment - Firm and Industry FE

	(1) Dummy ERP	(2) Dummy ERP	(3) Dummy ERP	(4) Dummy ERP	(5) Dummy ERP
MSA log rent index 1980	0.101	0.135*	0.115	0.096	0.146*
	(0.080)	(0.080)	(0.079)	(0.081)	(0.080)
MSA's log $\left(\frac{S}{U}\right)$ in 1980		-0.029			-0.024
		(0.032)			(0.034)
MSA Offshorability 1980			-0.081		-0.137
			(0.109)		(0.120)
MSA routine cognitive share 1980				0.005	0.007**
				(0.003)	(0.003)
log(Site's Size)	0.065***	0.065***	0.065***	0.065***	0.065***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
log(Site's Revenue)	0.075***	0.075***	0.075***	0.075***	0.075***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Housing Elasticity	0.007	0.008	0.008	0.006	0.007
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Firm FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
MSA Controls	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.1785	0.1785	0.1785	0.1786	0.1786
No. of Sites	59,828	59,828	59,828	59,828	59,828

Standard errors in parentheses. The dependent variable in all columns is a dummy indicating that ERP software is available in the establishment. Each observation (an establishment) is weighted by the probability weight from a match between the Aberdeen data and the 2015 County Business Pattern. Stars represent: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

F.1 Location Quotient

As a first pass, we consider a concentration measure that compares the distribution in a given MSA against the overall economy distribution. In particular, we consider that the degree of concentration of skill i in city j (λ_{ij}) is given by:

$$\lambda_{ij} = \frac{\frac{m_{ij}}{S_j}}{\frac{M_i}{\sum_{l=1}^{N} M_l}} \tag{114}$$

Intuitively, if a MSA is more concentrated in skill level i than the economy at large, this index's value would be above 1. Moreover, this measure has two additional benefits. First, by focusing on shares, it reduces the impact of the MSA's overall size on the analysis. Second, by comparing the region against the economy-wide distribution, it takes into account the potential changes in the national labor market. Consequently, it allows us to focus on the increase/decrease of concentration across regions as well as how it correlates to these regions' characteristics.

In our analysis, we consider two time periods – 1980 and 2015. Moreover, following Cortes et al. (2016), we divide the occupations in 4 groups: non-routine manual, routine manual, routine cognitive, and non-routine cognitive. We divide the regions in two groups around the median. As a first pass, we divide MSAs in terms of the size of its labor force, i.e., large vs. small. Similar results are obtained if we use the log rent index, i.e. cheap vs. expensive, as the measure to separate the MSAs. Results are presented in table 28.

Table 28: Simple Measure of Concentration across skill and city size groups

Panel A: 1980								
	Non-Routine Manual		*Routine Manual*		*Routine Cognitive*		*Non-Routine Cognitive*	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Large City	0.99	0.95	1.08	1.05	0.97	0.97	0.95	0.95
Small City	1.05*	1.03†	1.11	1.11	0.92**	0.91††	0.93	0.90
Panel B: 2015								
	Non-Routine Manual		*Routine Manual*		*Routine Cognitive*		*Non-Routine Cognitive*	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Large City	0.99	0.96	1.07	1.05	1.02	1.01	0.96	0.97
Small City	1.02	1.02	1.21**	1.19††	1.00	0.99	0.90*	0.89††

**,* represent significant at 1 and 5 % respectively in a t-test of means with unequal variances.
††,† represent significant at 1 and 5% respectively in a Wilcoxon rank-sum test of medians.

As we can see from table 28, in 1980, small cities had on average a higher concentration in non-routine manual jobs, a lower concentration in routine cognitive jobs, and were at par in routine manual and non-routine cognitive once compared to large cities. Differently, in 2015 we see small cities being on average more concentrated in routine manual jobs, less concentrated in non-routine cognitive jobs, and at par in routine cognitive and non-routine manual jobs. Taken as a whole, table 28 shows an increase in the concentration of routine cognitive and routine manual jobs in small cities, jointly with a decrease in non-routine manual and non-routine cognitive jobs, as expected from our theory.

