# Essays on Labor Markets

Christoph Hedtrich

TESI DOCTORAL UPF / year 2019

THESIS SUPERVISOR
Jan Eeckhout
Department Department d'Economia i Empresa

Universitat
Pompeu Fabra
Barcelona

*A Giulia.*

# Acknowledgements

Since day one in the PhD I had Jan Eeckhout advising and supporting me. If I needed anything, he would help. He guided me through my many attempts at research and gave helpful advice to put me back on track again and again. I always enjoyed our discussions and particularly cherished them during the initial stages of the PhD, as they helped me to start my own research projects.

Markus Poschke came to visit Pompeu Fabra during the early stages of my thesis, he had an open ear to discuss research and had helpful comments along the way. During the later period of my PhD Edouard Schaal came to Barcelona and pushed me towards finishing my projects. He always had thoughtful comments and helped me unasked in a time where I was lost, for which I am very thankful. Another guiding voice throughout the PhD was Roberto Pinheiro. We met through Jan for our joint project, but our discussions spread much further and are always entertaining and insightful.

During the writing of probably every PhD thesis there are many ups and downs. It was a pleasure to enjoy the good days with my friends, but they would also cheer me up during the bad days. I will mention only a few here by name, but there are many more to whom I am also very thankful. Florens and Alex lightened up my day many times with their thoughtful discussions about which stones they like to lift and throw around the beach. Adrian tried to introduce me to the pleasures of rock climbing, which was always fun and tiring, making for good breaks from the PhD. Another welcome break was snowboarding, for which I have to thank Adrià. He took me to the Pyrenees and taught me how to get down the hill in one piece. Back in Barcelona we shared many entertaining dinners with him and Ana. During my visit to London, it was Julen, Divya and Federica who made me feel at home. Tom, despite him leaving soon for Milan after my arrival in Barcelona, we met many times in Italy and around the world. He would always have an open ear and guiding words, and probably the most important: he would take me to football matches. A special mention must go to Korie, with whom I lived together for some time and would treat me to good food very very often. With Flavio I shared many coffees and discussions and I thank him in particular for helping this thesis to make its last steps to UPF.

Finally, the one person without whom I could not have done the PhD is Giulia. She was there enjoying the fun breaks with me, but also supporting me during all the stressful, nerve-racking times. When I lost my path and could not find my way to finish whatever task was at hand, she would always help me get back. For this and so much more, I am deeply grateful.

# Abstract

This thesis is composed of three articles. The first addresses the recent decline in worker mobility between jobs and links it to changes in demand for jobs along the skill distribution. I present a novel theoretical framework, where workers compete for and sort to jobs in a frictional labor market. Using the framework I analyze to what extent the recent decline in worker mobility can be attributed to changing demand for jobs. The second article analyzes how the advancement of Information and Communication Technology affects the sorting of workers and jobs across cities. I provide a theoretical framework allowing for rich interactions between technology usage, occupation choice and location choice of workers. I show that the productivity advancement of IT technology can rationalize the changes in sorting patterns of workers across cities. The last article documents gender pay gaps among professors at the University of California and shows that even after conditioning on research output women are paid less than men.

# Resum

Aquesta tesi està formada per tres articles. El primer estudia la recent disminució de la mobilitat laboral (en termes de canvis de feina) i la relaciona amb canvis en la demanda de feina per part de treballadors amb diferents nivells d'habilitat. Es proposa un nou marc teòric en el qual, dins d'un mercat laboral amb friccions, els treballadors competeixen entre ells i es distribueixen els llocs de treball. Amb aquest marc, s'analitza la mesura en la qual els canvis en la demanda de llocs de treball poden explicar la disminució recent de la mobilitat laboral. El segon article analitza com el progrés de les tecnologies de la informació i la comunicació afecta a la distribució de feines i treballadors entre les diverses ciutats. S'ofereix un marc teòric ric que permet molts tipus d'interaccions entre l'ús de la tecnologia, l'elecció d'una ocupació i la tria d'una ciutat per part dels treballadors. Es mostra que l'increment de la productivitat de les TIC pot racionalitzar els canvis observats en els patrons de distribució dels treballadors entre les diverses ciutats. El darrer article documenta l'existència de diferències de gènere en els salaris dels professors de la Universitst de Califòrnia i es mostra que, fins i tot quan es té en compte la diferència de producció científica, les dones reben una menor retribució que els homes.

# Contents

# List of Figures

# List of Tables

# Introduction

This thesis studies the labor market outcomes of heterogeneous workers. The first two chapters focus on the impact of recent technological change on workers. I highlight the competition between workers and how the sorting of workers to jobs and cities changes in response to new technologies. In the third chapter I study gender differences in pay among professors.

In the first chapter I study the relationship between the mobility of workers between employers and the polarization of employment into low and high skill job. Over the last two decades labor market dynamism, measured by flows of workers between employers, declined substantially in the US. During the same period employment polarized into low and high skill jobs. This chapter shows that the two trends are linked. First, I provide a framework to study employment and worker flows, where skill intensity of jobs and workers' skills are complements. I analyze within this framework the effects of routine-biased technological change and the increasing supply of college graduates on labor market flows. When routine-biased technological change displaces mid-skill jobs, it lowers the opportunity to move up to better jobs for low-skilled workers. Similarly, high skilled workers have less opportunity to take stepping stone jobs and are more likely to start employment further up the job ladder, reducing the frequency of transitions between employers. The rising share of college graduates puts further pressure on labor markets by increasing competition for jobs from top to bottom. In equilibrium workers trade down to jobs with lower skill intensity to gain employment, but find it harder to move up as they are competing with more highly educated workers. I quantitatively assess whether such mechanisms contribute to the fall in labor market dynamism, by estimating the model using data on labor market flows. I find that routine-biased technological change accounts for $40\%$ of the decline in job-to-job mobility.

In the second chapter I analyze the role of advancing Information and Communication Technology for the sorting of workers across cities. I present evidence showing that more expensive cities – measured by rental costs – have not only invested proportionately more in automation (measured by either IT budget per worker or investment in Enterprise Resource Planning software) but have also seen a higher decrease in the share of routine cognitive jobs (clerical workers and low-level white collar workers). I 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. I show that if computers are substitutes to middle skill workers – commonly known as the automation hypothesis – in equilibrium large and expensive cities invest more in computers and software, substituting middle skill workers with computers. Intuitively, in expensive cities, the relative benefit of substituting computers for routine cognitive workers is higher, since workers must be compensated for the high housing prices. Moreover, if the curvature of the production function is the same across skills, the model also delivers the thick tails in large cities' skill distributions presented by Eeckhout et al. (2014).

In the third chapter I study gender differences in pay among university professors. Using data from the University of California I compare the pay that male and female professors receive. I find a large gender gap in pay: women are on average paid 17% less. First, I analyze to what extent the pay gap can be explained by differences in pay across fields and institutions. Both institution of employment and field can not explain the gender gap in pay. However, many women work in lower paying positions which explains about half the gender gap in pay. Then I turn to the effect of research productivity, as measured by citations, on pay. While research productivity can predict differences in pay in general, it can only explain a small part of the gap in pay across men and women conditional on position.

# Chapter 1

# LABOR MARKET DYNAMISM AND JOB POLARIZATION

## 1.1 Introduction

In the last two decades the US labor market experienced a decline in labor market mobility and job polarization, that is a shift in employment away from mid skill jobs towards low and high skill jobs. While technology has been identified as an important driver of polarization in employment, the underlying causes for the decline in labor market mobility, as measured by job-finding rates on and off the job, are less clear.[1] In this paper, I argue

[1] At the same time, the decline in worker mobility raised concerns about the limited opportunities workers have to move to better jobs. See Moscarini and Postel-Vinay (2016) and Abel et al. (2018) for evidence of a "failing" job ladder in recent years

that the recent decline in worker mobility is driven by the displacement of mid-skill jobs and further intensified by the increasing supply of college graduates. The displacement of mid-skill jobs leaves low skilled workers with less opportunity to move up the job ladder. Thus, they are moving less between jobs. High skilled workers are less likely to find a stepping stone job in the middle of skill distribution and start out employment directly in higher skill jobs. As they start out employment further up the job ladder, they also move less between jobs. Such changes in the demand for skills have been accompanied by a large increase in the number of college educated workers. Additional high skilled workers intensify competition for high skilled jobs and in response workers trade down to lower skill jobs, that is competition trickles down the job ladder and intensifies at all types of jobs. The trickle down of competition makes it harder for everyone to move up the job ladder, leading to a further decline in job-to-job mobility.

First, I propose a novel theoretic framework that links the allocation of employment across jobs with worker mobility. The model embeds production with heterogeneous occupations and workers into a directed search model of the labor market. The setup highlights that, when workers' have a comparative advantage in some jobs, the division of production into occupations will depend both upon the relative productivity of occupations and the supply of skills. Furthermore, as workers compete with each other for jobs, the incentives for job search depend not only

Closely related, there is also concern about whether college graduates are increasingly employed in jobs that do not require a college degree, see for example Abel and Deitz (2014) for a discussion of the employment of college graduates in recent years.

upon the value of employment but also on the composition of the pool of applicants. Thus, the allocation of workers to jobs, their mobility between jobs and overall employment are determined jointly in equilibrium. To capture the key characteristics of the labor market the model incorporates search frictions, on-the-job search and endogenous termination of jobs. Furthermore, the model allows for two-sided heterogeneity and sorting. These features are essential to study labor market flows in the presence of rich heterogeneity and assortative matching, as observed in the data. The allocation of workers to jobs is not random, for example college graduates are more likely to work as managers, while high school graduates are more likely to work as waiters. Occupations with different levels of productivity coexist because they are imperfect substitutes. For instance, there are jobs as managers and waiters. In the model, the production of two waiters are perfect substitutes, but the output of waiters and managers are not. Thus, in equilibrium the relative price of output across occupations adjusts to ensure that job posting in all occupations is optimal.

Then, I proceed by applying the framework to study the recent experiences in the US labor market. To provide a quantitative assessment of the importance of technology for the decline in job-finding rates I estimate the main model parameters using labor market flow rates, separately for the late 1990s and the most recent years. For the estimation I group jobs based on two criteria: (1) whether the job's tasks are predominantly routine and (2) whether the job has mainly cognitive or manual skill requirements. Furthermore, I group workers based on their education level as a proxy for their skill level. Then, I fit the model to data on job-finding rates and vacancies. The model can capture well the observed distribution of job-finding rates by education-occupation group. Using the estimated parameters, I analyze to what extent routine-biased technological change explains the decline in job-to-job mobility. I find that by itself routine-biased technological change can explain approximately $40\%$ of the overall decline.

**Relation to Literature.** First, this article builds upon and contributes to the literature on the recent decline of labor market mobility. Davis and Haltiwanger (2014) and Hyatt and Spletzer (2013) provide empirical evidence for a decline in labor market mobility and argue that while composition shifts in the labor force are important, they can only explain 30-40% of the decline in mobility. Furthermore, they provide evidence that shifts in employment across industries has not been a driver of the decline as workers reallocate towards industries with traditionally higher turnover. In this study, I build upon their evidence, but focus upon a novel explanation of the decline in mobility. That is, changes in composition of the supply of and demand for skills have far-reaching equilibrium effects on labor markets.

Cairo (2013) studies the effect of increasing training costs on turnover in a random search model with large firms. She finds that increasing training costs, acting as a fixed cost to hiring that is subsequently lost when separating, decreases turnover. By increasing the cost of match formation the willingness to sustain matches under bad conditions increases and thus turnover declines. Fujita (2015) argues that increasing "turbulence" - a higher rate of skill loss at separation from employment - can explain lower turnover. The logic behind his finding is very similar to

Cairo (2013), but instead of an increase in the fixed cost of hiring there is an increase in the cost of separation. Both papers argue that their findings can explain a joint decline in job-finding and separation rates. In the descriptive analysis of labor market flows, however, I find that separations to non-employment conditional on a workers education level are increasing while job finding rates decline over the last two decades. This paper contributes to the findings of those papers by analyzing worker mobility in a framework with sorting and on-the-job search, two essential features of the data, and providing a rationale for declining worker mobility in the absence of changes in matching and separation of costs.

Engbom (2017) highlights aging and its interaction with firms hiring decisions and innovation as a force driving down labor demand and turnover. Mercan (2018) argues that the availability of information about workers has increased and thus allows tighter selection at the hiring stage, leading to fewer job-to-job moves. While these papers address potential explanations for the decline in mobility and employment, they do not address the sorting of workers to jobs and whether the decline in mobility is related to changes in sorting patterns. One exception is recent work by Eeckhout and Weng (2018) who study mobility and sorting. They focus on changes in the complementarity between workers' unobserved skills and jobs technology, but I focus on changes in demand for and supply of skills. While these papers study related questions they focus on different mechanisms and the importance of each mechanism for the decline in mobility is still an open question. Thus I consider them complimentary to this paper. The main contribution of my paper is to analyze worker mobility in a setting where there is not only sorting, but also competition between workers leading to rich equilibrium interaction between worker mobility and the demand for and supply of skills.

Second, this study also contributes to the literature on models with search frictions and sorting in the labor market. Barnichon and Zylberberg (2018) consider a setup of the labor market with similar features as in this paper and analyze employment by education level of workers over the business cycle. They find that highly-educated workers are downgrading towards low-skill jobs in downturns, which leads to more unemployment for workers with less education as high-skilled workers are preferentially hired. This paper is based on a similar job competition mechanism and they provide outside evidence that the mechanism is relevant for the allocation of workers to jobs. Though related, they do not focus on the trend in worker mobility and its possible causes. Furthermore, they do not include on-the-job search, which is at the core of this paper. Lise and Robin (2017) also study sorting over the business cycle, but use a random search framework that, in contrast, does not feature explicit competition at the hiring stage. While, they address only business cycles and I focus on trend changes in the labor market, it is also the key mechanisms of how sorting happens in the labor market that are different. I focus on competition between applicants and directed search, while in their framework sorting is entirely based on matching sets. By allowing for competition between workers at the hiring stage, I can address to what extent high skilled workers crowd out lower skilled workers from particular jobs and employment.

Third, the current article is also closely related to the literature on technological change, job polarization and wage inequality. Following the contributions by Goos and Manning (2003) and Autor et al. (2003) a large literature has analyzed how technology can explain job polarization and other labor market outcomes, for instance Acemoglu and Autor (2011), Goos et al. (2014) and Stokey (2016). Cortes et al. (2017) build on this literature and study a frictionless model of the labor market to analyze to what extent the declining labor force participation rate can be explained by technological factors. In this paper I proceed in a similar manner, but focus instead on the role of technology for job search both on and off the job. Beaudry et al. (2016) and Aum (2017) provide evidence that the supply of educated workers outpaced the demand for skilled workers since 2000. In this paper I find a similar pattern and will take into account both shifts in demand for jobs and the supply of educated workers. Aum et al. (2018) argue that the negative effect of "routinization" on aggregate productivity growth was not visible due the rise in productivity of the computer industry until the 1990s, which became a more important input across all industries over the same period. This is in line with the findings in Jaimovich and Siu (2012) and this paper, as the decline routine employment is concentrated in the period after 2000.

The remainder of the article is organized as follows. Section 1.2 provides a descriptive overview of the recent trends in worker mobility and employment. In Section 1.3 I lay out the theoretical framework. The structural estimation setup follows in Section 1.4, where I discuss identification and present the estimated parameters and model fit. In Section 1.5 I perform the decomposition of the decline in labor market flows using the estimated model. The last section offers concluding remarks.

## 1.2 Descriptive Evidence

### Data Sources and Sample Selection

The CPS Basic Monthly files for the period 1994 to 2017 are the main source of data. The raw data are provided by Sarah et al. (2018). Occupations are categorized based on their cognitive requirements and routine task content following Autor et al. (2003), see table 2.4 for an overview. The grouping into routine vs. non-routine jobs captures to what extent occupations are exposed to displacement by automation technology. The differentiation along cognitive skill requirements allows to distinguish jobs with high vs low cognitive ability requirements. I connect the jobs cognitive skill requirement to workers by using education levels as a proxy for cognitive skills. In the main analysis I use three groups for education levels: (1) at most a high school degree (2) some college, but not a full four year degree and (3) a four year college degree or more. In order to exclude individuals in education and close to retirement, I restrict the sample to individuals of age 25-45. All calculations use CPS sample weights.

### Decline in Worker Mobility and Job Polarization

In this section I present evidence for a trend decline in worker mobility and job polarization. Over the last two decades there was a substantial decline in job finding rates both on and off the job.

Figure 1.1 shows in panel 1.1a the job-to-job transition rate and in panel 1.1b the job finding rate from non-employment. The job-to-job

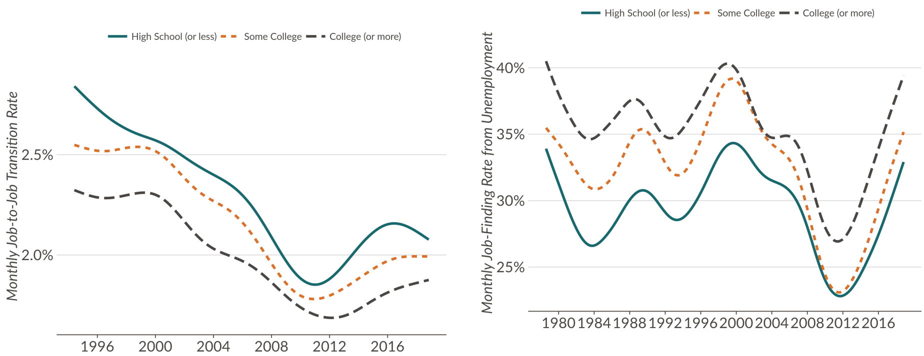
Table 1.1: Occupation Groups by Tasks

| 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 |
| Routine Cognitive | Sales and Related |
| | Office and Administrative Support |
| Routine Manual | Construction and Extraction |
| | Installation, Maintenance and Repair |
| | Production |
| | Transportation and Material Moving |
| Non-routine Manual | Service Occupations |

See Cortes et al. (2014) for details on classification and mapping to Census Occupation codes.

transition rate declined by over 20% between 1996 and 2016 for workers of all education levels. The decline in the switching rate between jobs has been remarkably common between workers of different education levels, which points towards broad based changes in the labor market. The job finding rate out of unemployment has declined somewhat. Again, the behavior over time is remarkably common for workers of different education levels. The trend decline in job finding rates can be driven by many factors related to the value of employment, costs of creating worker-employer re-

Figure 1.1: Job Finding Rates




(a) Job-to-Job Transition Rate. Own Calculations using CPS Basic Monthly Files. Trend calculated from monthly transition rates using HP Filter with smoothing parameter 129600.

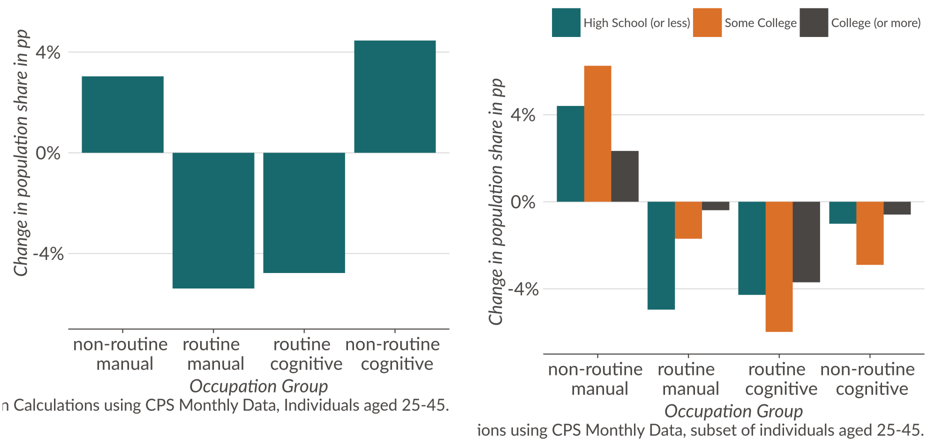
(b) Unemployment-Employment Transition Rate. Own Calculations using CPS Basic Monthly Files. Trend calculated from monthly transition rates using HP Filter with smoothing parameter 129600.

lationships and frictions in the labor market. In this paper, I focus on how changes in technology affect labor demand and in turn the distribution of potential jobs a worker can obtain. Particularly, I document that employment shifted away from mid skill (routine) employment towards low and high skill (non-routine) jobs. This trend has been called job polarization and a large literature following the contributions of Autor et al. (2003) and Goos and Manning (2003) has argued that routine biased technological change is behind such changes, but also that trade and off-shoring are other potential causes (Autor et al., 2016, Blinder and Krueger, 2013). Here I do not focus on the specific causes for changes in the composition of labor demand across jobs, but on its impacts on workers mobility and will henceforth combine those different mechanisms under the term "technology".

Figure 1.2: Change in Employment per Capita by Job Type: 1996-2016

(a) Aggregate

(b) Conditional on Education




In Figure 1.2a I show the change in employment per capita between 1996 and 2016 for each occupation group, as defined in table 2.4. Employment rose in non-routine jobs, while employment in routine jobs declined. The rise in non-routine employment took place both at the bottom and top of the wage distribution, while the decline in routine employment is situated in the middle of the wage distribution[2]. This trend has been

[2]The relative pay of these occupation groups has been widely documented. See appendix 1.6 for the weekly earnings of those occupation groups, calculated using the CPS outgoing rotation group.

termed “Job Polarization” by Goos and Manning (2003). This shift in composition of employment is closely connected to worker mobility. The different types of jobs form a job ladder that workers try to climb. However, the part of the ladder which is relevant for a worker depends upon her education level. For instance, for workers with at most a high school degree employment is concentrated in non-routine manual and routine jobs. For college educated workers instead employment is concentrated in routine and non-routine cognitive jobs[3]. Thus, as employment shifts away from mid skill jobs it becomes harder for workers with low education levels to move to better jobs, as they have low job finding rates at high skill jobs[4]. They can not easily move to high skill jobs, because they have to compete with college educated workers whose skills are likely more suited for such jobs. Thus, I argue that the opportunity to move up out of low skill employment have diminished for workers with low education levels. For workers with a college degree the decline in demand for mid-skill jobs, instead means that they have fewer opportunities to take a stepping stone job. Therefore, once they find employment they are on average more likely to be employed further up the job ladder, and thus they are less likely to move up. However, as workers compete with each other for jobs it is not only the demand for jobs, but also the supply of educated workers that is linked to mobility and employment. In Figure 1.2b the change in employment by occupation group is shown again, but conditional on a workers education level. There is a clearly distinct

[3]See appendix 1.6 for the distribution of employment by job type and education level of workers

[4]The differences in job-finding rates from non-employment to jobs by education and occupation are shown in section 1.4.

pattern in the cross-section compared to the aggregate. First, there does not seem to be an increase in employment in non-routine cognitive jobs. This is driven by the increase in supply of college graduates by over 10pp over the same time period, as shown in appendix 1.6. Therefore, conditional on a workers education level employment shifts towards low skill jobs. This suggests that the supply of college graduates outpaced demand for high-skill jobs which in turn puts pressure on labor markets from top to bottom. This interpretation is further corroborated by the evidence in Beaudry et al. (2016) and Aum (2017). For the main analysis in the paper I will therefore not only take into account potential changes in the demand for skills, but also in the supply of skills.

## 1.3 Framework

In this section I develop an equilibrium framework of the labor market incorporating skill heterogeneity across workers and technology differences across jobs. The framework allows for sorting and endogenous mobility of workers. Output from different occupations is aggregated into a final good with a finite elasticity of substitution. As I focus on stationary equilibria I drop time as a subscript.

