Multidimensional heterogeneity and matching in a frictional labor market - An application to polarization

Joanne Tan*

December 13, 2017

Abstract

Wage and job polarization are two defining phenomena that have characterized the US labor market from the 1980s to the late 2000s. Polarization has also differed distinctly between demographic subgroups, such as men and women. In addition, sorting patterns between workers and firms have changed over this period. This paper finds that technological change can account for a substantial part of these phenomena. It does so by constructing a structural model with two key ingredients: 1) directed search and 2) two-sided multidimensional heterogeneity. Estimation results show that the complementarity between cognitive skills and tasks increased relative to that between both interpersonal and manual skills and tasks. Complementarity between manual skills and tasks decreased substantially. This change in production technology induces polarization in the model and also generates improvement in sorting between workers and firms in the cognitive dimension relative to other dimensions, in line with the data. The model also accounts for a large part of the gender differences in polarization. Lastly, in this novel framework of directed search with multidimensional heterogeneity, I develop a definition of assortative matching and examine the conditions under which it is obtained.

Keywords: Polarization, technological change, multidimensional heterogeneity, gender gap

JEL codes: J3, J16, J21, J23, J24, J31

1 Introduction

The impact of technological change on the labor market is a highly pertinent issue. How has technological change impacted wages and employment? Who gains and loses from technological change? This paper assesses the role of technological change in the phenomenon of labor market polarization and the changes in sorting between workers and firms. Indeed, several papers1 have documented the existence of wage and job polarization in the US between the 1980s and the late 2000s. Wage polarization is defined as the wage growth in the bottom and upper tails of the wage distribution relative to the median. Job polarization is, in turn, defined as the growth in employment share at high and low-paying jobs relative to middle-paying jobs

* Author is contactable at joanne.tan@sciencespo.fr and joanne.tan@yale.edu. I thank Jean-Marc Robin, Fabien Postel-Vinay, Ilse Lindenlaub, Zsofia Barany, Florian Oswald and Julia Cage for their helpful comments. I also thank participants at the IZA-CEDEFOP conference on skills and skills mismatch, seminar participants at the Department of Economics at Sciences-Po, as well as participants at the 2017 SOLE, EEA-ESEM, SaM, WEAI and EALE conferences for their feedback. Lastly, I thank the Fox Fellowship at Yale and the Doctoral School of the Institut d’Etudes Politiques for funding provided. Any fault is my own.

1 Refer to Acemoglu and Autor (2011) and Autor et al. (2006a).
While wage and job polarization has been observed for the whole US workforce, differences exist between demographic subgroups, which have been hitherto overlooked in the literature. In this paper, I document the distinct polarization experiences of men and women in the US. Notably, I find that while both men and women undergo wage and job polarization, women have gained across the wage distribution relative to men and have also disproportionately increased their representation in high-ranking jobs. This finding is consistent across datasets, the service and non-service industries, as well as across education groups.

Also, during this time period, sorting between workers and firms evolved. Using educational attainment and educational requirement as the productive characteristic of workers and jobs respectively, highly-educated workers are increasingly working at jobs that require higher education. Apart from education, sorting between other productive characteristics of workers and firms has also changed. In particular, this paper delves into the sorting between the skills of workers and tasks of firms. It finds that sorting increased most between the cognitive skills and tasks of workers and firms, while sorting between interpersonal and manual skills and tasks improved by much less.

The paper jointly accounts for wage and job polarization without counterfactual predictions on sorting. It is also one of the first to address the differences in the polarization experiences between men and women\(^2\). Specifically, it examines the role of technological change in these phenomena.

The paper does so through the lens of a structural model. The model consists of two key ingredients: 1) multidimensional heterogeneity for workers and firms and 2) directed search. In particular, the model extends the directed search model with two-sided unidimensional heterogeneity of Shimer (2005) by introducing multidimensional heterogeneity. Under a plausible regularity condition and an assumption of symmetric strategies, I establish that unique equilibrium exists and is constrained efficient. I then take the model to the data.

In the empirical application, workers and firms are characterized by a vectors of skills and tasks respectively. Specifically, workers are defined by their levels of cognitive, interpersonal and manual skills and jobs by their cognitive, interpersonal and manual tasks. Upon estimating the model on data from the 1979 and 1997 cohorts of the National Longitudinal Survey of Youths (NLSY) and the O*NET, I find that production complementarities between cognitive skill and task increased substantially, followed by that between interpersonal skill and task. Production complementarity decreased the most between manual skill and task. This change improved sorting between workers and firms in the cognitive dimension, compared to other dimensions, in line with the data.

In addition, greater cognitive and interpersonal complementarity implies that jobs with high cognitive and interpersonal tasks are relatively more productive, while jobs consisting of high manual tasks become relatively less so. This decreases the return to filling the latter group of jobs most, since the dimension in which they are most endowed is now less productive. Jobs low in all three dimensions of tasks also experience a decrease in return to being filled, albeit by less. On the

\(^2\)Cerinay et al. (2017) and Ngai and Petrongolo (2017) show that women gained more in employment in high and low-ranked jobs in the service sector, but this paper documents gender differences in wage and job polarization, which hold in both the service and non-service industry.
contrary, jobs with high levels of cognitive and interpersonal tasks are more likely to be filled. Since jobs high in cognitive tasks are also those at the top of the job rank distribution while jobs low in all tasks are those at the bottom, the aforementioned differential changes result in job polarization.

Similarly, the rise in cognitive and interpersonal complementarities also increases the productivity of high cognitive and interpersonal skill workers while decreasing the productivity of workers less endowed with these skills. In particular, workers with low cognitive and interpersonal skills, and high manual skills experience the greatest decrease in productivity. This is because the skills in which they are better endowed are now of lower importance in the production. Since these workers are the middle earners and as the expected wages of workers reflect their productivity, wage polarization is hence obtained.

Therefore, a simple shift in production complementarities in favour of cognitive and interpersonal skills and tasks is able to produce both wage and job polarization, along with an improvement in sorting in the cognitive dimension relative to other dimensions. Directed search is a useful feature here as it creates a tight link between wage and job polarization in the model, thereby allowing wage and job polarization to occur jointly. Furthermore, with directed search, there are two channels through which job polarization can occur. The first is through the disappearance of middle-ranking vacancies and the second is through the fall in filling probabilities for middle-ranking vacancies, even if these are posted. Estimates presented in this paper suggest that the second channel is just as important the first in job polarization and should therefore not be ignored. Multidimensional heterogeneity is also useful, since it is the relative changes in complementarities that lead to polarization while improving sorting unambiguously in some dimensions relative to the others.

On top of being able to produce wage and job polarization as well as the sorting patterns in the data, the model provides some account of the distinct polarization experiences of men and women. These groups have been differently affected by technological change in part due to their disparate distributions of skills. Notably, the model generates the growth in women’s wages relative to men across all wage percentiles, along with the disproportionate increase in the share of women employed in high-ranking jobs. Two mechanisms account for this: i) the change in the skills distribution between women and men and ii) the change in production technology.

Indeed, since women gained in all skill dimensions relative to men and as complementarities in the cognitive dimension increased relative to the rest, their wages increased and they were assigned disproportionately to high cognitive task jobs. This then engenders the observed differences between men and women even as both groups polarize. In fact, despite precluding gender, the model accounts for 46.3 percent of the decrease in the gender wage gap, as well as 64.1 percent of the climb of women up job ranks on average. Certainly, other factors such as heterogeneous preferences and discrimination may also play a role. Nonetheless, the substantial explanatory power of the model suggests that the change in women’s skills and production technology play an important role. The differing impact of technology can also be investigated in a similar way for other demographic subgroups who have experienced distinct changes in skills.

This has been the channel through which existing papers generate job polarization, such as Acemoglu and Autor (2011) and Autor and Dorn (2013).
Apart from generating predictions in line with the data, the model also paves the way for an alternate characterization of assortative matching in the context of multidimensional heterogeneity. In its definition of assortative matching, this paper extends the notion of log-supermodularity of the matching function to the multidimensional case and finds that the sufficient condition for assortative matching is simply a multidimensional adaptation of the supermodularity of the production function. The model therefore lends itself readily to other empirical applications in any context where workers and firms are described by multidimensional characteristics.

In short, the contribution of this paper is threefold - 1) It provides a unified, intuitive and parsimonious explanation of the joint phenomena of wage polarization, job polarization and changes in sorting between workers and firms. 2) It documents and accounts for the different experiences of men and women in this regard. 3) It is the first to introduce multidimensional heterogeneity in the directed search framework, which is readily applicable to other empirical applications.

The rest of the paper proceeds as follows. Section 2 consists of the literature review. Section 3 presents the theoretical model and describes the equilibrium as well as the conditions under which it is obtained. Section 4 looks at comparative statics and the characterization of assortative matching. Section 5 discusses the relationship between the equilibrium assignment, wages and technological change. Section 6 presents the empirical application and estimation of the model. An assessment of how well the model accounts for the empirical phenomena is undertaken in Section 7. Section 8 conducts several counterfactual exercises and the last section concludes.

2 Literature Review

This paper is closely related to three strands of literature on: 1) wage and job polarization, 2) multidimensional heterogeneity and matching and 3) the impact of technological change on the gender gap.

The literature on polarization is well-known. Acemoglu and Autor (2011) and Autor et al. (2006a) document these phenomena in the US. Acemoglu and Autor (2011) also construct a stylized model to account for wage polarization. In their model, workers are categorized as high, medium or low-skill and jobs as high, medium or low task. The authors show, that when the return to high-skill labor in production increases, medium-skill workers’ wages decrease relative to that of high and low-skill workers. They also illustrate how job polarization can be generated by technological change that allows medium tasks to be performed by machines, thereby driving employment share in these tasks to zero.

Autor and Dorn (2013) also attempt to account for polarization, focusing in particular on the role of Routine-Biased Technical Change (RBTC), represented by the decline in the price of routine task-performing machines. They find that RBTC can account for the rise in low-skill employment in the service sector and hence, part of job polarization. Specifically, a fall in the price of machines induces lower wages for low-skill workers in the goods sector. Prices also decrease in the goods sector. Since goods and services are complements, wages rise for low-skill workers in the service sector,
prompting an inflow of low-skill workers. This then generates an increase in low-skill employment and wages in the service sector. Likewise, Goos et al. (2014) show how RBTC has been a key factor in job polarization both within and between industries.

This paper also seeks to account for wage and job polarization. While complementary to the aforementioned papers, this paper differs in the following two ways. In the aforementioned papers, job polarization in these papers are generated when middle-ranking jobs are replaced by machines, i.e. when the distribution of vacancies are hollowed out in the middle. On the contrary, the model presented in this paper is able to distinguish between two channels of job polarization: i) a disappearance of middle-ranking vacancies and ii) a decrease in the probability of these vacancies being filled, even if they are posted. This paper’s estimates suggest that the latter channel is also crucial. Secondly, this paper is able to make predictions on sorting in the various dimensions of heterogeneity, in line with the data⁴.

Yamaguchi (2012) and Lazear (2009) are among the first to characterize workers and firms by bundles of skills and tasks respectively. The former paper uses the Dictionary of Occupational Titles (DOT) to describe occupations by bundles of tasks, which then facilitates the estimation of returns to skills across occupations using the 1979 cohort of the National Longitudinal Survey of Youths (NLSY). In a similar vein, the latter paper presents a hitherto novel approach to firm-specific human capital, by characterizing human capital as bundles of general skills, with each skill element weighted differently across occupations. Likewise, this paper considers that workers and firms can be described by bundles of general skills and tasks.

Prior to the application of multidimensional matching to the labor market, the framework was first used in the study of the marriage market. For instance, Dupuy and Galichon (2014) examine matching in the marriage market when men and women are characterized by multiple traits. The issue of identifying match surplus when men and women are characterized by multidimensional traits is then studied in Galichon and Salanie (2014).

Lindenlaub (2017) presents the earliest application of multidimensional matching to the labor market. She considers the frictionless assignment of workers to firms when both are characterized by multidimensional traits and examines the conditions needed on the production function for Positive Assortative Matching (PAM) to occur. She then estimates the production function and examines the extent to which her model accounts for wage polarization in the data. Her model is indeed able to generate wage polarization, and does so with predictions on sorting that are in line with the data.

My paper is closest to Lindenlaub (2017) in its aims. Nonetheless, two stark differences exist between the two. Firstly, the model in this paper is a model of directed search with multidimensional heterogeneity and hence includes search frictions. Since search frictions exist and types are discrete, the characterization of assortative matching and the conditions under which it occurs differs from that in Lindenlaub (2017). Secondly, having included search frictions, the model presented in this

⁴In Acemoglu and Autor (2011) and Autor and Dorn (2013), middle-earning workers are displaced to low-skill jobs by machines, thereby implying less positive sorting between workers and firms. This does not seem to be the case in the data.
paper is able to account for both wage and job polarization, which are joint phenomena in the US. With a frictionless setup, not much can be said on job polarization unless the distribution of job postings were altered ex-ante to reproduce the phenomenon.

Two recent papers also deal with the issue of multidimensional matching, with very different setups and aims. In Lise and Postel-Vinay (2015), a model of random job search with multidimensional heterogeneity is constructed. Workers are allowed to perform on-the-job search and their skills are allowed to decay or accumulate according to the job they obtain. They aim to estimate the rate of appreciation and depreciation of workers’ skills over time and find that, for the NLSY 1979 cohort, skills depreciate and appreciate more rapidly in the manual dimension than in the interpersonal or cognitive dimensions. The next paper, Lindenlaub and Postel-Vinay (2016) again consider a model of random search with multidimensional heterogeneity. They categorize assortative matching in this framework and examine the extent to which welfare would be wrongly estimated if one were to assume unidimensional heterogeneity when skills are in fact multidimensional. While these two papers do include search frictions in the framework, I later argue that a directed search framework, by creating a tight link between sorting, wages and job filling probabilities is best-suited for the given empirical application.

The last strand of literature relates to how the gender gap has decreased due to i) changes in technology that have increased the return to skills in which women are better endowed and ii) the rise of the service sector that has seen a disproportionate participation of women. Examples of the first would be Black and Spitz-Oener (2011) and Cortes et al. (2017). Both papers argue, albeit differently, that technological change has increased the returns to women’s skills, thereby inducing a greater share of women in employment and narrowing the gender wage gap. Regarding the second sub-strand of literature, Ngai and Petrongolo (2017) argue that due to the marketization of home production as well as the burgeoning professional services sector, skilled women have been induced to enter the service sector en masse. This, they argue, has led to the narrowing of the gender wage and employment gap. These papers would then predict that women’s wages and employment share would be rising monotonically up the job and wage ranks. Yet, there is substantial empirical evidence, as will be later shown in this paper, that both men and women undergo polarization. While the model in this paper does not allow gender-specific productivity, it is able to generate the aforementioned differences between men and women based purely on the shifting distribution of skills and production complementarities. A small handful of papers, including Cerinay et al. (2017) and Ngai and Petrongolo (2017), document the rise in employment of women in high and low paying jobs in the service sector relative to men. This paper, however, records not only gender differences in employment polarization in the service sector, but also in the non-service sector. It also highlights the gender differences in wage polarization, which has been hitherto unaddressed.

The interested reader may refer to Rogerson et al. (2005) for an overview of labor search models.
3 The Model

3.1 Setup - workers and firms

There is a large number of risk-neutral first-time workers. At the start of the period, they enter the workforce. Workers' characteristics are described by $X = [x_1 \ x_2 \ \ldots \ x_K]^T$. I define these characteristics broadly as skills, where a skill is loosely defined as any worker attribute that enables the worker to perform his/her job. Let $\vartheta(X_i)$ be the number of type $X_i$ worker. There are $I$ discrete types of workers.

Similarly, jobs are defined by their job requirements $Y = [y_1 \ y_2 \ \ldots \ y_K]^T$. Let $\gamma(Y_j)$ be the measure of jobs with characteristics $Y_j$ and let $J$ be the number of different job types. The number of jobs in total is large. Since a single firm can only host one job, I use the terms firm and job interchangeably. For now, I assume that the distributions of jobs and workers are exogenously given. It should be noted that the dimensionality of heterogeneity for workers ($K_x$) and firms ($K_y$) need not be equal.

3.2 Job search

As in the standard directed search model, a firm hosting job type $Y_j$ posts a wage $w(X_i, Y_j)$ for $i = 1, \ldots, I$. Let $p(Y_j|X_i)$ be the probability that a worker applies to type $Y_j$ firm given that she has $X_i$ characteristics. In a given period, the worker can only make a single application to a firm. The expected number of $X_i$ type workers that a given firm of type $Y_j$ expects to receive is therefore $q(X_i, Y_j) = \frac{P(Y_j|X_i)\vartheta(X_i)}{\gamma(Y_j)}$.

3.3 Production

A one-to-one match between a worker of type $X_i$ and a firm of type $Y_j$ results in a the production of a quantity of the final good given by $\Phi(X_i, Y_j)$. Unmatched workers and firms produce nothing. The functional form of $\Phi(X_i, Y_j)$ is left unspecified for now.

3.4 Timeline

While the model is static, it nevertheless involves a sequence of decisions by agents as illustrated in Figure 1. Initially, all workers and firm types are observed. Each firm type $Y$ then posts a menu of wages, $w(X, Y)$, for every worker type in the economy. Upon the posting of wages, workers decide on their application strategies, given by $p(Y|X)$. A worker can make only a single application in this game. Since workers have no means of coordinating their applications, firms may receive 0 or $\geq 1$ applications. Firms who receive more than one application hire their most preferred worker, and produce $\Phi(X, Y)$. Unfilled firms and unemployed workers produce and earn nothing.

3.5 Whom do firms prefer?

There are two notions of ranking that should be distinguished. In the unidimensional context, with standard production technological specification, firms share a common ranking of preferences over workers, with all firms preferring workers with higher values of human capital characteristic.
On the contrary, when workers and firms are characterized by multidimensional heterogeneity, firms would not, in general, share the same preferences over workers\(^6\). Unless strong restrictions on the distribution of worker and firm types as well as the production function are imposed, a common ranking over workers by firms does not exist.

Nevertheless, one can show that, just as in the unidimensional case, firms prefer workers with whom they more productive\(^7\). As such, firm \(Y_j\) prefers worker \(X_i\) to \(X_i'\) only if \(\Phi(X_i, Y_j) > \Phi(X_i', Y_j)\). In addition, assume that there is no \(X_i \neq X_i'\) such that \(\Phi(X_i, Y_j) = \Phi(X_i', Y_j)\) at any firm \(Y_j\)\(^8\), such that each firm has a strict preference ranking over workers. This restriction will prove useful later when characterising the equilibrium. Next, following the notation of Shimer (2005), let \(Q(X_i, Y_j)\) be the expected number of workers preferred to worker \(X_i\) applying to firm \(Y_j\).

\[
Q(X_i, Y_j) = \sum_{i'} \mathbb{1}_{\{\Phi(X_i', Y_j) > \Phi(X_i, Y_j)\}} q(X_i', Y_j)
\]  

(1)

### 3.6 The worker’s problem

The probability that a worker of type \(X_i\) gets the job \(Y_j\) given he/she applies can be written as \(e^{-Q(X_i, Y_j)}(1-e^{-q(X_i, Y_j)})/q(X_i, Y_j)\). This is essentially the probability that no one preferred to \(X_i\) applies and that this individual \(X_i\) gets the job even if other individuals of the same type apply\(^9\). As such, the expected payoff of a worker from applying to firm \(Y_j\) is simply \(e^{-Q(X_i, Y_j)}(1-e^{-q(X_i, Y_j)})\) \(w(X_i, Y_j)\). As is standard in the directed search literature, the worker only applies with positive probability to a job \(Y_j\) if it offers her highest expected utility. Letting the maximum expected utility of worker \(X_i\) be \(U(X_i)\), \(p(Y_j|X_i) > 0\) only if

\[
U(X_i) = e^{-Q(X_i, Y_j)}(1-e^{-q(X_i, Y_j)}) w(X_i, Y_j)
\]  

(2)

### 3.7 The firm’s problem

The probability that a firm of type \(Y_j\) hires worker type \(X_i\) is \((1-e^{-q(X_i, Y_j)})e^{-Q(X_i, Y_j)}\). Hence, the expected profit of the firm can be written as

\[
\pi(Y_j) = \sum_{i=1}^I (1-e^{-q(X_i, Y_j)})e^{-Q(X_i, Y_j)} \{\Phi(X_i, Y_i) - w(X_i, Y_j)\}
\]  

(3)

The firm maximizes this with respect to \(w(X_i, Y_j)\) subject to equation (2) for the \(X_i\) types that it wishes to hire. Effectively, the firm chooses the application strategy \(p(Y_j|X_i)\) (and thus the expected

---

\(^6\)For example, how would a firm rank worker \(X_i = (0.7, 0.5)\) against worker \(X_i' = (0.9, 0.5)\) ?