Finally, figure 8 presents the density distribution of the simple concentration index for small and large cities across skill groups and time. While we observe that there is significant variance in this index across CMSAs, the overall message is the same as the one presented in table 28.

F.1.1 Ellison-Glaeser (1997) Index

We now adapt the concentration index presented by Ellison and Glaeser (1997) for the skill distribution context. Denote γ_i as the EG concentration index for skill i. To define this index, we first introduce some notation. Define s_{ij} as the share of workers of skill i in city j, i.e., $s_{ij} = \frac{m_{ij}}{M_i}$. Let x_j be the share of total employment in city j, i.e., $x_j = \frac{S_j}{\sum_{l=1}^{N} M_l}$. Then, our measure of spatial concentration of skill i is given by:

$$\gamma_i = \frac{\sum_j \left(s_{ij} - x_j\right)^2}{1 - \sum_j x_j^2} \tag{115}$$

According to Ellison and Glaeser (1997), there are several advantages in using this index. First, it is easy to compute with readily available data. Second, the scale of the index allows us to make comparisons with a no-agglomeration case in which the data is generated by the simple dartboard model of random location choices (in which case $E(\gamma_i) = 0$). Finally, the index is comparable across populations of different skill sizes. Notice that in this case, we have one index per skill group per year. Consequently, we are unable compare large and small cities. However, we are able to see if skill groups became more or less concentrated across cities over time.

Table 29: Ellison-Glaeser Index

	1980	**2015**	**% Change**
Non-Routine Manual	0.00063	0.00044	-0.29659
Routine Manual	0.00080	0.00068	-0.15094
Routine Cognitive	0.00011	0.00014	0.24356
Non-Routine Cognitive	0.00026	0.00029	0.11259

Results are presented in table 29. As we can see, manual occupations have seen a decline in concentration, whereas cognitive occupations have seen a (small) increase in concentration. These results complement the results regarding the location index, by indicating how concentration of each occupation group has changed across cities. While these results are generally in line with what we should expect given our model's outcomes, we are not able to precisely link them to city characteristics. In order to do that, in the next section we follow Oyer and Schaefer (2016) and adapt the Ellison and Glaeser (1997) to create a city's skill concentration index.

F.1.2 Oyer-Schaefer (2016) Index

We now consider an adapted version of the EG concentration index that we call the Oyer-Schaefer index (henceforth OS index). Hence, denote ζ_j the OS concentration index for city j. To define this index, we first introduce some notation. Define $\tilde{x}_i$ the overall share of workers of skill i in the economy, i.e. $\tilde{x}_i = \frac{M_i}{\sum_{l=1}^{N} M_l}$. Similarly, define $\tilde{s}_{ij}$ the share of workers of skill i in city j, i.e., $\tilde{s}_{ij} = \frac{m_{ij}}{S_j}$, where S_j is city j's labor force size. Then, the OS index is define as:

$$\zeta_j = \frac{S_j}{S_j - 1} \frac{\sum_i \left(\tilde{s}_{ij} - \tilde{x}_i\right)^2}{1 - \sum_i \tilde{x}_i^2} - \frac{1}{S_j - 1} \tag{116}$$

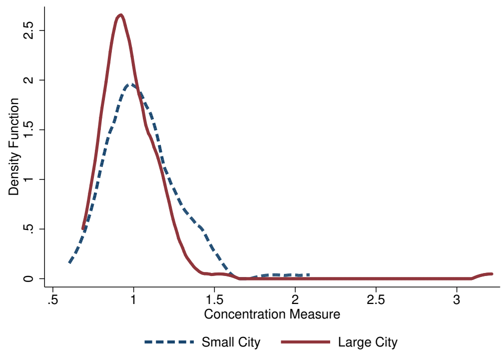
Differently from the EG index, in the OS index we are able to compare the degree of concentration across city sizes or across cities with different housing costs. Unfortunately, we are unable to pin down the source of the increase/decrease in within-city concentration. In particular, we are unable to tie the changes in concentration to changes in the shares of each particular skill group. In this sense, EG and OS indexes, while complementing each other, both have its weaknesses and do not give a complete picture of the changes in concentration.