**Agents and Technology.** Time $t$ is continuous. There is a measure one of risk-neutral workers in the economy. Workers differ in their level of skill $x = 1, \cdots, X$ which has an exogenous distribution $G(x)$. A worker is either unemployed and searching for a job or employed and searching for another job. The worker chooses search effort $s$ at cost $c(s)$, which is increasing and convex. The search effort cost on the job is multiplied by a constant $\phi_1$, capturing potential differences in the level of search costs on and off the job. Each unit of search effort translates into a proportional increase in the job finding rate. Workers also direct their job search, that is they observe the distribution of vacant jobs and choose to which vacancy to apply for. Among vacancies between they are indifferent workers potentially randomize. Furthermore, I assume that workers can not coordinate their applications, that is application strategies treat two vacancies with the same characteristics in the same way[5]. This assumption gives rise to matching frictions, as identical vacancies receive zero, one or many applications. This leaves some vacancies unfilled, while other vacancies have to turn away applicants.

There is a large measure of potential jobs. Each job chooses its occupation $y$ before entry. There are $y = 1, \dots, Y$ occupations, which are ordered by their skill intensity $y$. The productivity of labor $f(x, y)$ in a job of type $y$ depends both upon the workers skill $x$ and the jobs occupation $y$. Furthermore, flow output depends upon match-specific productivity $\epsilon$, which is redrawn at rate $\theta_y$ from the distribution $F_y(\epsilon)$. The flow revenue of a job of type $y$ employing a worker of type $x$ is flow output times price $p_y \epsilon f(x, t)$. The price of output $p_y$ of an occupation is determined in equilibrium. The allocation of workers to jobs in equilibrium will then strongly depend upon the properties of $f(x, y)$. The differences in productivity across jobs driven by $y$ and $\epsilon$ form a job ladder for workers, which will also depend upon the workers human capital level $x$ through

[5]See Shimer (2005a) for a discussion of this assumption and how it gives rise to matching frictions.

its impact on labor productivity $f(x, y)$. A new job opens by posting a vacancy at flow cost $k(y)$. The amount of entry of vacant jobs into the different occupations will be determined in equilibrium and the price of occupation output will adjust accordingly. The output of individual jobs within an occupation are perfect substitutes. Thus, occupation output follows $Q_y = \sum_x \int \epsilon f(x, y)e(x, y, \epsilon)\, d\epsilon$. The output of each occupation is turned into a single final good $Q_F$ by a CES aggregator with elasticity of substitution $\sigma$, that is $y_F = \left[\sum_y \omega_y Q_y^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$, where $\sum \omega_y = 1$. The production shares $\omega_y$ would allow for a situation where more productive jobs do not represent a larger share of employment, which occurs when their output represents only a small share of inputs in final goods production. The market for occupation output $Q_y$ is competitive. Therefore, the input costs of final goods producers will exhaust revenue. The final good is the numeraire $p_F = 1$.

**Labor Market Frictions and Search.** Meetings between workers and jobs are stochastic and are modeled by an urn ball matching function, most closely related to the static setup in Shimer (2005a)[6]. A model with similar features as Shimer (2005a) and the current setup was studied in Shi (2002). A worker applies for jobs sequentially, but many applications potentially arrive simultaneously at a job. Jobs hire their preferred candidate, as they can only hire one worker. In comparison to the standard setup job finding rates of workers do not simply depend upon the

[6]See Peters (1991) and Burdett et al. (2001) for micro foundations of the urn ball matching function. Shimer (2005a) extends this to a setting with two-sided heterogeneity where jobs rank workers.

overall tightness of the labor market, but also on the ranking among the set of applicants. In order to incorporate on-the-job search, which alters the outside option of a worker at time of hiring, I extend the type space. A worker is now described by a tuple $x, S_o$ where $S_o$ denotes the value of his outside option over unemployment. If there is no match-specific heterogeneity $B_y(x, S_o)$ denotes the set of workers ranked above worker $x, S_o$. However, the match-specific productivity $\epsilon$ is drawn in the moment when workers and jobs meet, therefore the set of better ranked workers will also depend upon $\epsilon$, that is $B_y(x, S_o, \epsilon)$. Now, I define job finding rate for a worker of type $x, S_o$ who sends an application to a job of type $y$. For that define a queue of workers $\lambda_y(x, S_o)$ as the effective number of searchers of type $(x, S_o)$ applying for type $y$ vacancies over the number of vacancies $v_y$. Then, we can also define the total queue of better ranked workers

$$\Lambda_y(x, S_o, \epsilon) = \sum_{(h', S'_o, \epsilon') \in B_y(x, S_o, \epsilon)} \lambda_y(x', S'_o) f_y(\epsilon').$$

The flow job finding rate at jobs of type $y$ for worker of type $(x, S_o)$ is then

$$\nu_y(x, S_o) = \int e^{-\Lambda_y(x, S_o, \epsilon)} \frac{1 - e^{-\lambda_y(x, S_o) f(\epsilon)}}{\lambda_y(x, S_o)} \{S(x, \epsilon, y) > S_o\} \, d\epsilon.$$

The filling rate for a job of type $y$ by a worker of type $(x, S_o)$ is then $\nu_y(x, S_o)\lambda_y(x, S_o)$, as the urn ball matching function exhibits aggregate returns to scale. The actual job finding rate for a worker not only depends upon the choice where to apply, potentially following a mixed strategy, but also her total search effort $s(x, S_o)$. Search effort translates one-to-one into job finding rates, that is the flow job finding rate conditional on applying for job $y$ is $s(x, S_o)$. Job separations happen at an exogenous rate $\delta$ and when a draw of match-specific productivity below the reservation threshold $\underline{\epsilon}_y(x)$ arrives, so the effective separation rate is $\delta + \lambda_y F_y(\underline{\epsilon}_y(x))$.

**Individual Decision Problems and Bellman Equations.** I denote the value of unemployment by $U(x)$, the value of a vacant job of type $y$ by $V(y)$, the value of a filled job by $J(x, S_o, \epsilon, y)$ and the value of employment for a worker in job $y$ by $E(x, S_o, \epsilon, y)$. Furthermore, I will denote deviations of values relative to outside options by hats, that is $\widehat{E}(x, S_o, \epsilon, y) = E(x, S_o, \epsilon, y) - S_o$. The surplus value of a match is defined as $S(x, S_o, \epsilon, y) = E(x, S_o, \epsilon, y) + J(x, S_o, \epsilon, y) - U(x) - V(y)$. The surplus value relative to the outside option is then $\widehat{S}(x, S_o, \epsilon, y) = S(x, S_o, \epsilon, y) - S_o$.

Workers choose how much to search and at which type of job. Vacant jobs choose which types of contracts to post. Contracts are complete and enforceable, that is jobs and workers commit to fulfilling the conditions of the contract. To describe a workers search decisions define the value of one unit of search effort spend on applications at job type $y$

$$W_y(x, S_o) = \int \nu_y(x, S_o, \epsilon) \widehat{E}(x, S_o, \epsilon, y) \, d\epsilon. \tag{1.1}$$

As workers freely choose to which type of job to apply to, they will only apply to a job of type $y$ if the application has at least as much value as their second best option.

$$W_y(x, S_o) \geq \max_{y'} W_{y'}(x, S_o) \ \perp \ \lambda_y(x, S_o) \geq 0, \tag{1.2}$$

where the two conditions hold with complementary slackness.

The workers search effort solves

$$\max_s sW(x, S_o) - c(s),$$

which has an interior solution $s \geq 0$ as $c(s)$ is increasing, monotone and convex.

The vacant jobs contract posting decision maximizes expected discounted profits. The expected discounted revenue of filling the job is the flow rate at which the job is filled times the total surplus value left after compensating the worker for his outside option. However, a job does not enjoy the remaining value $\widehat{S}$ by itself, but posts contract values $\widehat{E}$ under commitment that promise the worker a specific amount of the remaining value conditional on his characteristics. Following Shimer (2005a) I will formulate the decision problem of the vacant job as one of attracting queues of workers, instead of maximizing over contract values directly. The contract values will be defined implicitly. Using the workers indifference condition (1.2) we can write the vacant jobs problem as

$$\max_{\{\lambda_y(x,S_o)\}} \sum_{x,o} \int \mu_y(x, S_o)\widehat{S}(x, S_o, \epsilon, y)\, d\epsilon - \sum_{x,o} \lambda_y(x, S_o)W(x, S_o), \tag{1.3}$$

where $\lambda_y(x, S_o) \geq 0$. The corresponding set of first order conditions is

$$\begin{aligned} W(x, S_o) \geq & \int f_y(\epsilon)e^{-\lambda_y(x,S_o)f_y(\epsilon)}e^{-\Lambda_y(x,S_o,\epsilon)}\widehat{S}(x, S_o, \epsilon, y)\, d\epsilon \dots \quad (1.4) \\ & - \sum_{x',o'} \iint \mathbf{1}\{\widehat{S}' < \widehat{S}\} f_y(\epsilon)e^{-\Lambda_y(x',S_{o'},\epsilon')} \dots \\ & (1 - e^{-\lambda_y(x',S_{o'})f_y(\epsilon')})\widehat{S}(x', S_{o'}, \epsilon', y)\, d\epsilon\, d\epsilon' \\ \lambda_y(x, S_o) \geq & \ 0, \quad (1.5) \end{aligned}$$

where the two conditions hold with complementary slackness. If no application is attracted $\lambda_y(x, S_o) = 0$ any contract value below what equation 1.4 specifies could be offered, but this indeterminacy is without any consequence as no one applies. Replacing $W(x, S_o)$ with its definition in (1.2) one obtains a definition of expected contract values as a function of queue lengths. I assume that contracts are complete and enforceable, such that they can not only specify the value promised to the worker, but also on-the-job search and continuation decisions in case of match specific productivity shocks. Therefore, contracts will be specified to maximize the total value of the match. See Garibaldi et al. (2016) for a setup with a similar assumption. Contracts will maximize surplus, so we do not need to specify the value of the match separately for the worker and firm. It is sufficient to describe the joint surplus to describe allocations, as the surplus value does not depend on its split between worker and firm. The Bellman equations defining the values thus follow

$$rU(x) = \max_s b(x) + sW(s, 0) - c(s) \tag{1.6}$$

$$rV(y) = -k(y) + \sum_{x,o} \int \mu_y(x, S_o, \epsilon)\widehat{S}(x, S_o, \epsilon, y)\, d\epsilon \ldots \tag{1.7}$$

$$- \sum_{x,S_o} \lambda_y(x, S_o, \epsilon) W(x, S_o)$$

$$(r + \delta)S(x, \epsilon, y) = \max\{0, \max_s p_y \epsilon f(x, y) + sW(x, S) \ldots \tag{1.8}$$

$$- c(s) - b(x) - s(x, 0)W(x, 0) - c(s(x, 0))$$

$$+ \theta_y \int \max\{S(x, \epsilon', y) - S(x, \epsilon, y), 0\} dF_y(\epsilon'))\}. \tag{1.9}$$

Here I already set the derivative of the values with respect to time to zero, as I focus solely on stationary equilibria. See this sections appendix for omitted derivations.

**Distribution of Workers and Jobs.** Denote the unemployment rate of workers with skill $x$ by $u(x)$ and employment in job of type $y$ and match-specific productivity $\epsilon$ by $e(x, y, \epsilon)$. The hiring rate of a worker of type $(x, S_o)$ at a job of type $y$, while drawing $\epsilon$, is $s_y(x, S_o)\nu_y(x, S_o, \epsilon)$ and the rate of separation to unemployment of type $x$ workers at type $y$ jobs is denoted as $\delta_y(x) = \delta + \lambda_y F_y(\underline{\epsilon}(x, y))$ . Then the distribution of workers across unemployment $u(x)$ and jobs $e(x, y, \epsilon)$ evolves according to

$$\dot{u}(x) = -\, u(x) \sum_y \int s_y(x, 0)\nu_y(x, 0, \epsilon)\, d\epsilon + \sum_y \int \delta_y(x) e(x, \epsilon, y)\, d\epsilon \tag{1.10}$$

$$\dot{e}(x, \epsilon, y) = -\, e(x, \epsilon, y) \sum_{y'} \int s_{y'}(x, S')\nu_y(x, S', \epsilon)\, d\epsilon' - e(x, \epsilon, y)\delta_y(x) \tag{1.11}$$

$$+ \lambda_y f_y(\epsilon)\{S(x, \epsilon, y) > 0\} \int e(x, \epsilon', y)\, d\epsilon'$$

$$+ \int s_y(x, S')\nu_y(x, S', \epsilon)\, dH(x, S')$$

$$G(x) = u(x) + \sum_y \int e(x, \epsilon, y)\, d\epsilon \tag{1.12}$$

where $H(x, S')$ is the distribution of individuals by skill $x$ and current surplus value $S'$, $h(x, 0) = u(x)$ and $H(x, S') = u(x) + \sum_y \int e(x, \epsilon, y)\{S(x, \epsilon, y) \leq S'\}\, d\epsilon$ for $S' > 0$.

A stationary distribution satisfies the above law of motion with $\dot{u}(x) = 0$ and $\dot{e}(x, \epsilon, y) = 0$.

**Goods Market** The final good is a CES aggregate of the occupation output $y_F = \left[\sum_y \omega_y Q_y^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$. The intermediate inputs $Q_y$, the occupation level output, is bought at price $p_y$, which is taken as given by competitive final goods producers. Intermediate input demand follows

$$Q_y = \omega_y \left(\frac{p_y}{p_F}\right)^{-\sigma} Q_F. \tag{1.13}$$

The price index of the final good is

$$p_F = \left[\sum \omega_y p_y^{1-\sigma}\right]^{\frac{1}{1-\sigma}}. \tag{1.14}$$

Costinot and Vogel (2010) provide a model with a similar production structure, but analyze solely assignment between workers and jobs without frictions.

**Stationary Equilibrium**

**Definition 1.** *A pair* $\{U(y), V(y), S(x, \epsilon, y), u(x), e(x, \epsilon, y), \widehat{W}(x, S_o), \lambda_y(x, S_o)\}$ $\forall x, o, \epsilon$ *is a stationary equilibrium, if:*

1. *the workers indifference condition* (1.2) *holds;*
2. *the Bellman equations* (1.17)*,* (1.8) *and* (1.9) *hold;*
3. *there is free entry of jobs, that is* $V(y) = 0\, \forall y = 1, \ldots, Y$*;*
4. *the distribution of workers and jobs is constant over time, that is* (1.10) *holds with* $\dot{e} = 0$ *and* $\dot{u} = 0$*;*

5. $p_y$, *the price of occupation output, is such that* (1.13) *and* $Q_y = \sum_x \int \epsilon f(x,y) e(x,\epsilon,y)\, d\epsilon$ *hold;*

6. *the market for applications clears, that is* $s(x,S_o)h(x,S_o) = \sum_y \lambda_y(x,S_o) v_y \,\forall x, o$*;*

7. (1.4) *and* $\lambda_y(x,S_o) \geq 0$ *hold with complementary slackness.*

Computation of equilibria is implemented using the solver for mixed complementarity problems by Ferris and Munson (1999), see appendix 1.6 for details.

### 1.3.1 Examples: Sorting

In the model are several forces that drive sorting and in this section I give examples to clarify those mechanisms. First, I focus on the role of job output $f(x,y)$. Second, I will discuss the role of entry costs. Here sorting is defined as first order stochastic dominance.

**Definition 2.** *An allocation exhibits positive assortative matching (PAM), if the distribution over jobs $y$ for workers of type $x_2$ first order stochastically dominates that of workers with type $x_1$ when $x_1 < x_2$.*

1. *The conditional distribution of employment across jobs for a worker of type $x$ is* $\pi(y|x) = \frac{\sum_{j=1}^{y} \int e(x,\epsilon,j) d\epsilon}{\sum_{j=1}^{Y} \int e(x,\epsilon,j) d\epsilon}$

2. *An allocation exhibits PAM, if* $\pi(y|x_i) \leq \pi(y|x_{i'})$ $\forall i, i' > i \in 1, \cdots, X$ *with the inequality strict for at least one* $y \in \{1, \ldots, Y\}$ *and* $x_i < x_{i'}$*.*

Negative assortative matching (NAM) is defined analogously. Note that I define sorting globally across all pairs of workers. To clarify under which conditions sorting occurs it is useful to consider when no sorting occurs.

**Proposition 1.** *Assume match productivity $f(x, y)$ is log-modular, entry costs are independent of job type $k(y) = k_0$ and the distribution of $\epsilon$ is independent of job type. Then no sorting according to definition 2 occurs in a stationary equilibrium. Proof see appendix.*

The no sorting condition is the same as in the frictionless case when $k_0 \to 0$. In the frictionless limit there are no wage differences across jobs, but even in the case with frictions a similar condition holds in terms of surplus. Thus, the model presented in Shimer (2005a) by itself does not directly generalize to explain evidence from matched employer employee datasets highlighting differences in wages across jobs for similar workers. To explain such evidence, one needs to assume that firms have different entry costs in order to sustain surplus differences across jobs in equilibrium.

For the following examples, consider a simplified version of the model above where the only form of heterogeneity is in worker skill $x$ and job type $y$. There are two types of workers $x_L < x_H$ and two types of jobs $y_L < y_H$, and there is no match specific productivity variation. The home production value is $b(x) = \bar{b}$. In that case, surplus follows

$$(r+\delta)S(x, y) = p_y f(x, y) - b - s(x, 0)W(x, 0) + c(s(x, 0)) - s(x, S)W(x, S).$$

The production technology is

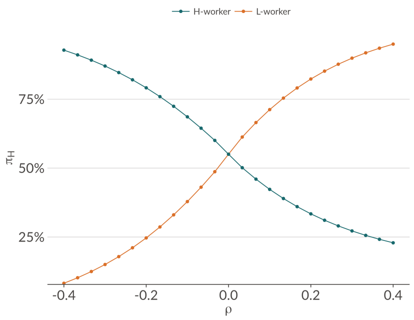
$$f(x, y) = [x^\rho + y^\rho]^{\frac{1}{\rho}},$$

where $\rho$ governs whether skill $x$ and job type $y$ are complementary. In general, the properties of surplus $S(x, y)$ determine sorting.Shimer (2005a) discusses some examples under which sorting arises. However, in this paper there is free entry and therefore higher surplus in some types of job are only sustainable to the extent that they reflect lower filling rates $\mu_y$ or higher posting cost $k$. In equilibrium additional entry will lead to a decrease in $p_y$ up until the free entry condition is satisfied. In the following I give examples in which sorting occurs.

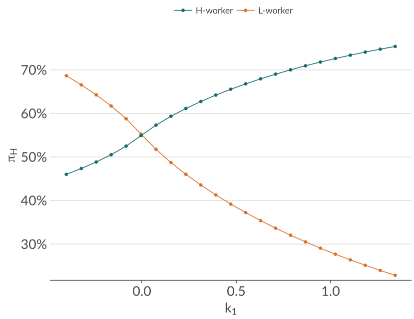
**Comparative Advantage in Production.** Assume $k(y) = k_0 \, \forall y$. Then, the conditions for sorting are the same, as in the frictionless limit $k_0 \to 0$. Costinot and Vogel (2010) show that in the frictionless assignment model sorting arises when $f(x, y)$ is log-supermodular, that is high skill workers have a comparative advantage in high skill intensive occupations. In the current example, the production function is log-supermodular if $\rho < 0$. In the two-type example we can summarize the distribution of workers across jobs, as the share of workers in high skill intensive jobs $\pi_H(x) = \frac{e(x,y_H)}{e(x,y_L)+e(x,y_H)}$. Figure 1.3a plots $\pi_H(x)$ for low and high skilled workers for various values of $\rho$ in a numerical example. The condition for PAM is satisfied if $\pi_H(x_H) > \pi_H(x_L)$. PAM occurs in equilibrium when $f(x, y)$ is log-supermodular. In this example with a CES production function, log-supermodularity holds when $\rho < 0$. When $\rho = 0$, there is no sorting and when $\rho > 0$ ($f(x, y)$ is log-submodular) the allocation exhibits NAM.

The reason that the condition for sorting is not stronger with frictions in the labor market relative to the frictionless case is that jobs select workers at the hiring stage. When they receive multiple applications, they hire the worker delivering the highest value to the firm, which coincides with the worker who provides the highest surplus. Therefore, when deciding which worker to hire the firm ranks according to the same criterion as in the frictionless case and sorting arises

Figure 1.3: Sorting with comparative advantage and heterogeneous entry costs.




(a) Employment Share in $y_H$ Jobs and $\rho$. Log-supermodular (-submodular) production function $\rho < 0$ ($\rho > 0$) implies PAM (NAM). Entry Cost $k(y) = k_0$.




(b) Employment share in $y_H$ jobs with heterogeneous entry cost $k(y) = k_0 y^{k_1}$. Increasing entry cost in job type $k_1 > 0$ implies PAM in absence of comparative advantage $\rho = 0$.

under the same conditions. However, there is mismatch. Some firms receive only applications by $L$ type workers, while others only receive applications by $H$ type workers. Therefore, sorting is not perfect as it would be in the frictionless case. Mismatch is sustained in equilibrium despite directed search, because firms post contracts conditional on worker heterogeneity rendering workers indifferent between applying at different jobs.

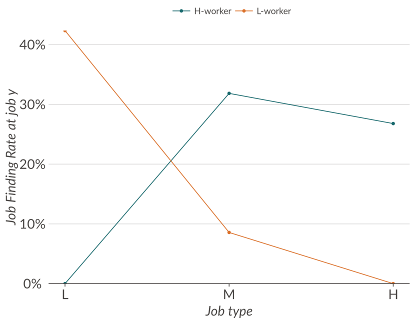
**Heterogeneous Entry Cost.** Differences in entry costs across occupations $y$ induce sorting, even when the production function is log-modular. The reason is that differences in entry costs are reflected in surplus values due to free entry. However, those differences are larger for more skilled workers even in absence of comparative advantage ($\rho = 0$). Consider the same setup, as in the previous example, but with $k(y) = k_0 y^{k_1}$. Figure 1.3b plots the share of employment in high skill jobs $\pi_H$ for low and skill workers. When high skill jobs are more costly to create, $k_1 > 0$, the equilibrium exhibits PAM even with a log-modular production function. When entry cost are increasing in $y$ the productivity advantage of $y_H$ jobs is not fully competed away due to entry. When $k_1 > 0$, the relative price of output of $y_H$ jobs is larger compared to an equilibrium with $k_1 = 0$. As the price of output for high type jobs does not fall as much, surplus can be supermodular without $f(x, y)$ being log-supermodular.

### 1.3.2 Examples: Allocation

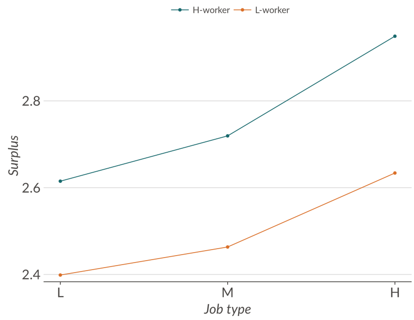
In this section I first show an example allocation to illustrate how job finding rates are affected by competition between workers, not just surplus value of jobs. Then I show how job-to-job mobility reacts to a displacement of mid-skill jobs. In this example I keep with the previous setup, but allow for heterogeneous match-specific productivity. The production function is chosen to be log-supermodular and $\omega_L < \omega_M < \omega_H$ and the posting cost is $k(y) = k_0\omega_y$. The allocation features positive assortative matching, as defined above.