\(^7\)Firms, in fact, prefer workers with whom they reap higher profits \(\Phi(X, Y)\) to \(w(X, Y)\). In equilibrium, however, posted wages are such that a more productive match is always a more profitable match. Firms therefore always prefer workers with whom they are more productive.

\(^8\)This assumption is not so implausible seeing as the weights on each skill in the production function, hitherto unestimated, could take on any value, and it would take no small amount of coincidence for workers \(X_i \neq X_i'\) to have different skills exactly offsetting each other such that \(\Phi(X_i, Y_j) = \Phi(X_i', Y_j)\).

\(^9\)How this expression is derived is presented in the Appendix.
queue length \(q(X_i, Y_j)\) of each worker. The firm’s problem thus rewrites as

\[
\pi(Y_j) = \sum_{i=1}^{I} (1 - e^{-q(X_i, Y_j)}) e^{-Q(X_i, Y_j)} \Phi(X_i, Y_j) - U(X_i)q(X_i, Y_j)
\]

### 3.8 Existence of the decentralized equilibrium

From Weierstrass’s theorem, an equilibrium exists if i) \(\pi(Y_j)\) is continuous in \(q(X_i, Y_j)\) and ii) \(q(X_i, Y_j)\) lies in a compact set. To see that the former condition holds, say that, without loss of generality, the firm \(Y_j\) assigns an index in ascending order of preference to each worker type such that \(X_1\) refers to \(Y_j\)'s least preferred worker type, \(X_I\), refers to \(Y_j\)'s most preferred worker and \(X_i\) refers to \(Y_j\)'s \(i\)th preferred worker. Hence,

\[
\pi(Y_j) = \sum_{i=1}^{I} (1 - e^{-q(X_i, Y_j)}) e^{-Q(X_i, Y_j)} \Phi(X_i, Y_j) - U(X_i)q(X_i, Y_j)
\]

such that \(Q(X_i, Y_j) = \sum_{i=1}^{I} q(X_i, Y_j)\). It is now evident that \(\pi(Y_j)\) is indeed continuous in \(q(X_i, Y_j)\). Concerning the latter condition, note firstly that \(q(X_i, Y_j) \geq 0\), since expected queue lengths cannot be negative. Secondly, note that the firm would never choose a \(q(X_i, Y_j)\) such that its expected profit from hiring a worker is negative. This implies that \(q(X_i, Y_j) \leq \frac{\Phi(X_i, Y_j)}{U(X_i)}\). Since \(q(X_i, Y_j) \in [0, \frac{1}{U(X_i)}]\) and \(\pi(Y_j)\) is continuous in \(q(X_i, Y_j)\), the decentralized equilibrium exists.

### 3.9 First order condition

The first order condition gives

\[
U(X_i) \geq e^{-q(X_i, Y_j)} e^{-Q(X_i, Y_j)} \Phi(X_i, Y_j) - \sum_{i'=1}^{I} (1 - e^{-q(X_i', Y_j)}) e^{-Q(X_i', Y_j)} \Phi(X_i', Y_j), \quad q(X_i, Y_j) \geq 0
\]

(4)

With complementary slackness. As such, unless the expected payoff offered by firm \(Y_j\) is equal to the highest that \(X_i\) can receive, no worker of type \(X_i\) applies to \(Y_j\). Substituting this into equation (2), the \(w(X_i, Y_j)\), when \(q(X_i, Y_j) \geq 0\), can be expressed as follows:

\[
w(X_i, Y_j) = \frac{q(X_i, Y_j)}{1 - e^{-q(X_i, Y_j)}} \left\{e^{-q(X_i, Y_j)} \Phi(X_i, Y_j) - \sum_{i'=1}^{I} (1 - e^{-q(X_i', Y_j)}) e^{-Q(X_i', Y_j)} \Phi(X_i', Y_j)\right\}
\]

(5)

\(\frac{e^{-Q(X_i', Y_j)}}{e^{-q(X_i', Y_j)}}\) is essentially the probability that no one better than \(X_i\) applies given that no one better than \(X_i\) applies, and the expression for \(w(X_i, Y_j)\) can thus be interpreted as being equivalent to the expected marginal output of \(X_i\) at job \(Y_j\). Unless there are strong restrictions placed on the distribution of worker types and the form of the production function, the inability to coordinate among themselves compels workers to adopt mixed job application strategies.

### 3.10 Summary of the decentralized equilibrium

In the equilibrium, the aggregate demand of each type \(X_i\), for \(i = 1, \ldots, I\) worker must equal the aggregate supply. The ‘demand’ of a worker type \(X_i\) by a given firm of type \(Y_j\) is given by the
queue lengths of each worker type \( q(X_i, Y_j) \) chosen by this firm. The equality between aggregate demand and supply can therefore be expressed as

\[
\sum_j \gamma(Y_j)q(X_i, Y_j) = \vartheta(X_i)
\]  

(6)

**Definition 1.** The decentralized equilibrium is characterized by workers’ application strategies \( p(Y_j|X_i) \), their queue lengths \( q(X_i, Y_j) \) and equilibrium wages \( w(X_i, Y_j) \), for \( i = 1, \ldots, I \) and \( j = 1, \ldots, J \) that satisfy equations 4, 5 and 6.

**Theorem 1.** The decentralized equilibrium exists and is unique. It is also constrained efficient.

The proof of Theorem 1 is in the Appendix. Note that uniqueness of the equilibrium here refers to the uniqueness of the equilibrium job application strategies, not the equilibrium assignment of workers to jobs. Indeed, if workers are allowed to adopt mixed strategies, the same application strategy profile can lead to several different equilibrium assignments. Hence, it is imperative to note that while workers’ equilibrium application strategies are unique, this may still produce different observable matches.

### 3.11 The constrained Social Planner’s problem

As the Social Planner’s perspective will prove useful in what follows, a quick exposition of the constrained Social Planner’s problem is given here. The Social Planner is constrained in the sense that she cannot distinguish between workers of the same type and hence can only issue the following order to all workers, “ If you are type \( X_i \), adopt application strategy \( p(Y_j|X_i) \)” for \( i = 1, \ldots, I \) and \( j = 1, \ldots, J \). Like the firm, the Social Planner effectively chooses the expected queue lengths \( q(X_i, Y_j) \) for each worker type at each firm. This is done to maximise total expected output,

\[
W = \sum_{j=1}^{J} \gamma(Y_j) \sum_i (1 - e^{-q(X_i, Y_j)}) e^{-Q(X_i, Y_j)} \Phi(X_i, Y_j)
\]

Which is simply the expected output of each firm \( Y_j \) with type \( X_i \), multiplied by the number of \( Y_j \) firms, summed over all worker and firm types in the economy. As in the decentralized case, the social planner tells the firm to prefer \( X_i \) to \( X_{i'} \) if \( \Phi(X_i, Y_j) > \Phi(X_{i'}, Y_j) \) for \( i = 1, \ldots, I \). Output is maximised subject to \( q(X_i, Y_j) \geq 0 \) as well as the resource constraint, which states that the expected number of workers of type \( X_i \) queuing at all firms equals to the total expected number of workers of type \( X_i \),

\[
\sum_{j=1}^{J} \gamma(Y_j)q(X_i, Y_j) = \vartheta(X_i) \quad \leftrightarrow \quad \sum_{j=1}^{J} p(Y_j|X_i) = 1
\]

By making use of the aforementioned definition \( q(X_i, Y_j) = \frac{p(Y_j|X_i)\vartheta(X_i)}{\gamma(Y_j)} \). Letting \( U(X_i)\vartheta(X_i) \) be the multiplier on this constraint, the first order condition gives

\[
U(X_i) \geq e^{-q(X_i,Y_j)}e^{-Q(X_i,Y_j)}\Phi(X_i,Y_j) - \sum_{i'} \mathbb{I}_{\{\Phi(X_i,Y_j) > \Phi(X_{i'},Y_j)\}} \left( 1 - e^{-q(X_{i'},Y_j)} \right) e^{-Q(X_{i'},Y_j)} \Phi(X_{i'},Y_j),
\]

\[
q(X_i,Y_j) \geq 0
\]
With complementary slackness, which gives the same assignment as in the decentralized case, where the shadow value on worker types is equivalent to their expected payoff.

Having studied the model and properties of its equilibrium, I now delve deeper into the theoretical model and consider how assortative matching can be characterized.

4 Comparative statics and Assortative matching

How does one define assortative matching in this context and under what conditions does it arise? Even in the unidimensional case, Shimer (2005) leaves this question unanswered. The strongest notion of Positive Assortative Matching (PAM) that can be considered is perfect PAM, under which worker and firm types have a correlation coefficient of 1. This essentially implies that a higher worker type is always matched to a higher firm type in equilibrium. Under a frictionless model with unidimensional heterogeneity, perfect PAM is obtained when the production function is supermodular. When search frictions are added, however, obtaining perfect PAM is slightly less straightforward.

In their directed search model with continuous types and non-type contingent wages, Eeckhout and Kircher (2010) show that root-n supermodularity is a sufficient condition for perfect Positive Assortative Matching (PAM). This condition is insufficient to generate perfect PAM in the unidimensional case with discrete types. In fact, even with an extreme complementarity, as in the Leontieff production function, perfect PAM does not arise in the unidimensional directed search case with discrete types. This is due to the fact that workers of the same type cannot coordinate their application strategies.

Additional complexity in characterizing PAM is created by multidimensional heterogeneity. In the frictionless multidimensional case with continuous types, Lindenlaub (2017) shows that if the cross-derivatives of the production function is a P-matrix, then the jacobian of the matching function is a P-matrix as well. The fact that the matching function is a P-matrix is an analogue of perfect PAM in the unidimensional case. This implies that positive sorting within the same dimension of skill and task exceeds that between different dimensions of skill and task. It also means that a pure matching exists between workers and firms. Nevertheless, unless the distributions of worker and firms are identical, the correlation between worker and firm types on each dimension of heterogeneity does not equal 1 in the multidimensional case. ¹⁰

In the case of directed search with discrete multidimensional types, a weaker notion of PAM must be considered. Before proceeding, one must first consider how equilibrium application strategies vary with productivity. This will in turn indicate how match probabilities, which are expressed in this model as \((1 - e^{-q(X,Y)}e^{-Q(X,Y)})\), vary with productivity. Given that the first order conditions give us no explicit function of application strategies in terms of productivity, I use the Implicit Function Theorem (IFT) to express the derivative of the expected queue lengths with respect to productivities. How match probabilities change with respect to productivities can then be derived.

¹⁰In her Appendix, she also extends the model of Eeckhout and Kircher (2010) to multidimensional types and finds that the same level of PAM is obtained under a condition that is the multidimensional corollary of their root-n supermodularity condition.
The detailed derivation of the following results, summarized in Theorem 2, can be found in the Appendix.

The first question to consider is how the expected queue length of a worker type changes with her productivity at a given firm type. In the Appendix, I show that

$$\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)} > 0$$

As would seem natural, an increase in the productivity of a worker type at a given firm type should increase the expected queue length, since it increases the marginal output (and hence the wage) of the worker. As the decentralized equilibrium is equivalent to that of the constrained Social Planner, the intuition of the result can also be understood from the following angle - the higher the productivity of a worker at a firm, the more the Social Planner increases the expected queue length since her aim is to increase the overall expected output.

Next, how do workers’ expected queue lengths vary with the productivities of other workers at a given firm who apply with positive probability? I find that

$$\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i', Y_j)} < 0 \text{ if } |i' - i| < 2, 0 \text{ otherwise}$$

The finding that a worker’s expected queue length decreases with other workers’ productivities at the firm is actually intuitive. Indeed, an increase in the productivity of a more preferred worker at the firm increases her application probability, hence lowering the probability of the lower ranked worker getting hired. This then discourages the worker from applying to some degree, thus lowering the expected queue length. The fact that the worker’s expected queue length only decreases with the productivity of the worker ranked one above her who applies with positive probability, but not the worker ranked two or above her is also intuitive. This is because so long as the worker ranked just one above her does apply to the firm, she will not get the job, regardless of what other higher-ranked workers do. As such, her expected queue length varies only with the worker ranked just one above her whom the firm also expects to attract with positive probability.

Next, why the expected queue length of a worker should vary with the productivity of the next best worker the firm seeks to attract is also intuitive. Recall that the wage of a worker is her expected marginal output. When the productivity of the next best worker increases, the expected marginal output and hence the wage of the worker decreases, thus lowering the expected queue length of the worker at that firm.

Lastly, I show that the expected queue length of a worker reacts more to a change in her own productivity than that of another worker at a given firm, a result that is also fairly intuitive.

$$|\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)}| \geq |\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i', Y_j)}|, \forall i' \neq i$$

**Theorem 2.** The Jacobian $\frac{\partial q}{\partial \Phi}$ is an irreducible diagonally dominant matrix. I.e. $i \frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)} > 0$,
ii) $\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i', Y_j)} < 0$, \forall |i' - i| < 2, and 0 otherwise, iii) $|\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i', Y_j)}| \geq |\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_{i'}, Y_j)}|$, \forall i' \neq i.

In words, Theorem 2 says the following: i) The equilibrium expected queue length for a given worker type at a given firm type is strictly increasing in the productivity of the worker at this firm type. ii) It is strictly decreasing in the productivity of the next less-preferred and the next better-preferred workers at this firm type and does not respond to the productivities of the rest of the workers at the firm. iii) Also, the expected queue length of a given worker type varies more with her own productivity at the firm than with the productivity of another worker type. The proof is in the Appendix.

**Corollary 1.** Theorem 2 implies that the probability that a given firm type is matched to a worker type is increasing in their joint productivity, and is weakly decreasing in the productivity of another worker type at that firm.

The proof for Corollary 1 is in the Appendix.

For future notational simplicity, denote the probability of a match occurring between a firm $Y$ and a worker $X$ as $\mu(X, Y)$.

**Theorem 3.** The Jacobian $\frac{\partial \mu}{\partial \Phi}$ is an irreducible diagonally dominant matrix. I.e. i) $\frac{\partial \mu(X,Y)}{\partial \Phi(X,Y)} > 0$, ii) $\frac{\partial \mu(X,Y)}{\partial \Phi(X,Y')} \leq 0$, iii) $\frac{\partial \mu(X,Y)}{\partial \Phi(X,Y')} > \left| \frac{\partial \mu(X,Y)}{\partial \Phi(X',Y')} \right|$.

In words, Theorem 3 states the following: i) The probability of a match occurring between $X$ and $Y$ is increasing in the productivity of that match. ii) The probability of a match occurring between $X$ and $Y$ is weakly decreasing the productivity of the match between $X'$ and $Y$, where $X' \neq X$. iii) The response of the probability of a match between $X$ and $Y$ to a change $\Phi(X, Y)$ is strictly greater in absolute terms than that to $\Phi(X', Y)$. The proof for Theorem 3 is in the Appendix.

Using the above result, the discussion can now move on to the conditions and characterization of assortative matching. At the beginning of this section, I argued that perfect PAM cannot be achieved in this model except under very restrictive conditions. Therefore, one could consider a weaker definition of assortative matching that is given by the log supermodularity of the matching function. This notion is also considered in Siow (2015) for the marriage market with unidimensional heterogeneity. The log supermodularity of the matching function is equivalent to it being totally positive of order 2 (TP2), as described in Siow (2015). This approach towards assortative matching is advantageous in the context of this paper since 1) matching probabilities between pairs of workers and firms are clearly defined in the model and 2) this notion of assortative matching is distribution-free.

First consider the unidimensional case. Denote the worker’s and firm’s unidimensional characteristics by $x$ and $y$ respectively. For ease of notation, denote probability of a match occurring between worker $x$ and firm $y$, $(1 - e^{-q(x,y)})e^{-Q(x,y)}$ as $\mu(x, y)$. Consider two pairs of workers and firms, $x, x', y$ and $y'$ where $x' > x$ and $y' > y$.

The log supermodularity of the matching function is given by
The above expression essentially states that the matching probability $\mu(x, y)$ exhibits log-supermodularity. As is well-known, log-supermodularity of the matching function implies that the matching displays First Order Stochastic Dominance (FOSD). The testable empirical implication of this is that $E[y]$ is increasing in $x$.

Under what condition does $m1$ hold? Consider the production function $\Phi(x, y)$. It is supermodular if

$$\Phi(x, y) + \Phi(x', y') > \Phi(x, y') + \Phi(x', y)$$

(M1)

Theorem 4. When worker and firm types are unidimensional, the sufficient condition for $m1$ is simply $p1$.

This is shown in the Appendix.

Having just shown the conditions under which this notion of assortative matching holds in the unidimensional world, this section now moves on to discuss the assortative matching in the multidimensional case. For simplicity, let us consider the case of two-dimensional heterogeneity, even if the results can be easily extended to higher levels of dimensionality.

Consider any 2 worker types $X = (x_1, x_2)$ and $X' = (x'_1, x_2)$ as well as two firm types $Y = (y_1, y_2)$ and $Y' = (y'_1, y'_2)$, where $x'_1 > x_1$, $y'_1 > y_1$ and no restriction is placed on $y_2$ and $y'_2$. The log-supermodularity of the matching function is said to exist in the first dimension when

$$\mu(X, Y)\mu(X', Y') > \mu(X', Y)\mu(X, Y')$$

(M1)

In words, this essentially means that given the same $x_2$, the product of the probabilities of a worker with a higher $x_1$ being matched to a firm with a higher $y_1$ and of a worker with a lower $x_1$ to a firm with a lower $y_1$ is higher than that of the opposite happening. A testable implication is that $E[y_1|x_2]$ should be increasing in $x_1$.

At this point, one may question why $x_2$ should be restricted to be the same between the worker pair while $y_2$ and $y'_2$ should be allowed to differ between the firm pair. This decision can be easily justified. Adopt the perspective of the worker. From the worker’s perspective, her type $X$ is exogenous, while the choice of the firm is endogenous. Indeed, the worker’s type determines the type of firm to which she is assigned. The tradeoff between $y_1$ and $y_2$ is thus inherent to the decision of the worker. As such, restricting the $y_2$s of the pair of firms while comparing the $y_1$s would be a highly unnatural imposition. However, the comparison still necessitates some discipline since there is no clear way of determining the direction of sorting if the pair of workers differed in both dimensions. In simple terms, one cannot compare apples with oranges. Therefore, it becomes necessary to restrict the $x_2$s of the workers to be equal. This is a reasonable restriction, since worker types are anyway exogenous.
Likewise, to characterize sorting in the second dimension, consider any 2 worker types \( X = (x_1, x_2) \) and \( X' = (x'_1, x'_2) \) as well as two firm types \( Y = (y_1, y_2) \) and \( Y' = (y'_1, y'_2) \), where \( x'_2 > x_2 \) and where \( y'_2 > y_2 \). No restriction is placed on \( y_1 \) and \( y'_1 \). Log supermodularity of the matching function holds in the second dimension when

\[
\mu(X, Y), \mu(X'', Y'') > \mu(X', Y), \mu(X', Y') \quad (M2)
\]

A testable implication for the above statement is that \( E[y_2|x_1] \) increases in \( x_2 \).