Table 30 presents the results for 1980 and 2015. As we can see, in both periods, small cities are consistently more concentrated than large cities, although there is also more variance of concentration across small cities. Furthermore, while both small and large cities have seen a reduction in concentration over time, the reduction has been on average larger at large cities.

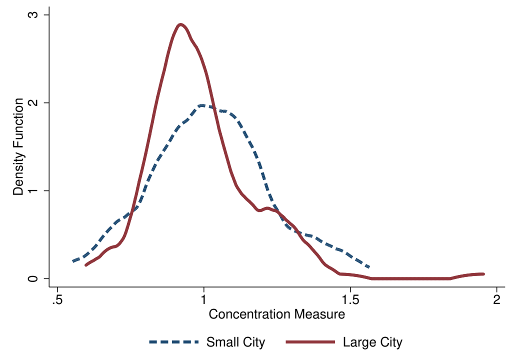
Finally, we present the changes in the density distribution of the OS index in figure 9. As we can see, the distribution of the OS index became more concentrated as we move from 1980 to 2015.

Table 30: OS Index across city sizes and time

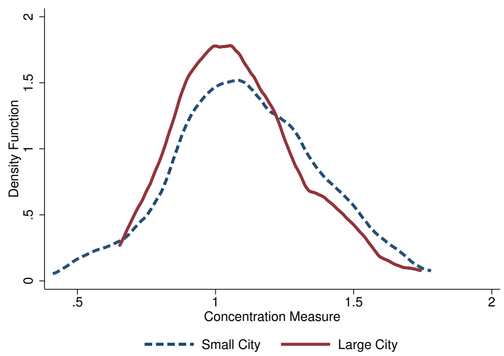
	Mean	Median	St. Dev.	Min	Max
Panel A: 1980					
Large City	0.01193	0.00551	0.01732	0.00012	0.10032
Small City	0.01879	0.00965	0.02132	0.00037	0.11660
Panel B: 2015					
	Mean	Median	St. Dev.	Min	Max
Large City	0.00896	0.00406	0.01156	0.00014	0.06074
Small City	0.01835	0.01259	0.01738	0.00003	0.10652