Panel 1.4a shows the flow job finding rate of L,H type workers at L,M,H type jobs. Low type workers only find jobs at L,M type jobs, while high type workers find jobs only at M,H type jobs. Workers segregate to the tails of the skill distribution, but mix in the middle. However, surplus is increasing in job type for both low and high skilled workers, as shown in Panel 1.4b. In equilibrium, low skilled workers do not apply at high type jobs, because the contract offered to them does not compensate them enough for the increased competition by high skilled worker, which lowers their job-finding rate. At mid-skill jobs the productivity advantage of high skill workers is not as large, thus the posted contracts

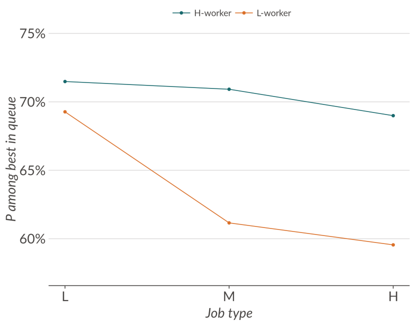
Figure 1.4: Job Finding Rates, Competition and Directed Search



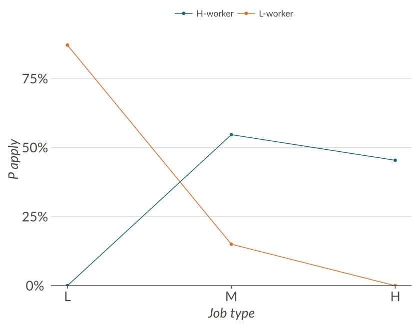

(a) Flow job finding rate of L,H type workers at L,M,H type jobs: $s_y(x,0)\nu_y(x,0)$




(b) Surplus: $S(x,\epsilon=1,y)$




(c) Flow rate of applications at which the worker is the best applicant: $E_\epsilon e^{-\Lambda_y(x,0)}$




(d) Flow rate at which worker applies for job: $s_y(x,0)$

optimally offer sufficient value to also attract low type workers. On the other hand, high type workers require too much compensation in order to be attracted for low type jobs, as their option value of searching is larger. Panel 1.4c shows the rate at which a workers application meets a job and is among the best applicants. High type workers are more likely to be among the best applicants at all jobs. However, the probability to be among the best applicants decreases in job type, as the share of high type applicants increases. The increased competition lowers job finding rates relatively more for low-skilled workers, as they are more frequently sent to the back of the queue. Finally, in equilibrium workers apply to different jobs at different rates. Panel 1.5 shows the flow rate of applications for each type of job. Low skill workers apply predominantly for low skill jobs and high skill workers mix relatively evenly between mid and high skill jobs. Here I focused on workers who are looking for jobs from unemployment, but the exact same mechanism applies for all searchers independent of employment status. Once employed workers continue searching for jobs and move up the job ladder, that is low skill workers stochastically move to mid skill jobs and high skill workers move to high skill jobs.

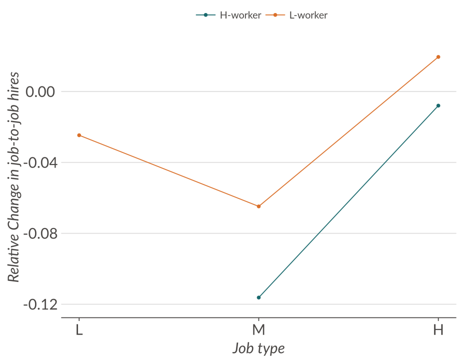



Figure 1.5: Change in job-to-job hires in response to 7% decline in $\omega_M$.

Consider the displacement of mid-skill jobs driven by a decline in $\omega_M$. Job-to-job hires into mid-skill jobs decline as expected because the share of mid skill vacancies decreases. Overall job-to-job hires decline. Workers as a response decrease their search intensity as it became harder to find a job and redirect their search, which is the reason for a subdued response in job-to-job hires at low and high skill jobs. In appendix 1.6 I show evidence of similar changes in job-to-job hires in the data. This indicates, that the displacement of mid-skill jobs is a potential driver of the decline in job-to-job mobility.

## 1.4 Estimation

### 1.4.1 Setup

The goal of the estimation is to identify the structural parameters governing production and matching in the economy. The model parameters are estimated by Indirect Inference following Gourieroux et al. (1993). I pick a set of moments $m$ to identify the model parameters $\theta$. The estimation procedure minimizes the weighted square distance between model $m(\theta)$ and data moments $\bar{m}$ by choosing parameters[7].

$$\min_{\theta} \quad (\bar{m} - m(\theta))'\Omega(\bar{m} - m(\theta)) \tag{1.15}$$

where $\Omega$ is a weighting matrix. The estimation is done separately for the period 1995-1997 and 2015-2017, while treating each allocation as stationary. Discounting is large and the half-life of distributions is short, because labor market flow rates are large. Thus, treating allocations as approximately stationary does not result in large errors. In practice the model parameters are estimated following the approach in Chernozhukov and Hong (2003). The simulation of model parameters by a Markov Chain Monte Carlo (MCMC) method is done using the Differential Evolution Markov Chain (DEMC) approached developed in ter

[7]The model equilibrium is solved for by using the "PATH" solver (Ferris and Munson, 1999). Furthermore, the model implied stationary distribution is used for calculating model moments

Braak and Vrugt (2008). The DEMC allows to efficiently simulate from highly correlated parameter distributions and achieve fast convergence.

**Moments and Identification** To estimate the model parameters I mainly use moments on labor market flows. The reason for not using wage moments is that the theory does not specify a unique wage contract. As wage contracts are not unique the model is consistent with a wide range of observed wage moments and therefore additional assumptions would be needed to use information from wages.

To be estimated are the production function $f(x, y)$, the entry cost $k(y)$, the distribution of match-specific productivity shocks $F(\epsilon)$, the arrival rate of productivity shocks $\theta$ and the search cost parameters $\eta$ and $\phi_1$. The production function is parameterized as

$$\log(f(x, y)) = \alpha_y + \beta_x + \gamma xy \tag{1.16}$$

The worker type $x$ and job type $y$ are specified as uniform spaced points in $(1, 2)$. The parameters $alpha_y$, $\beta_x$ and $\gamma$ are to be estimated. The comparative advantage of workers in different types of jobs is governed by $\gamma$, which can be identified from flows of workers by type $x$ to jobs $y$. I use the flow rate of unemployed workers by education level to jobs by occupation group to identify $\gamma$. The parameter $\beta_x$ governs the relative productivity of workers $x$ and thus can be identified by their relative job finding rates. The occupation level productivity shifter $\alpha_y$ affects the level of employment by job type and can be identified by the job finding rate by job type $y$. The entry cost $k(y)$ affects the surplus value of jobs. Thus, as the surplus value of jobs implies a ranking of jobs in terms of continuation value, the observed job-to-job mobility between job types $y$ identifies $k(y)$. I use the job-to-job hires at a particular job type to identify the entry cost parameters. A similar strategy to rank jobs has been implemented by

Bagger and Lentz (2014), who uses the share of hires from other employers out of all hires. The match specific productivity distribution is parameterized as a two point distribution with equal weight on both points. To be estimated is the distance between the two points $\Delta\epsilon$. We identify the match-specific productivity dispersion by matching the share of job-to-job moves that result in a move down the job ladder in terms of $y$. The arrival rate of shocks to match-specific productivity $\theta$ is identified from job-to-job mobility at high tenures. The search cost parameters $\eta$ and $\phi_1$ are disciplined by job-to-job mobility relative to job finding rates out of unemployment.

The home production value $b(x)$ is set following Hall and Milgrom (2008) and Shimer (2005b) who set the home production value proportional to the average wage. I set $b(x) = 0.7E_y f(x, y)$, where $E_y p_y f(x, y)$ is the average revenue productivity of employed workers of type $x$. A similar strategy for setting the value of home production was used in Lise and Robin (2017). The elasticity of substitution $\sigma$ of occupation level output is set to 3, a value within the range of empirical studies. Table 1.3 summarizes the parameters that are set based on external targets.

## Parameter Estimates

The estimated parameters are summarized in table 1.4a, 1.4b and 1.4c. The production function estimates are summarized in 1.4a for both 1996 and 2016. The productivity shifter $\alpha_y$ across occupations shows the expected ordering, increasing from non-routine manual up to non-routine cognitive occupations. The change over time of the productivity shifter $\alpha_y$ is also in line with expectations, it decreases for routine occupations, while productivity in non-routine occupations increases, but to a much lesser extent. Furthermore, I estimate a (small) positive $\gamma$, which means that the output is weakly log-supermodular in workers

## Table 1.2: Targeted Moments

### (a) Unemployment-Employment Transition Rate

| Education | Occupation | Model 96 | Data 96 | Model 16 | Data 16 | Δ Model | Δ Data |
|---|---|---|---|---|---|---|---|
| | non-routine manual | 6.5 | 6.7 | 8.1 | 7.9 | 1.6 | 1.2 |
| High School | routine | 21.6 | 20.8 | 17.3 | 16.5 | -4.3 | -4.3 |
| | non-routine cognitive | 0.0 | 1.5 | 0.0 | 1.7 | 0.0 | 0.1 |
| | non-routine manual | 4.4 | 5.5 | 7.6 | 7.3 | 3.2 | 1.9 |
| Some College | routine | 23.1 | 21.7 | 18.7 | 17.2 | -4.4 | -4.5 |
| | non-routine cognitive | 6.0 | 7.3 | 4.6 | 6.1 | -1.4 | -1.3 |
| | non-routine manual | 0.0 | 2.9 | 1.6 | 3.7 | 1.6 | 0.8 |
| College | routine | 14.5 | 14.5 | 10.6 | 10.0 | -3.9 | -4.6 |
| | non-routine cognitive | 21.4 | 19.4 | 21.7 | 20.6 | 0.3 | 1.2 |

### (b) Job-to-Job Hires by Occupation

| | Model 96 | Data 96 | Model 16 | Data 16 | Δ Model | Δ Data |
|---|---|---|---|---|---|---|
| non-routine manual | 0.13 | 0.32 | 0.15 | 0.38 | 0.02 | 0.06 |
| routine | 1.11 | 1.36 | 0.53 | 0.84 | -0.57 | -0.52 |
| non-routine cognitive | 1.07 | 0.78 | 1.17 | 0.82 | 0.1 | 0.04 |

### (c) Job-to-Job Moves down Ladder and Decline by Tenure

| | Model 96 | Data 96 | Model 16 | Data 16 | Δ Model | Δ Data |
|---|---|---|---|---|---|---|
| $\frac{j2j_{n'<n}}{j2j}$ | 13.47 | 12.23 | 9.64 | 12.19 | -3.84 | -0.04 |
| $\frac{j2j_{t\in(2,4)}}{j2j_{t\in(0,2)}}$ | 15.35 | 17.55 | 16.23 | 18.65 | 0.88 | 1.09 |

*Notes:* Own Calculations using CPS Basic Monthly Files and Tenure Supplements.

skill $x$ and job type $y$. The estimated value of $\gamma$ hardly changes between the periods. The productivity advantage of college education $\beta_{SC}$ and $\beta_C$ stays roughly constant over time, that is the productivity advantage of higher education that is independent of job type.

Table 1.4b shows the estimated posting cost parameter $k_y$ and the search cost parameters of workers $\phi_1$ and $\eta$. The posting cost parameters decrease somewhat

Table 1.3: External Targets

| | Parameter | Value | Source |
|---|---|---|---|
| Home Production | $b$ | 70% avg productivity | Shimer (2005b)<br>Hall and Milgrom (2008) |
| Posting Cost | $\bar{k}$ | see Table 4 b) | Vacancy Index<br>by Barnichon (2010) |
| EOS Occupations | $\sigma$ | 3 | $\sigma \in [0.3, 10]$<br>Eden and Gaggl (2018) |

and most for routine occupations. For both periods the entry cost parameters are increasing in job type, lowest for non-routine manual jobs and highest for non-routine cognitive jobs. This is consistent with a common job ladder for all worker types and highlights the importance of entry costs for sorting alongside productivity differences across workers. The search cost parameters $\phi_1$ is estimated to be slightly negative, meaning on-the-job search is more efficient than unemployed search. However the difference is small and not statistically significant in both periods. The curvature $\eta$ rises from 4.8 to 5, but is estimated with substantial noise and therefore the difference is not statistically significant.

Table 1.4c shows the estimated productivity dispersion $\Delta\epsilon$ of the match specific productivity shocks and their arrival rate $\theta$. The estimated productivity dispersion between good and bad matches is estimated to be approximately 33% for 1996 and 27% for 2016, indicating a decline in match specific productivity dispersion. The arrival rate of shocks declines from 0.1, on average match specific productivity is redrawn every 10 months, down to 0.06. Those results indicate that match productivity, unrelated to the worker and job characteristics, is becoming less dispersed over time and more persistent.

Table 1.4: Parameter Estimates

(a) Production Function

$$\log(f(x,y)) = \alpha_y + \beta_x + \gamma xy$$

| | $\alpha_{NRM}$ | $\alpha_R$ | $\alpha_{NRC}$ | $\beta_{SC}$ | $\beta_C$ | $\gamma$ |
|---|---|---|---|---|---|---|
| 1996 | -1.8 | 0.44 | 0.36 | 0.17 | 0.29 | 0.014 |
| | (0.033) | (0.048) | (0.058) | (0.0077) | (0.011) | (0.0011) |
| 2016 | -1.6 | -0.34 | 0.39 | 0.15 | 0.33 | 0.016 |
| | (0.017) | (0.014) | (0.015) | (0.0072) | (0.0057) | (0.00052) |

(b) Posting and Search Cost

$$c^e(s) = \phi_1 \frac{s^\eta}{\eta}$$

| | $log(k_{NRM})$ | $log(k_R)$ | $log(k_{NRC})$ | $log(\phi_1)$ | $\eta$ |
|---|---|---|---|---|---|
| 1996 | -0.13 | 0.046 | 0.34 | -0.023 | 4.8 |
| | (0.039) | (0.04) | (0.042) | (0.085) | (0.23) |
| 2016 | -0.22 | -0.18 | 0.27 | -0.11 | 5.0 |
| | (0.014) | (0.014) | (0.019) | (0.041) | (0.16) |

(c) Productivity Dispersion $\Delta\epsilon$ and arrival rate of shocks $\lambda$

| | $log(\Delta\epsilon)$ | $log(\theta)$ |
|---|---|---|
| 1996 | -1.1 | -2.3 |
| | (0.034) | (0.042) |
| 2016 | -1.3 | -2.8 |
| | (0.0047) | (0.02) |

# 1.5 Results

## Mobility Decomposition

To evaluate the importance of technological change, particularly routine-biased technological change, for the decline in labor market mobility I use the estimated model to perform a decomposition of the decline in job-to-job mobility. The model based based decomposition allows me to take into account the rich equilibrium interactions between workers and jobs.

While the main focus is the relative decline in productivity in routine occupations, it is important to also take into account the changes in the supply of college educated labor as the demand for and supply of skills jointly determine mobility rates. The supply of skills has an important effect on mobility, because the rising share of college graduates rises implies that lower skilled workers are more likely to compete with higher skilled workers for jobs and thus their opportunities to move up to better jobs are potentially diminished.

Table 1.5: Decomposition: Job-to-Job Transition Rates

| | $\Delta$ Data in % | $\Delta$ Model in % | $\Delta$ $\alpha$ | $\Delta\alpha$ & $k$ | $\Delta$ $\alpha$ & $k$ & $G(x)$ |
|---|---|---|---|---|---|
| j2j | -18.9 | -19.6 | -7.8 | -6.0 | -4.7 |

Table 1.5 summarizes the results regarding the average job-to-job transition rate in the economy. The first two column show the change in the job-to-job transition rate in the data and model. The model can capture the decline in mobility well. The third column $\Delta\alpha$ shows the change in the job-to-job transition rate when only the productivity level across occupations would change. By itself, the change in productivity can account for roughly $40\%$ of the decline in job-to-job transitions. Taking together the change in productivity and the relative change in entry costs, in column $\Delta\alpha\&k$, we can account for $30\%$ of the overall decline in job-to-job transitions. The decline in entry costs for all job types leads to less transitions between employers, as differences in value across jobs decline moves are less frequent. The increasing share of college graduates actually mitigates the decline in aggregate job-to-job transitions. While these results are indicative that both technology and shifts in skill supply are important for determining worker

mobility, it is illustrative to perform the same decomposition conditional on the education level of workers.

Table 1.6: Decomposition: Job-to-Job Transition Rates by Education Level

| | $\Delta$ Data in % | $\Delta$ Model in % | $\Delta$ $\alpha$ | $\Delta\alpha$ & $k$ | $\Delta$ $\alpha$ & $k$ & $G(x)$ |
|---|---|---|---|---|---|
| High School | -22.0 | -14.0 | 4.6 | 2.7 | 1.9 |
| Some College | -17.3 | -17.8 | -13.8 | -11.1 | -2.7 |
| College | -17.4 | -25.4 | -11.8 | -10.7 | -9.5 |

In table 1.6 I show the results of the decomposition in the decline in job-to-job mobility conditional on the education level of workers. The results show that changes in technology and skill supply have heterogeneous effects across the skill distribution. First, the shift in productivity lowers job-to-job transition rates particularly for mid and high skill workers. However, lower skill workers are actually moving more frequently between jobs. As competition for routine jobs intensifies due to technological change, they are more likely to start employment in non-routine manual jobs right out of unemployment. Therefore, they are more frequently trying to move to better paying jobs, albeit higher competition. Second, the shift in entry costs mitigates the decline in mobility for mid and high skill workers. For lower skill workers it actually leads to less mobility. An increase in the supply of college graduates increases the mobility of workers, particularly for those with some college education. When the share of college graduates rises, they diminish the opportunities for lower skilled workers they compete with. Thus, the competition between workers trickles down from top to bottom. However, there are countervailing forces on job-to-job mobility. Workers sort down right out of unemployment, thus starting out further down the ladder and therefore switch jobs more often once employed. However, the competition by higher skilled workers also makes it harder to move to a different job. In this instance, for workers with some college education the net effect on job-to-job mobility is positive, that is the impact of sorting down out of unemployment leads to more job-to-job transitions despite more competition. There is an additional effect of the rising college share, it leads actually to more entry, as the average worker is becoming more skilled the return to posting a vacancy rises and therefore more jobs are posted. With more vacancies posted it becomes easier for all workers to find a job.

Overall, the demand shifts captured by $\alpha$ can account for over half the decline in job-to-job mobility by workers with some college education and almost half of the decline for college educated workers. However, it can not account for the decline in mobility by workers with at most a High School degree.

## 1.6 Conclusion

A growing literature documents job polarization and declining worker mobility. My analysis suggests that these two phenomena are linked. To study the phenomena, I propose a theoretical framework of the labor market with two sided heterogeneity, search frictions and on-the-job search where the demand for occupations is endogenous. I apply this framework to study the recent decline in job-to-job mobility and find that routine-biased technological change not only gives rise to job polarization, but also shortens the job ladders of workers. With shorter job ladders, workers move less often between jobs and therefore mobility declines. The shifts in demand for labor across jobs can account for $40\%$ of the total decline in job-to-job mobility, where workers without a college degree are affected the most. The results indicate, that to understand recent trends in the labor market it is important to consider underlying changes in demand for and supply of skills. Those shifts mater above and beyond composition, it is their equilibrium interactions that are important to understand the observed trends.

The framework presented in the paper has many possible applications as it provides an appealing way to take into account the role of two-sided heterogeneity and sorting for labor market outcomes. For example, studying sorting based on unobserved heterogeneity using matched employer-employee datasets is a fruitful application of the framework. Card et al. (2013) indicated that sorting based on unobserved heterogeneity may have contributed to rising wage inequality in Germany. However, interpretations of the common two-way fixed effects approach pioneered by Abowd et al. (1999) in terms of sorting are difficult. The model presented here is a useful tool to interpret the findings from reduced form wage regressions. A first step in that direction was taken by Abowd et al. (2014) who apply the static assignment model of Shimer (2005a) directly to US administrative data. However, it is important to take into account entry of jobs and individual labor market dynamics to account for sorting.

Another line of future research would extend the framework to study business cycle dynamics in sorting, employment and wages. This would enrich our understanding of how different types of workers are affected by aggregate transitory fluctuations and what mechanisms they use to insure against income risk.

# Appendices

## Value Functions

**Surplus.** Define the value of unemployment and a vacant job

$$rU(x) = \max_s b(x) + sW(s,0) - c(s) \tag{1.17}$$

$$rV(y) = -k(y) + \sum_{x,o} \int \mu_y(x, S_o, \epsilon) J(x, S_o, \epsilon, y)\, d\epsilon \tag{1.18}$$

A contract specifies a sequence of payments $w$ depending on the worker type $x$, job type $y$, match specific productivity $\epsilon$ and the outside option of the worker

$S_o$ when the worker was hired. Additionally the contract specifies a transfer $P(x, S_o, \epsilon, y)$ between worker and job in case the worker makes a job-to-job move.

$$
\begin{aligned}
rE(x, S_o, \epsilon, y) =& \max_s w(x, S_o, \epsilon, y) + \lambda_y \int \Delta E(x, S_o, \epsilon', y) dF_y(\epsilon') \ldots \\
& - \delta(E - U) + sW(x, E + P - U) - c(s) \\
W_{y'}(x, S_o) =& \int \nu_y(x, S_o, \epsilon)[E(x, S_o, \epsilon', y') - S_o - U] \, d\epsilon' \\
W(x, S_o) =& \max_{y'} W_{y'}(x, S_o) \\
s =& \sum_{y'} s_{y'} \\
rJ(x, S_o, \epsilon, y) =& p_y \epsilon f(x, y) - w(x, S_o, \epsilon, y) + \lambda_y \int \Delta E(x, S_o, \epsilon', y) dF_y(\epsilon') \ldots \\
& - \delta(J - V) \ldots \\
& - \sum_{y}^{\prime} s_{y'} \nu_{y'}(x, E - P - U, \epsilon)[\min\{E' - E, J\} \ldots \\
& + P(x, S_o, \epsilon', y)].
\end{aligned}
$$

I assume that contracts are complete and can be enforced. Therefore, the contract will maximize the total value of the match, as any contract that does not is Pareto dominated.

Now, I specify which penalty schedule $P(x, S_o, \epsilon, y)$ maximizes total match value. To that end separate the set of alternative jobs $\{y', \epsilon'\}$ into two non-overlapping sets: (1) jobs whose total value is lower than that of the current match, (2) jobs whose total value is at least as large as that of the current match. For jobs in the first set, any application presents a net loss in terms of total private value, because the maximum possible amount of contract value offered to the worker can not compensate for the loss of value for the job owner. For jobs in the second set, applications are valuable because the worker will be able to compensate the job owner for his losses. Specify the penalty for a job-to-job move as $P(x, S_o, \epsilon, y) = [\min\{E(x, S_o, \epsilon', y') - E(x, S_o, \epsilon, y), J(x, S_o, \epsilon, y)\}$. It follows that applications to jobs which offer $E' \leq E + J$ offer no value to the worker. Therefore, she will only apply to jobs which offer $E' > E + J$. Note that in equilibrium the value of a vacancy is $V(y) = 0$. It follows, that

$$
\begin{aligned}
rJ(x, S_o, \epsilon, y) =& p_y \epsilon f(x, y) - w(x, S_o, \epsilon, y) \dots \\
& + \lambda_y \int \Delta E(x, S_o, \epsilon', y) dF_y(\epsilon') - \delta(J - V) \\
r[E(x, S_o, \epsilon, y) + J(x, S_o, \epsilon, y)] =& p_y \epsilon f(x, y) \dots \\
& + \lambda_y \int \Delta[E(x, S_o, \epsilon', y) + J(x, S_o, \epsilon', y)] dF_y(\epsilon') \dots \\
& - \delta(E + J - U - V) \dots \\
& - c(s^*) + s^* W(x, J + E - U) \\
S(x, \epsilon, y) =& E(x, S_o, \epsilon, y) + J(x, S_o, \epsilon, y) - U(x) \\
rS(x, \epsilon, y) =& p_y \epsilon f(x, y) + \lambda_y \int \Delta S(x, \epsilon', y) dF_y(\epsilon') \dots \\
& + \max_s sW(x, S) - c(s) \dots \\
& - \delta S(x, \epsilon, y) \dots \\
& - b(x) - s^U W(x, 0) + c(s^U)
\end{aligned}
$$

Surplus is independent of the current surplus split, because the gain from on-the-job search to the worker is only whatever the new job offers above and beyond the total match value of the current match. This is achieved by setting the penalty for a job-to-job move equal to the loss for the job owner. This contract maximizes surplus value, because the worker already maximizes his private value and the jobs valuation of the match is independent of on-the-job search as the job is compensated for any loss.