So far, we have assumed that \( K_x = K_y = 2 \). However, what if the dimensions of worker and firm heterogeneity differed? For instance, what if workers were characterized by two dimensional heterogeneity while firms were characterized by three? The equal dimensionality case can be easily extended. Say for instance that one was interested in whether assortative matching existed between these? For instance, what if workers were characterized by two dimensional heterogeneity while firms were characterized by three? The equal dimensionality case can be easily extended. Say for instance that one was interested in whether assortative matching existed between the 1st dimension of \( X \) and the 3rd dimension of \( Y \). Then, consider any two pairs of workers \( X = (x_1, x_2) \) and \( X = (x'_1, x'_2) \) and two pairs of firms \( Y = (y_1, y_2, y_3) \) and \( Y' = (y'_1, y'_2, y'_3) \), where \( x'_1 > x_1, y'_1 > y_1 \) and no restrictions are placed on the other dimensions of \( Y \). Assortative matching holds between the 1st dimension of \( X \) and the 3rd dimension of \( Y \) when

\[
\mu(X, Y), \mu(X', Y') > \mu(X', Y), \mu(X', Y')
\]

Where \( \mu(X, Y) \) is defined as before. A testable implication of the above statement is that \( E[y_2|x_1] \) is increasing in \( x_1 \).

Under what conditions is log-supermodularity of the matching function obtained? The next part shows that the supermodularity of the production function is in fact a sufficient condition. However, how supermodularity is defined in the multidimensional context differs slightly from how it is defined in the unidimensional case. Consider once again 2 worker types \( X = (x_1, x_2) \) and \( X' = (x'_1, x'_2) \) as well as two firm types \( Y = (y_1, y_2) \) and \( Y' = (y'_1, y'_2) \), where \( x'_1 > x_1, y'_1 > y_1 \) and no restriction is placed on \( y_2 \) and \( y'_2 \). The production function is supermodular in the first dimension if

\[
\Phi(X, Y) + \Phi(X', Y') > \Phi(X, Y') + \Phi(X', Y) \quad (P1)
\]

This essentially means total output is larger if the higher \( x_1 \) worker is matched to the higher \( y_1 \) firm, conditional on \( x_2 \).

Similarly, consider any 2 worker types \( X = (x_1, x_2) \) and \( X'' = (x_1, x'_2) \) as well as two firm types \( Y = (y_1, y_2) \) and \( Y'' = (y''_1, y''_2) \), where \( x''_2 > x_2 \) and where \( y''_2 > y_2 \). No restriction is placed on \( y_1 \) and \( y''_1 \). Supermodularity in the second dimension can be written as

\[
\Phi(X, Y) + \Phi(X'', Y'') > \Phi(X, Y'') + \Phi(X'', Y) \quad (P2)
\]

**Theorem 5.** A sufficient condition for \( M1 \) is \( P1 \). Likewise, a sufficient condition for \( M2 \) is \( P2 \).

The proof for the above is shown in the Appendix.
What do P1 and P2 mean concretely? To see what these conditions demand, consider the following example of a production function which meets these conditions\textsuperscript{11}.

\[ \Phi(X, Y) = \beta_1 x_1 y_1 + \beta_2 x_2 y_2 \]

In this simple specification, it is easy to see that all that is needed for supermodularity in dimension 1 is that \( \beta_1 > 0 \). Likewise, all that is needed for supermodularity in dimension 2 is that \( \beta_2 > 0 \).

Having explored the properties of sorting in the multidimensional framework, the next section examines how sorting responds to changes in production technology in the presence of multidimensional heterogeneity and how this lends the model easily to the empirical application of wage and job polarization.

5 The impact of technological change on the equilibrium assignment, sorting and wages

In what follows, it would be useful to adopt the perspective of the constrained Social Planner. When production technology changes, one may expect equilibrium sorting to change, since there may be a different output-maximizing assignment. This section demonstrates how the model presented in this paper can generate changes in sorting patterns, as well as wage and job polarization.

From the previous section, the expected queue length, as well as the probability of a match occurring between a given worker and firm type increase in their match productivity. How does the expected wage of a worker respond to an increase in her productivity at a given firm type? From the perspective of the Social Planner, the shadow value of worker types are given by their expected wage. Naturally, one would reasonably expect the shadow values of a worker to be increasing in the productivity of a worker at any given firm to which she applies with positive probability. From equation (4), the expected wage \( U(X) \) is given by the equilibrium queue lengths. Deriving the term with respect to \( \Phi(X, Y) \) proves fairly tedious. However, numerical simulations show that this is clearly the case. This will prove useful when studying the model’s account of polarization.

The catalyst for shifts in sorting, wage and job polarization that I consider is a change in production technology. For an illustration of how this could arise, consider a change from \( \Phi(X, Y) = x_1 y_1 + x_2 y_2 \) to \( \Phi(X, Y) = 1.5x_1 y_1 + 0.5x_2 y_2 \). Production complementarity between \( x_1 \) and \( y_1 \) therefore increases while that between \( x_2 \) and \( y_2 \) decreases. The productivity of matches between high \( x_1 \) and \( y_1 \) firms should improve substantially relative to that between other matches. The comparative static exercises above suggests this shift in productivity should alter both the assignment probabilities and expected wages in equilibrium. Adopt the perspective of the Social Planner. When the above change in production occurs, the Social Planner’s assignment decision is altered in three ways:

1. There is greater returns to assigning higher \( x_1 \) types to high \( y_1 \) types, which induces stronger positive sorting in the first dimension of heterogeneity.

\textsuperscript{11}A variation of this specification will be later used in the estimation.
2. There are higher returns to filling high $y_1$ jobs. The probability of high $y_1$ firms being filled should therefore rise.

3. Since high $x_1$ workers are now more productive, they should be valued more by the Social Planner. Hence, the shadow value of high $x_1$ workers, which is the expected wage, should increase.

The opposite occurs in the second dimension.

To show how sorting improves in the first dimension while deteriorating in the second dimension, consider the coefficients from the two regressions below. As previously mentioned, a testable implication of PAM in dimension $k$ is that $E[y_k|X_k']$ increases in $x_k$, where $k' \neq k$. Hence, $\theta_{11}$ is expected to increase and $\theta_{22}$ to decrease as complementarity rises in the first dimension.

\[
\begin{align*}
y_1 &= \theta_{10} + \theta_{11}x_1 + \theta_{12}x_2 + \epsilon_1 \\
y_2 &= \theta_{20} + \theta_{21}x_1 + \theta_{22}x_2 + \epsilon_2
\end{align*}
\]

Figure 2 shows how the coefficients $\theta_{11}$ and $\theta_{22}$ evolve as $\beta_1$ increases in $\Phi(X,Y) = \beta_1 x_1 y_1 + \beta_2 x_2 y_2$. It is clear that while sorting in the first dimension increases, it deteriorates in the second dimension, as expected.

Apart from changes in sorting, expected payoffs and filling probabilities are also altered. Figure 3 illustrates the convexifying of wages in skill 1 and a concavifying of wages in skill 2, implying that high $x_1$ workers gain relative to high $x_2$ workers in wages. This is in line with the expectation that high $x_1$ workers become more highly valued under the new technology, thereby leading to an increase in their expected payoff. Similarly, as shown in Figure 4, the probability of firms being filled convexifies in $y_1$ and concavifies in $y_2$. This corresponds to high $y_1$ being filled with greater probability under the new production technology, as the returns from doing so increases.

To see how this can lead to wage and job polarization, imagine 4 broad categories of workers: i) high $x_1$ and $x_2$ ($H_1H_2$), ii) high $x_1$ and low $x_2$ ($H_1L_2$), iii) high $x_2$ and low $x_1$ ($L_1H_2$) and iv) low $x_1$ and $x_2$ ($L_1L_2$). The first category of workers consists of the highest earners and the fourth contains the lowest earners. When the production technology changes as before, the first and second groups of workers gain the most while the third group of workers loses the most, since the dimension of skill in which they were most endowed is now valued far less by the Social Planner. Likewise, consider 4 broad categories of firms: i) high $y_1$ and $y_2$ ($H_1H_2$), ii) high $y_1$ and low $y_2$ ($H_1L_2$), iii) high $y_2$ and low $y_1$ ($L_1H_2$) and iv) low $y_1$ and $y_2$ ($L_1L_2$). Under the change in production, the third group of firms loses out the most, since filling these jobs are now less a priority for the Social Planner than before. These changes are illustrated in Figure 5. Figure 5a shows how the expected wage differs for the four broad groups of workers - $H_1H_2$ (in red), $H_1L_2$ (in green), $L_1H_2$ (in blue) and $L_1L_2$ (in purple). The x-axis measures the expected wage of workers under the old production technology while the y-axis gives the expected wage of workers under the new production technology. Scatterplots fall on the 45 degree line, therefore, if expected wages do not change between the old and new technology...
for a given worker type. From the figure, it is clear that the $L_1H_2$ types of workers experience a dip in expected wages relative to other types of workers as production technology shifts. Figure 5b illustrates how the probability of being filled is altered for the four broad groups of firms. On the x-axis, firms are ranked based on their expected profit under the old production technology. On the y-axis, the probability with which each firm type is filled, $Pr(filled)$ is given. The figure displays the $Pr(filled)$ for the four broad groups of firms - $H_1H_2$ (in red), $H_1L_2$ (in green), $L_1H_2$ (in blue) and $L_1L_2$ (in purple), under the old (triangle plots) and new (circle plots) production technology.

How the above changes lead to wage and job polarization is shown in figure 6. In figure 6a, the change in expected wage over wage percentiles between the old and new production technology is displayed. The change in production complementarity clearly leads to wage polarization. The labels at the top of the graph indicate the mean values of $X$ for each wage quartile. The biggest dip in wages occurs in the second quartile, where workers have a rather high mean $x_2$ of 1.7 and low mean $x_1$ of 0.4, which is consistent with the observations from figure 5a. The change in production complementarity also leads to job polarization, as displayed in figure 6b. On the x-axis, jobs are ranked by their mean wage under the old production technology, as is conventionally done in the literature. On the y-axis, the change in filling probability for each firm type is indicated. Since there is an equal number of firms of each type in this illustrative example, the change in filling probability would mirror the change in the expected number of filled jobs of each type. The labels at the top of the graph give the mean $Y$ for each job rank quartile. It is apparent that the second quartile of firms, with on average high $y_2$ and low $y_1$, face a drop in filling probabilities relative to the rest. Job polarization is hence obtained.

As such, a simple change in complementarity has been shown, in this example, to generate an improvement in sorting in the dimension in which it has increased, along with a deterioration in sorting in the dimension in which it has decreased. Wage and job polarization can also be generated by this same change in production complementarity. A close link is therefore established between changes in sorting patterns, wage and job polarization in the model. This property will prove useful when the model is brought to the data in the next section.

How does the model of multidimensional heterogeneity and directed search add to i) the frictionless undimensional model of Acemoglu and Autor (2011), Autor and Dorn (2013) and Autor et al. (2006b), or ii) the multidimensional frictionless model of Lindenlaub (2017) has yet to be addressed. In the former, the authors segment workers into three groups and demonstrate how an increase in the productivity for the highest skill group can lead to wage polarization and the displacement of middle skill workers to low-rank jobs. They also illustrate how job polarization can occur, when middle tasks are taken over by machines. The 1D frictionless case differs from on 2 main points: i) sorting between workers and firms would be predicted to deteriorate or remain unaltered$^{12}$ and ii) job polarization only occurs via the hollowing out of middle-rank vacancies.

---

$^{12}$If worker types and firm types are discrete, the shift of middle workers to lower-ranked jobs implies a deterioration of sorting. If worker and firm types are continuous, a shift of middle workers to lower-ranked jobs has no impact on sorting as workers and firms remain perfectly positively correlated.
As seen from the example exercise above, this is not the case in the model presented here. Sorting between workers and firms in different dimensions changes distinctly with production complementarities, as shown above. This allows for the study of trends in sorting in different dimensions of heterogeneity, in particular the trade-offs between sorting in each of these dimensions. Furthermore, as shown above, job polarization is obtained even if the distribution of vacancies posted remains fixed. This point is further discussed below.

Concerning why directed search is useful in the application to polarization, consider the frictionless model with multidimensional heterogeneity as presented in Lindenlaub (2017). The frictionless multidimensional model is able to produce wage polarization by a similar shift in complementarities. Moreover, it is able to generate predictions on sorting in line with the data, since the same trade-offs in sorting between different dimensions are present. Yet, it is silent on job polarization, as all jobs are filled and all workers are employed. Indeed the only way to obtain job polarization here would be if the initial distribution of vacancies were altered such that middle paying vacancies disappeared, just as in Acemoglu and Autor (2011), and Autor and Dorn (2013). On the contrary, in the model of multidimensional heterogeneity with directed search, job polarization can still arise even if the distribution of vacancies were to remain constant, since the probability that jobs are filled are in fact altered when production technology changes. Hence, in this model there are two channels through which job polarization can occur: i) change in the initial distribution of vacancies and ii) change in the production function.

To summarize, this subsection shows how changes in production technology can generate wage polarization and further, that job polarization can be produced as well, even when the initial distributions of workers and firms do not change. It has also illustrated how the shift in production complementarities generates clear trends in sorting patterns, due to trade-offs between sorting in the different dimensions of heterogeneity. This lends the model very easily to the subsequent empirical application.

6 The Empirical Application

The NLSY and O*NET are used to construct three dimensional vector of worker skills and job tasks - cognitive, interpersonal and manual. These three dimensions were chosen since they represent, arguably, three very different aspects of heterogeneity that broadly characterizes workers and jobs. Moreover, the productivity returns on these three dimensions could be expected to have evolved differently over time, with the advent of computers and ICT. These three aspects have also been considered in Lise and Postel-Vinay (2015) in their examination of how workers accumulate these skills on the job.

The NLSY dataset is well-known. It consists of panel surveys of individuals from two cohorts - 1997 and 1979. Individuals are interviewed annually or bi-annually up to the present day. I use information from both cohorts. As mentioned previously, the O*NET contains numerous descriptors of skills and knowledge sets required in each occupation, with information provided by analysts and worker surveys. I use the 15.0 installment of O*NET, which was released in 2010. An older version
might have been more appropriate for the 1979 NLSY cohort, since job tasks may change over time, but the older versions of O*NET contain information on far fewer jobs since the O*NET was only put in place from 2000 and were hence not used. In light of this, the assumption has to be made that if job tasks did change over time, they evolved in the same way across all occupations.

6.1 Construction of worker and firm characteristics

ONET contains 55 descriptors related to the level of cognitive, 32 related to interpersonal tasks and 60 related to manual tasks for each job. From the NLSY cohorts, I use the individual’s ASVAB test scores, her and parents’ educational attainment levels to describe cognitive skills. Scores on Rosenberg esteem scale and Rotter-Locus of control, as well as records of youth delinquency are then used to describe interpersonal skills. To measure manual skills, I look at their scores on the auto-information and object assembly part of the ASVAB test. I then apply Principal Components Analysis (PCA) separately on each set of descriptors to reduce them down to two dimensions of characteristics for workers and firms. I normalize each of these characteristics, such that their measures lie between 0 and 1. One should note that since these tests in the NLSY do vary somewhat between cohorts, it could be misleading to rank workers across cohorts. The construction of worker skills via PCA and the subsequent normalization are therefore done within cohorts and not between cohorts. It is also important to state that the measured skills of NLSY respondents do not change over time, since they only take the above tests once and their highest educational attainment is used when they leave school for good.

To illustrate what the vectors of workers and firm characteristics looks like concretely, consider the example of a respondent from the NLSY 1979 cohort. This is a person with $X = (x_c, x_i, x_m) = (0.6, 0.6, 0.8)$. She worked as a Funeral Director in her first job after leaving school which has tasks $Y = (y_c, y_i, y_m) = (0.6, 0.8, 0.6)$. The composition of tasks is unsurprising, since the job of a funeral director would involve moderately high cognitive tasks like planning and managing, as well as high interpersonal tasks, since she must deal with grieving clients. It also involves moderate manual tasks since she may have to, for instance, involve herself in hearse arrangements.

Table 1 displays the mean and standard deviations of each dimension of $X$ for both cohorts. There are 4193 respondents in the 1979 cohort and 4047 respondents in the 1997 cohort with full information on their $X$ vectors. The mean values of $x_c$, $x_i$ and $x_m$ and their standard deviations are given for both cohorts, as well as for men and women separately. As previously mentioned, the tests vary between cohorts and this likely explains, in part, the differences in means and standard deviations across cohorts. Nevertheless, two observations should be made. Firstly, there is substantial variation in all dimensions of skills within cohorts, with the largest standard deviation observed in $x_c$. Secondly, while women were slightly less skilled in all dimensions than men on average in NLSY 1979, they became more skilled than men in the cognitive and interpersonal dimensions than men in NLSY 1997.

---

14This includes a battery of tests that measures cognitive ability administered by the US army.
15The Rosenberg esteem scale measures an individual’s self-esteem while the Rotter-Locus of control measures a person’s self-motivation and determination through a series of survey questions.
6.2 Multidimensional heterogeneity and existing classifications of workers and firms

At this point, it would be useful to explore how the characterization of workers as bundles of skills and jobs as bundles of tasks differs from that in the existing literature. In this subsection, I consider how $x_c$, $x_i$ and $x_m$ vary with the standard high, medium and low skill categorization of workers. I also examine the relation of $y_c$, $y_i$ and $y_m$ to the routine and non-routine occupation classification established in previous papers.

Explanations of recent changes in wages and employment trends have largely centred around RBTC. In this literature, jobs are divided in two blocs - routine and non-routine. Routine jobs are characterized by codifiable rules while non-routine jobs involve more flexibility and adaptability. These two broad groups can then be further sub-divided into non-routine and routine cognitive jobs as well as non-routine and routine manual jobs. This classification is adopted, for instance, in Autor et al. (2003). In this paper, the authors argue that, since computer capital substitutes for labor in routine tasks and complements labor in non-routine tasks, the recent fall in computer prices should increase the demand for routine labor relative to non-routine labor.

Likewise, the abstract, routine and manual job classification in Acemoglu and Autor (2011) and Autor and Dorn (2013) correspond very closely to the previous classification. Abstract jobs coincide with non-routine cognitive, manual jobs with non-routine manual and routine jobs with both routine cognitive and routine manual. Broadly, they show, abstract jobs are analogous to managerial, professional and technical occupations, routine cognitive jobs to sales, clerical and administrative occupations, routine manual jobs to production, repair and operative occupations, and non-routine manual jobs to service occupations.

How does the description of jobs as bundles of cognitive, interpersonal and manual tasks correspond to the above classification of jobs? Table 3 shows the mean and standard of cognitive, manual and interpersonal tasks for each of the 4 broad groups of jobs calculated from the O*NET data. As expected, the mean cognitive task is highest in non-routine and routine cognitive jobs, while the mean manual task is largest in non-routine and routine manual jobs. Apart from this, however, it is evident that there each of the 4 broad job types involve substantial amounts of all 3 tasks. Furthermore, columns 5 to 7 show that there is quite a bit of variation in the three tasks within the broad groups, even when compared to the overall standard deviations, as indicated in columns 5 to 7 of the last row. This suggests that the characterization of jobs as bundles of tasks does not fully overlap with the routine/non-routine description in the above literature. Instead, the former appears to be complementary to the latter, while adding an extra layer of detail.

Similarly, the characterization of workers by vectors of $x_c$, $x_i$ and $x_m$ is distinct from the high and low skill classification of workers employed for instance in Autor and Dorn (2013). Taking, as these authors do, high skill workers to be the college-educated and low-skill workers to be the non college-educated, table 3 shows the mean and standard deviation of $x_c$, $x_i$ and $x_m$ for each of the two broad groups in the 1979 NLSY cohort as well as for the population as a whole. While the mean
of \( x_c \) is higher for college-educated and the standard deviation is lower, as expected, there is still substantial variation in \( x_c \) within the non-college group, as well as in \( x_i \) and \( x_m \) in both groups.

From the substantial variation in the dimensions of heterogeneity within the broad categories of workers and firms, it seems that the multidimensional heterogeneity considered in this paper is complementary to the description of workers and firms employed by the existing literature.