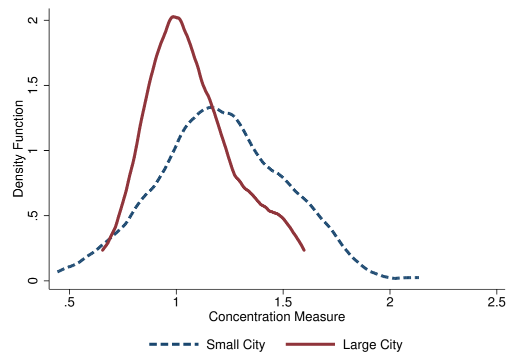
(a) Non-Routine Manual: 1980



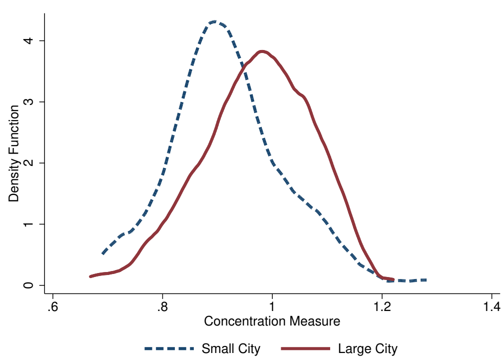
(b) Non-Routine Manual: 2015



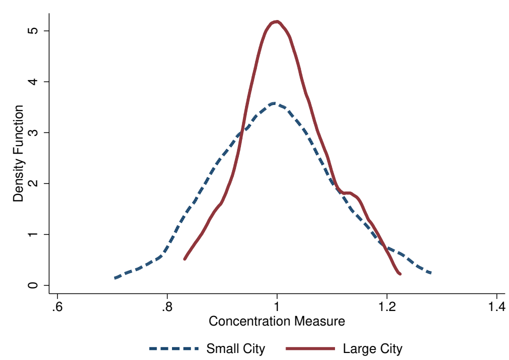
(c) Routine Manual: 1980



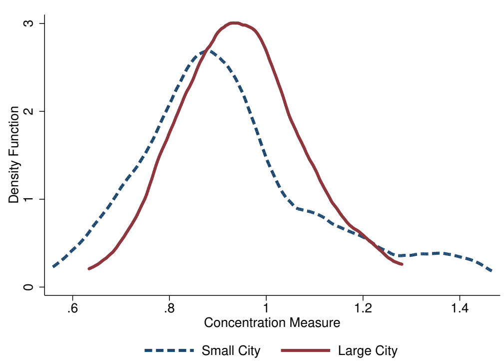
(d) Routine Manual: 2015



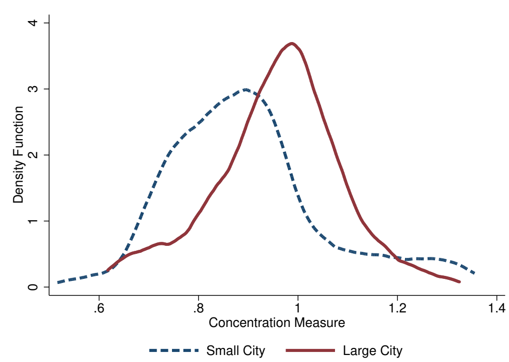
(e) Routine Cognitive: 1980



(f) Routine Cognitive: 2015

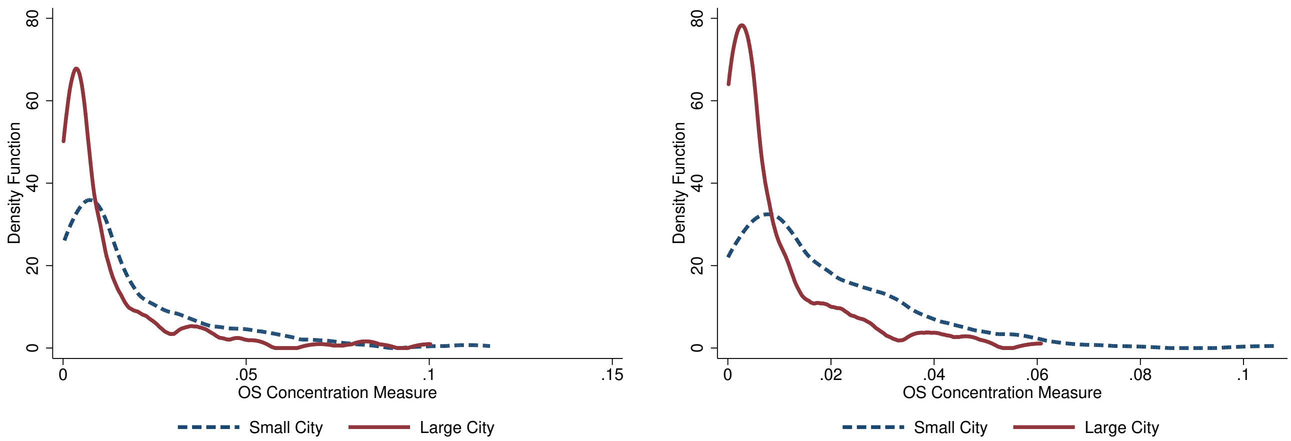


(g) Non-Routine Cognitive: 1980



(h) Non-Routine Cognitive: 2015

Figure 8: Skill Distribution across city sizes and time



(a) 1980

(b) 2015

Figure 9: Distribution of OS index across city sizes and time

G Estimation Details

G.1 Standard Errors

The estimator $\hat{\theta}$ solves

$$min_{\theta} \sum_i \omega_i \left(\frac{\bar{m}_i - m_i(\theta)}{\bar{m}_i} \right)^2 .$$

The variance covariance matrix of the estimator is

$$\hat{V} = (\hat{M}'\Omega\hat{M})^{-1}\hat{M}'\Omega\hat{\Sigma}\Omega\hat{M}(\hat{M}'\Omega\hat{M})^{-1} \tag{117}$$

where $\hat{\Sigma}$ is the variance covariance matrix of the moments m_i. $\hat{M}$ is the Jacobian of the moments with respect to the parameters. And Ω is the weight matrix, here $\Omega = diag(1)$.