**Contract Posting.** Each vacant job posts contract values $E(x, S_o, \epsilon, y)$. The flow value of a vacancy follows

$$rV(y) = -k(y) + \sum_{x,o} \int \mu_y(x, S_o, \epsilon) J(x, S_o, \epsilon, y)\, d\epsilon$$

Replacing $J = S - (E - U)$ and using the workers indifference condition

$$W(x, S_o) = \max_{y'} \nu_y(x, S_o, \epsilon')[E(x, S_o, \epsilon', y') - S_o - U]\, d\epsilon',$$

we can replace also $E - U$ and write the value of a vacancy as

$$rV(y) = -k(y) + \sum_{x,o} \int \mu_y(x, S_o, \epsilon) S(x, \epsilon, y)\, d\epsilon - \sum_{x,o} \lambda_y(x, S_o) W(x, S_o).$$

I used $\mu_y = \lambda_y \nu_y$ to simplify the equation.

## Proposition 1

*Proof.* To proof proposition 1, one needs to verify no sorting occurs if $f(x, y)$ is log-modular and entry costs are independent of job type $k(y) = k \geq 0$. For simplicity we also assume $\theta$ and match-specific productivity distribution $F(\epsilon)$ are independent of job type.

Guess that surplus is independent of job-type, that is

$$S(x, \epsilon, y) = S(x, \epsilon, y')\, \forall y, y' = 1, \ldots, Y.$$

Optimal contract posting (1.4) and worker indifference (1.2) then imply that the same contract values are posted for all jobs $y$ and that all jobs have the same job finding rate per unit of search effort.

The free entry condition is

$$k = \sum_{x,S_o} \int \mu_y(x, S_o, \epsilon) \widehat{S}(x, S_o, \epsilon, y) - \lambda_y(x, S_o) W(x, S_o)$$

As surplus is the same across jobs, it follows that also the job filling rate (and queue length $\lambda$) are independent of job type.

As $f(x,y)$ is log-modular we can write it as $f(x,y) = f_1(x)f_2(y)$.

We need to show that $e(x,y,\epsilon) = e_1(x)e_2(y,\epsilon)$. Denote the number of matches created of a type $m(x,S_o,y)$ and following the definition of $\mu$ it holds that

$$m(x,S_o,y) = \mu_y(x,S_o)v_y.$$

Thus the ratio of matches created across jobs

$$\frac{m(x,S_o,y)}{m(x,S_o,y')} = \frac{v_y}{v_{y'}},$$

is the same as the ratio of vacancies, because job filling rates are the same. Thus the distribution of inflow into employment is independent of worker type. There is no sorting in hiring. Then sorting could still occur if for example low skill workers are more likely to separate from high skill jobs than high skill jobs or vice versa. However, as $S(x,y,\epsilon) = S(x,y',\epsilon)$ it directly follows that separation rates are independent of job type. That is workers of different types might separate at different rates from jobs, but they do so independent of a jobs type. Thus, employment rates differ across jobs and workers, but they are independent of each other in the sense that $e(x,y) = e_1(x)e_2(y)$. Output in each occupation follows

$$Q_y = e_2(y)f_2(y)\sum_x \int e_1(x,\epsilon)f_1(x)\,d\epsilon.$$

Thus, the relative price of output across job types follows

$$\frac{p_y}{p_{y'}} = \left(\frac{e_2(y)f_2(y)\omega_{y'}}{e_2(y')f_2(y')\omega_y}\right)^{-\frac{1}{\sigma}}, \tag{1.19}$$

which holds because $v_y$ adjusts. Search effort satisfies the definition of $\lambda_y$ and market clearing $\sum_y \lambda_y(x,S_o)v_y = s(x,S_o)h(x,S_o)$. Thus all equilibrium conditions are satisfied and no sorting occurs.

$\square$

# Computation

In order to solve the system of equations that defines a stationary equilibrium I use the PATH Solver (Ferris and Munson, 1999). The distribution is solved for at any given guess of parameters. The stationary distribution is a solution (a zero) to the linear system (1.10) with the constraint that the supply of workers is exhausted.

## Standard Errors

$$\hat{\theta} = \arg\min_{\theta} \sum_{i} \omega_i \left( \frac{\bar{m}_i - m_i(\theta)}{\bar{m}_i} \right)^2 \tag{1.20}$$

then variance covariance matrix of the estimates $\hat{V}$ is

$$\hat{V} = (\hat{M}'\Omega\hat{M})^{-1}\hat{M}'\Omega\hat{\Sigma}\Omega\hat{M}(\hat{M}'\Omega\hat{M})^{-1} \tag{1.21}$$

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 $\Omega$ is the weight matrix, here $\Omega = diag(\frac{\omega_i}{\bar{m}_i})$

# Additional Descriptive Statistics

In this section I provide some additional labor market statistics related to the main evidence in the paper.

**Occupation and Education Composition and Job-to-Job Transitions.** I perform a shift-share analysis to show that the decline in job-to-job mobility is not simply driven by a reallocation of employment to jobs with lower mobility rates. Consider the following decomposition of the job-to-job transition rate

$$\text{j2j}_t = \sum_{x,y} \text{j2j}_{t,x,y} \pi_{t,x,y},$$

where $\pi_{t,x,y}$ denotes the share of type $x$ workers in jobs $y$ in period $t$. I perform a simple descriptive decomposition, that is I hold the conditional job-to-job mobility rates constant and only let the employment shares vary over time. The share of variation that is explained solely by the shift in employment shares I will attribute to a pure composition shift.

| j2j | 1996 | 2016 | rel. change |
|---|---|---|---|
| Actual | 2.46 | 2.01 | -18.3 |
| Education and Occupation Share | 2.46 | 2.42 | -1.6 |

Table 1.7: Job-to-job Mobility actual vs only composition shift in education and occupations

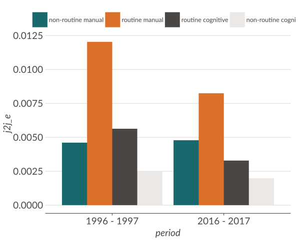
Table 1.7 shows the aggregate job-to-job mobility in the CPS sample, as described in section 1.2, for 1996 and 2016. Then the second row of the table compares this, with the job-to-job mobility rate that would have been observed if job-to-job mobility conditional on education and occupation would have remained constant. The changes in composition can hardly explain any part of the decline, the actual decline is $-18.3\%$ while pure composition explains only $-1.3\%$.

**Job Ladders.** Here I want to illustrate that (1) workers of different education levels move to the different types of jobs at different rates and (2) that job ladders hollowed out in the middle.

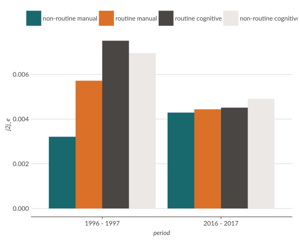
In Panel 1.6a the job-to-job transition rate of High School Graduates split up by the occupation group of the destination occupation is shown. The most likely destination occupation for high-school graduates are routine manual jobs, followed by routine cognitive jobs. From 1996-2016 there has been a substantial drop in such moves to routine jobs. This is consistent with the main hypothesis of the paper. Panel 1.6b shows a similar picture for workers with Some College education. Routine Jobs are still frequently the destination, but moves to non-routine cognitive jobs are also frequent in contrast to lower educated workers. The decline in job-to-job moves for workers with some college education was concentrated in routine jobs, but also there are fewer moves to high-skill non-routine cognitive jobs. At the same time moves to lower skill jobs actually increased. This is consistent with (a) the decline in routine employment and (b) more competition by college graduates. Finally, for College graduates job-to-job moves are mostly to non-routine cognitive jobs. While job-to-job moves to routine jobs also decline, these make up only a small share. The decline in job-to-job moves to high skill jobs are potentially driven by changes in sorting directly out of unemployment and changes in competition. Both mechanisms are captured in the theoretic framework.

Figure 1.6: Job-to-Job Moves by destination Occupation

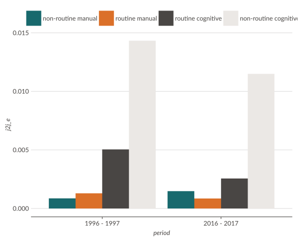
(a) High School Graduates




(b) Some College




(c) College




In table 1.8 the change in job-to-job moves originating in the row occupation group and moving to the column occupation group are shown. Note that job-to-job transitions to routine jobs have declined substantially from all occupations. The decline in job-to-job hires to routine jobs make up 67% of the decline in job-to-job moves originating from non-routine manual jobs. Indicating a strong decline in moves up the job ladder. For job-to-job transitions originating in routine jobs, basically the whole decline in job-to-job moves is concentrated in moves to routine jobs. For job-to-job transitions from non-routine cognitive jobs there is a decline in the moves towards routine jobs, indicating that workers in high skill jobs might take those routine jobs as insurance to not be unemployed and over time this option is diminished. Overall the evidence points towards that the declining demand for routine jobs is closely related to the decline in job-to-job mobility.

(a) Change 1996-2016

| | NRM | RM | RC | NRC | total |
|---|---|---|---|---|---|
| NRM | -0.15 | -0.18 | -0.07 | 0.03 | -0.37 |
| RM | 0.02 | -0.47 | -0.01 | -0.04 | -0.49 |
| RC | 0.06 | -0.0 | -0.49 | -0.05 | -0.48 |
| NRC | 0.0 | -0.07 | -0.16 | -0.05 | -0.27 |

Table 1.8: Change in Monthly Job-to-Job Transition Rate from row to column occupation group. Data Source: CPS Basic Monthly Files. Own Calculations.

Table 1.9 shows the job-to-job transition rate between and within occupation groups, from row to column.

**Wage Premia** The average wage premium relative to "High School Graduates", where wages are measured as weekly earnings, in the Outgoing Rotation Group of the CPS, is shown in figure 1.8. The "College" Premium is relatively stable over the last 20 years, while it increases slightly for workers with a full year degree or more education, it decreases somewhat for workers who went

(a) 1995-1997

| | NRM | RM | RC | NRC | total |
|---|---|---|---|---|---|
| NRM | 1.7 | 0.47 | 0.38 | 0.31 | 2.86 |
| RM | 0.2 | 2.09 | 0.21 | 0.21 | 2.72 |
| RC | 0.16 | 0.25 | 1.53 | 0.5 | 2.45 |
| NRC | 0.09 | 0.15 | 0.32 | 1.39 | 1.95 |

(b) 2015-2017

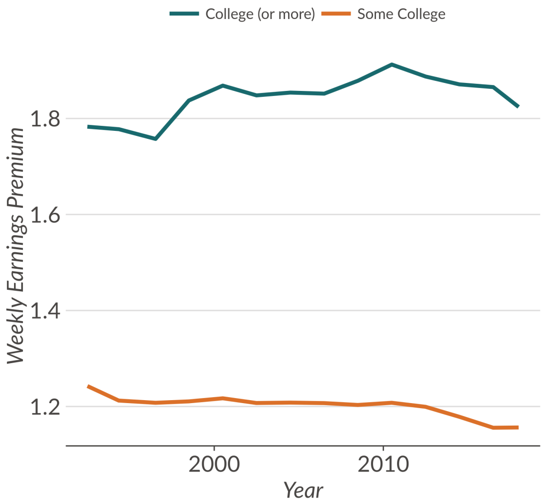
| | NRM | RM | RC | NRC | total |
|---|---|---|---|---|---|
| NRM | 1.55 | 0.29 | 0.31 | 0.34 | 2.48 |
| RM | 0.22 | 1.63 | 0.21 | 0.17 | 2.23 |
| RC | 0.23 | 0.25 | 1.04 | 0.45 | 1.97 |
| NRC | 0.09 | 0.08 | 0.16 | 1.35 | 1.68 |

Table 1.9: Monthly Job-to-Job Transition Rate from row to column occupation group. Data Source: CPS Basic Monthly Files. Own Calculations.

to college but did not finish a 4 year degree. Figure 1.7 shows the wage premium of the occupation groups defined in section 1.2 relative to the average pay in "non-routine manual" occupations. The pay premium of routine occupations decreased from roughly 50% to 40%, while that of non-routine cognitive occupations stayed roughly constant at about 110%.

**Employment outflows.** Figure 1.9 shows the separation rates out of employment to non-employment conditional on the education level of workers. The separation rates increased somewhat, particularly for lower skilled workers. This is in contrast with findings in Fujita (2015) and Cairo (2013), however their analysis focuses on a different time period and different definitions of separation

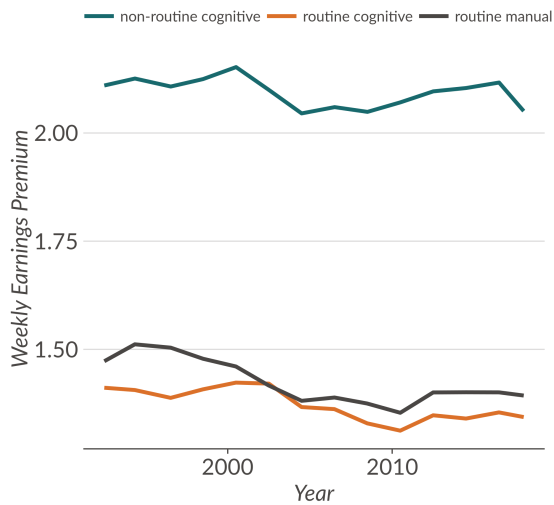
Figure 1.7: Wage Premium of Education



ns using CPS Outgoing Rotation Group, Individuals aged 25-45.

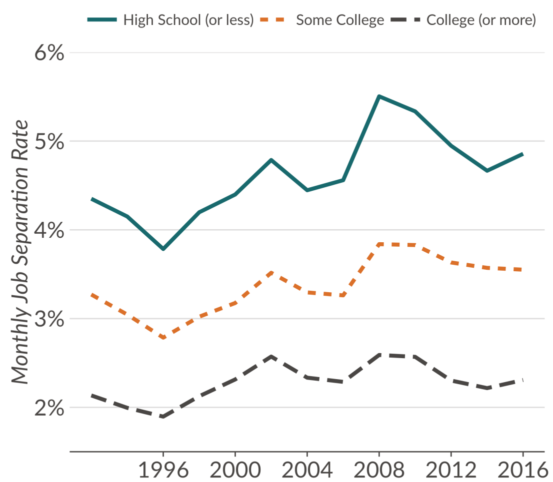
rates, job destruction and separations to unemployment respectively, while also focusing on average rates in the economy.

Figure 1.8: Wage Premium - Occupation Groups




ns using CPS Outgoing Rotation Group, Individuals aged 25-45.

Figure 1.9: Job Separation Rates




n Calculations using CPS Monthly Data, Individuals aged 25-45.

# Chapter 2

# AUTOMATION AND WAGE INEQUALITY ACROSS SPACE

based on joint work with Jan Eeckhout and Roberto Pinheiro

## 2.1 Introduction

Wage inequality has risen sharply since the early 1980s. In particular, the 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 skilled workers, thus leading to higher wages.

[1] See Acemoglu and Autor (2011).

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

At the same time, wage inequality has increased as a result of job polarization. Technological change, but now automation in particular, has resulted in the "disappearing middle" of the income distribution. The process of automation has directly substituted capital for labor in tasks previously performed by moderately skilled workers. Machines are most likely to displace jobs that are intensive in routine tasks. This affects both the low skilled with the highest wages most (manufacturing and operative occupations), and the high skilled most with the lowest wages (clerical and administrative occupations). Rather than an increase in gap between the high and the low skilled, automation leads to a disappearing middle, where only the lowest and highest wages increase. This is a very distinct from the increase in the college premium.

An open question is which of the technological forces that lead to the different outcomes – the rising college premium or job polarization – is at work and how. To investigate these distinct drivers of 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 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 MSA (Metropolitan Statistical Area). 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 a 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 a accurate measure of investment in technology, being possibly used to either automate away routine tasks or complement non-routine cognitive tasks. Moreover, the 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)), the introduction of ERP software have as its main goal the replacement of clerical work.[4]

[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 num-

Our empirical results show that large and more expensive cities invest more in technology, measured either by the total IT budget per worker as well as 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 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 that can 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 choice. Heterogeneously skilled citizens earn a living based on a competitive wage and choose housing in a competitive housing market. Under perfect mobility, their location choice makes 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). 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

---

ber of licenses, for example), as well as the fact that we have information on software installation for 10% of the establishments in our sample.

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 – as the introduction of ERP software.

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 gives rise an increase in the gap between skilled and unskilled workers.

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 et al. (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 beyond these facts by providing a mechanism that can explain the economic mechanism. Moreover, we use a direct measure of technology, namely the price of investment in technological capital. We have data on the use of technology at the establishment level. This establishment level data is unique. Acemoglu and Restrepo (2017) also analyzes 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 7 sections. Sections 2 and 3 present our model and

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

theoretical results, as well as some simple numerical exercises. Section 4 and 5 describe the data and empirical results, respectively. Section 6 estimates an extended version of the model that includes occupational choice and a housing supply sector. It also shows preliminary counterfactual experiments. Finally, section 7 concludes the paper. All proofs are presented in the Appendix.

## 2.2 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

[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.

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: $A_j F(m_{1j}, ..., m_{Ij}, k_j)$. 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). 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:

$$\sum_{j=1}^{J} C_j m_{ij} = M_i, \ \forall i \qquad \sum_{i=1}^{I} h_{ij} m_{ij} = H_j, \ \forall j. \tag{2.1}$$

where $C_j$ denotes the number of cities with TFP $A_j$.

## 2.3 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)

$$\begin{aligned} \max_{\{c_{ij},h_{ij}\}} u(c_{ij}, h_{ij}) &= c_{ij}^{1-\alpha} h_{ij}^{\alpha} \\ \text{s.t. } c_{ij} + p_j h_{ij} &\leq w_{ij} \end{aligned} \tag{2.2}$$

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:

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{w_{ij}}{p_j^{\alpha}} \implies w_{ij} = U_i p_j^{\alpha} \frac{1}{\alpha^{\alpha} (1-\alpha)^{1-\alpha}}, \tag{2.3}$$

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:

$$\frac{w_{ij}}{w_{ij'}} = \left( \frac{p_j}{p_{j'}} \right)^{\alpha}. \tag{2.4}$$

*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:

$$\max_{m_{ij}, \forall i} A_j F(m_{1j}, ..., m_{Ij}, k_j) - \sum_{i=1}^{I} w_{ij} m_{ij} - r k_j, \tag{2.5}$$

subject to the constraint that $m_{ij} \geq 0$ and $k \geq 0$. 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]

[7] In what follows, the non-negativity constraint on $m_{ij}$ and $k_j$ are dropped. This is

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:

$$A_j F(m_{1j}, ..., m_{Ij}, k_j) = A_j \left( \sum_{i=1}^{I} m_{ij}^{\gamma_i} x_i + k_j x_k \right) \tag{2.6}$$

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 $\gamma_i$, 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.

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

We now proceed by introducing varying degrees of substitutability (or complementarity) 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 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 3.** *Consider the following technologies:*

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

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

*where* $\gamma_2 < \theta$

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

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.

*complements.*

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

*where* $\gamma_3 > \theta$

### 2.3.1 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:

$$\begin{aligned}
&(m_{1j}): \quad A_j\gamma_1 m_{1j}^{\gamma_1-1}x_1 = w_{1j}, \ \ \forall j \in J; \\
&(m_{2j}): \quad A_j\tfrac{\gamma_2}{\theta}\left( m_{2j}^{\theta}x_2 + k_j^{\theta}x_k \right)^{\frac{\gamma}{\theta}-1}\theta m_{2j}^{\theta-1}x_2 = w_{2j}, \ \ \forall j \in J; \\
&(k_j): \quad A_j\tfrac{\gamma_2}{\theta}\left( m_{2j}^{\theta}x_2 + k_j^{\theta}x_k \right)^{\frac{\gamma_2}{\theta}-1}\theta k_j^{\theta-1}x_k = r, \ \ \forall j \in J; \\
&(m_{3j}): \quad A_j\gamma_3 m_{3j}^{\gamma_3-1}x_3 = w_{3j}, \ \ \forall j \in J;
\end{aligned} \tag{2.7}$$

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:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1-1}}M_1}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1-1}}\right\}} \tag{2.8}$$

and

$$m_{31} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_3-1}}M_3}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma_3-1}}\right\}} \tag{2.9}$$

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:

$$m_{21} = \frac{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}} \frac{k_1}{k_2}}{\left[1 + \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}} \frac{k_1}{k_2}\right]} M_2 \tag{2.10}$$

and

$$k_2 = \frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{2.11}$$

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. 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 2** (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 3 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.

**Proposition 3** (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

[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.

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.*

### 2.3.2 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:

$$\begin{aligned}
&(m_{1j}) : A_j \gamma_1 m_{1j}^{\gamma_1 - 1} x_1 = w_{1j} \\
&(m_{2j}) : A_j \gamma_2 m_{2j}^{\gamma_2 - 1} x_2 = w_{2j} \\
&(m_{3j}) : A_j \gamma_3 \left( m_{3j}^{\theta} x_3 + k_j^{\theta} x_k \right)^{\frac{\gamma_3}{\theta} - 1} m_{3j}^{\theta - 1} x_3 = w_{3j} \\
&(k_j) : A_j \gamma_3 \left( m_{3j}^{\theta} x_3 + k_j^{\theta} x_k \right)^{\frac{\gamma_3}{\theta} - 1} k_j^{\theta - 1} x_k = r
\end{aligned} \tag{2.12}$$

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:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1 - 1}} M_1}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1 - 1}}\right\}} \tag{2.13}$$

and

$$m_{21} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_2 - 1}} M_2}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_2 - 1}}\right\}} \tag{2.14}$$

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:

$$m_{31} = \frac{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}} \frac{k_1}{k_2}}{\left[1 + \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta-1}} \frac{k_1}{k_2}\right]} M_3 \tag{2.15}$$

and

$$k_2 = \frac{M_3 x_3^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{2.16}$$

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. 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 4** (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 5 shows, the city with higher TFP also demands more capital.

**Proposition 5** (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.

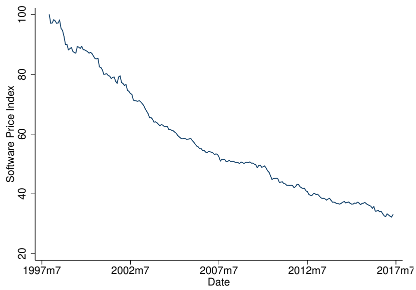
**Corollary 1** (SBTC and demand for high skill)**.** *Assume* $\gamma_3 > \theta$. $A_i > A_j \Rightarrow m_{3i} > m_{3j}$.

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.

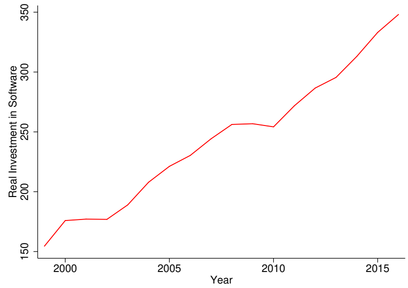
**Theorem 3.** *Assume* $\gamma_i \equiv \gamma$, $\forall i \in \{1, 2, 3\}$ *and* $\gamma > \theta$. *If* $A_1 > A_2$, *we have that city 1's skill distribution F.O.S.D. city 2's skill distribution.*

Differently from the case of Automation, **SBTC does not imply that the high-TFP city is larger**. In the appendix, 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.

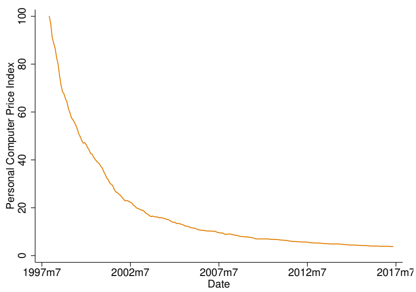
In the next section, we simulate the model in order to get a better understanding of the model's mechanisms and how changes in the parameters may affect the two regions' labor markets. We focus on two parameter changes that are related to the observed evolution of computing power prices over the last 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 2.1 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 2.1. 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.