### 6.3 The estimation samples

Following Lindenlaub (2017), I include in the estimation sample of all men and women aged between 27 and 29 from the two NLSY cohorts who are working full-time and have left school\(^{16}\). I exclude those with missing information on any dimension of Y or X. Hourly wages, provided by the NLSY, are deflated by prices. As is often done, individuals with real hourly wages of below a dollar or above a hundred dollars are trimmed off. Since individuals may appear several times with different job assignments in 27 to 29 age period, I count these as separate observations. This is so as to increase the number of observations in the estimation sample at little cost.

The age group of 27 to 29 is considered firstly because the 1997 cohort respondents are fairly young and secondly because the time periods considered are ideal for studying polarization. Specifically, the 1979 cohort was aged 27 to 29 in the years between 1984 and 1993, while the 1997 cohort was aged 27 to 29 in the years between 2007 and 2013. This leaves a nice interval of a decade and a half between the two cohorts during which polarization is known to have occurred. A sample size of 5216 and 6550 is obtained for the 1979 and 1997 cohorts respectively.

Recall also that ours is a model of discrete types. In addition, the distribution of firm types, also discrete, must be estimated. I therefore place each dimension of worker skill into 3 bins and each dimension of job task into 2 bins, thereby creating 27 distinct worker types and 8 firm types. This reduces the number of firm type parameters to estimate, albeit possibly at the cost of model fit later.

Tables 4 and 5 show how workers’ skills correlate with job tasks for both NLSY cohorts, using the newly discretized measures. Consider the correlation matrix for the 1979 cohort, in table 4. Correlation between skill and task in the cognitive dimension, with \( \text{corr}(x_c, y_c) = 0.25 \), followed by the interpersonal dimension with \( \text{corr}(x_i, y_i) = 0.052 \). Correlation is negative on the manual dimension, with \( \text{corr}(x_m, y_m) = -0.08 \). For the 1997 cohort, correlation between cognitive skill and task is also the most strongly positive, with \( \text{corr}(x_c, y_c) = 0.34 \). It is slightly positive in between interpersonal skill and task, with \( \text{corr}(x_i, y_i) = 0.09 \). Correlation between manual skill and task is negative, with \( \text{corr}(x_m, y_m) = -0.05 \).

From the same two tables, one can also remark that while all three dimensions of workers’ skills are moderately positively correlated, the cognitive and interpersonal tasks of jobs are positively

---

\(^{16}\)Unlike Lindenlaub (2017), however, I use a broader range of years, thereby covering all individuals’ full-time work spells between age 27 and 29. In addition, the Xs of workers are constructed differently for \( x_c \) and \( x_m \). Lindenlaub (2017) matches workers’ education and degrees to O*NET occupations, then uses characteristics from O*NET to define workers’ \( x_c \) and \( x_m \).
correlated, and strongly negatively correlated with manual tasks. This suggests that jobs demand more specialization along the cognitive and interpersonal dimensions versus the manual dimensions than workers’ skills can muster. Put simply, workers tend to be jack of all trades, master of none, while jobs consist of either high cognitive and interpersonal tasks or high manual tasks.

Why workers tend to be jack of all trades is perhaps not so surprising, since there is some scientific literature on the positive link between interpersonal, manual and cognitive skills. For instance, the psychological literature on grit and achievement test scores in Duckworth et al. (2011) demonstrate that performance on cognitive tests is positively associated with internal motivation, which is included in our measure of interpersonal skill. In addition, Izard et al. (2001) demonstrate that emotional knowledge is positively linked to social behavior and academic competence. There is also literature on the positive link between psychomotor skills of the sort used in our manual skill measure and cognitive ability. For instance, Smits-Engelsman and Hill (2012) find that individuals with lower IQ tend to have poorer motor skills, even if some variation exists. Given the psychological literature then, it is therefore not so puzzling why some level of positive correlation between all dimensions of skills exist.

Table 6 shows the sorting regression coefficients for the two cohorts. From the combined evidence of tables 4, 5 and 6, it seems that sorting has increased most in the cognitive dimension and, to a much lesser extent, in the interpersonal and manual dimensions. Though sorting changes may suggest that complementaries have increased most in the cognitive dimension, second in the manual dimension and least in the interpersonal dimension, it is in fact premature to say so, since the distribution of $X$ and $Y$ has not remained constant across cohorts. While the distribution of posted $Y$ types has yet to be estimated, we can already see from Tables 4 and 5 that the correlation between $x_c$ and $x_i$ has increased somewhat between cohorts, while that between $x_c$ and $x_m$ has increased. These changes could have been either due to differences in the tests across the cohorts or due to actual changes in skills acquisition patterns between cohorts. Regardless, the estimation of the model will give the production parameters while taking into account the changes in worker and job distributions. Extra details on the assignment of $X$ and $Y$ for both cohorts are provided in the Online Appendix.

Having studied the characteristics of the estimation subsample, I now move on discuss the estimation procedure. To begin, one must consider what is observed and what has to be estimated. In the data, one does not observe the posted wages that are not accepted by any worker, the production function and the number of each type of job that are posted. What one can observe are the accepted wages, existing worker types $X$ and the types $Y$ of the jobs to which they are assigned.

The estimation of the original Shimer (2005) model has been done in Abowd et al. (2017), but two striking differences exist between their estimation procedure and that discussed here: i) With matched employer-employee data, Abowd et al. (2017) were unable to observe not only their workers’ and firms’ types, but were also unable to observe workers who were unemployed. They therefore had to estimate worker and firm types following the wage decomposition method in J. et al. (1999) before estimating the production function. ii) They focus exclusively on the context of unidimensional heterogeneity, which greatly simplifies their estimation of the production function.
6.4 The estimation strategy

There are two sets of parameters to be estimated: i) the production parameters and ii) the distribution of posted job types. Regarding the latter, while all the possible job types are given in the O*NET, it is impossible to identify the $\gamma(Y_j)$s for the $Y_j$ that did not hire any NLSY respondent in the data. Therefore, estimation can only be carried out for the $\gamma(Y_j)$s of the jobs that were actually observed to have hired someone. Concerning the former, a non-parametric estimation of the production function is not possible since not all worker types are matched to all firm types in the data. Therefore, estimation can only be carried out for the $\gamma(Y_j)$s of the jobs that were actually observed to have hired someone. Concerning the former, a non-parametric estimation of the production function is not possible since not all worker types are matched to all firm types in the data. Therefore, I adopt the following simple specification:

$$
\Phi(X, Y) = \beta_c x_c y_c + \beta_i x_i y_i + \beta_m x_m y_m + c_0
$$

(7)

Where $\beta_c$, $\beta_i$ and $\beta_m$ measure the degree of intra-dimensional complementarity between cognitive, interpersonal, manual skills and job tasks respectively. As already seen in the section on technological change and sorting, the $\beta$s alone suffice to produce wage and job polarization as well as changes in sorting and I therefore adopt, for simplicity, this parsimonious specification. $c_0$ represents the average productivity level in the economy.

The model is estimated by the Simulated Method of Moments (SMM). As previously mentioned, since types are discrete in my model, I segregate each X dimension into 3 bins and each Y dimension into 2 bins. This gives 27 worker types and 8 firm types. As such there are 15 parameters in total to be estimated. While identification is tricky to prove, the argument that identification is indeed achieved proceeds as follows. First assume that the $\gamma(Y_j)$s are known. Then the production parameters can be identified from the sorting patterns in the data. In particular, for $k = \{c, i, m\}$, $\beta_k$ is identified coefficient on $x_k$ in regression of $y_k$ on all X dimensions as well as the pure correlations between $x_k$ and $y_k$. In addition, coefficients from the regression of log wage on all Y dimensions are used for identification.

$$
\log(w) = \eta_c y_c + \eta_i y_i + \eta_m y_m + \text{constant} + \epsilon
$$

Why sorting coefficients should identify the $\beta$s is obvious in light of the previous section on how production technology change affects sorting. Ceteris paribus, an increase in $\beta_k$ should increase the sorting coefficient in that dimension relative to the sorting coefficients in dimension $k'$, where $k' \neq k$. The intersection of the all three sorting coefficients in the data should therefore pinpoint the values for the $\beta$s. Moreover, an increase in $\beta_k$ should intuitively increase $\eta_k$. $c_0$ is identified by the estimated intercept as well as the mean empirical wage.

Given the production parameters, the probability that a job type $Y_j$ is filled can be constructed. Recall that this is given by $\sum_i \mu(X_i, Y_j) = \sum_i (1 - e^{-\phi(X_i, Y_j)}) e^{-Q(X_i, Y_j)}$. Denote the number of observed matches between $X_i$ and $Y_j$ by $m(X_i, Y_j)$. Observe that $m(X_i, Y_j) = \gamma(Y_j) \sum_i \mu(X_i, Y_j)$. Since $m(X_i, Y_j)$ is observed in the data and $\mu(X_i, Y_j)$ can be constructed once production parameters are known, $\gamma(Y_j)$ is given by

$$
\gamma(Y_j) = \frac{\sum_i m(X_i, Y_j)}{\sum_i \mu(X_i, Y_j)}
$$

The $\gamma(Y_j)$s are thus identified by the observed number of worker matches with $Y_j$ in the data - $\sum_i m(X_i, Y_j)$. Evidently, however, $\gamma(Y_j)$ can only be identified for $Y_j$ that are filled by at least 1
worker in the dataset.

While the estimation strategy is presented in a two-step way, in practice both the production parameters and the $\gamma(Y_j)$s are estimated simultaneously. The 12 estimators therefore jointly minimize the 20 aforementioned moments. In order to find the global minimum, I use the Differential Evolution (DE) algorithm in R. Briefly, the DE algorithm is an evolutionary algorithm that finds the global minimum by differential mutation. A population of initial guesses for the parameters is first posited and these are mutated randomly to form successive generations of parameter guess. If a later generation of guesses outperforms its predecessor, the predecessor is discarded. If not, another new generation of guesses is generated from the predecessor. This continues until the global minimum is obtained or until the time limit is reached. The interested reader can refer to Ardia et al. (2016) and Ardia et al. (2011) for detailed information on the algorithm. To at least provide concrete evidence on local identification, I show that the loss function is indeed minimized at the estimated values of the parameters. This is shown in Figures 18 to 21 for the estimation on both NLSY cohorts.

6.5 Estimation results

Table 7 provides the estimation results for the production function parameters for the two NLSY cohorts. As can be gleaned from the table, both $\beta_c$ and $\beta_i$ have increased substantially between cohorts while $\beta_m$ has deterioriated significantly. From the table, $\beta_c$ rises from 0.393 to 1.356, while $\beta_i$ increases from 0.05 to 0.593. $\beta_m$, on the other hand, declines substantially, from 0.468 to 0.053 between cohorts. $\omega_0$ decreases very slightly, from 2.307 to 2.168. To get of sense of what these magnitudes imply, consider the following example. Say, for instance, that a worker of type $X = (1, 1, 1)$ was matched to a job of type $Y = (0.5, 0.5, 0.5)$. What is the gain in productivity as the worker’s match improves in each dimension of heterogeneity? Under the old technology, the worker would have increased her productivity by 7.12 percent if she had been matched to $Y = (1, 0.5, 0.5)$, 0.91 percent if matched to $Y = (0.5, 1, 0.5)$ and 8.48 percent if matched to $Y = (0.5, 0.5, 1)$. Under the new technology however, the productivity gain would have been 21.2 percent, 9.27 percent and 0.84 percent respectively.

Using the intuitions gleaned from Section 5, the estimated change in production technology should suffice to generate wage polarization. Moreover, it certainly explains why sorting on the cognitive dimension improved between the cohorts, since sorting patterns served as targeted moments. Moments used to identify the production parameters for both cohorts are also reasonably well-matched, as shown in Tables 8, 9, 10 and 11. At this juncture, it is important to note that wage percentiles were not targeted in the estimation. As such, the model is ex-ante not compelled to generate wage polarization.

More will be said on the estimated distribution of vacancy types in the next section. Recall that the number of filled job types observed was used to identify the $\gamma(Y_j)$s. Hence, the model is expected to fit job polarization in the data. Tables 12 and 13 show the quality of model fit to the number of filled job types, which is indeed very satisfactory.
7 Cross-cohort comparison: polarization and gender differences

Having viewed the estimation results, this section now examines the extent to which the model can account for the observed changes across cohorts. Before this, however, it is instructive to study these trends in the estimation datasets.

7.1 Wage and job polarization in the NLSY

Given the now extensive literature on wage and job polarization, it is reassuring to observe that these phenomena also exist in the NLSY. Figure 7 displays the wage polarization graph for the NLSY and cross-checks it with data from the March round of the CPS, which consists of a larger sample than the CPS. Both these graphs were constructed by gathering women and men aged 27 to 29 in from 1984 to 1992 and from 2007 to 2013, calculating the wage percentiles for each time period and then taking the change in wage percentiles across the time period. Unlike in Acemoglu and Autor (2011), I do not normalize the change to 0 at the median, which would preserve the same U-shape but mask the stark differences between different segments of the population. In particular, it is apparent from Figure 7 that women gain across the wage distribution relative to men, even as both men and women’s wages polarize. This point is missed in Acemoglu and Autor (2011), since changes in the median wage percentiles are normalized to 0 for men and women.

Likewise, while both men and women polarize in terms of jobs, they do so in differing ways. This is shown in Figure 8a, which displays the change in employment share across job ranks between cohorts, for men and women separately as well as for the population as a whole. Job ranks are obtained from the IPUMS-USA, which provides a ranking of occupations based on their average wage in 1980. This is in line with what is generally used in the literature. The occupation ranks are merged with the NLSY datasets using the occupation codes. Since the NLSY datasets are of moderate size, the employment shares for men, women and the whole population are then calculated for broader occupation groups. Furthermore, a robustness check is provided by the CPS data, as shown in Figure 8b. As there are many more observations in the CPS data, employment shares are calculated for more minute job categories, which explains the differences in share magnitudes. From these figures, it is evident that women actually gain more in the higher-ranked jobs while men gain more in lower-ranked jobs, even as both experience a dip in employment share in middling jobs. This empirical finding has yet to garner significant attention despite also being inadvertently documented in Acemoglu and Autor (2011).

7.2 Gender differences in wage and job polarization - plausible explanations

What could explain the differences in between men and women in their polarization experiences? While few papers in the literature have drawn a link between polarization and gender differences, two inter-linked strands of literature could provide some explanation for the above findings - i) structural change and ii) the rising importance of women’s skills. This subsection attempts

\[\text{Otherwise there would be too many zeroes and a comparison across cohorts would be difficult.}\]
to shed light on the validity of these two accounts and argues that a skills and tasks framework, such as the one presented in this paper, more adequately accounts for the empirical findings.

The structural change literature, including Ngai and Petrongolo (2017) and Barany and Siegel (2017), argue that women have gained in market hours and wages relative to men due to the rise of the service economy. This, in turn, has led to the marketization of the home production and allowed for the en masse movement of both high and low-skilled into the service industry workforce. Such a change could possibly account for the rise in employment share at the top and bottom of the job distribution over time in the service sector. In order to verify this account, I examine whether the same gender differences in job and wage polarization occur within both the service and non-service sector. If so, then the structural change explanation falls short, since it predicts that these gender differences should occur only in the service sector. As shown in Figures 9 and 10, wage and job polarization occur both in the service and non-service sectors. More importantly, women still gain more than men in terms of wages and employment in high-ranked jobs within both sectors. Therefore, the structural change narrative seems unable to wholly account for these phenomena.

The second strand of literature, including Cortes et al. (2017) and Black and Spitz-Oener (2011), argue that women, who are dominant in cognitive and interpersonal skills, benefit more from developments in ICT than men. This argument does seem to play out in the data. Figure 11 shows the change in the share of women in each decile of worker skill and job task. In terms of skill, the representation of women has increased in the upper deciles of all skill dimensions, as has the representation of women in the upper deciles of cognitive and interpersonal task deciles. On the contrary, the representation of women has decreased in the lower deciles of manual job task. This therefore seems to lend weight to this literature. However, the growing importance of women's skills cannot account for why both women and men polarized, as it basically predicts that women should gain in wages in a monotonically increasing way along the wage distribution, and that women should increase their representation exclusively in high-ranking jobs. This is clearly not the case in the data.

One should also note this improvement across all X dimensions among women does not just hold within the estimation sample and also cannot be wholly explained by the changes in educational attainment. As shown in Figure 12, even within the estimation and non-estimation samples, the increase in representation of women in higher X deciles holds. This then suggests that positive selection cannot fully account for this change. Similarly, the gain of women in all dimensions of X also holds within education groups, suggesting that the increase in educational attainment of women is not the only factor at work. As such, any attempt to get to the bottom of gender differences in X would have to go beyond selection and look at differential skill accumulation by gender. While this issue is beyond the scope of this paper, one can look to existing literature such as Bertrand and Pan (2013) and Bassok et al. (2016), which could suggest that the formalization of early childhood education could advantage girls over boys. This could explain why women have gained in skills relative to men between cohorts.
7.3 The model’s account of polarization and gender differences

While Section 5 has already presented how the model could produce wage and job polarization, this subsection examines how the estimation results, when fed back into the model can jointly account for wage and job polarization as well as the gender differences that exist. Note that neither the wage percentiles nor the gender differences in wages or jobs were targeted moments. In fact, the model is gender-blind. Only job polarization for the whole population was targeted since the distribution of filled jobs was used to identify the $\gamma(Y_j)$s.

Figure 13 shows wage polarization for men, women and the whole workforce generated by the model (full lines) and compared to the data (dashed lines). It is apparent that the model not only predicts wage polarization, but also predicts the gain for women in wages across the distribution, despite not having targeted any wage percentile or gender-specific moments. Nevertheless, the model clearly overpredicts the rise in wages for top earners and underpredicts the the rise for middle and low earners. Given the magnitude of increase in $\beta_c$ as well as the parsimony of the empirical specification, this is perhaps unsurprising.

Another measure of model fit to empirical wage polarization is given in Table 14. As seen from the table, the model over-estimates the growth in the 90th log wage percentile relative to the 50th percentile for the whole population, as well as for men and women separately. The model also over-predicts the decline 50th percentile relative to the 10th. This is perhaps unsurprising, as $\beta_c$ rose very sharply, especially relative to $\beta_m$, between cohorts. Given that wage percentiles were not targeted and the relative parsimony of the empirical specification, the model’s account of wage polarization is reasonable, even if other omitted factors are clearly important.

Figure 14a presents the model’s prediction on job polarization for men, women and the workforce as a whole. Again, the model not only predicts job polarization but also the disproportionate gain by women at the top of the job distribution relative to men. Nonetheless, while the model fits the targeted overall job polarization well, it performs less well in predicting job polarization for men and women separately. In particular, the model underpredicts the extent of job polarization for women, especially their rise in employment in high-rank jobs. Another assessment of fit is given by table 15. As expected, the model fits the data fairly well for the whole population, but clearly underpredicts the rise in the share of women employed in jobs ranked above the 70th percentile relative to jobs ranked between the 30th and 70th percentile (i.e. middling jobs). It also underpredicts the decline in the employment share of women in middling jobs relative to jobs ranked below the 30th percentile. The opposite is true for men. Nevertheless, the model is successful in capturing the essentials.

Figure 14b displays the fit of job polarization predicted by the model (in red) to that in the data (in blue). As expected, the model fits the data well in this aspect. The orange line in Figure 14b shows the change in the distribution of posted job types between cohorts which is given by the estimated $\gamma(Y_j)$s in the model. The share of posted job types seems to have grown disproportionately in lower-ranked jobs, even as middle-ranked jobs declined the most. Why this is so is beyond the scope of this paper, since vacancy creation is precluded from the model. Nonetheless, it is clear
that the distribution of posted job types does not alone suffice in accounting for job polarization. Judging from the distance between the red and orange curves, the change in production parameters is the key driver of job polarization across cohorts.

How is the model able to generate gender differences in wage and job polarization? Recall Figure 11a, which presented the gains of women in the estimation sample relative to men in all dimensions of \( X \). This, along with the large increase in returns to complementarity on \( x_c \) relative to \( x_i \) and \( x_m \), explains why women polarize in wages and yet also gain in wages across the wage distribution relative to men. As for why women go to higher-ranked jobs disproportionately while still facing job polarization, consider the increase of \( x_c \) of women relative to men coupled with the rise in \( C \) complementarity. These changes imply that women are sent disproportionately to higher \( y_c \) jobs, which are highest-ranked. However, since women of other types still exist, the same forces work to engender job polarization for women just as for men.