G.2 List of Softwares

Table 31: Software assignment to Occupations

Occupation Category	Software
routine cognitive	Entreprise Resource Planning
	Document Management
	Supply Chain Management
	Human Resource
non-routine cognitive	Entreprise Management
	Business Intelligence
	Datawarehouse
	Development
	Workflow

References

Acemoglu, D. and D. Autor (2011): "Skills, tasks and technologies: Implications for employment and earnings," *Handbook of labor economics*, 4, 1043–1171.

Acemoglu, D. and P. Restrepo (2017): "Robots and jobs: Evidence from US labor markets," *NBER working paper*.

Albouy, D. (2012): "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," .

Autor, D. and D. Dorn (2013): "The growth of low-skill service jobs and the polarization of the US labor market," *The American Economic Review*, 103, 1553–1597.

Autor, D. H. (2019): "Work of the Past, Work of the Future," *AEA Papers and Proceedings*, 109, 1–32.

Baum-Snow, N. and R. Pavan (2013): "Inequality and city size," *Review of Economics and Statistics*, 95, 1535–1548.

Beaudry, P., M. Doms, and E. Lewis (2010): "Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas," *Journal of Political Economy*, 118, 988–1036.

Bessen, J. (2002): "Technology adoption costs and productivity growth: the transition to information technology," *Review of Economic dynamics*, 5, 443–469.

Bloom, N., L. Garicano, R. Sadun, and J. Van Reenen (2014): "The distinct effects of information technology and communication technology on firm organization," *Management Science*, 60, 2859–2885.

Byrne, D., S. Oliner, and D. Sichel (2017): "Prices of High-Tech Products, Mismeasurement, and Pace of Innovation," .

Byrne, D. M. and C. A. Corrado (2017): "ICT Prices and ICT Services: What do they tell us about Productivity and Technology?" .

Cortes, G. M., N. Jaimovich, C. J. Nekarda, and H. E. Siu (2014): "The micro and macro of disappearing routine jobs: A flows approach," Tech. rep., National Bureau of Economic Research.

Cortes, G. M., N. Jaimovich, and H. E. Siu (2016): "Disappearing Routine Jobs: Who, How, and Why?" .

Davis, M. A. and F. Ortalo-Magné (2011): "Household expenditures, wages, rents," *Review of Economic Dynamics*, 14, 248–261.

Doms, M. and E. Lewis (2006): "Labor supply and personal computer adoption," Tech. rep., Federal Reserve Bank of Philadelphia.

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

Ellison, G. and E. L. Glaeser (1997): "Geographic concentration in US manufacturing industries: a dartboard approach," *Journal of political economy*, 105, 889–927.

French, E. and C. Taber (2011): "Chapter 6 - Identification of Models of the Labor Market," Elsevier, vol. 4 of *Handbook of Labor Economics*, 537 – 617.

Ganong, P. and D. Shoag (2017): "Why has regional income convergence in the US declined?" *Journal of Urban Economics*, 102, 76–90.

Goldin, C. D. and L. F. Katz (2009): *The race between education and technology*, Harvard University Press.

Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect inference," *Journal of applied econometrics*, 8.

Gyourko, J., A. Saiz, and A. Summers (2008): "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 45, 693–729.

Jovanovic, B. and P. L. Rousseau (2005): "General purpose technologies," *Handbook of economic growth*, 1, 1181–1224.

Katz, L. F. and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107, 35–78.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000): "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 68, 1029–1053.

Lucas, R. E. and E. Rossi-Hansberg (2002): "On the internal structure of cities," *Econometrica*, 70, 1445–1476.

Oyer, P. and S. Schaefer (2016): "Firm/Employee Matching: An Industry Study of US Lawyers," *ILR Review*, 69, 378–404.

Saiz, A. (2010): "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 125, 1253–1296.