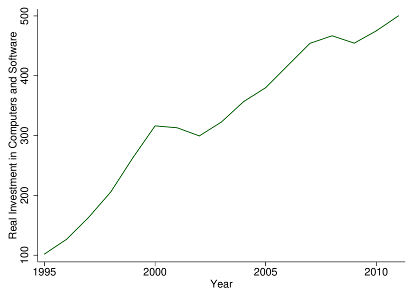

(a) Software's Price Index: 1997–2017
**Source:** Bureau of Labor Statistics




(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 2.1: Price Index and Real Investment in Technology

### 2.3.3 Numerical Example

#### 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 $\gamma$ as well. Moreover, we follow Davis and Ortalo-Magné (2011) and set $\alpha = 0.24$. Finally, we must specify values for both $\theta$ 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 2.1. We assume that these parameters are fixed over time in our numerical exercise.

Table 2.1: Maintained Parameters – from Eeckhout et al. (2014)

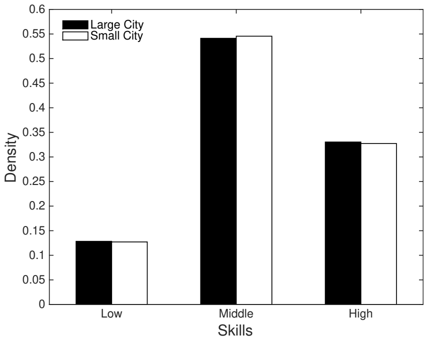
| $\gamma$ | $\theta$ | $A_1$ | $A_2$ | $y_1$ | $y_2$ | $y_3$ | $H_i$ | $\alpha$ |
|---|---|---|---|---|---|---|---|---|
| 0.8 | 0.85 | 19,118 | 9,065 | 0.3189 | 1 | 1.4733 | 62,559,000 | 0.24 |

We then consider two periods in time: 1995 and 2015. We consider changes in the size and 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 2.2.

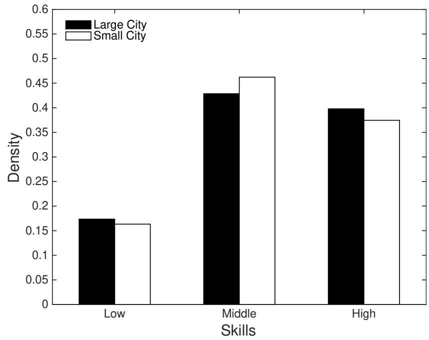
Table 2.2: Adjusted Parameters - Experiments

| | $y_k$ | $r$ | $M_1$ | $M_2$ | $M_3$ |
|---|---|---|---|---|---|
| **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 |

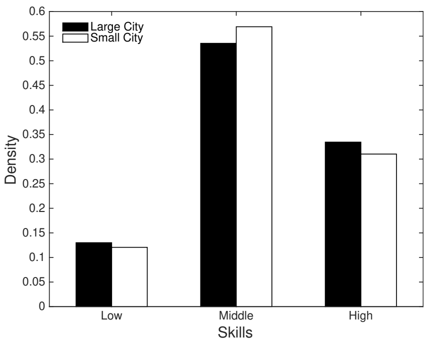
Results are presented in figure 2.2 and table 2.3. As we can see from figures 2.2(a) and 2.2(b) and table 2.3's panel B, between 1995 and 2015, city 1 not only became even bigger than city 2, but it also became more unequal – the proportion of mid-skilled workers went down significantly more in city 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 2.2, 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 2.2(c) and in table 2.3 **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 2.2(d) and in table 2.3 **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.



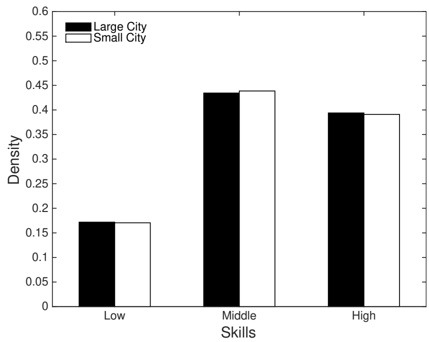

(a) Skill Distribution – 1995




(b) Skill Distribution – 2015




(c) Skill Distribution – Population Constant




(d) Distribution – Technology Constant

Figure 2.2: Skill Distribution across cities – 1995 vs. 2015

Table 2.3: 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% |

## 2.4 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 an MSA we use is the 2000 Combined Metropolitan Statistical Areas (CMSA) by the Census for all MSAs that are part of an CMSA and 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 a 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 2.4 shows the classification into groups by task components and the corresponding titles of occupation groups in the Census 2010 Occupation Classification system[10].

Table 2.5 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.

**Technology Data**

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 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

[10] See Cortes et al. (2014) for the mapping 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 2.4: 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 |

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 2.6

[12]National Establishment Time-Series (NETS) and the County Business Pattern

Table 2.5: Descriptive Statistics

| | 1980<br>mean<br>(st. dev.) | 2015<br>mean<br>(st. dev.) |
|---|---|---|
| *MSA's Occupation Shares* | | |
| Non-Routine Cognitive | 34.6% | 45.3% |
| | (3.95) | (5.46) |
| Non-Routine Manual | 9.9% | 14.8% |
| | (2.43) | (2.38) |
| Routine Cognitive | 29.8% | 22.9% |
| | (2.12) | (1.96) |
| Routine Manual | 25.3% | 16.7% |
| | (4.71) | (3.08) |
| *MSA's Rent and Size* | | |
| log rent index | 0.01 | 0.01 |
| | (0.13) | (0.23) |
| Employment in 000s | 861.61 | 1535.77 |
| | (1049.25) | (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.

(CBP), for example.

shows that, compared to the National Establishment Time-Series (NETS), our sample covers on average 53% of the MSA's employed labor force. Moreover, while table 2.6 shows that our sample covers on average only 13% of the MSA's establishments, table 2.7 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 2.8).[13] Third, it is an easily interpretable continuous variable, i.e., it does not suffer of 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,

[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.

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 2.6, 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. 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 2.5. However, we present the results for ERP measures in the appendix. While results are understandably weaker for ERP – due to smaller sample size is a coarser measure – they are still qualitatively similar to the ones presented in section 2.5.

Finally, in terms of geographical coverage and summary statistics, figure 2.3a shows the geographical dispersion of IT budget per worker across the country in 2015. First of all, corroborating the results presented in table 2.6, 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 2.9 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 2.9. 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 2.9'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 2.9. As we can see in row 1 of table 2.9, there is significant variation in IT Budget per worker across MSAs.

Table 2.6: 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 2.7: Coverage Ci Aberdeen relative to NETS by Establishment Size

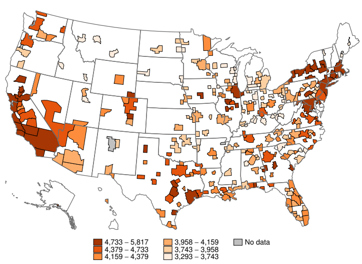
| | Mean | S.D. | p10 | p25 | p50 | p75 | p90 | N |
|---|---|---|---|---|---|---|---|---|
| **IT Budget Sample** | | | | | | | | |
| 1 to 4 Employees | 3% | 1% | 2% | 3% | 3% | 4% | 4% | 272 |
| 5 to 9 Employees | 27% | 4% | 22% | 25% | 27% | 29% | 31% | 272 |
| 10 to 19 Employees | 56% | 7% | 50% | 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 |

## 2.5 Empirical Evidence

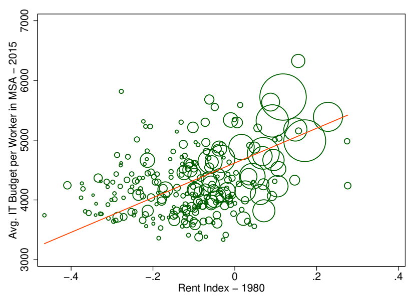
In this section we describe our evidence regarding the adoption of automation technology and the occupational composition of cities. We focus on the two

Table 2.8: 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) Avg. IT Budget per worker




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

Figure 2.3: Geographical distribution of IT across CMSAs – 2015

main predictions of the theory: (1) locations with higher housing costs should implement automation technology at higher rates and (2) locations with higher

Table 2.9: 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 |

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 2.4, in this section we focus on IT budget per worker as our key variable on technology investment. In the appendix, 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 2.10 shows the results for MSA-level linear regression models of the average IT budget per worker on the local price index and the share of routine cognitive workers in 1980, weighted by MSA size. Columns 1 and 2 indicate that, when considered separately, both the current local price level and the past share of routine-cognitive workers help to explain a substantial amount of variation in IT budget per worker. In the first specification, a one standard deviation increase in local price index (an increase of 19.6% in the local price index) is associated with an increase of $564.88 in the MSA's average IT budget per worker. This magnitude corresponds to an increase of 10.3% on the average IT budget per worker. In the second specification, a one standard deviation increase in the 1980's share of routine-cognitive workers (an increase of 10% in the local share of routine cognitive jobs) is associated with an increase of \$522.67 in the MSA's average IT budget per worker. This magnitude corresponds to an increase of 9.53% on the average IT budget per worker. However, when considering both variables jointly, the effect of the lagged routine-cognitive share conditional on the local price level shrinks substantially. In fact, the marginal effect of one standard deviation increase in the 1980's local share of routine cognitive jobs declined from 9.53% to 6.6% between specifications (2) and (3) (a decline of 2.92 percentage points). Differently, the decline in the marginal effect of one standard deviation increase in local price index from specifications (1) and (3) was just of 1.3 percentage points (from 10.3% to 9%). Moreover, once we introduce 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 effect of the 1980's share of routine-cognitive workers on IT budget per worker is no longer statistically significant. Differently, the effect of local prices is still highly significant. Finally, in the appendix we present alternative specifications for the regressions presented in table 2.10, in which we replace the local price index by the MSA's size. Results are qualitatively similar.

However, results presented in table 2.10 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 2.11 we run establishment-level linear regression models of the establishment's IT budget per worker 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.[14] Results presented in table 2.10, where we restrict our sample to es-

[14]While 55% of the multi-establishment firms have all their establishments in the same MSA (11,788 firms), the remaining 45% (9,237 firms) have establishments distributed

Table 2.10: IT budget per worker

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | log(IT) | log(IT) | log(IT) | log(IT) |
| 1980's log rent index | 0.772*** | | 0.679*** | 0.522** |
| | (0.187) | | (0.168) | (0.165) |
| 1980's Share Routine-Cognitive | | 3.265** | 2.284** | 1.073 |
| | | (0.996) | (0.722) | (0.936) |
| 1980's Offshorability Index | | | | 0.733* |
| | | | | (0.360) |
| Adj. $R^2$ | 0.19 | 0.10 | 0.23 | 0.26 |
| Observations | 261 | 261 | 261 | 261 |

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Each observation (an MSA) is weighted by its employment in 2015

tablishments 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 worker, 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 13.7% in the local price index) is associated with an increase in the establishment's average IT budget per worker of about $66.50. This magnitude corresponds to an increase of 2.3% increase on 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. 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 2.11. Finally, neither the coefficient of the share of routine-cognitive workers in 1980 nor the coefficient of the

across MSAs with significant differences in local prices.

MSA's average degree of offshorability of the local jobs in 1980 are statistically significant.

Table 2.11: IT Investment by Establishment - Firm and Industry FE

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | log(IT) | log(IT) | log(IT) | log(IT) |
| 1980's log rent index | 0.187*** | 0.169*** | 0.181*** | 0.170*** |
| | (0.026) | (0.031) | (0.027) | (0.031) |
| log(Site's Size) | -0.054*** | -0.054*** | -0.054*** | -0.054*** |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| log(Site's Revenue) | 2.242*** | 2.242*** | 2.242*** | 2.242*** |
| | (0.038) | (0.038) | (0.038) | (0.038) |
| 1980's Offshorability Index | | 0.069 | | 0.054 |
| | | (0.053) | | (0.073) |
| 1980's Share Routine-Cognitive | | | 0.002 | 0.001 |
| | | | (0.002) | (0.002) |
| Firm FE | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Adj. $R^2$ | 0.75 | 0.75 | 0.75 | 0.75 |
| No. of Sites | 196,586 | 196,586 | 196,586 | 196,586 |
| No. of Firms | 61,571 | 61,571 | 61,571 | 61,571 |

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

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 2.12: Change in routine-cognitive share, 1980-2015

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $\Delta$ RC | $\Delta$ RC | $\Delta$ RC | $\Delta$ RC |
| 1980's log rent index | -0.093*** | | -0.061** | -0.041* |
| | (0.029) | | (0.026) | (0.023) |
| 1980's Share Routine-Cognitive | | -0.872*** | -0.783*** | -0.622*** |
| | | (0.082) | (0.075) | (0.132) |
| 1980's Offshorability Index | | | | -0.097** |
| | | | | (0.048) |
| Adj. $R^2$ | 0.189 | 0.473 | 0.550 | 0.580 |
| Observations | 261 | 261 | 261 | 261 |

$\Delta$ RC: Change in share of workers employed in routine-cognitve occupation.
Standard errors in parentheses
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$
Each observation (an MSA) is weighted by its employment in 2015.

Table 2.12 presents the results of linear regressions of the change in the routine-cognitive share of MSAs between 1980 and 2015 on its 1980 level and the 1980's local rent index. Again, columns 1 and 2 present the bivariate specifications, column 3 the model with both covariates, and column 4 includes the average offshorability of jobs in the MSA. The first column 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 1.2 percentage point larger drop in the routine-cognitive share over 1980-2015. Thus, the most expensive places have about a 7 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. Column 2 presents the bivariate specification with 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 2.8 percentage points in the local share of routine-cognitive jobs) is associated with a 2.5 percentage point larger drop subsequently. In column 3, the results for the multivariate regression are presented. Both 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 0.8 percentage point and the effect of a one standard deviation higher initial share drops to 2.2 percentage points, respectively. Finally, we do 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 column 1, while the effect of the share of routine-cognitive workers in 1980 drops by 30%. 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.

### 2.5.1 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").

### 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 $j$ ($\lambda_{ij}$) is given by:

$$\lambda_{ij} = \frac{\frac{m_{ij}}{S_j}}{\frac{M_i}{\sum_{l=1}^{N} M_l}} \tag{2.17}$$

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 2.13.

Table 2.13: 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 2.13, 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 2.13 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 2.4 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 2.13.

### Ellison-Glaeser (1997) Index

We now adapt the concentration index presented by Ellison and Glaeser (1997) for the skill distribution context. Denote $\gamma_i$ 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:

$$\gamma_i = \frac{\sum_j (s_{ij} - x_j)^2}{1 - \sum_j x_j^2} \tag{2.18}$$

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 2.14: 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 2.14. 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.

### 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 $\zeta_j$ 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:

$$\zeta_j = \frac{S_j}{S_j - 1} \frac{\sum_i (\tilde{s}_{ij} - \tilde{x}_i)^2}{1 - \sum_i \tilde{x}_i^2} - \frac{1}{S_j - 1} \tag{2.19}$$

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 2.15 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 2.5. As we can see, the distribution of the OS index became more concentrated as we move from 1980 to 2015.

## 2.6 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,

Table 2.15: OS Index across city sizes and time

| | Mean | Median | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| **Panel A: 1980** | | | | | |
| | Mean | Median | St. Dev. | Min | Max |
| 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 |

the production function allows for generic substitution patterns between capital and labor across occupations. Finally, individuals are heterogenous in their skill, which can differ across occupations, and in their preferences for locations. Workers choose location and occupation jointly, thus locations biased towards a certain type of job might 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.

## 2.6.1 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

$$F(\mathbf{m}_j, \mathbf{k}_j, \mathbf{A}_j) = A_j \left\{ \sum_i \left[ m_{ij}^{\gamma_i} A_{l,ij} + k_{ij}^{\gamma_i} A_{k,i} \right]^{\frac{\lambda}{\gamma_i}} \right\}^{\frac{\beta}{\lambda}}, \tag{2.20}$$

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.

*Workers* are heterogeneous in their skills and preferences for locations. They consume the final good and housing, where housing must be consumed in the same city as the workplace. Preferences over consumption and housing follow

$$u(c, h) = c^{1-\alpha}(h - \underline{h})^{\alpha}, \tag{2.21}$$

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

$$c + p_j h \leq w_j^*, \tag{2.22}$$

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 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. 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

$$V_j(\mathbf{s})\varepsilon_j = a_j \max_i a_i \frac{(w_{i,j}(\mathbf{s}) - p_j \underline{h})}{p_j^{\alpha}} \varepsilon_j, \tag{2.23}$$

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 general amenity $a_j$ of location $j$ is commonly enjoyed by all workers and $\varepsilon_j$ represents idiosyncratic tastes for different locations. The distribution of idiosyncratic tastes is i.i.d. across locations and individuals, following a Frechet distribution with scale parameter $\tau$. The location parameter is normalized to 1. The share of workers choosing a location $j$, then follows

$$P(j|\mathbf{s}) = \frac{V_j(\mathbf{s})^{\tau}}{\sum_j V_j(\mathbf{s})^{\tau}}. \tag{2.24}$$

The skill distribution in each location is $P(\mathbf{s}|j) = \frac{P(j|\mathbf{s})G(\mathbf{s})}{P(j)}$ where $P(j) = \int \cdots \int P(j|\mathbf{s})dG(\mathbf{s})$.

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

$$p(H) = \bar{p}_j H^{\epsilon_p}. \tag{2.25}$$

In an equilibrium housing supply adjusts such that the housing amount demanded by workers equals that supplied.

### 2.6.2 Identification

We shortly describe which parameters we estimate and how those can be identified. The main goal is to identify the parameters of the production function, the utility function parameters, housing supply and the distribution of skills.

1. Relative productivity of labor by occupation and location $A_{l,ij}$: The demand for labor depends directly on its productivity. Therefore, we can identify the relative productivity of labor in an occupation relative to a reference occupation for each location from its relative demand.

2. Productivity of IT capital by occupation $A_{k,i}$: The productivity of capital can inferred from the quantity of usage in output units.

3. Elasticity of substitution of IT capital and labor $\gamma_i$ can be inferred from the joint demand for labor and capital and its variation with respect to house prices, as it shifts the relative price of labor and capital.

4. Amenity of Jobs $a_i$ can be identified up to a normalizing constant from wages and employment in an occupation. An occupation that has high employment, but low wages tends to have a high amenity.

5. Utility function parameters $\alpha$ and $\underline{h}$ can be identified from spending shares. Rewriting the housing demand equation one obtains

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

As spending shares and the ratio of wages to house prices are directly observable in the data, the utility function parameters are directly identified.

6. Common amenities of locations $a_j$ can be identified from the wages, city size and local house prices. A location with high house prices, low wages, but a large population must offer benefits that are not due to work. Such amenities are captured by $a_j$.

7. Housing supply shifter $\bar{p}_j$ and $\epsilon_p$: The housing supply shifter is identified from the level of house prices. The elasticity of housing supply can not be identified without additional data, we fix its value following Saiz (2010).

8. Skill distribution $G(\mathbf{s})$ can be identified under parameteric restrictions from higher order moments of the wage distribution. Nonparametric identification would fail partially. We parameterize the distribution of skills as a product of Beta distributions $B(\alpha_i, \beta_i)$ with support over $[\underline{x_i}, \overline{x_i}]$ and normalize $\alpha_i = \beta_i = 1$ and set $\frac{\overline{x_i}}{\underline{x_i}} = 5$.

### 2.6.3 Moments, Fit and Estimates

We estimate a parameterized version of the model by fitting a set of moments. We calculate the model solution with 3 cities and all moments are calculated directly without simulation. The data sources are the same as in the previous sections. We construct city level measures of employment and IT usage from the data. The IT capital variables 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 2.29 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 set of moments we target is shown in table 2.16. Overall the fit is very good, most moments are fit almost exactly. Almost half of workers are employed in non-routine cognitive occupations, while just over 20 % are employed in either routine-manual or routine-cognitive occupations. Next, we consider the co-variation of employment by occupation category and house prices using the following regression

$$\log(m_i) = b_0 + b_p^m \log(p) \tag{2.26}$$

The parameter $b_p^m$ approximates the elasticity of employment in an occupation with respect to house prices. We find that routine-manual jobs are relatively unlikely to sort towards high price locations, while both non-routine manual jobs and cognitive jobs are more likely to appear in expensive cities. This relationship is well captured by the model.

Essential to our exercise however is the joint allocation of capital and labor. The aggregate capital stock $\int pk$ is the sum of all units of capital across all locations, which we calculate by occupation category. Next, we consider the

Table 2.16: Moments 2015 and Model Fit

| | Data | Model |
|---|---|---|
| Share employed in RM | 0.23 | 0.22 |
| | (0.00021) | |
| Share employed in RC | 0.22 | 0.21 |
| | (8.1e-5) | |
| Share employed in NRC | 0.44 | 0.42 |
| | (0.00026) | |
| $b_p^m$ NRM | 2.7 | 2.7 |
| | (0.15) | |
| $b_p^m$ RM | 1.1 | 1.1 |
| | (0.21) | |
| $b_p^m$ RC | 2.0 | 2.0 |
| | (0.21) | |
| $b_p^m$ NRC | 2.1 | 2.1 |
| | (0.25) | |
| $\int \log(pk)$ RC | 0.9 | 0.89 |
| | (0.0043) | |
| $\int \log(pk)$ NRC | 1.0 | 1.0 |
| | (0.0047) | |
| $b_p^k$ RC | 2.1 | 2.1 |
| | (0.27) | |
| $b_p^k$ NRC | 1.9 | 1.9 |
| | (0.26) | |
| $\bar{l}og(w)$ NRM | 2.1 | 2.1 |
| | (3.6e-7) | |
| $\bar{l}og(w)$ RM | 2.5 | 2.5 |
| | (2.6e-7) | |
| $\bar{l}og(w)$ RC | 2.6 | 2.6 |
| | (4.3e-7) | |
| $\bar{l}og(w)$ NRC | 3.1 | 3.1 |
| | (3.1e-7) | |

Table 2.17: Worker Productivity

$\log(A_{l;i,j)} = b_0 + b_1 log(A_j)$

| | $b_0$ | $b_1$ |
|---|---|---|
| RM | 1.9 | -8.2 |
| | (0.0052) | (0.26) |
| RC | 2.0 | -4.0 |
| | (0.014) | (0.15) |
| NRC | 3.3 | -4.5 |
| | (0.016) | (0.24) |

elasticity of capital usage with respect to house prices

$$\log(k_i) = b_0 + b_p^k \log(p), \tag{2.27}$$

which is approximated by $b_p^k$. This elasticity is at a similar range as that of labor, and the model fits the respective elasticities exactly. Finally, we consider the average wage by occupation. There is a clear ranking in terms of average wages, where non-routine manual jobs are at the lower end of the wage distribution, while routine jobs can be considered middle wage jobs and non-routine cognitive jobs are high paying.

Next, we consider the parameterization of parameters and their estimates. The productivity of jobs relative to that of non-routine manual jobs is parameterized as a log-linear function of city TFP. The parameter estimates for $b_0$ and $b_1$ represent the relationship of employment shares and relative elasticities with respect to house prices.

IT capital productivity and its elasticity of substitution with routine-cognitive and non-routine cognitive labor respectively are presented in table 2.18 The estimated elasticity of substitution of labor and IT capital is larger in routine-cognitive, relative to non-routine cognitive occupations. Routine type jobs are

Table 2.18: Capital Productivity $A_{k;i}$ and Elasticity of Substitution with labor $\frac{1}{1-\gamma_i}$

| | $x_k$ | $\gamma$ |
|---|---|---|
| RC | 0.59 | 0.56 |
| | (0.03) | (0.21) |
| NRC | 0.47 | 0.41 |
| | (0.2) | (0.15) |

Table 2.19: Job Amenity

| | $a_i$ |
|---|---|
| RM | 0.85 |
| | (0.023) |
| RC | 0.49 |
| | (0.013) |
| NRC | 0.34 |
| | (0.011) |

expected to be easier to automate relative compared to non-routine jobs, thus the estimates actually reflect and support that categorization.