Hence, the model provides a reasonable account of the joint phenomena of wage and job polarization as well as the gender differences between men and women based solely on 2 channels: 1) changes in the distribution of \( X \) and \( Y \) and 2) changes in production technology. This is despite the model being relatively parsimonious as well as not having targeted wage percentiles or any gender-specific moments.

8 Counterfactuals

What would have happened to wage and job polarization, as well as the gender gap, had the distribution of worker types remained constant or had production technology not changed? How much of the actual reduction in the gender gap can be accounted for by the model and through what channels? Also, to what extent would total output be increased if workers became more specialist or generalist in their distributions of skills? This section attempts to address these through various counterfactual exercises.

8.1 The relative importance of ex-ante distributions and production technology

How would sorting, wages and job polarization look had production technology not changed, or had the distributions of workers and vacancies had remained the same? From Section 5, the model is expected to still generate wage and job polarization even with worker and vacancy distributions constant, so long as production technology changes. Yet, without the stark changes in skills by gender, one might expect gender differences to be understated. In addition, there would be no reason, \( a \ priori \), for there to be any polarization if only ex-ante distributions changed.

As such, I first consider the following 3 counterfactual scenarios:

1. When production parameters are allowed to change while the distributions of workers and firms are held at 1979 cohort level.
2. When production parameters are allowed to change while the distributions of workers and firms are held at 1997 cohort level.

3. When production parameters are held at the 1979 estimated levels while the distributions of workers and firms are allowed to vary between the cohorts.

These above counterfactuals have also been carried out in Lindenlaub (2017) for wage polarization. She finds that the change in production complementarities alone are able to account for wage polarization in her model, while the distributions of firms and workers are not. Here, the exercise is conducted for both wage polarization, job polarization and the related gender differences. Furthermore, since 3 dimensions of heterogeneity are considered, I later decompose the impact of complementarity changes in each dimension on both polarization and gender differences.

Figures 15 and 16 display wage and job polarization under the first 3 counterfactual scenarios as well as that predicted by the actual model. From Figures 15b and 15c, it is clear that even if the distributions of \( X \) and \( Y \) remained constant at 1979 or 1997 cohorts, the change in production complementarities is sufficient to generate wage polarization. This is in line with the counterfactuals conducted in Lindenlaub (2017). Given the simulated 2D exercise done in Section 5, this is finding is unsurprising. When only the distribution of \( Y \) and \( X \) changes, as shown in Figure 15d, wage polarization does not occur. Regarding gender differences, given that the women had not yet experienced a large increase in all dimensions of \( X \) in the 1979 cohort, the change in women’s wages does not dominate that of men as much in Figure 15b compared to Figure 15c.

Just as for wage polarization, all that is needed for job polarization is the change in production complementarities regardless of whether the distributions of \( X \) and \( Y \) are kept at the 1979 (Figure 16b) or 1997 (Figure 16c) cohort levels. Yet, women gain more relative to men in high-rank jobs with the production technology change under the 1997 distribution of \( X \) than the 1979 distribution, since their cognitive skills outstripped that of men in the later cohort. When the distribution of \( X \) and \( Y \) changes but production parameters are kept fixed, the growth of low-rank jobs far exceeds that of high-rank jobs. This is in line with the change in estimated \( \gamma(Y_j) \)'s, which has been presented in Figure 14b.

While production technology has been shown to be the main driver of polarization, one might reasonably question the extent to which complementarity changes are crucial to this result. As such, in the next set of counterfactual exercises, \( \beta_c, \beta_i \) and \( \beta_m \) are varied sequentially in order to assess the effect of each on wage and job polarization. These results are presented in Tables 16 and 17. From the tables, the changes in each of the \( \beta \)'s play important roles in wage and/or job polarization. It is therefore not merely the rise in \( \beta_c \) that drives polarization, but also the increase in \( \beta_i \) and decline in \( \beta_m \). This lends some weight to inclusion of all 3 dimensions of heterogeneity.

### 8.2 Changes in gender gaps as explained by the model

In the data, it is evident that women gained relative to men in wages as well as in job ranks even if both polarized. How much of the narrowing gender gap in terms of wages and job ranks does the model predict? As noted previously, there is no gender in the model, and any differences in gender...
outcomes must be due to differences in skills, tasks or production technology. The residual of what is left unexplained, should therefore be due to other factors, such as discrimination or heterogeneous preferences, which are precluded from the model.

Row 1 of Table 18 shows the mean wage difference between men and women in the data as well as in the full model. Recall that in the model, any gender differences that exist are solely due to X and the subsequently assigned Y. Comparing columns 1 and 2 of row 1 in 18, the average wage gap between men and women predicted by the model for the 1979 cohort is already much smaller than that in the data. Similarly, while the gender gap in the data still exists for the 1997 NLSY, the model predicts that women should have earned, on average, more than men in the 1997 cohort. Thus, other factors, apart from the distributions of the 3 skills and tasks as well as production technology, seem to be depressing women’s wages relative to men, across both cohorts.

Table 19 displays the change in the gender wage gap in the data and the model, as well as the 3 aforementioned counterfactual scenarios. While the gender wage gap decreased significantly in the data by 11.4 percentage points, the model accounts for 46.3 percent of the decrease in the gender wage gap, as shown in row 2 of the table. Had production parameters changed while X and Y kept at 1979 levels, the gender gap would have widened, since women in the sample were slightly less skilled than men on all dimensions. On the other hand, if X and Y had been kept at 1997 levels while production parameters hadn’t changed, the gender wage gap would have declined by 3.7 percentage points, a third of that in the data. This is perhaps unsurprising, since the distribution of X for women changed advantageously relative to men. For the same reason, as given by row 5, when the distributions of X and Y evolve without any change in production, the gender gap decreases 1.5 percentage points. If the assumptions of the model are believed, the fact that the model overpredicts the wages of women relative to that of men is indicative of other influences affecting the gender wage gap apart from the three dimensions of worker skills and firm tasks as well as production technology.

To what extent does the model account for the narrowing of the job rank between men and women? Table 20 gives the difference in average job rank (in percentiles) of men and women in the model and predicted by the data, while Table 21 presents the change in job rank gap between cohorts in the data, the full model and the 3 counterfactual scenarios. As can be gleaned from the tables, the model actually predicts 64.1 percent of the relative increase job ranks of women between cohorts. While in the data, women on average climbed 11.9 percentiles of job rank between cohorts, they rose by 7.6 in the model. The model’s ability to generate the large part of the increase in average job rank for women stems from its close account of women’s climb in cognitive tasks, which again is due to 1) the substantial increase in women’s cognitive skill relative to men and 2) the large rise in complementarity between cognitive skill and task. Had the distribution of skills been fixed at the 1979 level, the job rank gap would have widened. The same is true if only the distributions of skills and tasks had changed, as seen from the last row of Table 21.

Lastly, $\beta_c$, $\beta_i$ and $\beta_m$ are varied sequentially, so as to ascertain the impact of each on the gender wage and job rank gaps. As displayed in table 22, the rise of $\beta_c$ appears to be the main driver of the narrowing of the average gender wage gap, accounting for 0.032 out of the 0.05 increase their average log wage. The same is true for the narrowing of average gender job rank gap. This
is perhaps unsurprising, as women did become increasingly skilled in $x_c$ relative to men and since $\beta_c$ rose substantially between cohorts. Nevertheless, all $3 \beta$s contribute to the narrowing of wage and/or job gaps.

In short, despite precluding gender, the model is actually able to account for a large fraction of the decrease in the gender wage and job rank gap. Of course, skills, tasks and technology cannot explain the entirety of the changes in women’s labor market outcomes relative to men’s, but the fact that the simple mechanisms of the model have such explanatory power suggests that these factors are an important part of the story.

8.3 Generalist or specialist workers

With multidimensional heterogeneity, there would be richer implications of changes in worker skills. For instance, an increase in one dimension of workers’ skills could have different welfare implications than a smaller rise in all dimensions of workers’ skills. Considering how the overall weight on the cognitive dimension in production has increased, one might reasonably expect that a shift towards more cognitive specialists would raise total expected output more than a shift towards generalist workers, particularly under the new production technology. This is illustrated in Figure 17, which compares the evolution of total expected output under the production technologies of both cohorts when workers become either more specialized in cognitive skill or more skilled in all dimensions.

As expected, under the old production technology, the productivity difference when worker’s cognitive skills increase (in light blue, dashed) compared to when workers’ skills increase in equal proportion (in light green, dashed) is much less than that under the new production technology (in dark blue and dark green respectively). An economy with better cognitive specialists rather than better generalists, therefore, enjoys a greater increase in welfare from the baseline.

Evidently, the above is merely a partial equilibrium exercise and the distribution of vacancy types would have to be endogenized before obtaining a complete picture of the welfare effects of changes in the distribution of worker skills. Nonetheless, the above exercise illustrates the different implications of various shifts in worker skills when worker types are multidimensional.

9 Conclusion

This paper has sought to account for wage and job polarization, as well as the associated gender differences, through the lens of a model of directed search with multidimensional heterogeneity. It shows, under plausible regularity conditions, that the equilibrium is unique and exists. In addition, it has developed a characterization of assortative matching in this context and the conditions under which it occurs. Despite the inclusion of multidimensional heterogeneity, the model remains tractable. Moreover, it is able to make predictions on how changes in production complementarities affect how workers sort into jobs, on wages and the probability that different job types are filled.

\footnote{Refer to Bertrand (2011).}
Because of these predictions, the model lends itself easily to the empirical application of wage and job polarization.

The model is taken to the data, using the NLSY and O*NET databases to construct vectors of worker skills and job tasks. Three dimensions of worker skills and job tasks are considered - cognitive, interpersonal and manual. Between cohorts, four sets of trends have occurred: 1) improvements in sorting in the cognitive relative to the interpersonal and manual dimensions, 2) wage polarization, 3) job polarization and 4) gains in women relative to men in wages and job ranks even as both genders polarized. I examine the extent to which the model can account for these phenomena. Upon estimating the model by SMM, I then check the model fit and its predictions. I find that the model can in fact account for the four trends despite not having targeted any wage percentile moments and being gender-blind. While other factors could certainly have been at play in generating polarization and the changes in gender wage and job gaps, the mechanisms in the model, namely skills, tasks and production technology, suffice to account for a large part of the picture.

Having already created a ready framework for examining multidimensional heterogeneity, one could apply the model to another empirical setup, for instance with workers characterized by education and experience and with jobs characterized by occupation and industry. Lastly, an analysis of the welfare effects of changing workers’ skills would also be informative. While a partial equilibrium exercise has been conducted here, for any welfare analysis to be comprehensive, workers and firms types should be endogenized in this multidimensional framework. This will be left for future research.
Table 1: X vectors across cohorts

Summary statistics: X vectors

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>mean($x_c$)</th>
<th>mean($x_i$)</th>
<th>mean($x_m$)</th>
<th>sd($x_c$)</th>
<th>sd($x_i$)</th>
<th>sd($x_m$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979 (All)</td>
<td>0.460</td>
<td>0.636</td>
<td>0.566</td>
<td>0.324</td>
<td>0.182</td>
<td>0.192</td>
</tr>
<tr>
<td>1997 (All)</td>
<td>0.603</td>
<td>0.796</td>
<td>0.408</td>
<td>0.242</td>
<td>0.213</td>
<td>0.144</td>
</tr>
<tr>
<td>1979 (Men)</td>
<td>0.463</td>
<td>0.647</td>
<td>0.669</td>
<td>0.327</td>
<td>0.179</td>
<td>0.190</td>
</tr>
<tr>
<td>1979 (Women)</td>
<td>0.457</td>
<td>0.626</td>
<td>0.473</td>
<td>0.322</td>
<td>0.184</td>
<td>0.140</td>
</tr>
<tr>
<td>1997 (Men)</td>
<td>0.582</td>
<td>0.771</td>
<td>0.440</td>
<td>0.250</td>
<td>0.227</td>
<td>0.158</td>
</tr>
<tr>
<td>1997 (Women)</td>
<td>0.625</td>
<td>0.822</td>
<td>0.374</td>
<td>0.231</td>
<td>0.195</td>
<td>0.119</td>
</tr>
</tbody>
</table>

Table shows the mean and standard deviation of $x_c$, $x_i$ and $x_m$ individuals in the 1979 and 1997 NLSY cohorts. This is done for the population as a whole, as well as for men and women separately.

Table 2: Job classification: a comparison

Job types and their tasks

<table>
<thead>
<tr>
<th>Job types</th>
<th>mean($y_c$)</th>
<th>mean($y_i$)</th>
<th>mean($y_m$)</th>
<th>sd($y_c$)</th>
<th>sd($y_i$)</th>
<th>sd($y_m$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-routine Cognitive</td>
<td>0.570</td>
<td>0.631</td>
<td>0.255</td>
<td>0.117</td>
<td>0.137</td>
<td>0.172</td>
</tr>
<tr>
<td>Routine Cognitive</td>
<td>0.382</td>
<td>0.612</td>
<td>0.221</td>
<td>0.104</td>
<td>0.116</td>
<td>0.148</td>
</tr>
<tr>
<td>Routine Manual</td>
<td>0.308</td>
<td>0.430</td>
<td>0.686</td>
<td>0.099</td>
<td>0.098</td>
<td>0.144</td>
</tr>
<tr>
<td>Non-routine Manual</td>
<td>0.314</td>
<td>0.593</td>
<td>0.533</td>
<td>0.126</td>
<td>0.143</td>
<td>0.178</td>
</tr>
<tr>
<td>All</td>
<td>0.427</td>
<td>0.563</td>
<td>0.431</td>
<td>0.167</td>
<td>0.153</td>
<td>0.255</td>
</tr>
</tbody>
</table>

Table shows the mean and standard deviation of $y_c$, $y_i$ and $y_m$ for each of the broad job categories used by Acemoglu and Autor (2011), Autor et al. (2003) and Autor and Dorn (2013). These values are obtained by using the mapping of broad occupation categories to the 4 job types given in Acemoglu and Autor (2011), then calculating the mean and standard deviation of the three task measures for each of these broad categories using the O*NET data.

Table 3: Worker classification: a comparison

Worker types and their skills

<table>
<thead>
<tr>
<th>Worker types</th>
<th>mean($x_c$)</th>
<th>mean($x_i$)</th>
<th>mean($x_m$)</th>
<th>sd($x_c$)</th>
<th>sd($x_i$)</th>
<th>sd($x_m$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-college</td>
<td>0.269</td>
<td>0.602</td>
<td>0.531</td>
<td>0.168</td>
<td>0.180</td>
<td>0.199</td>
</tr>
<tr>
<td>College</td>
<td>0.898</td>
<td>0.704</td>
<td>0.627</td>
<td>0.067</td>
<td>0.169</td>
<td>0.171</td>
</tr>
<tr>
<td>All</td>
<td>0.450</td>
<td>0.621</td>
<td>0.544</td>
<td>0.320</td>
<td>0.184</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Table shows the mean and standard deviation of $x_c$, $x_i$ and $x_m$ for college and non college-educated, as well as for the population as a whole, in the NLSY 1979 cohort.
Table 4: Correlations between X and Y for 1979 cohort (estimation sample) (5216 obs)

<table>
<thead>
<tr>
<th></th>
<th>$x_c$</th>
<th>$x_i$</th>
<th>$x_m$</th>
<th>$y_c$</th>
<th>$y_i$</th>
<th>$y_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_c$</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_i$</td>
<td>0.236</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_m$</td>
<td>0.320</td>
<td>0.193</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_c$</td>
<td>0.251</td>
<td>0.107</td>
<td>0.167</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_i$</td>
<td>0.120</td>
<td>0.052</td>
<td>0.002</td>
<td>0.349</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$y_m$</td>
<td>-0.199</td>
<td>-0.064</td>
<td>-0.085</td>
<td>-0.670</td>
<td>-0.428</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Worker and firm heterogeneity are denoted as follows: $x_c$ - (y$_c$) - cognitive skill (task), $x_i$ - (y$_i$) - interpersonal skill (task), $x_m$ - (y$_m$) - manual skill (task)

Table 5: Correlations between X and Y for 1997 cohort (estimation sample) (6550 obs)

<table>
<thead>
<tr>
<th></th>
<th>$x_c$</th>
<th>$x_i$</th>
<th>$x_m$</th>
<th>$y_c$</th>
<th>$y_i$</th>
<th>$y_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_c$</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_i$</td>
<td>0.179</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_m$</td>
<td>0.339</td>
<td>-0.042</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_c$</td>
<td>0.384</td>
<td>0.133</td>
<td>0.173</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_i$</td>
<td>0.245</td>
<td>0.093</td>
<td>-0.015</td>
<td>0.435</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$y_m$</td>
<td>-0.328</td>
<td>-0.158</td>
<td>-0.048</td>
<td>-0.441</td>
<td>-0.444</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Worker and firm heterogeneity are denoted as follows: $x_c$ - (y$_c$) - cognitive skill (task), $x_i$ - (y$_i$) - interpersonal skill (task), $x_m$ - (y$_m$) - manual skill (task)
Table 6: Changes in sorting between cohorts

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>y_c</td>
<td>-0.262</td>
<td>-0.176</td>
<td>0.273</td>
</tr>
<tr>
<td>x_c</td>
<td>-0.129</td>
<td>0.0766</td>
<td>-0.112</td>
</tr>
<tr>
<td>x_i</td>
<td>0.0236</td>
<td>0.0184</td>
<td>-0.00894</td>
</tr>
<tr>
<td>x_m</td>
<td>0.0553</td>
<td>-0.0267</td>
<td>-0.0126</td>
</tr>
<tr>
<td>cohort 1997 \times x_c</td>
<td><strong>0.0860</strong></td>
<td>0.0906</td>
<td>-0.0897</td>
</tr>
<tr>
<td>cohort 1997 \times x_i</td>
<td>0.0209</td>
<td><strong>0.0058</strong></td>
<td>-0.0502</td>
</tr>
<tr>
<td>cohort 1997 \times x_m</td>
<td>-0.0206</td>
<td>-0.0379</td>
<td><strong>0.0495</strong></td>
</tr>
<tr>
<td>Constant</td>
<td>0.163</td>
<td>0.446</td>
<td>0.650</td>
</tr>
</tbody>
</table>

Observations 11766 11766 11766

Worker and firm heterogeneity are denoted as follows: $x_c$ ($y_c$) - cognitive skill (task), $x_i$ ($y_i$) - interpersonal skill (task), $x_m$ ($y_m$) - manual skill (task). Cohort 1997 is a dummy equaling 1 if the respondent is from the 1997 NLSY cohort.

Table 7: Production parameter estimates

<table>
<thead>
<tr>
<th>Production Parameters</th>
<th>1979 cohort (all)</th>
<th>1997 cohort (all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_c$</td>
<td>0.598 (0.020)</td>
<td>1.356 (0.175)</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>0.031 (0.015)</td>
<td>0.593 (0.423)</td>
</tr>
<tr>
<td>$\beta_m$</td>
<td>0.306 (0.137)</td>
<td>0.053 (0.162)</td>
</tr>
<tr>
<td>$c_0$</td>
<td>2.32 (0.009)</td>
<td>2.168 (0.047)</td>
</tr>
</tbody>
</table>

Standard errors of estimates are presented in parentheses.
Table 8: Model fit to sorting moments 1979 cohort

Model fit for 1979 cohort

<table>
<thead>
<tr>
<th>Matched moments</th>
<th>Empirical</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{corr}_c)</td>
<td>0.252</td>
<td>0.356</td>
</tr>
<tr>
<td>(\text{coef}_c)</td>
<td>0.193</td>
<td>0.249</td>
</tr>
<tr>
<td>(\text{corr}_i)</td>
<td>0.049</td>
<td>0.135</td>
</tr>
<tr>
<td>(\text{coef}_i)</td>
<td>0.026</td>
<td>0.095</td>
</tr>
<tr>
<td>(\text{corr}_m)</td>
<td>-0.084</td>
<td>-0.165</td>
</tr>
<tr>
<td>(\text{coef}_m)</td>
<td>-0.018</td>
<td>-0.032</td>
</tr>
</tbody>
</table>

\(\text{coef}_k\) for \(k \in \{c, i, m\}\) refers to the coefficient of the regression of \(y_k\) on \(x_k\), controlling for \(x_{k'}\), where \(k' \neq k\). \(\text{corr}_k\) for \(k \in \{c, i, m\}\) refers to the correlation coefficient between \(x_k\) and \(y_k\).