Finally, to jointly account for employment and wages by job type we estimate an amenity by job type, see table 2.19

### 2.6.4 Results

We consider an experiment where the price of information technology capital increases. 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.

We calculate within the model the elasticity of employment shares $\pi_i$ with respect to IT prices. In the data, we found that routine-cognitive workers sort away from expensive locations, while non-routine cognitive workers increasingly concentrate there. Furthermore, aggregate employment shares in routine-cognitive jobs fell, while employment in non-routine cognitive jobs rose.

Table 2.20: Elasticity of Employment Shares by Occupation with respect to IT prices

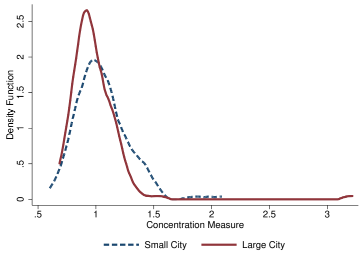
$\frac{d\pi_i/\pi_i}{dr/r}$

| | RC | NRC |
|---|---|---|
| Aggregate | -0.3 | 0.085 |
| Cheap City | -0.22 | 0.00019 |
| Expensive City | -0.49 | 0.31 |

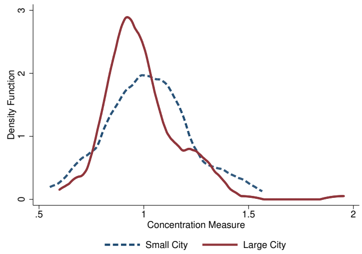
Table 2.20 shows the elasticity of employment shares with respect to IT prices. We find that economy wide, a $1\%$ increase in IT prices leads to a $0.3\%$ decline in employment in routine-cognitive occupations, while it leads to an $0.085\%$ increase in employment in non-routine cognitive jobs. We also compare the relative behavior of cheap and expensive cities. What we find is that cheap cities the employment share of routine cognitive jobs falls by less than in expensive cities. While, for non-routine cognitive jobs the opposite holds. Interestingly, the change of non-routine cognitive employment in cheap cities is almost zero, consistent with a rise and concentration of high skill employment in expensive urban areas. Thus, a fall in IT prices over time is consistent with the models implications in terms of employment patterns across jobs and locations.

## 2.7 Conclusion

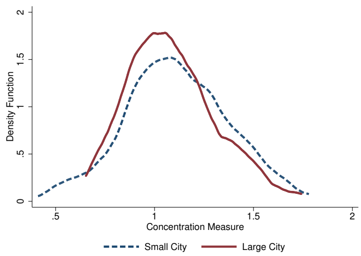
In this paper, we show that the substitution of routine jobs and tasks with machines, computers, and software has not happened evenly in space. In fact, the relative benefit of replacing middle-skill workers that perform routine tasks by computers and software depend on the cost of hiring a worker in 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.



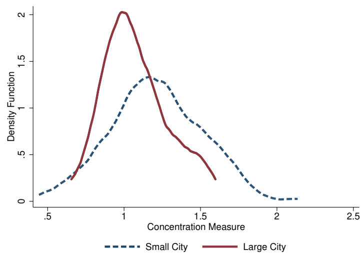

(a) Non-Routine Manual: 1980



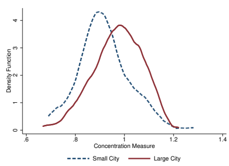

(b) Non-Routine Manual: 2015



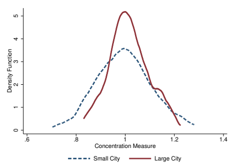

(c) Routine Manual: 1980



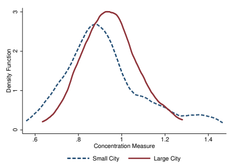

(d) Routine Manual: 2015



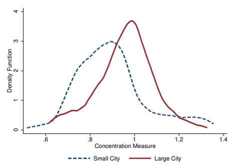

(e) Routine Cognitive: 1980




(f) Routine Cognitive: 2015

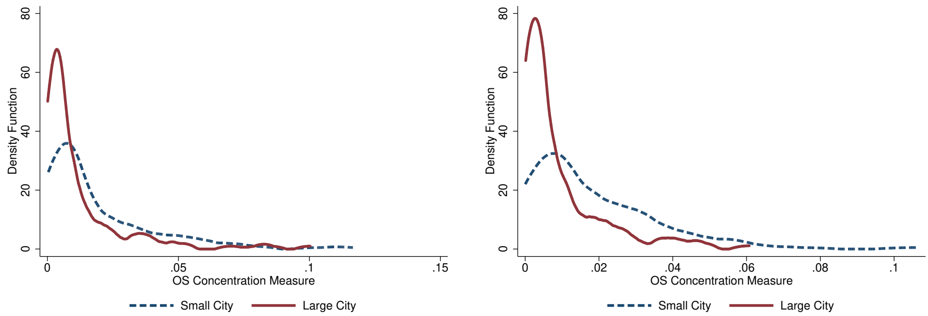



(g) Non-Routine Cognitive: 1980




(h) Non-Routine Cognitive: 2015

Figure 2.4: Skill Distribution across city sizes and time




(a) 1980

(b) 2015

Figure 2.5: Distribution of OS index across city sizes and time

# Appendices

# Theory - Preliminary Steps - Automation

## Preliminaries

### 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:

$$F_j(m_{1j}, m_{2j}, m_{3j}, k_j) = A_j \left[ m_{1j}^{\gamma_1} x_1 + \left( m_{2j}^{\theta} x_2 + k_j^{\theta} x_k \right)^{\frac{\gamma_2}{\theta}} + m_{3j}^{\gamma_3} x_3 \right] \tag{2.28}$$

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:

$$\frac{w_{ij}}{w_{ij'}} = \left( \frac{p_j}{p_{j'}} \right)^{\alpha}, \quad \forall i \in \{1, 2, 3\} \text{ and } \forall j \in \{1, 2\} \tag{2.29}$$

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

$$m_{11} = \frac{\left[ \frac{A_2}{A_1} \left( \frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_1 - 1}} M_1}{1 + \left[ \frac{A_2}{A_1} \left( \frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_1 - 1}}} \tag{2.30}$$

Similarly, for skill 3:

$$m_{31} = \frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3 - 1}} M_3}{1 + \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3 - 1}}} \tag{2.31}$$

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

$$\left(\frac{m_{21}}{m_{22}}\right) = \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta - 1}} \frac{k_1}{k_2} \tag{2.32}$$

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

$$m_{21} = \left\{\frac{1}{x_2}\left[\left(\frac{r}{A_1 \gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k\right]\right\}^{\frac{1}{\theta}} k_1 \tag{2.33}$$

Similarly, for $(k_2)$, we have:

$$m_{22} = \left\{\frac{1}{x_2}\left[\left(\frac{r}{A_2 \gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2 - \theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k\right]\right\}^{\frac{1}{\theta}} k_2 \tag{2.34}$$

Dividing (2.33) by (2.34)and substituting (2.32), we have:

$$\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta - 1}} = \left\{\frac{\left[\left(\frac{r}{A_1 \gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k\right]}{\left[\left(\frac{r}{A_2 \gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2 - \theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k\right]}\right\} \tag{2.35}$$

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:

$$M_2 x_2^{\frac{1}{\theta}} = \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_1 + \dots \tag{2.36}$$
$$\left[ \left( \frac{r}{A_2 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_2$$

Substituting (2.35) and manipulating, we have:

$$k_2 = \frac{M_2 x_2^{\frac{1}{\theta}} - \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_1}{\left( \frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta}} \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \tag{2.37}$$

Substituting (2.37) into (2.36) and manipulating, we have:

$$\left\{ \frac{M_2 x_2^{\frac{1}{\theta}} - \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}} k_1}{\left( \frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta}} \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} \right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} = \tag{2.38}$$
$$= \left( \frac{A_2}{A_1} \right)^{\frac{\theta}{\gamma_2 - \theta}} \left( \frac{p_1}{p_2} \right)^{\frac{\alpha\theta}{1-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} + \left( \frac{r}{A_2 \gamma_2 x_k} \right)^{\frac{\theta}{\theta - \gamma_2}} \left[ 1 - \left( \frac{p_1}{p_2} \right)^{\frac{\alpha\theta}{1-\theta}} \right] x_k$$

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:

$$\left\{\begin{array}{c} \left(m_{21}^{\theta}x_2 + k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}} m_{21}^{\theta}x_2 - \\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{22}^{\theta}x_2 + k_2^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}} m_{22}^{\theta}x_2 \end{array}\right\} =$$
$$\left\{\begin{array}{c} \left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right] \\ + \left(\frac{M_3}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3} x_3\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_3}{\gamma_3-1}}\right] \end{array}\right\} \tag{2.39}$$

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

$$\left(m_{22}^{\theta}x_2 + k_2^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}} = \left(\frac{p_2}{p_1}\right)^{\alpha}\left(m_{21}^{\theta}x_2 + k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}} \dots \tag{2.40}$$
$$\left(\frac{m_{21}}{m_{22}}\right)^{\theta-1}\left(\frac{A_1}{A_2}\right)$$

Substituting (2.40) into (2.39) and rearranging, we have:

$$\left\{ \left[1 - \left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_2-m_{21}}{m_{21}}\right]\left(m_{21}^{\theta}x_2 + k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}} m_{21}^{\theta}x_2 \right\} =$$
$$\left\{\begin{array}{c} \left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1} x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right] \\ + \left(\frac{M_3}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3} x_3\left[\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_3}{\gamma_3-1}}\right] \end{array}\right\} \tag{2.41}$$

But then, from equation (2.33), we have that:

$$m_{21}^{\theta}x_2 = \left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\theta)}{\gamma_2-\theta}} - k_1^{\theta}x_k \tag{2.42}$$

Similarly, from $(k_1)$, we have:

$$\left(m_{21}^{\theta}x_2 + k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}} = \left(\frac{r}{A_1\gamma_2 x_k}\right)k_1^{1-\theta} \tag{2.43}$$

Then, from (2.42) and (2.43), we have:

$$\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}m_{21}^{\theta}x_2=\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\gamma_2}{\gamma_2-\theta}}k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2-\theta}}-\frac{r}{A_1\gamma_2}k_1 \quad (2.44)$$

Substituting equation (2.37) into (2.32) and manipulating, we have:

$$\frac{M_2-m_{21}}{m_{21}}=\frac{M_2x_2^{\frac{1}{\theta}}-k_1\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}{k_1\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}} \quad (2.45)$$

Consequently:

$$\left[1-\left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_2-m_{21}}{m_{21}}\right]=\frac{\left\{\begin{array}{c}\left(1+\left(\frac{p_1}{p_2}\right)^{1-\alpha}\right)k_1\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}\\-\left(\frac{p_1}{p_2}\right)^{1-\alpha}M_2x_2^{\frac{1}{\theta}}\end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}} \quad (2.46)$$

Then, from equations (2.44) and (2.46), we have that:

$$\left[1-\left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_2-m_{21}}{m_{21}}\right]\left(m_{21}^{\theta}x_2+k_1^{\theta}x_k\right)^{\frac{\gamma_2-\theta}{\theta}}m_{21}^{\theta}x_2= \frac{\left\{\begin{array}{c}\left(1+\left(\frac{p_1}{p_2}\right)^{1-\alpha}\right)k_1\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}\\-\left(\frac{p_1}{p_2}\right)^{1-\alpha}M_2x_2^{\frac{1}{\theta}}\end{array}\right\}}{\begin{array}{c}k_1\left[\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}\\\times\left\{\left(\frac{r}{A_1\gamma_2x_k}\right)^{\frac{\gamma_2}{\gamma_2-\theta}}k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2-\theta}}-\frac{r}{A_1\gamma_2}k_1\right\}\end{array}} \quad (2.47)$$

Notice that the LHS of equation (2.47) is the same of the one of equation

(2.41). Substituting it back, we have:

$$\frac{\left\{\begin{array}{c}\left(1+\left(\frac{p_1}{p_2}\right)^{1-\alpha}\right)k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}\\ -\left(\frac{p_1}{p_2}\right)^{1-\alpha}M_2x_2^{\frac{1}{\theta}}\end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}}k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}}-x_k\right]^{\frac{1}{\theta}}}\times\left\{\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\gamma_2}{\gamma_2-\theta}}k_1^{\frac{\gamma_2(1-\theta)}{\gamma_2-\theta}}-\frac{r}{A_1\gamma_2}k_1\right\}=$$

$$\left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\ +\left(\frac{M_3}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_3-1}}}\right)^{\gamma_3}x_3\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_3}{\gamma_3-1}}\right]\end{array}\right\} \tag{2.48}$$

Finally, notice that equations (2.48) and (2.38) generate a system with two equations and two unknowns ($k_1$ and $\frac{p_1}{p_2}$):

$$\left\{\begin{array}{l}\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_2x_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\}= \qquad (F.1)\\ \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\}\\ \left\{\frac{M_2x_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}}= \qquad (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\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.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.:

$$\frac{m_{i1}}{m_{11}+m_{21}+m_{31}} = \frac{m_{i2}}{m_{12}+m_{22}+m_{32}} \tag{2.49}$$

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 (2.44), 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 (2.43), 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:

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

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

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}}{\left( \frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta}} \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} < k_1 \left( \frac{p_2}{p_1} \right)^{\left(1 + \frac{\alpha\theta}{1-\theta}\right)} \tag{2.50}$$

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

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}}{\left( \frac{p_1}{p_2} \right)^{\frac{\alpha}{1-\theta}} \left[ \left( \frac{r}{A_1 \gamma_2 x_k} \right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k \right]^{\frac{1}{\theta}}} > \left( \frac{p_2}{p_1} \right)^{\frac{\alpha\theta}{1-\theta} \times \frac{\theta - \gamma_2}{\theta(1-\gamma_2)}} k_1 \tag{2.51}$$

Then, notice that:

$$1 + \frac{\alpha\theta}{1-\theta} - \frac{\alpha\theta}{1-\theta} \times \frac{\theta - \gamma_2}{\theta(1-\gamma_2)} = 1 + \frac{\alpha\theta}{1-\theta} \left[ 1 - \frac{\theta - \gamma_2}{\theta(1-\gamma_2)} \right] = \quad \dots \tag{2.52}$$

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

Therefore the exponent at $\frac{p_2}{p_1}$ is higher at the RHS of (2.50). 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 (2.50) and (2.51) 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:

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

Since $p_1 < p_2$, $\gamma_1 < 1$, and $\gamma_3 < 1$. Therefore, the LHS of $(F.1)$ must also be 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 $\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}$:

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[ \left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[ \left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}}} > k_1 \left(\frac{p_2}{p_1}\right)^{\left(1+\frac{\alpha\theta}{1-\theta}\right)} \tag{2.53}$$

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

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[ \left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[ \left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k \right]^{\frac{1}{\theta}}} < \left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta} \times \frac{\theta-\gamma_2}{\theta(1-\gamma_2)}} k_1 \tag{2.54}$$

Since:

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

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 (2.53) and (2.54) give us a contradiction. Therefore, we have that $p_1 = p_2 \Leftrightarrow A_1 = A_2$. ■

# 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:

$$F_j(m_{1j},m_{2j},m_{3j},k_j)=A_j\left[m_{1j}^{\gamma_1}x_1+\left(m_{3j}^{\theta}x_3+k_j^{\theta}x_k\right)^{\frac{\gamma_3}{\theta}}+m_{2j}^{\gamma_2}x_2\right] \tag{2.55}$$

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:

$$\frac{w_{ij}}{w_{ij'}} = \left(\frac{p_j}{p_{j'}}\right)^{\alpha}, \ \ \forall i \in \{1,2,3\} \text{ and } \forall j \in \{1,2\} \tag{2.56}$$

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

$$m_{11} = \frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1 - 1}} M_1}{1 + \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1 - 1}}} \tag{2.57}$$

Similarly, for skill 2:

$$m_{21} = \frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2 - 1}} M_2}{1 + \left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2 - 1}}} \tag{2.58}$$

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

$$\left(\frac{m_{31}}{m_{32}}\right) = \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{\theta - 1}} \frac{k_1}{k_2} \tag{2.59}$$

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

$$m_{31} = \left\{\frac{1}{x_3}\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3 - \theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k\right]\right\}^{\frac{1}{\theta}} k_1 \tag{2.60}$$

Similarly, for $(k_2)$, we have:

$$m_{32} = \left\{\frac{1}{x_3}\left[\left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3 - \theta}} k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k\right]\right\}^{\frac{1}{\theta}} k_2 \tag{2.61}$$

Dividing (2.33) by (2.61)and substituting (2.59), we have:

$$\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta - 1}} = \left\{\frac{\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3 - \theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k\right]}{\left[\left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3 - \theta}} k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3 - \theta}} - x_k\right]}\right\} \tag{2.62}$$

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:

$$M_3 x_3^{\frac{1}{\theta}} = \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1 + \left[\left(\frac{r}{A_2\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_2^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_2 \tag{2.63}$$

Substituting (2.62) and manipulating, we have:

$$k_2 = \frac{M_3 x_3^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \tag{2.64}$$

Substituting (2.64) into (2.63) and manipulating, we have:

$$\left\{\frac{M_3 x_3^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} \left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}}\right\}^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} = \tag{2.65}$$

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

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:

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

$$\left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\+\left(\frac{M_2}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2}x_2\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_2}{\gamma_2-1}}\right]\end{array}\right\}\tag{2.66}$$

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

$$\left(m_{32}^{\theta}x_3+k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}}=\left(\frac{p_2}{p_1}\right)^{\alpha}\left(m_{31}^{\theta}x_3+k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}}\times\ldots\tag{2.67}$$
$$\left(\frac{m_{31}}{m_{32}}\right)^{\theta-1}\times\left(\frac{A_1}{A_2}\right)$$

Substituting (2.67) into (2.66) and rearranging, we have:

$$\left\{\left[1-\left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_3-m_{31}}{m_{31}}\right]\left(m_{31}^{\theta}x_3+k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}}m_{31}^{\theta}x_3\right\}=$$

$$\left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\+\left(\frac{M_2}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2}x_2\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_2}{\gamma_2-1}}\right]\end{array}\right\}\tag{2.68}$$

But then, from equation (2.60), we have that:

$$m_{31}^{\theta}x_3=\left(\frac{r}{A_1\gamma_3x_k}\right)^{\frac{\theta}{\gamma_3-\theta}}k_1^{\frac{\theta(1-\theta)}{\gamma_3-\theta}}-k_1^{\theta}x_k\tag{2.69}$$

Similarly, from $(k_1)$, we have:

$$\left(m_{31}^{\theta}x_3+k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}}=\left(\frac{r}{A_1\gamma_3x_k}\right)k_1^{1-\theta}\tag{2.70}$$

Then, from (2.69) and (2.70), we have:

$$\left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 = \left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\gamma_3}{\gamma_3-\theta}} k_1^{\frac{\gamma_3(1-\theta)}{\gamma_3-\theta}} - \frac{r}{A_1\gamma_3}k_1 \quad (2.71)$$

Substituting equation (2.64) into (2.59) and manipulating, we have:

$$\frac{M_3 - m_{31}}{m_{31}} = \frac{M_3 x_3^{\frac{1}{\theta}} - k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \quad (2.72)$$

Consequently:

$$\left[1 - \left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_3 - m_{31}}{m_{31}}\right] = \frac{\left\{\begin{array}{c}\left(1 + \left(\frac{p_1}{p_2}\right)^{1-\alpha}\right) k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} \\ - \left(\frac{p_1}{p_2}\right)^{1-\alpha} M_3 x_3^{\frac{1}{\theta}}\end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \quad (2.73)$$

Then, from equations (2.71) and (2.73), we have that:

$$\left[1 - \left(\frac{p_1}{p_2}\right)^{1-\alpha}\frac{M_3 - m_{31}}{m_{31}}\right]\left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 =$$
$$\frac{\left\{\begin{array}{c}\left(1 + \left(\frac{p_1}{p_2}\right)^{1-\alpha}\right) k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}} \\ - \left(\frac{p_1}{p_2}\right)^{1-\alpha} M_3 x_3^{\frac{1}{\theta}}\end{array}\right\}}{k_1\left[\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\theta}{\gamma_3-\theta}} k_1^{\frac{\theta(1-\gamma_3)}{\gamma_3-\theta}} - x_k\right]^{\frac{1}{\theta}}} \times \left\{\left(\frac{r}{A_1\gamma_3 x_k}\right)^{\frac{\gamma_3}{\gamma_3-\theta}} k_1^{\frac{\gamma_3(1-\theta)}{\gamma_3-\theta}} - \frac{r}{A_1\gamma_3}k_1\right\} \quad (2.74)$$

Notice that the LHS of equation (2.74) is the same of the one of equation

(2.68). Substituting it back, we have:

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

$$\left\{\begin{array}{c}\left(\frac{M_1}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_1-1}}}\right)^{\gamma_1}x_1\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_1}{\gamma_1-1}}\right]\\ +\left(\frac{M_2}{1+\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_2-1}}}\right)^{\gamma_2}x_2\left[\frac{A_2}{A_1}\frac{p_1}{p_2}-\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_2}{\gamma_2-1}}\right]\end{array}\right\} \tag{2.75}$$

Finally, notice that equations (2.75) and (2.65) generate a system with two equations and two unknowns ($k_1$ and $\frac{p_1}{p_2}$):

$$
\begin{cases}
\dfrac{\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_3x_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}}} \\
\quad\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\}= \qquad (F.1)\\
\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_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\} \\
\left\{\dfrac{M_3x_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}} = \qquad (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{cases}
$$

## 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* $\frac{m_{i1}}{m_{i2}} = \textit{constant}, \forall i \in \{1,2,3\}$.