Table 9: Model fit to wage regression parameters - 1979 cohort

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_c)</td>
<td>0.479</td>
<td>0.390</td>
</tr>
<tr>
<td>(y_i)</td>
<td>0.025</td>
<td>0.020</td>
</tr>
<tr>
<td>(y_m)</td>
<td>0.032</td>
<td>0.134</td>
</tr>
<tr>
<td>Constant</td>
<td>2.170</td>
<td>2.189</td>
</tr>
</tbody>
</table>

\(t\) statistics in parentheses

Table 10: Model fit to sorting moments 1997 cohort

Model fit for 1997 cohort

<table>
<thead>
<tr>
<th>Matched moments</th>
<th>Empirical</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{corr}_c)</td>
<td>0.379</td>
<td>0.336</td>
</tr>
<tr>
<td>(\text{coef}_c)</td>
<td>0.317</td>
<td>0.413</td>
</tr>
<tr>
<td>(\text{corr}_i)</td>
<td>0.090</td>
<td>0.053</td>
</tr>
<tr>
<td>(\text{coef}_i)</td>
<td>0.035</td>
<td>0.072</td>
</tr>
<tr>
<td>(\text{corr}_m)</td>
<td>-0.040</td>
<td>-0.049</td>
</tr>
<tr>
<td>(\text{coef}_m)</td>
<td>0.061</td>
<td>0.010</td>
</tr>
</tbody>
</table>

\(\text{coef}_k\) for \(k \in \{c, i, m\}\) refers to the coefficient of the regression of \(y_k\) on \(x_k\), controlling for \(x_{k'}\), where \(k' \neq k\). \(\text{corr}_k\) for \(k \in \{c, i, m\}\) refers to the correlation coefficient between \(x_k\) and \(y_k\).
Table 11: Model fit to wage regression parameters - 1997 cohort

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Model</td>
<td></td>
</tr>
<tr>
<td>$y_c$</td>
<td>0.772</td>
<td>0.765</td>
</tr>
<tr>
<td>$y_i$</td>
<td>-0.132</td>
<td>-0.113</td>
</tr>
<tr>
<td>$y_m$</td>
<td>-0.031</td>
<td>-0.021</td>
</tr>
<tr>
<td>Constant</td>
<td>2.303</td>
<td>2.290</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses

Table 12: Distribution of filled job types - model fit for 1979 cohort

Model fit for filled job types - 1979 cohort

<table>
<thead>
<tr>
<th>Y types</th>
<th>Empirical</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_c = 0.5, y_i = 0.5, y_m = 0.5$</td>
<td>0.048</td>
<td>0.057</td>
</tr>
<tr>
<td>$y_c = 0.5, y_i = 0.5, y_m = 1$</td>
<td>0.213</td>
<td>0.202</td>
</tr>
<tr>
<td>$y_c = 0.5, y_i = 1, y_m = 0.5$</td>
<td>0.051</td>
<td>0.070</td>
</tr>
<tr>
<td>$y_c = 0.5, y_i = 1, y_m = 1$</td>
<td>0.109</td>
<td>0.107</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 0.5, y_m = 0.5$</td>
<td>0.107</td>
<td>0.103</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 0.5, y_m = 1$</td>
<td>0.050</td>
<td>0.041</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 1, y_m = 0.5$</td>
<td>0.410</td>
<td>0.402</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 1, y_m = 1$</td>
<td>0.011</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Table 13: Distribution of filled job types - model fit for 1997 cohort

Model fit for filled job types - 1997 cohort

<table>
<thead>
<tr>
<th>Y types</th>
<th>Empirical</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_c = 0.5, y_i = 0.5, y_m = 0.5$</td>
<td>0.045</td>
<td>0.043</td>
</tr>
<tr>
<td>$y_c = 0.5, y_i = 0.5, y_m = 1$</td>
<td>0.300</td>
<td>0.303</td>
</tr>
<tr>
<td>$y_c = 0.5, y_i = 1, y_m = 0.5$</td>
<td>0.114</td>
<td>0.111</td>
</tr>
<tr>
<td>$y_c = 0.5, y_i = 1, y_m = 1$</td>
<td>0.045</td>
<td>0.041</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 0.5, y_m = 0.5$</td>
<td>0.095</td>
<td>0.100</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 0.5, y_m = 1$</td>
<td>0.029</td>
<td>0.028</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 1, y_m = 0.5$</td>
<td>0.271</td>
<td>0.272</td>
</tr>
<tr>
<td>$y_c = 1, y_i = 1, y_m = 1$</td>
<td>0.099</td>
<td>0.101</td>
</tr>
</tbody>
</table>
Table 14: An account of wage polarization - model vs data

Wage polarization: Model vs Data

<table>
<thead>
<tr>
<th>Wage change ratio</th>
<th>Model</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \log(w)<em>{90} - \Delta \log(w)</em>{50}$</td>
<td>0.284</td>
<td>0.067</td>
</tr>
<tr>
<td>$\Delta \log(w)<em>{50} - \Delta \log(w)</em>{10}$</td>
<td>-0.081</td>
<td>-0.064</td>
</tr>
<tr>
<td>$\Delta (w)<em>{F90} - \Delta (w)</em>{F50}$</td>
<td>0.287</td>
<td>0.015</td>
</tr>
<tr>
<td>$\Delta (w)<em>{F50} - \Delta (w)</em>{F10}$</td>
<td>-0.087</td>
<td>-0.048</td>
</tr>
<tr>
<td>$\Delta (w)<em>{M90} - \Delta (w)</em>{M50}$</td>
<td>0.274</td>
<td>0.117</td>
</tr>
<tr>
<td>$\Delta (w)<em>{M50} - \Delta (w)</em>{M10}$</td>
<td>-0.008</td>
<td>-0.048</td>
</tr>
</tbody>
</table>

The table compares the following statistics given by the model and the data: i) $\Delta (\log(w)_{90} - \log(w)_{50})$ - the change in the difference between the 90th and 50th log wage percentiles between cohorts and ii) $\Delta (\log(w)_{50} - \log(w)_{10})$ - the change in the difference between the 50th and 10th log wage percentiles between cohorts.

Table 15: An account of job polarization - model vs data

Job polarization - model vs data

<table>
<thead>
<tr>
<th>Employment change</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta (emp_{&gt;70} - emp_{&gt;30&amp;&lt;70})$</td>
<td>0.481</td>
<td>0.334</td>
</tr>
<tr>
<td>$\Delta (emp_{&gt;30&amp;&lt;70} - emp_{&lt;30})$</td>
<td>-0.937</td>
<td>-0.769</td>
</tr>
<tr>
<td>$\Delta (emp_{F&gt;70} - emp_{F&gt;30&amp;&lt;70})$</td>
<td>1.120</td>
<td>0.310</td>
</tr>
<tr>
<td>$\Delta (emp_{F&gt;30&amp;&lt;70} - emp_{F&lt;30})$</td>
<td>-0.423</td>
<td>-0.130</td>
</tr>
<tr>
<td>$\Delta (emp_{M&gt;70} - emp_{M&gt;30&amp;&lt;70})$</td>
<td>0.004</td>
<td>0.371</td>
</tr>
<tr>
<td>$\Delta (emp_{M&gt;30&amp;&lt;70} - emp_{M&lt;30})$</td>
<td>-1.315</td>
<td>-1.165</td>
</tr>
</tbody>
</table>

The table shows i) $\Delta (emp_{>70} - emp_{>30\&<70})$ - the change in the difference between the share of employment above the 70th percentile and between the 30th and 70th percentile of job ranks and ii) $\Delta (emp_{>30\&<70} - emp_{<30})$ - the change in the difference between the share of employment above the 70th percentile and between the 30th and 70th percentile of job ranks. Effects are studied for the whole sample, as well as separately for men (rows 5 and 6) and women (rows 3 and 4).
Table 16: How sequential changes in production parameters contribute to wage polarization

Sequential effects of production parameter changes on wage polarization

<table>
<thead>
<tr>
<th></th>
<th>(1): $\beta_c$</th>
<th>(2): $(1) + \beta_i$</th>
<th>(3): $(2) + \beta_m$</th>
<th>(4): $(3) + f_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta(\log(w)<em>{90} - \log(w)</em>{50})$</td>
<td>0.085</td>
<td>0.218</td>
<td>0.256</td>
<td>0.238</td>
</tr>
<tr>
<td>$\Delta(\log(w)<em>{50} - \log(w)</em>{10})$</td>
<td>0.049</td>
<td>-0.012</td>
<td>-0.014</td>
<td>-0.009</td>
</tr>
<tr>
<td>$\Delta(\log(w)<em>{F90} - \log(w)</em>{F50})$</td>
<td>0.111</td>
<td>0.209</td>
<td>0.306</td>
<td>0.285</td>
</tr>
<tr>
<td>$\Delta(\log(w)<em>{F50} - \log(w)</em>{F10})$</td>
<td>0.019</td>
<td>-0.010</td>
<td>-0.035</td>
<td>-0.017</td>
</tr>
<tr>
<td>$\Delta(\log(w)<em>{M90} - \log(w)</em>{M50})$</td>
<td>0.081</td>
<td>0.273</td>
<td>0.315</td>
<td>0.311</td>
</tr>
<tr>
<td>$\Delta(\log(w)<em>{M50} - \log(w)</em>{M10})$</td>
<td>0.048</td>
<td>-0.027</td>
<td>-0.012</td>
<td>-0.001</td>
</tr>
</tbody>
</table>

The table shows a counterfactual exercise in which production parameters $\beta_c$, $\beta_i$, and $\beta_m$ are allowed to change from their 1979 to 1997 cohort values sequentially for the 1997 cohort. The impact of the change of each production parameter on i) $\Delta(\log(w)_{90} - \log(w)_{50})$ - the change in the difference between the 90th and 50th log wage percentiles between cohorts and ii) $\Delta(\log(w)_{50} - \log(w)_{10})$ - the change in the difference between the 50th and 10th log wage percentiles between cohorts, are recorded. Effects are studied for the whole sample, as well as separately for men (rows 5 and 6) and women (rows 3 and 4).

Table 17: How sequential changes in production parameters contribute to job polarization

Sequential effects of production parameter changes on job polarization

<table>
<thead>
<tr>
<th></th>
<th>(1): $\beta_c$</th>
<th>(2): $(1) + \beta_i$</th>
<th>(3): $(2) + \beta_m$</th>
<th>(4): $(3) + f_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta(\text{emp}<em>{\geq 70} - \text{emp}</em>{&gt;30&amp;&lt;70})$</td>
<td>0.283</td>
<td>0.289</td>
<td>0.313</td>
<td>0.334</td>
</tr>
<tr>
<td>$\Delta(\text{emp}<em>{&gt;30&amp;&lt;70} - \text{emp}</em>{&lt;30})$</td>
<td>-0.883</td>
<td>-0.841</td>
<td>-0.811</td>
<td>-0.769</td>
</tr>
<tr>
<td>$\Delta(\text{emp}<em>{F,70} - \text{emp}</em>{F,&gt;30&amp;&lt;70})$</td>
<td>-0.036</td>
<td>0.096</td>
<td>0.271</td>
<td>0.310</td>
</tr>
<tr>
<td>$\Delta(\text{emp}<em>{F,&gt;30&amp;&lt;70} - \text{emp}</em>{F,&lt;30})$</td>
<td>-0.261</td>
<td>-0.124</td>
<td>-0.173</td>
<td>-0.130</td>
</tr>
<tr>
<td>$\Delta(\text{emp}<em>{M,70} - \text{emp}</em>{M,&gt;30&amp;&lt;70})$</td>
<td>0.442</td>
<td>0.389</td>
<td>0.392</td>
<td>0.371</td>
</tr>
<tr>
<td>$\Delta(\text{emp}<em>{M,&gt;30&amp;&lt;70} - \text{emp}</em>{M,&lt;30})$</td>
<td>-1.129</td>
<td>-1.357</td>
<td>-1.159</td>
<td>-1.165</td>
</tr>
</tbody>
</table>

The table shows a counterfactual exercise in which production parameters $\beta_c$, $\beta_i$, $\beta_m$ and $f_0$ are allowed to change from their 1979 to 1997 cohort values sequentially for the 1997 cohort. The impact of the change of each production parameter on i) $\Delta(\text{emp}_{\geq 70} - \text{emp}_{>30\&<70})$ - the change in the difference between the share of employment above the 70th percentile of job ranks and between the 30th and 70th percentile of job ranks and ii) $\Delta(\text{emp}_{>30\&<70} - \text{emp}_{<30})$ - the change in the difference between the share of employment above the 70th percentile of job ranks and between the 30th and 70th percentile of job ranks, are recorded. Effects are studied for the whole sample, as well as separately for men (rows 5 and 6) and women (rows 3 and 4).

Table 18: Differences between model predictions and data on the gender wage gap for both cohorts

Gender wage gap - empirics vs model prediction

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Full Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[\ln(wage)</td>
<td>\text{female}] - E[\ln(wage)</td>
<td>\text{male}]$ in 1979</td>
</tr>
<tr>
<td>$E[\ln(wage)</td>
<td>\text{female}] - E[\ln(wage)</td>
<td>\text{male}]$ in 1997</td>
</tr>
</tbody>
</table>

Table shows the average log wage difference between men and women given by the data and the full model.
Table 19: Model predictions and counterfactuals on changes in the gender wage gap

<table>
<thead>
<tr>
<th>Change in $ln(wage)$ gap</th>
<th>$\Delta$ Model wage gap</th>
<th>$\Delta$ Empirical wage gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>0.114</td>
<td>1</td>
</tr>
<tr>
<td>Full Model</td>
<td>0.053</td>
<td>0.463</td>
</tr>
<tr>
<td>Counterfactual 1</td>
<td>-0.046</td>
<td>-0.402</td>
</tr>
<tr>
<td>Counterfactual 2</td>
<td>0.037</td>
<td>0.329</td>
</tr>
<tr>
<td>Counterfactual 3</td>
<td>0.015</td>
<td>0.134</td>
</tr>
</tbody>
</table>

Column 1 of the table shows the change in the average log wage gap between men and women given by the data, model and counterfactuals. The ratio of the change in wage gap between the model’s predictions and the data is given in the second column.

Table 20: Differences between model predictions and data on the gender job rank gap for both cohorts

<table>
<thead>
<tr>
<th>Gender job rank gap - empirics vs model prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
</tr>
<tr>
<td>-2.281</td>
</tr>
<tr>
<td>Full Model</td>
</tr>
<tr>
<td>-2.092</td>
</tr>
<tr>
<td>Data</td>
</tr>
<tr>
<td>9.584</td>
</tr>
<tr>
<td>Full Model</td>
</tr>
<tr>
<td>5.517</td>
</tr>
</tbody>
</table>

Table shows the average job rank difference between men and women given by the data and the full model.

Table 21: Model predictions and counterfactuals on changes in the gender job rank gap

<table>
<thead>
<tr>
<th>Change in job rank gap</th>
<th>$\Delta$ Model job rank gap</th>
<th>$\Delta$ Empirical job rank gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>11.86</td>
<td>1</td>
</tr>
<tr>
<td>Full Model</td>
<td>7.609</td>
<td>0.641</td>
</tr>
<tr>
<td>Counterfactual 1</td>
<td>-1.204</td>
<td>-0.102</td>
</tr>
<tr>
<td>Counterfactual 2</td>
<td>8.485</td>
<td>0.715</td>
</tr>
<tr>
<td>Counterfactual 3</td>
<td>-6.597</td>
<td>-0.556</td>
</tr>
</tbody>
</table>

Column 1 of the table shows the change in the average job rank gap between men and women given by the data, model and counterfactuals. The ratio of the change in job rank gap between the model’s predictions and the data is given in the second column.

Table 22: How individual changes in production parameters affect the gender wage and job gaps

<table>
<thead>
<tr>
<th>Sequential effects of production parameter changes on gender gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$ wage gap</td>
</tr>
<tr>
<td>$(1): \beta_c$</td>
</tr>
<tr>
<td>$(2): (1)+\beta_i$</td>
</tr>
<tr>
<td>$(3): (2)+\beta_m$</td>
</tr>
<tr>
<td>$(4): (3)+f_0$</td>
</tr>
<tr>
<td>$\Delta$ job rank gap</td>
</tr>
<tr>
<td>3.961</td>
</tr>
<tr>
<td>6.212</td>
</tr>
<tr>
<td>6.823</td>
</tr>
<tr>
<td>7.609</td>
</tr>
</tbody>
</table>

The table shows a counterfactual exercise in which production parameters $\beta_c$, $\beta_i$, $\beta_m$ and $f_0$ are allowed to change from their 1979 to 1997 cohort values sequentially for the 1997 cohort. The impact of the change of each production parameter on i) $\Delta$ wage gap - the change in average log wage gap between cohorts and ii) $\Delta$ job rank gap - the change in average job rank gap between cohorts is then recorded.
Figure 1: Timeline of the model

Firms post wages. Each firm type $Y_j$ posts wages $w(X_i, Y_j)$.

Each worker $X_i$ decides on application probability $p(Y_j|X_i)$ to all firms. A single application per worker is made.

Firm may receive $\geq 1$ or 0 applicants. If former, it hires its most preferred worker.

Production begins. Matched $X_i$ worker and $Y_j$ firm produce $\Phi(X_i, Y_j)$.

Unmatched workers and firms produce nothing.
Figure 2: Changes in sorting as $\beta_1$ increases

(a) Coefficient on $x_1$ from regression of $y_1$ on $X$

(b) Coefficient on $x_2$ from regression of $y_2$ on $X$

For Figure 2a: $y_1$ is regressed on $x_1$ and $x_2$. The coefficient on $x_1$ is plotted with respect to $\beta_1$. For Figure 2b: $y_2$ is regressed on $x_1$ and $x_2$. The coefficient on $x_2$ is plotted with respect to $\beta_1$.

Figure 3: Change in wage over each dimension of skill under old and new production technologies in the two dimension case

(a) $E[wage|x_1]$ under old and new technologies

(b) $E[wage|x_2]$ under old and new technologies

Figure 3a plots the expected wage with respect to $x_1$ under the old (in blue triangles) and new (in red circles) technologies. Figure 3b plots the expected wage with respect to $x_2$ under the old (in blue triangles) and new (in red circles) technologies. Local polynomial fitted lines are drawn for both figures, thereby illustrating the convexifying of expected wage in $x_1$ and the concavifying of expected wages in $x_2$ as the production technology changes.
Figure 4: Change in employment share under old and new production technologies in the two dimension case

(a) $Pr(filled|y_1)$ under old and new technologies
(b) $Pr(filled|y_2)$ under old and new technologies

Figure 4a plots the probability of jobs being filled with respect to $y_1$ under the old (in blue triangles) and new (in red circles) technologies. Figure 4b plots the probability of jobs being filled with respect to $y_2$ under the old (in blue triangles) and new (in red circles) technologies. Local polynomial fitted lines are drawn for both figures, thereby illustrating the convexifying of the filling probability in $y_1$ and the concavifying of filling probability in $y_2$ as the production technology changes.

Figure 5: Wages and filling probabilities under the old and new technologies in the two dimension case

(a) $E[wage]$ under old and new technologies
(b) $Pr(filled)$ under old and new technologies

Figure 5a and 5b plot the change in $E[wage]$ of worker types and the filling probabilities of firm types respectively under the old and new technologies. Worker types are categorized into 4 broad categories based on whether their values of $x_1$ and $x_2$ are above or below the population mean. A worker is therefore $H_1H_2$ if her $x_1$ and $x_2$ are above the mean, $H_1L_2$ if her $x_1$ is above the mean, while $x_2$ is below the mean, $L_1H_2$ if her $x_1$ is below the mean and $x_2$ is above the mean, and $L_1L_2$ if both $x_1$ and $x_2$ are below the mean. The same categorization of firm types is done based on their $y_1$ and $y_2$ values.
Figure 6: Polarization when $\Phi(X, Y)$ changes from $x_1y_1 + x_2y_2$ to $1.5x_1y_1 + 0.5x_2y_2$

For Figure 6a: The changes in $E[\text{wage}]$ across percentiles between the old and new technologies are plotted respectively. The coloured labels above the graph show the mean $x_1$ and $x_2$ values in each broad segment of the graph. For Figure 6b: Job types are ranked according to their expected profits under the old production technology. The change in the filling probabilities between the old and new technologies is then plotted. Since there are an equal number of job types in this simple example, the change in filling probabilities is equivalent to the change in employment share. The coloured labels above the graph show the mean $y_1$ and $y_2$ values in each broad segment of the graph.

Figure 7: Wage polarization in the NLSY and CPS data (All)

(a) Wage polarization in NLSY
(b) Wage polarization in CPS

Wage polarization graphs are constructed by taking the log wage percentiles for the older cohort and subtracting them from the log wage percentiles of the more recent cohort. This is done for the estimation population as a whole, as well as separately for men and women.
For Figure 8a: Due to the smaller sample size of the NLSY, the same 8 broad categories of jobs as in the estimation exercise are considered. Data on job ranks from the IPUMS-USA are used to rank these categories in ascending order. The employment shares in each job category are calculated for the 1979 cohort and then subtracted from that for the 1997 cohort. This change in employment share is then plotted after some local polynomial smoothing. For Figure 8b: Occupations from the CPS are also ranked using IPUMS-USA and then grouped into 20 bins. The average employment share in each bin is then calculated for the 1984 to 1993 period and then subtracted from that for the 2007 to 2013 period. The change in employment share is plotted after some local polynomial smoothing.
Figure 9: Wage and job polarization in the NLSY and CPS data (Services)

For Figures 9a and 9b: Wage polarization graphs are constructed by taking the log wage percentiles for the older cohort and subtracting them from the log wage percentiles of the more recent cohort. This is done for the estimation population working in the Service sector, as well as separately for men and women in the Service sector. For Figure 9c: Due to the smaller sample size of the NLSY, the same 8 broad categories of jobs as in the estimation exercise are considered. Data on job ranks from the IPUMS-USA are used to rank these categories in ascending order. The employment share in each job category is calculated for the 1979 cohort and then subtracted from that for the 1997 cohort. This change in employment share for the Service sector is then plotted after some local polynomial smoothing. For Figure 9d: Occupations from the CPS are also ranked using IPUMS-USA and then grouped into 20 bins. The average employment share in each bin is calculated for the 1984 to 1993 period and then subtracted from that for the 2007 to 2013 period. The change in employment share is plotted after some local polynomial smoothing for the Service sector.
For Figures 10a and 10b: Wage polarization graphs are constructed by taking the log wage percentiles for the older cohort and subtracting them from the log wage percentiles of the more recent cohort. This is done for the estimation population working in the Non-Service sector, as well as separately for men and women in the Non-Service sector. For Figure 10c: Due to the smaller sample size of the NLSY, the same 8 broad categories of jobs as in the estimation exercise are considered. Data on job ranks from the IPUMS-USA are used to rank these categories in ascending order. The employment share in each job category is calculated for the 1979 cohort and then subtracted from that for the 1997 cohort. This change in employment share for the Non-Service sector is then plotted after some local polynomial smoothing. For Figure 10d: Occupations from the CPS are also ranked using IPUMS-USA and then grouped into 20 bins. The average employment share in each bin is calculated for the 1984 to 1993 period and then subtracted from that for the 2007 to 2013 period. The change in employment share is plotted after some local polynomial smoothing for the Non-Service sector.
Figure 11: Change in share of women over $X$ and $Y$ deciles across NLSY cohorts

For Figure 11a: Workers are ranked in each dimension of heterogeneity by deciles. The share of women in each $x_k$ decile is computed for the 1979 and 1997 NLSY cohorts. For Figure 11b: Filled jobs are ranked in each dimension of heterogeneity by deciles. The share of women employed in each $y_k$ decile is then computed for each cohort.
Figure 12: Change in $X$ for women across NLSY cohorts within sample and education

For Figure 12a: Workers are split into two groups - in and out of the estimation sample. Within each group, workers are ranked in each dimension of $X$ by deciles. The share of women in each $x_k$ decile is computed for the 1979 and 1997 NLSY cohorts. The difference in shares for women between cohorts is then taken and plotted in the bar graph above.

For Figure 12b: Workers are split into two groups - college and non-college graduates. Within each education group, workers are ranked in each dimension of $X$ by deciles. The share of women in each $x_k$ decile is computed for the 1979 and 1997 NLSY cohorts. The difference in shares for women between cohorts is then taken and plotted in the bar graph above.
To construct the above graph, the estimated parameters for the NLSY 1979 and NLSY 1997 cohorts are fed back into model and the predicted wage percentiles for both cohorts are computed for the whole population, then for men and women separately. The predicted percentiles for the 1979 cohort are then subtracted from the predicted percentiles of the 1997 cohort. This is done for the estimation population as a whole, as well as separately for men and women.

For Figure 14a: The estimated parameters for the NLSY 1979 and NLSY 1997 cohorts are fed back into model and the share of filled jobs of each type for both cohorts are computed for the whole population, then for men and women separately. The difference in the predicted share of ranked filled job types between cohorts is then taken, both for the estimation population as a whole, as well as separately for men and women. A local polynomial smoothening is then applied. For Figure 14b: Plotted here is the change in share of filled jobs (ranked) in the data (in blue) and predicted by model (in red). The estimated change in share of vacancy types between cohorts is also graphed (in orange).
Figure 15: Wage polarization in the model - counterfactuals

For Figure 15a: Production parameters as well as the distribution of \( X \) and \( Y \) are allowed to vary between cohorts. For Figure 15b: The distributions of \( X \) and \( Y \) are held at the 1979 cohort level while production parameters are allowed to vary. For Figure 15c: The distributions of \( X \) and \( Y \) are held at the 1997 cohort level while production parameters are allowed to vary. For Figure 15d: The distributions of \( X \) and \( Y \) are allowed to vary between cohorts while production parameters are kept fixed.
Figure 16: Job polarization in the model - counterfactuals

(a) Full model
(b) Distributions held at 1979 level
(c) Distributions held at 1997 level
(d) Only distributions changed

For Figure 16a: Production parameters as well as the distribution of $X$ and $Y$ are allowed to vary between cohorts. For Figure 16b: The distributions of $X$ and $Y$ are held at the 1979 cohort level while production parameters are allowed to vary. For Figure 16c: The distributions of $X$ and $Y$ are held at the 1997 cohort level while production parameters are allowed to vary. For Figure 16d: The distributions of $X$ and $Y$ are allowed to vary between cohorts while production parameters are kept fixed.
Using the estimated production parameters for both cohorts, the total expected output is plotted as skills increase. Three scenarios are considered: i) when only $x_c$ increases by the amount indicated by values, ii) when $x_c$ and $x_i$ increase by the amount indicated by values divided by 2, and iii) when $x_c$, $x_i$, and $x_m$ increase together by the amount indicated by values divided by 3.
References


10 Appendix

Proof of uniqueness of the decentralized equilibrium

To show that the decentralized equilibrium is unique, I show that \( \pi(Y_j) \) is strictly concave in \( q(X_i, Y_j) \). Consider some firm type \( Y_j \). Recall that worker types are indexed by \( i \) and that firms have subjective preferences over worker types. Let firm \( Y_j \) give each worker type an index in order of preference such that worker type \( X_i \) refers to firm \( Y_j \)'s \( l \)th least preferred worker. The first and second derivatives of \( \pi(Y_j) \) can therefore be expressed as follows

\[
\frac{\partial \pi(Y_j)}{\partial q(X_i, Y_j)} = e^{-q(X_i, Y_j)}e^{-Q(X_i, Y_j)}\Phi(X_i, Y_j) - \sum_{l=1}^{I} (1 - e^{-q(X_{i'}, Y_j)})e^{-Q(X_{i'}, Y_j)}\Phi(X_{i'}, Y_j) - U(X_i)
\]

\[
\sum_{l=1}^{I} e^{-Q(X_{i'}, Y_j)}[\Phi(X_{i'}, Y_j) - \Phi(X_{i'-1}, Y_j)]
\]

\[
\frac{\partial^2 \pi(Y_j)}{\partial q(X_i, Y_j)_2} = \begin{pmatrix}
\frac{\partial^2 \pi(Y_j)}{\partial q(X_1, Y_j)^2} & \frac{\partial^2 \pi(Y_j)}{\partial q(X_1, Y_j)\partial q(X_2, Y_j)} & \cdots & \frac{\partial^2 \pi(Y_j)}{\partial q(X_1, Y_j)\partial q(X_I, Y_j)} \\
\frac{\partial^2 \pi(Y_j)}{\partial q(X_2, Y_j)\partial q(X_1, Y_j)} & \frac{\partial^2 \pi(Y_j)}{\partial q(X_2, Y_j)^2} & \cdots & \frac{\partial^2 \pi(Y_j)}{\partial q(X_2, Y_j)\partial q(X_I, Y_j)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 \pi(Y_j)}{\partial q(X_I, Y_j)\partial q(X_1, Y_j)} & \frac{\partial^2 \pi(Y_j)}{\partial q(X_I, Y_j)\partial q(X_2, Y_j)} & \cdots & \frac{\partial^2 \pi(Y_j)}{\partial q(X_I, Y_j)^2}
\end{pmatrix}
\]

Substituting in the expressions for the second derivatives, the quadratic form is

\[
-\sum_{l=1}^{I} [e^{-Q(X_{i'}, Y_j)}(\Phi(X_i, Y_j) - \Phi(X_{i'-1}, Y_j))] \left( \sum_{n=1}^{I} \rho_n \right)^2 < 0
\]

Where \( \rho_1...\rho_I \) is an arbitrary vector of non-zero numbers. \( \pi(Y_j) \) is therefore strictly concave in \( q(X_i, Y_j) \) and the decentralized equilibrium is hence unique.

To see that the above holds, consider a simple example of a firm type \( Y_j \) in an economy with only 4 worker types each with skills vector \( X_i, \ i = 1, 2, 3, 4 \). Assume \textit{wlog} that firm \( Y_j \) prefers the worker types in the following order: 2, 1, 3, 4 such that type 2 is \( Y_j \)'s most preferred worker and type 4 is its least preferred worker. The quadratic form is thus obtained from the following:

\[
Q = \begin{pmatrix}
\rho_1 & \rho_2 & \rho_3 & \rho_4 \\
\frac{\partial^2 \pi(Y_j)}{\partial q(X_1, Y_j)^2} & \frac{\partial^2 \pi(Y_j)}{\partial q(X_1, Y_j)\partial q(X_2, Y_j)} & \cdots & \frac{\partial^2 \pi(Y_j)}{\partial q(X_1, Y_j)\partial q(X_I, Y_j)} \\
\frac{\partial^2 \pi(Y_j)}{\partial q(X_2, Y_j)\partial q(X_1, Y_j)} & \frac{\partial^2 \pi(Y_j)}{\partial q(X_2, Y_j)^2} & \cdots & \frac{\partial^2 \pi(Y_j)}{\partial q(X_2, Y_j)\partial q(X_I, Y_j)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 \pi(Y_j)}{\partial q(X_I, Y_j)\partial q(X_1, Y_j)} & \frac{\partial^2 \pi(Y_j)}{\partial q(X_I, Y_j)\partial q(X_2, Y_j)} & \cdots & \frac{\partial^2 \pi(Y_j)}{\partial q(X_I, Y_j)^2}
\end{pmatrix} \begin{pmatrix}
\rho_1 \\
\rho_2 \\
\rho_3 \\
\rho_4
\end{pmatrix}
\]
Which gives

\[ Q = \sum_{i=1}^{4} \rho_i^2 \left( \frac{\partial^2 \pi(Y_j)}{\partial q(X_i, Y_j)} \right)^2 + 2 \sum_{i=1}^{3} \sum_{i' > i} \rho_{i'} \frac{\partial^2 \pi(Y_j)}{\partial q(X_i, Y_j) \partial q(X_{i'}, Y_j)} \]

Substituting the fact that

\[ \frac{\partial^2 \pi(Y_j)}{\partial q(X_i, Y_j) \partial q(X_{i'}, Y_j)} = \begin{cases} \frac{\partial^2 \pi(Y_j)}{\partial q(X_i, Y_j)} & \text{if } \Phi(X_{i'}, Y_j) > \Phi(X_i, Y_j) \\ \frac{\partial^2 \pi(Y_j)}{\partial q(X_{i'}, Y_j)} & \text{if } \Phi(X_{i'}, Y_j) < \Phi(X_i, Y_j) \end{cases} \]

and rearranging, one obtains

\[ Q = \frac{\partial^2 \pi(Y_j)}{\partial q(X_4, Y_j)^2} \left[ \rho_4^2 + 2(\rho_1 \rho_4 + \rho_2 \rho_4 + \rho_3 \rho_4) \right] + \frac{\partial^2 \pi(Y_j)}{\partial q(X_3, Y_j)^2} \left[ \rho_3^2 + 2(\rho_1 \rho_3 + \rho_2 \rho_3) \right] + \frac{\partial^2 \pi(Y_j)}{\partial q(X_2, Y_j)^2} \left[ \rho_1^2 + 2 \rho_1 \rho_2 \right] \]

Then, substituting in

\[ \frac{\partial^2 \pi(Y_j)}{\partial q(X_l, Y_j)^2} = -\sum_{l'=1}^{l} e^{-Q(X_{l'}, Y_j)} [\Phi(X_{l'}, Y_j) - \Phi(X_{l'-1}, Y_j)] \]

Where I let \( l \) be the preference index assigned by \( Y_j \) to the 4 worker types, such that \( l = 1 \) refers to worker type \( i = 4 \), \( l = 2 \) refers to worker type \( i = 3 \), \( l = 3 \) refers to worker type \( i = 1 \) and \( l = 4 \) refers to worker type \( i = 2 \). Rearranging, one gets

\[ Q = -\sum_{l=1}^{4} e^{-Q(X_l, Y_j)} (\Phi(X_l, Y_j) - \Phi(X_{l-1}, Y_j)) \left( \sum_{n=1}^{4} \rho_i \right)^2 < 0 \]

**Proof of Theorem 2**

Consider the firm \( Y_j \). In what follows, let index \( i \) be firm \( Y_j \)'s ranking over workers, such that \( i = 1 \) refers to \( Y_j \)'s least preferred worker type and \( i = I \) refers to \( Y_j \)'s most preferred worker type. **WLOG**, assume that firm \( Y_j \) seeks to attract non-zero queue lengths from all worker types. As such,
the First Order Conditions (FOCs) for this firm’s problem are

\[ e^{-q(X_1, Y_j)}e^{-Q(X_1, Y_j)}\Phi(X_1, Y_j) - U(X_1) = 0 \]
\[ e^{-q(X_2, Y_j)}e^{-Q(X_2, Y_j)}\Phi(X_2, Y_j) - (1 - e^{-q(X_2, Y_j)})e^{-Q(X_2, Y_j)}\Phi(X_1, Y_j) - U(X_2) = 0 \]
\[ \vdots \]
\[ e^{-q(X_i, Y_j)}e^{-Q(X_i, Y_j)}\Phi(X_i, Y_j) - \sum_{i'=1}^{i-1} (1 - e^{-q(X_{i'}, Y_j)})e^{-Q(X_{i'}, Y_j)}\Phi(X_{i'}, Y_j) - U(X_i) = 0 \]
\[ \vdots \]
\[ e^{-q(X_l, Y_j)}e^{-Q(X_l, Y_j)}\Phi(X_l, Y_j) - \sum_{i'=1}^{l-1} (1 - e^{-q(X_{i'}, Y_j)})e^{-Q(X_{i'}, Y_j)}\Phi(X_{i'}, Y_j) - U(X_l) = 0 \]

Where \( e^{-Q(X_l, Y_j)} = 1 \). From the FOCs, it is evident that the equilibrium queue lengths of workers depends not only on their own productivities at firm \( Y_j \) but also on the queue lengths and productivities of other worker types at firm \( Y_j \). The aim is to show that each worker’s queue length is a function of her productivity and that of other workers. For notational simplicity, express the above first order conditions as follows:

\[ g_1(q(X_1, Y_j), ..., q(X_l, Y_j), \Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) = 0 \]
\[ g_2(q(X_1, Y_j), ..., q(X_l, Y_j), \Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) = 0 \]
\[ \vdots \]
\[ g_l(q(X_1, Y_j), ..., q(X_l, Y_j), \Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) = 0 \]
\[ \vdots \]
\[ g_m(q(X_1, Y_j), ..., q(X_l, Y_j), \Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) = 0 \]

Can the IFT be used to show that the following exists?

\[ q(X_1, Y_j) = f_1(\Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) \]
\[ q(X_2, Y_j) = f_2(\Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) \]
\[ \vdots \]
\[ q(X_i, Y_j) = f_i(\Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) \]
\[ \vdots \]
\[ q(X_l, Y_j) = f_l(\Phi(X_1, Y_j), ..., \Phi(X_l, Y_j)) \]

Two conditions have to be satisfied for the IFT to be applicable. Firstly, \( g_1 \) to \( g_m \) must be continuous and differentiable around the neighbourhood of the \( \Phi \)s and the optimum queue lengths. This is indeed true. Secondly, the determinant of the Jacobian matrix \( \frac{\partial g}{\partial q} \) has to be non-zero. I show the second condition holds via Mathematical Induction (MI). Consider the case when firm \( Y_j \) expects non-zero queue length from only 1 worker type. Let \[ |B_1| = \left| \frac{\partial g_1}{\partial q(X_1, Y_j)} \right| = -e^{-q(X_1, Y_j)}\Phi(X_1, Y_j) < 0. \]
Consider $|B_2|$, when firm $Y_j$ expects non-zero queue length from 2 worker types.

$$|B_2| = \det \left( \begin{array}{ccc} \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_n}{\partial q(X_1, Y_j)} & \frac{\partial q_n}{\partial q(X_2, Y_j)} & \cdots & \frac{\partial q_n}{\partial q(X_n, Y_j)} \end{array} \right)$$

$$= \left[ e^{-q(X_1, Y_j)}e^{-Q(X_1, Y_j)}\Phi(X_1, Y_j) \right] \left[ e^{-q(X_2, Y_j)}e^{-Q(X_2, Y_j)}\Phi(X_2, Y_j) - (1 - e^{-q(X_1, Y_j)})e^{-Q(X_1, Y_j)}\Phi(X_1, Y_j) \right]$$

$$- \left[ e^{-q(X_1, Y_j)}e^{-Q(X_1, Y_j)}\Phi(X_1, Y_j) \right]^2$$

$$= \left[ e^{-q(X_1, Y_j)}e^{-Q(X_1, Y_j)}\Phi(X_1, Y_j) \right] \left[ e^{-q(X_2, Y_j)}\Phi(X_2, Y_j) - \Phi(X_1, Y_j) \right]$$

$$> 0$$

I posit that $|B_n| < 0$ for $n$ odd and $|B_{n'}| > 0$ for $n'$ even. Suppose that $|B_n| < 0$ for (wlog) $n$ odd, where $|B_n| = \det \left( \begin{array}{ccc} \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_n}{\partial q(X_1, Y_j)} & \frac{\partial q_n}{\partial q(X_2, Y_j)} & \cdots & \frac{\partial q_n}{\partial q(X_n, Y_j)} \end{array} \right)$ . Now show that $|B_{n+1}| > 0$.

$$|B_{n+1}| = \det \left( \begin{array}{ccc} \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_n}{\partial q(X_1, Y_j)} & \frac{\partial q_n}{\partial q(X_2, Y_j)} & \cdots & \frac{\partial q_n}{\partial q(X_n, Y_j)} \end{array} \right) \left( \begin{array}{ccc} \frac{\partial q_{n+1}}{\partial q(X_1, Y_j)} & \frac{\partial q_{n+1}}{\partial q(X_2, Y_j)} & \cdots \\ \frac{\partial q_{n+1}}{\partial q(X_1, Y_j)} & \frac{\partial q_{n+1}}{\partial q(X_2, Y_j)} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_{n+1}}{\partial q(X_1, Y_j)} & \frac{\partial q_{n+1}}{\partial q(X_2, Y_j)} & \cdots & \frac{\partial q_{n+1}}{\partial q(X_n, Y_j)} \end{array} \right)$$

Some algebra gives

$$|B_{n+1}| = \left( e^{-q(X_n, Y_j)}e^{-Q(X_n, Y_j)}\Phi(X_n, Y_j) - \sum_{n' = 1}^{n-1} (1 - e^{-q(X_n', Y_j)})e^{-Q(X_n', Y_j)}\Phi(X_n', Y_j) \right)$$

$$- e^{-q(X_{n+1}, Y_j)}e^{-Q(X_{n+1}, Y_j)}\Phi(X_{n+1}, Y_j) + \sum_{n' = 1}^{n} (1 - e^{-q(X_n', Y_j)})e^{-Q(X_n', Y_j)}\Phi(X_n', Y_j) \right) |B_n|$$

$$> 0$$

Since both the term in the large bracket and $|B_n|$ are negative. As such, by MI, $|B_n| < 0$ for all $n$ odd and $|B_{n'}| > 0$ for all $n'$ even. Thus, \frac{\partial q}{\partial q} has a non-zero determinant.