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

$$\frac{m_{i1}}{m_{11}+m_{21}+m_{31}} = \frac{m_{i2}}{m_{12}+m_{22}+m_{32}} \tag{2.76}$$

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 (2.62), we have $k_1 < k_2$. But then, from equation (2.59), we obtain $m_{31} < m_{32}$. Finally, from the RHS of (2.39), 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 (2.62) and (2.59), the LHS of (2.66) 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 (2.62), we have $k_1 > k_2$. Similarly, from (2.59), we obtain $m_{31} > m_{32}$. But then, from (2.66), 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(2.66) $< 0$. 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 (2.59), we have:

$$\frac{m_{31}}{m_{32}} = \frac{k_1}{k_2} \tag{2.77}$$

From (2.62) , we have

$$\frac{k_1}{k_2} = \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{2.78}$$

Combining (2.77) and (2.78), we have:

$$\frac{m_{31}}{m_{32}} = \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{2.79}$$

But then, from LHS(2.66), substituting (2.77) and (2.79) given $p_1 = p_2$, we have:

$$\left\{ \begin{array}{c} \left(m_{31}^{\theta}x_3 + k_1^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta}x_3 - \\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3 \end{array} \right\} = \ldots \tag{2.80}$$
$$\left[\left(\frac{A_1}{A_2}\right)^{\frac{\gamma_3}{1-\gamma_3}} - \frac{A_2}{A_1}\right]\left(m_{32}^{\theta}x_3 + k_2^{\theta}x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta}x_3$$

while the RHS(2.66) gives us:

$$\left\{ \begin{array}{c} \left(\frac{M_1}{1+\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_1}}}\right)^{\gamma_1} x_1\left[\frac{A_2}{A_1} - \left(\frac{A_1}{A_2}\right)^{\frac{\gamma_1}{1-\gamma_1}}\right] \\ +\left(\frac{M_2}{1+\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_2}}}\right)^{\gamma_2} x_2\left[\frac{A_2}{A_1} - \left(\frac{A_1}{A_2}\right)^{\frac{\gamma_2}{1-\gamma_2}}\right] \end{array} \right\} \tag{2.81}$$

Then, consider the case in which $A_1 > A_2$. From (2.80), we have that LHS(2.66)$>0$, while (2.81) gives us RHS(2.66)$< 0$. Similarly, if $A_1 < A_2$, (2.80) gives us LHS(2.66)$< 0$ while (2.81) gives us RHS(2.66)$> 0$. Consequently, (2.66) is only satisfied if $A_1 = A_2$, concluding our proof.■

# Proofs

### Proof of Proposition 2

*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 (2.50) 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:

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}} > \left(\frac{p_2}{p_1}\right)^{\frac{\alpha\theta}{1-\theta}\times\frac{\theta-\gamma_2}{\theta(1-\gamma_2)}} k_1 \quad (2.82)$$

Given (2.52) 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, (2.50) and (2.82) 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 (2.35), we have:

$$\left(\frac{p_1}{p_2}\right)^{\frac{\alpha\theta}{\theta-1}} = \left[\frac{\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}{\left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}\right] \tag{2.83}$$

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

$$\left[\frac{\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}{\left(\frac{r}{A_2\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_2^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k}\right] < 1 \tag{2.84}$$

Rearranging it:

$$\left(\frac{k_1}{k_2}\right)^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} < \left(\frac{A_1}{A_2}\right)^{\frac{\theta}{\gamma_2-\theta}} \tag{2.85}$$

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:

$$\frac{A_2}{A_1}\frac{p_1}{p_2} - \left[\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2}\right]^{\frac{\gamma_i}{1-\gamma_i}}, \ \forall i \in \{1,3\} \tag{2.86}$$

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 (2.86) 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 (2.50) 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:

$$\left\{\frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}\right\} > \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\gamma_2}}\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}\times\frac{\gamma_2-\theta}{(1-\gamma_2)}} k_1 \tag{2.87}$$

Then, we have that:

$$\frac{\text{RHS(2.50)}}{\text{RHS(2.87)}} = \left(\frac{p_2}{p_1}\right)^{1+\frac{\alpha\theta}{1-\theta}\left[1-\frac{\gamma_2-\theta}{\theta(1-\gamma_2)}\right]}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_2}} \tag{2.88}$$

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

$$\frac{\text{RHS(2.50)}}{\text{RHS(2.87)}} = \left(\frac{p_2}{p_1}\right)^{1+\frac{\gamma_2\alpha}{(1-\gamma_2)}}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_2}} = \left\{\left(\frac{p_2}{p_1}\right)^{1-\gamma_2(1-\alpha)}\frac{A_1}{A_2}\right\}^{\frac{1}{1-\gamma_2}} \tag{2.89}$$

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:

$$\left(\frac{p_2}{p_1}\right)^{1-\gamma_2(1-\alpha)}\frac{A_1}{A_2} < \left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2} < 1 \tag{2.90}$$

where the last inequality comes from our assumption for the contradiction. Then, since $\frac{1}{1-\gamma_2} > 0$, we have $\frac{\text{RHS}(2.50)}{\text{RHS}(2.87)} < 1$. But then inequalities (2.50) and (2.87) cannot both be satisfied and we have a contradiction. $\square$

**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:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1 - 1}} M_1}{\left\{1 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma_1 - 1}}\right\}} = \frac{\left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma_1}} M_1}{\left\{1 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma_1}}\right\}} \tag{2.91}$$

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 $m_{31} > m_{32}$. $\square$

**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 (2.45):

$$\frac{M_2 x_2^{\frac{1}{\theta}} - k_1 \left[\left(\frac{r}{A_1 \gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k\right]^{\frac{1}{\theta}}}{\left[\left(\frac{r}{A_1 \gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2 - \theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2 - \theta}} - x_k\right]^{\frac{1}{\theta}}} > k_1 \tag{2.92}$$

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

$$\left\{\frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}}\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}\right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} = \tag{2.93}$$

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

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:

$$\left\{\frac{M_2 x_2^{\frac{1}{\theta}} - \left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}} k_1}{\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}\right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} > \dots \tag{2.94}$$

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

Now, from (2.92) we have that, since $\gamma_2 < \theta$:

$$\left\{\frac{M_2 x_2^{\frac{1}{\theta}} - k_1\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}{\left[\left(\frac{r}{A_1\gamma_2 x_k}\right)^{\frac{\theta}{\gamma_2-\theta}} k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} - x_k\right]^{\frac{1}{\theta}}}\right\}^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} < k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} \tag{2.95}$$

Now, substituting (2.95) into (2.94), we have:

$$k_1^{\frac{\theta(1-\gamma_2)}{\gamma_2-\theta}} > \dots \tag{2.96}$$

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

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

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 (2.45) 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:

$$\frac{pdf_{11}}{pdf_{12}} > \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma}} \times \frac{1}{\left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma}}} \tag{2.97}$$

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 4

*Proof.* Towards a contradiction, assume that $A_1 > A_2$ and $p_2 > p_1$. Then, from (2.62), 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.:

$$\frac{k_1}{k_2} > \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{2.98}$$

From equation (2.59), we have:

$$\frac{m_{31}}{m_{32}} > \frac{k_1}{k_2} \Rightarrow \frac{m_{31}}{m_{32}} > \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_3}} \tag{2.99}$$

Then, from LHS(2.66), substituting (2.98) and (2.99), we have:

$$\left\{ \begin{array}{c} \left(m_{31}^{\theta} x_3 + k_1^{\theta} x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{31}^{\theta} x_3 - \\ -\frac{A_2}{A_1}\frac{p_1}{p_2}\left(m_{32}^{\theta} x_3 + k_2^{\theta} x_k\right)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta} x_3 \end{array} \right\} > \left[\left(\frac{A_1}{A_2}\right)^{\frac{\gamma_3}{1-\gamma_3}} - \frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)\right] (m_{32}^{\theta} x_3 + k_2^{\theta} x_k)^{\frac{\gamma_3-\theta}{\theta}} m_{32}^{\theta} x_3 > 0 \tag{2.100}$$

While from RHS(2.66), 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(2.39)$< 0$, 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 5

*Proof.* Without loss of generality, assume that $A_1 > A_2$. From proposition 4 we have that $A_1 > A_2 \Rightarrow p_1 > p_2$. From (2.64) 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 (2.59), 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:

$$\frac{m_{31}}{m_{32}} > \left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right]^{\frac{1}{1-\gamma_3}} \tag{2.101}$$

Let's consider two cases:

**Case 1:** $\left[\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)^{\alpha}\right] \geq 1$ – In this case, equation (2.101) already implies that $m_{31} \geq m_{32}$. From (2.59) and $\theta < 1$, we have that:

$$\frac{k_1}{k_2} \geq \left(\frac{p_1}{p_2}\right)^{\frac{\alpha}{1-\theta}} > 1 \tag{2.102}$$

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(2.66), we have that:

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

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(2.66)$>$ 0.

But then, from (2.68), 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:

$$\frac{m_{31}}{m_{32}} > \left(\frac{p_1}{p_2}\right)^{1-\alpha} > 1 \tag{2.103}$$

From (2.103) and (2.59), we have:

$$\left(\frac{p_2}{p_1}\right)^{\frac{\alpha}{1-\theta}}\frac{k_1}{k_2} > \left(\frac{p_1}{p_2}\right)^{1-\alpha} \Rightarrow \frac{k_1}{k_2} > \left(\frac{p_1}{p_2}\right)^{1+\frac{\alpha\theta}{1-\theta}} \tag{2.104}$$

Consequently, (2.104) implies that $k_1 > k_2$, concluding our proof. □

# Proof of Corollary 1

*Proof.* Proof of proposition 5 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(2.39)$> 0$. Following the same steps presented in the proof of proposition 5, we can conclude that $m_{3i} > m_{3j}$. $\square$

# Proof of Theorem 3

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

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

Rearranging and simplifying it, we obtain:

$$m_{31}m_{12} - m_{32}m_{11} + m_{31}m_{22} - m_{32}m_{21} \leq 0 \tag{2.105}$$

From equations (2.57) and (2.58) and labor market clearing conditions, we have:

$$m_{11} = \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}} m_{12} \text{ and } m_{21} = \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}} m_{22} \tag{2.106}$$

As a result, we have:

$$m_{31}m_{12} - m_{32}m_{11} = m_{32}m_{12}\left\{\frac{m_{31}}{m_{32}} - \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}}\right\} > 0 \tag{2.107}$$

and

$$m_{31}m_{22} - m_{32}m_{21} = m_{32}m_{22}\left\{\frac{m_{31}}{m_{32}} - \left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}}\right\} > 0 \tag{2.108}$$

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 (2.101). Consequently, equations (2.105), (2.107), and (2.108) jointly show a contradiction. As a result, $pdf_{31} > pdf_{32}$.

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

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

Rearranging and simplifying it, we obtain:

$$m_{12}m_{21} - m_{22}m_{11} + m_{32}m_{21} - m_{31}m_{22} \leq 0 \tag{2.109}$$

From (2.106), after some manipulations, we have:

$$m_{12}m_{21} - m_{22}m_{11} = 0 \tag{2.110}$$

and

$$m_{32}m_{21} - m_{31}m_{22} = m_{32}m_{22}\left\{\left[\frac{A_1}{A_2}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}} - \frac{m_{31}}{m_{32}}\right\} < 0 \tag{2.111}$$

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 (2.101). Consequently, equations (2.109), (2.110), and (2.111) jointly show a contradiction. As a result, $pdf_{21} < pdf_{22}$.

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

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

Rearranging and simplifying it, we obtain:

$$m_{11}m_{22} - m_{12}m_{21} + m_{32}m_{11} - m_{31}m_{12} \leq 0 \tag{2.112}$$

In equation (2.110), we already showed that $m_{11}m_{22} - m_{12}m_{21} = 0$. Then, from (2.106) and (2.101), we have:

$$m_{32}m_{11} - m_{31}m_{12} = m_{32}m_{12} \left\{ \left[ \frac{A_1}{A_2} \left( \frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{1-\gamma}} - \frac{m_{31}}{m_{32}} \right\} < 0 \quad (2.113)$$

Consequently, equations (2.112), (2.110), and (2.113) jointly show a contradiction. As a result, $pdf_{11} < pdf_{12}$, concluding our proof that $pdf_1$ F.O.S.D. $pdf_2$. $\square$

# Empirical Evidence - 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.

## Data Coverage

As discussed in section 2.4 and presented in table 2.6, 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 2.21, even after controlling for establishment size, MSA average coverage is above 30% only for establishments that have 250 employees or more. Finally, table 2.22 shows that employment coverage is below 30% in all industry sectors.

Table 2.21: 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 |

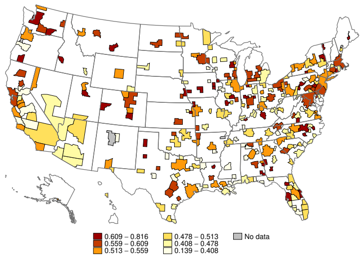
Table 2.22: Ci Coverage relative to NETS: Employment by Industry

| | **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 |

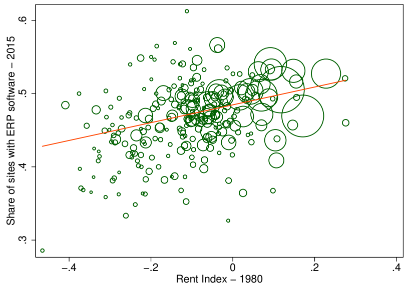
Nonetheless, table 2.23 shows that there is a lot of dispersion in the ERP shares across MSAs even in 2015, when we should expect already a more widespread use of technology. As we can see, we have at least some information on 272 MSAs across the country. Moreover, we can see that, while on average about 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 2.6b, 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 2.6a 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 2.23: 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 2.6: Geographical distribution of ERP across CMSAs – 2015

## Empirical Evidence

Table 2.24 shows the results of a linear regression of the index of ERP usage on the local price index and the share of routine cognitive workers in 1980. Columns 1 and 2 indicate that, when considered separately, both the current local price level and the past share of routine-cognitive workers can explain a substantial amount of variation in ERP adoption. In the first specification, a $10\%$ higher local price index (about half a standard deviation) is associated with a $0.8\%$ increase in the share of sites that use ERP. The most expensive places in our sample have about $8\%$ more sites that use ERP, relative to the cheapest places. In the second specification, a $1\%$ higher the share of routine-cognitive workers in 1980 (about half a standard deviation) is associated with a $0.05\%$ higher share of sites with ERP. However, when considering both variables jointly the effect of the lagged routine-cognitive share conditional on the local price level shrinks substantially and turns out insignificant. Yet, the coefficient on the price level is stable across the specifications. These results indicate that the usage of ERP is more widespread in MSAs with a higher local price level, but conditional on the current price level the past concentration of routine-cognitive workers does not predict the change in concentration. This is in line with the theoretic prediction that cities with high living costs invest more in automation technology.

Table 2.24: Enterprise Resource Planning Software

| | (1) | (2) | (3) |
|---|---|---|---|
| | ERP 2015 | ERP 2015 | ERP 2015 |
| log rent index | 0.0858*** | | 0.0777*** |
| | (0.0153) | | (0.0180) |
| routine cognitive share 1980 | | 0.508*** | 0.205 |
| | | (0.133) | (0.137) |
| Observations | 253 | 253 | 253 |
| $R^2$ | 0.230 | 0.082 | 0.242 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Each observation (an MSA) is weighted by its employment in 2015.

## Robustness Check - Weighting

In the main analysis of the data we weighted each MSA by its employment. This may raise concerns that we are relying too much on large cities to inform our estimates. Therefore, we consider a different weighting scheme to check for robustness. One may argue that we should compare our previous results to results without any weighting. However, our estimates of the ERP share are very noisy for some MSAs due to small sample size. Therefore we use the inverse of the standard error of the ERP index as weights. We calculate the standard error

$\sigma_{ERP}$ according to the following formula

$$\hat{\sigma}_{ERP} = \sqrt{\frac{\hat{\pi}_{ERP}(1 - \hat{\pi}_{ERP})}{N}} \tag{2.114}$$

where $\hat{\pi}_{ERP}$ is the empirical share of ERP sites in a location and $N$ the number of sites.

In order to examine the differences in empirical results between this weighting scheme and the original weighting by size table 2.25 and 2.26 replicate the corresponding estimations of the main section with the new weighting scheme. The results in table 2.24 and 2.25 show the results regarding ERP adoption with the two weighting schemes. We find that the parameter estimates are very similar.

Table 2.25: Entreprise Resource Planning Software - weights $\hat{\sigma}^{-1}_{ERP}$

| | (1) | (2) | (3) |
|---|---|---|---|
| | ERP 2015 | ERP 2015 | ERP 2015 |
| log rent index | 0.0769*** | | 0.0650*** |
| | (0.0141) | | (0.0158) |
| routine cognitive share 1980 | | 0.420*** | 0.225* |
| | | (0.110) | (0.112) |
| Observations | 253 | 253 | 253 |
| $R^2$ | 0.114 | 0.060 | 0.128 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Each observation (an MSA) is weighted by the inverse of the estimated standard error of the ERP measure.

Comparing the results regarding the change in the routine-cognitive share of employment, table 2.12 and 2.26, we find again the same pattern. Results are very similar, albeit parameter estimates on rent being slightly smaller in absolute size. From this exercise we conclude that results were not simplz driven by very large MSAs.

Table 2.26: Change in routine-cognitive share, 1980-2015 $\hat{\sigma}_{ERP}^{-1}$

| | (1) | (2) | (3) |
|---|---|---|---|
| | $\Delta$ rout-cog | $\Delta$ rout-cog | $\Delta$ rout-cog |
| log rent index | -0.0700*** | | -0.0326*** |
| | (0.00903) | | (0.00902) |
| routine cognitive share 1980 | | -0.801*** | -0.703*** |
| | | (0.0466) | (0.0426) |
| Observations | 253 | 253 | 253 |
| $R^2$ | 0.240 | 0.558 | 0.601 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Each observation (an MSA) is weighted by the inverse of the estimated standard error of the ERP measure.

Finally, in the next appendix we replace the log rent index by log employment. As presented in section 2, the model delivers an equilibrium in which not only the more productive city is more expensive, but it is also bigger. As we can see from tables 2.27 and 2.28, results are qualitatively the same as the ones presented here.

# Empirical Evidence - City Size

## 2.7.1 Size

In the theory there is strong relationship between size and productivity of locations, which should also lead to a similar house price productivity relationship. In this section we consider, whether it holds that size has a similar relationship as house prices with automation technology adoption and the change in occupational shares.

Table 2.27: Entreprise Resource Planning Software

| | (1) | (2) | (3) |
|---|---|---|---|
| | ERP 2015 | ERP 2015 | ERP 2015 |
| log employment | 0.0143*** | | 0.0135*** |
| | (0.00259) | | (0.00331) |
| routine cognitive share 1980 | | 0.508*** | 0.106 |
| | | (0.133) | (0.166) |
| Observations | 253 | 253 | 253 |
| $R^2$ | 0.261 | 0.082 | 0.264 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Each observation (an MSA) is weighted by its employment in 2015.

Qualitatively the relationship between rent and ERP adoption, as well as the change in the routine-cognitive share is the same as with size.

Table 2.28: Change in routine-cognitive share, 1980-2015

| | (1) | (2) | (3) |
|---|---|---|---|
| | $\Delta$ rout-cog | $\Delta$ rout-cog | $\Delta$ rout-cog |
| log employment | -0.00941*** | | -0.00417 |
| | (0.00238) | | (0.00219) |
| routine cognitive share 1980 | | -0.831*** | -0.707*** |
| | | (0.0765) | (0.0948) |
| Observations | 253 | 253 | 253 |
| $R^2$ | 0.252 | 0.490 | 0.529 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Each observation (an MSA) is weighted by its employment in 2015.

# Estimation Details

## 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

$$\hat{V} = (\hat{M}'\Omega\hat{M})^{-1}\hat{M}'\Omega\hat{\Sigma}\Omega\hat{M}(\hat{M}'\Omega\hat{M})^{-1} \tag{2.115}$$

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 $\Omega$ is the weight matrix, here $\Omega = diag(1)$.

## List of Softwares

Table 2.29: 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 |

# 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.

## High-TFP city is smaller (the "Boulder" case)

Consider the following parameter values:

In this case, we obtain the following equilibrium prices and quantities:

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,

## Example 1: Parameters

| $x_1$ | $x_2$ | $x_3$ | $A_1$ | $A_2$ | $H$ | $x_k$ | r |
|---|---|---|---|---|---|---|---|
| 0.3189 | 1 | 1.4733 | 19118 | 19000 | 62559000 | 1.333 | 635.58 |
| $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\theta$ | $\alpha$ | $M_1$ | $M_2$ | $M_3$ |
| 0.8 | 0.8 | 0.82 | 0.5 | 0.24 | 15836150 | 66973717 | 40745094 |

## Example 1: Equilibrium outcomes

| | $m_{1j}$ | $m_{2j}$ | $m_{3j}$ | $S_j$ | $p_j$ |
|---|---|---|---|---|---|
| **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 |

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 1.

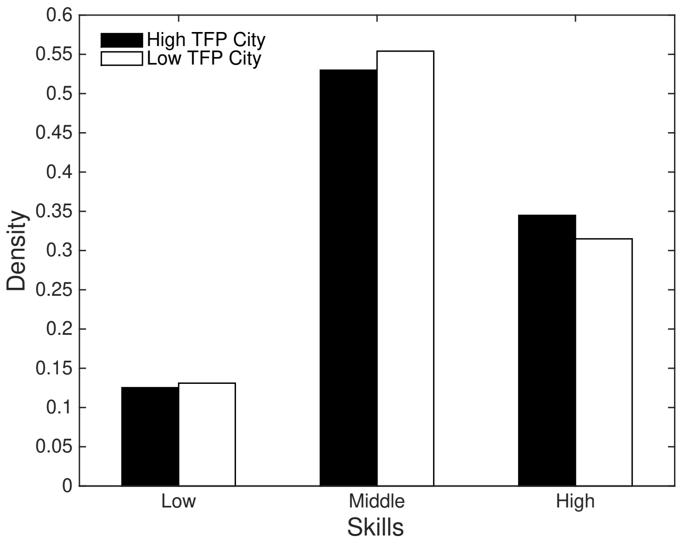



Figure 2.7: Skill Distribution: High vs. Low TFP cities - Example 1

### High-TFP city is larger (the “NYC” case)

Consider the following parameter values:

In order to make a simple comparison, the parameters are the same of Example 1, apart from a higher $A_1$. In this case, we obtain the following equilibrium prices and quantities:

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 2.

### Example 2: Parameters

| $x_1$ | $x_2$ | $x_3$ | $A_1$ | $A_2$ | $H$ | $x_k$ | r |
|---|---|---|---|---|---|---|---|
| 0.3189 | 1 | 1.4733 | 21118 | 19000 | 62559000 | 1.333 | 635.58 |
| $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\theta$ | $\alpha$ | $M_1$ | $M_2$ | $M_3$ |
| 0.8 | 0.8 | 0.82 | 0.5 | 0.24 | 15836150 | 66973717 | 40745094 |

### Example 2: Equilibrium outcomes

| | $m_{1j}$ | $m_{2j}$ | $m_{3j}$ | $S_j$ | $p_j$ |
|---|---|---|---|---|---|
| **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 |

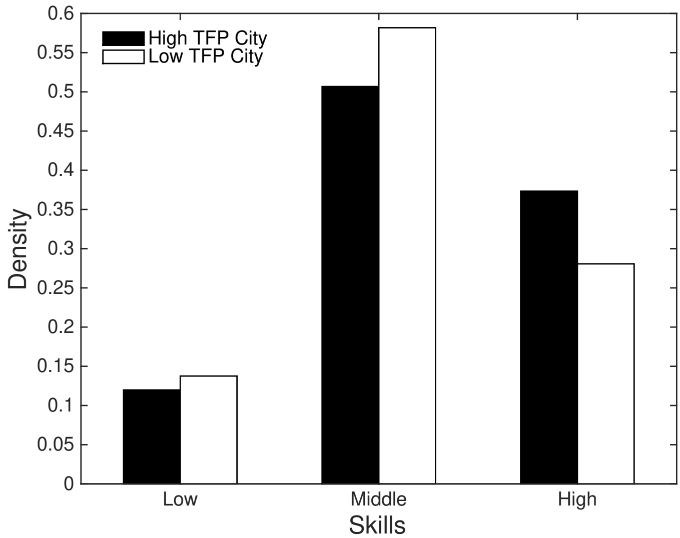



Figure 2.8: Skill Distribution: High vs. Low TFP cities - Example 2

# Chapter 3

# GENDER GAP IN EARNINGS AMONG PROFESSORS

## 3.1 Introduction

In this paper I study the extent of the gender pay gap among professors at the University of California (UC). The academic environment provides a setting with a high skill labor market and readily observable productivity measures of research output. Thus, the setting allows to decompose to what extent differences in pay can be attributed to differences in productivity. The main question this study addresses to what extent women and men are paid differently in similar jobs with similar productivity. I find that women are consistently paid less than men, which holds across major fields of study and UC campuses. After controlling for productivity and position type, a substantial gender gap remains. Interestingly, the fields with the lowest share of female professors (Physics, Math and Engineering) show relatively small gender gaps in pay, which disappear after controlling for research productivity and type of position. In terms of returns to research productivity I find no differences between men and women.

The present study relates to a large and expanding literature examining gender gaps in a variety of labor markets and on different outcomes. For example, Azmat and Ferrer (2017) study young lawyers and find that women are on average less productive and that differences in productivity can explain a substantial part in differences in pay. The study is closely related and complementary as it focuses on a different segment of the labor market. Weisshaar (2017) analyses the career trajectories of assistant professors in Sociology, Computer Science, and English. She finds that the gender gap in promotion to tenure, that women are less likely to receive tenure, can be partially explained by productivity. However, that the department context does not explain the gender gap in tenure. Furthermore, she finds that women, once they receive tenure they do so on average in lower prestige departments. Overall, the results point towards differential treatment of men and women in the tenure evaluation process. In this study, I analyze the gender gap in earnings of professors instead of promotion. One piece of evidence for differential treatment of men and women is by Sheltzer and Smith (2014), who show that male faculty members tend to employ fewer women than female faculty members in the life-sciences across US academic institutions. This points towards that departments or fields that have a higher share of women might provide better opportunities for women. A hypothesis I will address for the case of gender gaps in pay. Bandiera et al. (2016) report substantial gaps in earnings between men and women at LSE. The results are in several features quite similar to the findings in the present study. One major concern of the literature studying gender gaps in academic environments is the under-representation of women, see e.g. Ceci et al. (2014) or Ginther and Kahn (2004). Reuben et al. (2014) show that women are discriminated against in the hiring for a mathematical task. Providing more information shrinks the gap, but even under full information about past performance the gap does not close fully.

They find that the bias is related to stereotypes as measured by an Implicit Association test. Such stereotypes give a rationale for the finding of gender gaps in pay, if stereotypes are biased against women in an academic setting.

In the following I will first present the data used in the study and then present the results regarding the gender pay gap.

## 3.2 Data

**University of California Salary Records** are made public on an annual basis. I use the records for the years 2013 - 2017. The records include the name of the employee, job title, pay (split up into subcategories) and some further information. I subset the records to professors based on job title. For those individuals I estimate their gender based on their first name using the service *gender.io*. In the rest of the article I will refer to the estimated gender based on first name simply as ”gender” for brevity.

**The Microsoft Academic Graph (MAG)** is a database on academic articles and patents (Sinha et al., 2015). It includes information on individual articles, authors and citations/references. Additionally, the MAG also provides a classification into fields of study. I construct the list of authors who were ever affiliated to the University of California according to their affiliation on papers that are found in the MAG. Then I match this list of authors to the University of California salary records. For the matching I use the machine-learning algorithm developed in Bilenko (2006), which I train using the names of authors[1]. Table 3.1 show that the sample coverage in terms of matching the MAG and the public salary records is very good, but almost a third of observations could not

[1] The python library implementing the algorithm can be found at `https://github.com/dedupeio/dedupe.git`.

have their gender inferred based on first name. I drop those observations from the analysis.

Table 3.1: Coverage

| N | Sample |
|---|---|
| 583 | Citation Info missing |
| 34810 | Gender Info missing |
| 59221 | Matched Sample |

In order to provide an overview of the data, I provide basic statistics in terms of pay, citations and share of female professors by UC campus and major field of study. Table 3.2 shows that the share of female professors varies between 30 % at UC Santa Barbara and 44 % at UC San Francisco. The average pay across campuses varies substantially from around 130.000 US$ at UC Riverside to 235.000 US$ at UC Los Angeles. There is similar variation in citation counts per professor across these institutions. One outlier seems to be UC Merced, however the campus was only established in 2005, which explains why it is still small compared to the other campuses.

Table 3.2: University of California Campuses

| UC Campus | Female Share | Avg. Pay in US-$ | Avg. Citation Count | Observations |
|---|---|---|---|---|
| Santa Barbara | 0.30 | 159158 | 1592 | 2775 |
| Berkeley | 0.31 | 160817 | 1433 | 6033 |
| Riverside | 0.31 | 126194 | 637 | 2584 |
| San Diego | 0.32 | 211688 | 1790 | 7666 |
| Los Angeles | 0.33 | 235937 | 1390 | 12681 |
| Irvine | 0.33 | 192639 | 1180 | 5691 |
| Davis | 0.36 | 182952 | 859 | 9725 |
| Santa Cruz | 0.37 | 133960 | 1258 | 1863 |
| Merced | 0.39 | 119898 | 284 | 441 |
| San Francisco | 0.44 | 225914 | 1888 | 9762 |

Table 3.3 show that across the major fields of study there is large variation in the share of female professors. Traditionally male-dominated fields like Physics and Engineering have a share of females professors at only 18 %, while at the other end of the distribution Sociology actually has a female share of 54 %. This shows that on average professor positions are male dominated across almost all fields and campuses.

Table 3.3: Major Fields of Study

| Field | Female Share | Avg. Pay in US-$ | Avg. Citation Count | Observations |
|---|---|---|---|---|
| Engineering | 0.18 | 169427 | 849 | 1047 |
| Physics | 0.18 | 177855 | 1809 | 3325 |
| Mathematics | 0.21 | 151157 | 898 | 2512 |
| Computer science | 0.23 | 167107 | 2297 | 2701 |
| Chemistry | 0.25 | 172801 | 1521 | 2516 |
| Economics | 0.29 | 225218 | 1230 | 1740 |
| Environmental science | 0.31 | 101873 | 57 | 36 |
| Biology | 0.32 | 191004 | 2459 | 13517 |
| Materials science | 0.32 | 172450 | 396 | 531 |
| Geology | 0.32 | 154614 | 1404 | 739 |
| Business | 0.33 | 207575 | 398 | 553 |
| History | 0.33 | 135792 | 17 | 114 |
| Medicine | 0.39 | 254111 | 1117 | 18232 |
| Philosophy | 0.42 | 135227 | 169 | 1370 |
| Political science | 0.44 | 160095 | 325 | 4073 |
| Geography | 0.47 | 127352 | 81 | 519 |
| Psychology | 0.49 | 165233 | 1261 | 3032 |
| Art | 0.51 | 130093 | 13 | 1247 |
| Sociology | 0.54 | 132951 | 137 | 1417 |

Furthermore, it seems to be the case that fields with high male share also have relatively high pay. However, for example medicine has a high share of females but tops the income distribution. Across campuses the picture is even less clear. To what extent those differences can also explain differences in pay across men and women will be explored in the next section.

## 3.3 Gender Gap in Annual Earnings

In this section I describe the gender gap in pay among professors at the University of California. First I present the raw distribution of total pay and citations by gender.

Figure 3.1: Distribution of Annual Pay in US-$

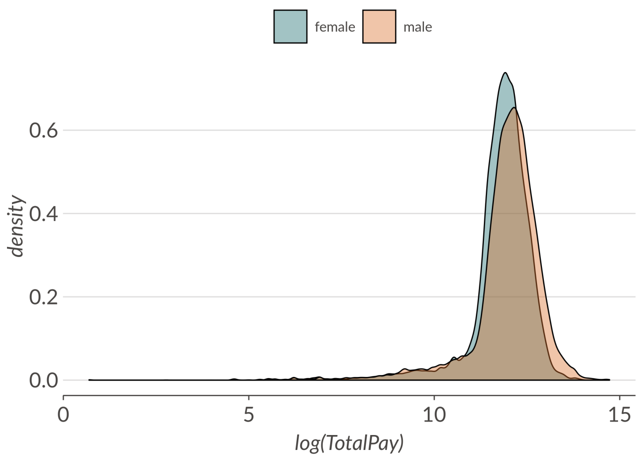



Figure 3.1 shows the raw distribution of the log of annual pay by gender. There is substantial variation in pay within both males and females, but the distribution of earnings of men first-order stochastically dominates that of women. Thus, there is a gender gap in earnings: men on average out-earn women.

I will now summarize the average gender gap and show whether differences in sorting to fields, institutions and/or research output can explain gender gaps in pay. Table 3.4 summarizes the gender gap in annual earnings by regressions of the log of *earnings* on a set of controls. Column (1) shows the raw difference in earnings between men and women: female professors at UC campuses earn on average about 16% less then their male colleagues. Interestingly, the difference can not be explained by sorting of women to low paying fields or low paying campuses. Columns (2) and (3) control for the major field of study of the professor and the UC campus at which they are employed, but the gender gap in earnings can not be explained by those factors. In column (4) I control for the type of position the professor holds. This can explain about half of the gap in earnings between men and women. That is, at a similar position women on average earn about 8% less than their male counterparts. The difference in earnings can not be explained by differences in the number of citations as the estimated gender gap conditional on citations is still over 7%. That is even after controlling for the type of position, field, institution and a measure of research productivity a substantial gender gap in earnings remains.

Table 3.4: Gender gap in annual pay among UC professors

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | log(TotalPay) | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| female | $-0.160^{***}$ | $-0.160^{***}$ | $-0.157^{***}$ | $-0.080^{***}$ | $-0.072^{***}$ |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.008) |
| log(1 + Cite) | | | | | $0.028^{***}$ |
| | | | | | (0.002) |
| Cite == 0 | | | | | $0.040^{***}$ |
| | | | | | (0.013) |
| Fixed effects: Field | No | Yes | Yes | Yes | Yes |
| Fixed effects: Institution | No | No | Yes | Yes | Yes |
| Fixed effects: Position | No | No | No | Yes | Yes |
| Observations | 59,221 | 59,221 | 59,221 | 59,221 | 59,221 |
| $R^2$ | 0.007 | 0.033 | 0.044 | 0.175 | 0.179 |
| *Note:* | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ | | | | |

The analysis in table 3.4 established that the gender gap in pay is substantial and can not be explained by (1) field of research, (2) differences in pay across campuses and (3) research output as measured by citations. Conditional on the position a professor holds, the pay gap on average is about half of the overall gap. In other words, about half of the pay gap between male and female professors can be attributed to the different position men and women hold.

### 3.3.1 Differences across Campuses and Fields

In order to investigate, whether the average effects previously established are common across campuses and fields I will now estimate the gender gap in pay by field and institution of employment. I repeat the previous analysis separately by UC campus and field. In interest of brevity I only present estimated gender gaps, raw and after controlling for position, citations and field/institution. This corresponds to columns (1) and (5) in table 3.4.

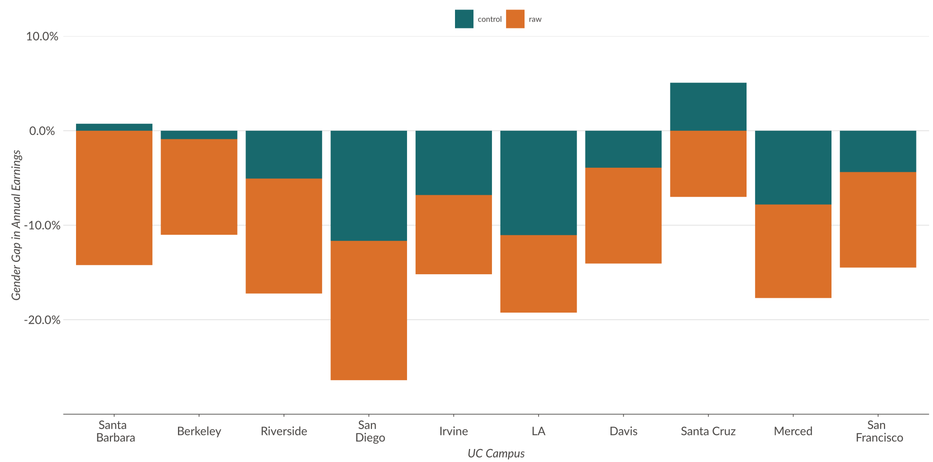



Figure 3.2: Gender gap in earnings by UC campus. Controls: Position, citations and field of study. Ordered by female share.

Figure 3.2 shows the difference in earnings among men and women by UC campus. In orange the raw gap is shown, the green bar shows the gap in earnings after controlling for position, citations and field of study. The raw gender earnings gap is substantial and negative, male professors out-earn their female counterparts consistently across all UC campuses, which are ordered by the share of female professors. However, the earnings gap after controls reaches approximately zero for Berkeley and Santa Barbara, while it turns positive for the Santa Cruz campus. The largest gap in earnings are to be found at the LA and San Diego campuses, both before and after controls are applied.

Across fields the picture is somewhat unexpected, with traditionally male dominated fields like Physics, Math and Engineering showing small differences in earnings between men and women, both before and after applying controls. An extreme case is "Material Sciences", where the gap before controls reaches over 40%, however after controls "Material Sciences" looks quite similar in terms of gender gap to Business and Medicine. Another field with a large difference in gender gaps in earnings before and after controls is Economics. However, it remains that in most cases even after controlling for productivity and position women earn substantially less then men at UC campuses. Relating to the findings of Sheltzer and Smith (2014), that male faculty tend to hire relatively more men, a similar bias seems not to be present in pay setting as there is no relationship between the gender gap in earnings and the share of female professors at a campus or in a field. However, that may simply be the case because the relevant measure would be the interaction between field and campus. I address this in table 3.5. There is no apparent relationship between the share of females in a field of study at a specific campus and the estimated residual pay gap at the same unit of observation.

Table 3.5: Residualized gender gap in earnings and share of females in Field X Campus

| | *Dependent variable:* |
|---|---|
| | 'Residual Gender gap' |
| 'Female share' | −0.163 |
| | (0.126) |
| Constant | 0.011 |
| | (0.047) |
| Fixed effects: Position | Yes |
| Control Citations: | Yes |
| Observations | 173 |
| $R^2$ | 0.010 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |
| | Weighted by square root of observations in FieldXCampus. |

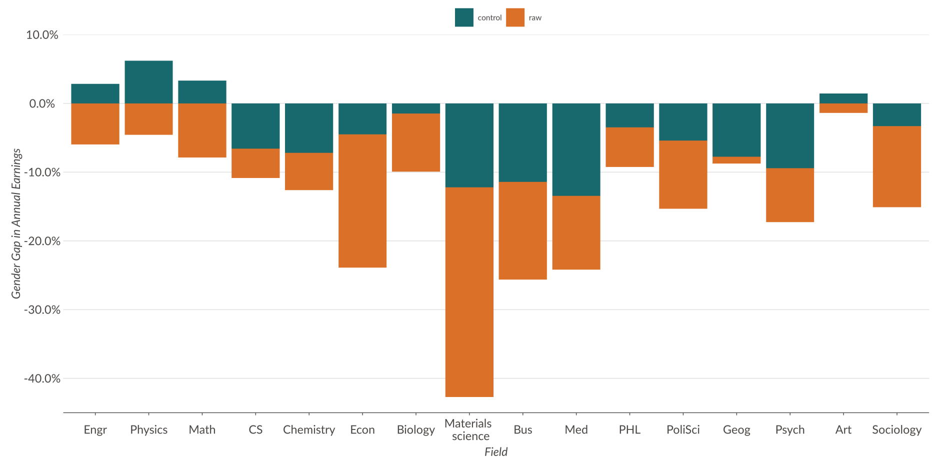



Figure 3.3: Gender gap in earnings by field of study. Controls: Position, citations and UC campus. Ordered by female share.

### 3.3.2 Returns to research productivity

As the overall gap can not be explained by differences in productivity, I will now analyze whether the returns to productivity are different for men and women. In table 3.6 I show the estimated returns to citations. Column (1) shows the average estimate, while column (2) allows for differential returns to citations for men and women. The point estimate for the difference in returns between men and women is close to zero and relatively precisely estimated. This suggests that the gain in pay from research productivity is similar for men and women at UC campuses.

## 3.4 Conclusion

Women earn substantially less than men as professors at the University of California. This holds even after controlling for field, position and research productivity. On the positive side, gender gaps shrink substantially by taking into account productivity and position (by ca. $50\%$). However, this also means there is likely room for improvement in pay setting to make academic jobs more attractive for women and which in turn could help mitigate under-representation. At the same time, the estimated returns to citations are similar for men and women. Which suggests that research output is treated similarly for men and women in pay setting. However, this may be the case because we consider an easily observable measure of productivity. Other factors, that are more subjective, likely have more room for differential treatment.

Table 3.6: Gender gap in returns to citations among UC professors

| | *Dependent variable:* | |
|---|---|---|
| | log(TotalPay) | |
| | (1) | (2) |
| female | $-0.072^{***}$ | $-0.069^{***}$ |
| | (0.008) | (0.020) |
| log(1 + Cite) | $0.028^{***}$ | $0.029^{***}$ |
| | (0.002) | (0.002) |
| Cite == 0 | $0.040^{***}$ | 0.024 |
| | (0.013) | (0.017) |
| femaleTRUE:log(1 + Cite) | | $-0.002$ |
| | | (0.003) |
| femaleTRUE:Cite == 0 | | 0.034 |
| | | (0.027) |
| Fixed effects: Field | Yes | Yes |
| Fixed effects: Institution | Yes | Yes |
| Fixed effects: Position | Yes | Yes |
| Observations | 59,221 | 59,221 |
| $R^2$ | 0.179 | 0.180 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

# Bibliography


Abel, J. and Deitz, R. (2014). Do the benefits of college still outweigh the costs?

Abel, J. R., Florida, R., and Gabe, T. M. (2018). Can low-wage workers find better jobs?

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Abowd, J. M., Kramarz, F., Pérez-Duarte, S., and Schmutte, I. M. (2014). Sorting between and within industries: A testable model of assortative matching. Technical report, National Bureau of Economic Research.

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

Acemoglu, D. and Restrepo, P. (2017). Robots and jobs: Evidence from us labor markets. *NBER working paper*, (w23285).

Aum, S. (2017). The rise of software and skill demand reversal. Technical report, Technical report.

Aum, S., Lee, S. Y. T., and Shin, Y. (2018). Computerizing industries and routinizing jobs: Explaining trends in aggregate productivity. *Journal of Monetary Economics*.

Autor, D. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *The American Economic Review*, 103(5):1553–1597.

Autor, D. et al. (2019). *Work of the past, work of the future*. National Bureau of Economic Research.

Autor, D. H., Dorn, D., and Hanson, G. H. (2016). The china shock: Learning from labor-market adjustment to large changes in trade. *Annual Review of Economics*, 8:205–240.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333.

Azmat, G. and Ferrer, R. (2017). Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, 125(5):1306–1355.

Bagger, J. and Lentz, R. (2014). An empirical model of wage dispersion with sorting. *The Review of Economic Studies*.

Bandiera, O., Rana, S. A., and Xu, G. (2016). The gender and ethnicity earnings gap at lse. Technical report.

Barnichon, R. (2010). Building a composite help-wanted index. *Economics Letters*, 109(3):175–178.

Barnichon, R. and Zylberberg, Y. (2018). Under-employment and the trickle-down of unemployment. *American Economic Journal: Macroeconomics*.

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

Beaudry, P., Green, D. A., and Sand, B. M. (2016). The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, 34(S1):S199–S247.

Bilenko, M. Y. (2006). *Learnable similarity functions and their application to record linkage and clustering*. PhD thesis.

Blinder, A. S. and Krueger, A. B. (2013). Alternative measures of offshorability: a survey approach. *Journal of Labor Economics*, 31(S1):S97–S128.

Bloom, N., Garicano, L., Sadun, R., and Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60(12):2859–2885.

Burdett, K., Shi, S., and Wright, R. (2001). Pricing and matching with frictions. *Journal of Political Economy*, 109(5):1060–1085.

Byrne, D., Oliner, S., and Sichel, D. (2017). Prices of high-tech products, mismeasurement, and pace of innovation.

Byrne, D. M. and Corrado, C. A. (2017). Ict prices and ict services: What do they tell us about productivity and technology?

Cairo, I. (2013). The slowdown in business employment dynamics: The role of changing skill demands. *Unpublished manuscript, Universitat Pompeu Fabra*.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3):967–1015.

Ceci, S. J., Ginther, D. K., Kahn, S., and Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 15(3):75–141.

Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

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

Cortes, G. M., Jaimovich, N., and Siu, H. E. (2016). Disappearing routine jobs: Who, how, and why?

Cortes, G. M., Jaimovich, N., and Siu, H. E. (2017). Disappearing routine jobs: Who, how, and why? *Journal of Monetary Economics*, 91:69–87.

Costinot, A. and Vogel, J. (2010). Matching and inequality in the world economy. *Journal of Political Economy*, 118(4):747–786.

Davis, M. A. and Ortalo-Magné, F. (2011). Household expenditures, wages, rents. *Review of Economic Dynamics*, 14(2):248–261.

Davis, S. J. and Haltiwanger, J. (2014). Labor market fluidity and economic performance.

Doms, M. and Lewis, E. (2006). Labor supply and personal computer adoption. Technical report, Federal Reserve Bank of Philadelphia.

Eden, M. and Gaggl, P. (2018). On the welfare implications of automation. *Review of Economic Dynamics*, 29:15–43.

Eeckhout, J., Pinheiro, R., and Schmidheiny, K. (2014). Spatial Sorting. *Journal of Political Economy*, 122(3):554–620.

Eeckhout, J. and Weng, X. (2018). The technological origins of the decline in labor market dynamism.

Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of political economy*, 105(5):889–927.

Engbom, N. (2017). Firm and worker dynamics in an aging labor market. Technical report.

Ferris, M. C. and Munson, T. S. (1999). Interfaces to path 3.0: Design, implementation and usage. *Computational Optimization and Applications*, 12(1-3):207–227.

Fujita, S. (2015). Declining labor turnover and turbulence. (15-29).

Garibaldi, P., Moen, E. R., and Sommervoll, D. E. (2016). Competitive on-the-job search. *Review of Economic Dynamics*, 19:88–107.

Ginther, D. K. and Kahn, S. (2004). Women in economics: moving up or falling off the academic career ladder? *Journal of Economic perspectives*, 18(3):193–214.

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

Goos, M. and Manning, A. (2003). Mcjobs and macjobs: the growing polarisation of jobs in the uk. In *The labour market under New Labour*, pages 70–85. Springer.

Goos, M., Manning, A., and Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, 104(8):2509–26.

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, 8(S1).

Hall, R. E. and Milgrom, P. R. (2008). The limited influence of unemployment on the wage bargain. *American economic review*, 98(4):1653–74.

Hyatt, H. R. and Spletzer, J. R. (2013). The recent decline in employment dynamics. *IZA Journal of Labor Economics*, 2(1):5.

Jaimovich, N. and Siu, H. E. (2012). The trend is the cycle: Job polarization and jobless recoveries. Technical report, National Bureau of Economic Research.

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

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V., and Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5):1029–1053.

Lise, J. and Robin, J.-M. (2017). The macrodynamics of sorting between workers and firms. *American Economic Review*, 107(4):1104–35.

Lucas, R. E. and Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica*, 70(4):1445–1476.

Mercan, A. Y. (2018). *Essays on Macroeconomics and Labor Markets*. PhD thesis, UC Berkeley.

Moscarini, G. and Postel-Vinay, F. (2016). Did the job ladder fail after the great recession? *Journal of Labor Economics*, 34(S1):S55–S93.

Oyer, P. and Schaefer, S. (2016). Firm/employee matching: An industry study of us lawyers. *ILR Review*, 69(2):378–404.

Peters, M. (1991). Ex ante price offers in matching games non-steady states. *Econometrica: Journal of the Econometric Society*, pages 1425–1454.

Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair womenś careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 125(3):1253–1296.

Sarah, F., King, M., Rodgers, R., Ruggles, S., and Warren, J. R. (2018). Integrated public use microdata series, current population survey: Version 6.0 [dataset].

Sheltzer, J. M. and Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences*, 111(28):10107–10112.

Shi, S. (2002). A directed search model of inequality with heterogeneous skills and skill-biased technology. *The Review of Economic Studies*, 69(2):467–491.

Shimer, R. (2005a). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025.

Shimer, R. (2005b). The cyclicality of hires, separations, and job-to-job transitions. *Review-Federal Reserve Bank of St Louis*, 87(4):493.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-j. P., and Wang, K. (2015). An overview of microsoft academic service (mas) and applications.

In *Proceedings of the 24th international conference on world wide web*, pages 243–246. ACM.

Stokey, N. L. (2016). Technology, skill and the wage structure. Technical report, National Bureau of Economic Research.

ter Braak, C. J. and Vrugt, J. A. (2008). Differential evolution markov chain with snooker updater and fewer chains. *Statistics and Computing*, 18(4):435–446.

Weisshaar, K. (2017). Publish and perish? an assessment of gender gaps in promotion to tenure in academia. *Social Forces*, 96(2):529–560.