Since both conditions are satisfied, the IFT states that the set of functions $f_1, \ldots, f_I$ exists. Although the explicit expressions cannot be obtained, the IFT, along with Cramer’s rule allows us to express the following

$$\frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)} = - \det \left( \begin{array}{ccc} \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_n}{\partial q(X_1, Y_j)} & \frac{\partial q_n}{\partial q(X_2, Y_j)} & \cdots & \frac{\partial q_n}{\partial q(X_n, Y_j)} \end{array} \right)$$

$$\det \left( \begin{array}{ccc} \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \frac{\partial q_1}{\partial q(X_1, Y_j)} & \frac{\partial q_1}{\partial q(X_2, Y_j)} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_n}{\partial q(X_1, Y_j)} & \frac{\partial q_n}{\partial q(X_2, Y_j)} & \cdots & \frac{\partial q_n}{\partial q(X_n, Y_j)} \end{array} \right) \left( \begin{array}{ccc} \frac{\partial q_{n+1}}{\partial q(X_1, Y_j)} & \frac{\partial q_{n+1}}{\partial q(X_2, Y_j)} & \cdots \\ \frac{\partial q_{n+1}}{\partial q(X_1, Y_j)} & \frac{\partial q_{n+1}}{\partial q(X_2, Y_j)} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_{n+1}}{\partial q(X_1, Y_j)} & \frac{\partial q_{n+1}}{\partial q(X_2, Y_j)} & \cdots & \frac{\partial q_{n+1}}{\partial q(X_n, Y_j)} \end{array} \right)$$

60
Where the denominator is actually $|B_I|$ and the numerator is just the determinant of $B_I$ but with the $i$th column replaced by $\frac{\partial q_i}{\partial q(X_{i-1},Y_j)}$, ..., $\frac{\partial q_i}{\partial q(X_{i+1},Y_j)}$ instead. Since we already know that the denominator > 0 when $I$ is even and < 0 when $I$ is odd, let us focus on the sign of the numerator.

Consider first the sign of $\frac{\partial q(Y_i)}{\partial \Phi(X_i,Y_j)}$, which describes how the equilibrium queue length of $X_i$ at $Y_j$ changes as the productivity of $X_i$ at $Y_j$ increases. From the above, this denominator of this object is $|B_I|$, while the numerator is $\det \begin{pmatrix} 
\frac{\partial q_1}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i-1},Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+1},Y_j)} \\
\frac{\partial q_2}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_2}{\partial q(X_{i-1},Y_j)} & \cdots & \frac{\partial q_2}{\partial q(X_{i+1},Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_i}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_i}{\partial q(X_{i-1},Y_j)} & \cdots & 0 \\
\frac{\partial q_{i+1}}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i-1},Y_j)} & \cdots & 0 \\
\end{pmatrix} e^{-q(X_i,Y_j)}$. Using MI,

I will prove that this object is positive when $I$ is odd and negative when $I$ is even.

**Step 1** First the case where firm $Y_j$ expects a non-zero queue length from the least preferred $i$ worker types. **WLOG** assume that $i$ is odd. The numerator is thus

$$|A_i| = \det \begin{pmatrix} 
\frac{\partial q_1}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i-1},Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+1},Y_j)} \\
\frac{\partial q_2}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_2}{\partial q(X_{i-1},Y_j)} & \cdots & \frac{\partial q_2}{\partial q(X_{i+1},Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_i}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_i}{\partial q(X_{i-1},Y_j)} & \cdots & 0 \\
\frac{\partial q_{i+1}}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i-1},Y_j)} & \cdots & 0 \\
\end{pmatrix} e^{-q(X_i,Y_j)} |B_{i-1}| > 0 \quad \text{Since } |B_{i-1}| > 0.
$$

Now assume that firm $Y_j$ expects non-zero queue lengths from all of the least preferred $i + 1$ worker types. Then we have

$$|A_{i+1}| = \det \begin{pmatrix} 
\frac{\partial q_1}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i-1},Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+1},Y_j)} \\
\frac{\partial q_2}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_2}{\partial q(X_{i-1},Y_j)} & \cdots & \frac{\partial q_2}{\partial q(X_{i+1},Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_i}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_i}{\partial q(X_{i-1},Y_j)} & \cdots & 0 \\
\frac{\partial q_{i+1}}{\partial q(X_1,Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i-1},Y_j)} & \cdots & 0 \\
\end{pmatrix} e^{-q(X_i,Y_j)} |B_{i-1}| > 0 \quad \text{Since } |B_{i-1}| > 0.
$$

Step 2 Assume that $|A_{i+n}| < 0$ where **WLOG** $n$ is some even number such that $i + n$ is odd.
Step 3 Show that $|A_{i+n+1}| > 0$.

$$|A_{i+n+1}| = \det \begin{pmatrix} \frac{\partial q_1}{\partial q(Y_i)} & \cdots & \frac{\partial q_1}{\partial q(Y_{i-1})} & \frac{\partial q_1}{\partial q(Y_{i+\gamma})} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial q_{i+n}}{\partial q(Y_i)} & \cdots & \frac{\partial q_{i+n}}{\partial q(Y_{i-1})} & \frac{\partial q_{i+n}}{\partial q(Y_{i+\gamma})} \\ \frac{\partial q_{i+n+1}}{\partial q(Y_i)} & \cdots & \frac{\partial q_{i+n+1}}{\partial q(Y_{i-1})} & \frac{\partial q_{i+n+1}}{\partial q(Y_{i+\gamma})} \end{pmatrix} = \det \begin{pmatrix} \frac{\partial q_1}{\partial q(Y_i)} & \cdots & 0 & \cdots & \frac{\partial q_1}{\partial q(Y_{i+n+1})} & \frac{\partial q_1}{\partial q(Y_{i+\gamma})} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial q_{i+n}}{\partial q(Y_i)} & \cdots & \frac{\partial q_{i+n}}{\partial q(Y_{i-1})} & (1-e^{-q(Y_i,Y_j)})e^{-Q(Y_i,Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial q(Y_{i+\gamma})} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial q_{i+n+1}}{\partial q(Y_i)} & \cdots & \frac{\partial q_{i+n+1}}{\partial q(Y_{i-1})} & (1-e^{-q(Y_i,Y_j)})e^{-Q(Y_i,Y_j)} & \cdots & \frac{\partial q_{i+n+1}}{\partial q(Y_{i+\gamma})} \end{pmatrix} = -e^{Q(Y_{i+n+1},Y_j)}[(\Phi(Y_{i+n+1},Y_j) - \Phi(Y_{i+n},Y_j)) |A_{i+n}| < 0 \quad \forall i > 0 \quad \text{after some manipulation}$$

Hence, by MI, it is shown that $A_I > 0$ for $I$ odd and $A_I < 0$ for $I$ even. Therefore, it can be concluded that

$$\frac{\partial q(Y_i,Y_j)}{\partial \Phi(Y_i,Y_j)} = -\det \begin{pmatrix} \frac{\partial q_1}{\partial \Phi(Y_i,Y_j)} & \cdots & \frac{\partial q_1}{\partial \Phi(Y_{i-1},Y_j)} \\ \vdots & \ddots & \vdots \\ \frac{\partial q_{i+n}}{\partial \Phi(Y_i,Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial \Phi(Y_{i-1},Y_j)} \\ \frac{\partial q_{i+n+1}}{\partial \Phi(Y_i,Y_j)} & \cdots & \frac{\partial q_{i+n+1}}{\partial \Phi(Y_{i-1},Y_j)} \end{pmatrix} > 0 \quad \forall i \quad \text{and } I$$

This is because the numerator and denominator always take opposite signs which, pre-multiplied by a minus, gives a positive number. Otherwise put, the higher the productivity of worker $X_i$ at firm $Y_j$, the higher the expected equilibrium queue length of this worker type at this firm.

Now I show that $\frac{\partial q(Y_i,Y_j)}{\partial \Phi(Y_{i'},Y_j)} < 0$, $\forall |i' - i| < 2$ and 0 otherwise as well as $|\frac{\partial q(Y_i,Y_j)}{\partial \Phi(Y_{i'},Y_j)}| > |\frac{\partial q(Y_i,Y_j)}{\partial \Phi(Y_{i'},Y_j)}|$, $\forall i' \neq i$.

First, I prove that $\frac{\partial q(Y_i,Y_j)}{\partial \Phi(Y_{i-1},Y_j)} < 0$, using MI. WLOG, assume that $i$ is odd. From the previous working, note that the expression takes the form
Recall that the denominator is positive when \( i \) is even and negative when \( i \) is odd. What remains unknown is the sign of the numerator.

**Step 1** Consider the case where \( i \) the best worker type that firm \( Y_j \) attracts with positive probability. The numerator then writes as

\[
|A_i| = \det \begin{bmatrix}
\frac{\partial g_i}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & \frac{\partial g_i}{\partial \Phi(X_{i-1}, Y_j)} \\
\vdots & \ddots & \vdots & \vdots \\
\frac{\partial g_i}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & \frac{\partial g_i}{\partial \Phi(X_{i-1}, Y_j)} \\
\frac{\partial g_i}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & \frac{\partial g_i}{\partial \Phi(X_{i-1}, Y_j)}
\end{bmatrix} = \det \begin{bmatrix}
\frac{\partial g_i}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & 0 \\
\frac{\partial g_i}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & e^{-q(X_{i-1}, Y_j)}e^{-Q(X_{i-1}, Y_j)} \\
\frac{\partial g_i}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & -(1 - e^{-q(X_{i-1}, Y_j)})e^{-Q(X_{i-1}, Y_j)} \\
\frac{\partial g_i}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & e^{-Q(X_{i-1}, Y_j)}
\end{bmatrix}
\]

\[
= -e^{-Q(X_{i-1}, Y_j)}B_{i-1} < 0 \quad \text{after some manipulation}
\]

Now consider the sign of \( |A_{i+1}| \), which is the case where the best worker firm \( Y_j \) attracts with positive probability is the worker ranked just above worker \( i \).

\[
|A_{i+1}| = \det \begin{bmatrix}
\frac{\partial g_{i+1}}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_{i+1}}{\partial q(X_{i-1}, Y_j)} & \frac{\partial g_{i+1}}{\partial \Phi(X_{i-1}, Y_j)} \\
\vdots & \ddots & \vdots & \vdots \\
\frac{\partial g_{i+1}}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_{i+1}}{\partial q(X_{i-1}, Y_j)} & \frac{\partial g_{i+1}}{\partial \Phi(X_{i-1}, Y_j)} \\
\frac{\partial g_{i+1}}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_{i+1}}{\partial q(X_{i-1}, Y_j)} & \frac{\partial g_{i+1}}{\partial \Phi(X_{i-1}, Y_j)}
\end{bmatrix} = \det \begin{bmatrix}
\frac{\partial g_{i+1}}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_{i+1}}{\partial q(X_{i-1}, Y_j)} & 0 \\
\frac{\partial g_{i+1}}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_{i+1}}{\partial q(X_{i-1}, Y_j)} & e^{-q(X_{i-1}, Y_j)}e^{-Q(X_{i-1}, Y_j)} \\
\frac{\partial g_{i+1}}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_{i+1}}{\partial q(X_{i-1}, Y_j)} & -(1 - e^{-q(X_{i-1}, Y_j)})e^{-Q(X_{i-1}, Y_j)} \\
\frac{\partial g_{i+1}}{\partial q(X_i, Y_j)} & \cdots & \frac{\partial g_{i+1}}{\partial q(X_{i-1}, Y_j)} & e^{-Q(X_{i-1}, Y_j)}
\end{bmatrix}
\]

\[
= -e^{-Q(X_{i-1}, Y_j)}e^{-q(X_{i+1})}[\Phi(X_{i+1}, Y_j) - \Phi(X_i, Y_j)]B_{i-1} > 0 \quad \text{after some manipulation}
\]
Step Two  WLOG, let \( n \) be an even number. \( i + n \) is thus odd. Assume for some \( i + n \) that \(|A_{i+n}| < 0\).

Step Three  Prove that \(|A_{i+n+1}| > 0\).

\[
|A_{i+n+1}| = \det \begin{bmatrix}
\frac{\partial q_1}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_1}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_1}{\partial q(X_{i-1}, Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial q_i}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial q_i}{\partial g(X_{i-1}, Y_1)} & \frac{\partial q_i}{\partial g(X_{i-1}, Y_j)} & \cdots & \frac{\partial q_i}{\partial g(X_{i+n+1}, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial q_{i+n}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_{i+n}}{\partial q(X_{i-1}, Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial g_{i+n}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i+n}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i+n}}{\partial g(X_{i-1}, Y_j)} & \cdots & \frac{\partial g_{i+n}}{\partial g(X_{i+n+1}, Y_j)} \\
\frac{\partial g_{i+n+1}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial g_{i+n+1}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial g_{i+n+1}}{\partial q(X_{i-1}, Y_j)} & \cdots & \frac{\partial g_{i+n+1}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial g_{i+n+1}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i+n+1}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i+n+1}}{\partial g(X_{i-1}, Y_j)} & \cdots & \frac{\partial g_{i+n+1}}{\partial g(X_{i+n+1}, Y_j)} \\
\end{bmatrix}
\]

\[
= \det \begin{bmatrix}
\frac{\partial q_1}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_1}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_1}{\partial q(X_{i-1}, Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial g_i}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_1)} & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & \cdots & \frac{\partial g_i}{\partial q(X_{i+n+1}, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial q_{i+n}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_{i+n}}{\partial q(X_{i-1}, Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial g_{i+n}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i+n}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i+n}}{\partial g(X_{i-1}, Y_j)} & \cdots & \frac{\partial g_{i+n}}{\partial g(X_{i+n+1}, Y_j)} \\
\frac{\partial g_{i+n+1}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial g_{i+n+1}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial g_{i+n+1}}{\partial q(X_{i-1}, Y_j)} & \cdots & \frac{\partial g_{i+n+1}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial g_{i+n+1}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i+n+1}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i+n+1}}{\partial g(X_{i-1}, Y_j)} & \cdots & \frac{\partial g_{i+n+1}}{\partial g(X_{i+n+1}, Y_j)} \\
\end{bmatrix}
\]

\[
= -|A_{i+n}| e^{-Q(X_{i+n+1}, Y_j)} [\Phi(X_{i+n+1}, Y_j) - \Phi(X_{i+n}, Y_j)] > 0 \quad \text{after some manipulation}
\]

Therefore, as shown by MI, the \(|A_i|\) is negative for \(I\) odd and is positive for \(I\) even. Since the denominator \(|B_i|\) is also negative for \(I\) odd and positive for \(I\) even, one can conclude that \(\frac{\partial q(X_i, Y_i)}{\partial q(X_{i-k}, Y_j)} < 0, \forall i \text{ and } I\).

Next, I show that \(\frac{\partial q(X_i, Y_i)}{\partial q(X_{i-k}, Y_j)} = 0, \forall k > 1\).

Step 1  Let \(k\) be any arbitrary number greater than \(1\). Let us first consider the case where \(i\) is the most preferred worker whom firm \(Y_j\) attracts with some positive probability. WLOG, let \(i\) be odd.

\[
|A_i| = \det \begin{bmatrix}
\frac{\partial q_1}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_1}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_1}{\partial q(X_{i-1}, Y_j)} \\
\vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_i}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial q_i}{\partial g(X_{i-1}, Y_1)} & \frac{\partial q_i}{\partial g(X_{i-1}, Y_j)} \\
\frac{\partial q_{i-k}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_{i-k}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_{i-k}}{\partial q(X_{i-1}, Y_j)} \\
\frac{\partial g_{i-k}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i-k}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i-k}}{\partial g(X_{i-1}, Y_j)} \\
\frac{\partial q_{i-k-1}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_{i-k-1}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_{i-k-1}}{\partial q(X_{i-1}, Y_j)} \\
\frac{\partial g_{i-k-1}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i-k-1}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i-k-1}}{\partial g(X_{i-1}, Y_j)} \\
\end{bmatrix}
\]

\[
= \det \begin{bmatrix}
\frac{\partial q_1}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_1}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_1}{\partial q(X_{i-1}, Y_j)} & 0 \\
\frac{\partial g_i}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial g_i}{\partial q(X_{i-1}, Y_1)} & \frac{\partial g_i}{\partial q(X_{i-1}, Y_j)} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\frac{\partial q_{i-k}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_{i-k}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_{i-k}}{\partial q(X_{i-1}, Y_j)} & \cdots \\
\frac{\partial g_{i-k}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i-k}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i-k}}{\partial g(X_{i-1}, Y_j)} & \cdots \\
\frac{\partial q_{i-k-1}}{\partial q(X_1, Y_1)} & \cdots & \frac{\partial q_{i-k-1}}{\partial q(X_{i-1}, Y_1)} & \frac{\partial q_{i-k-1}}{\partial q(X_{i-1}, Y_j)} & \cdots \\
\frac{\partial g_{i-k-1}}{\partial g(X_1, Y_1)} & \cdots & \frac{\partial g_{i-k-1}}{\partial g(X_{i-1}, Y_1)} & \frac{\partial g_{i-k-1}}{\partial g(X_{i-1}, Y_j)} & \cdots \\
\end{bmatrix}
\]

\[
= 0 \quad \text{Since the last two rows of the matrix are collinear.}
\]

Step 2  Assume that \(|A_{i+n}| = 0\) where WLOG \(n\) is some even number such that \(i + n\) is odd.
Step 3  Show that $|A_{i+n+1}| = 0$.

$$
|A_{i+n+1}| = \det \begin{pmatrix}
\frac{\partial q_1}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+n+1}, Y_j)} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial q_{i+n}}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial q_{i+1}}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial q_{i+1}}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+n+1}, Y_j)}
\end{pmatrix}

= -e^{Q(X_{i+n+1}, Y_j)} [\Phi(X_{i+n+1}, Y_j) - \Phi(X_{i+n+1}, Y_j)] \cdot |A_{i+n}| = 0

<0

As such, by MI, it has been proven that $\frac{\partial q(X_i, Y_j)}{\partial q(X_{i+n}, Y_j)} = 0, \forall k > 1$.

Likewise, it can be shown that $\frac{\partial q(X_i, Y_j)}{\partial q(X_{i+n+1}, Y_j)} < 0$ and $\frac{\partial q(X_i, Y_j)}{\partial q(X_{i+k}, Y_j)} = 0, \forall k > 1$. Let us start with proving the former.

Step 1  WLOG let $i$ be odd such that $i + 1$ is even. Consider the case where $i + 1$ is the most preferred type that firm $Y_j$ attracts.

$$
|A_{i+1}| = \det \begin{pmatrix}
\frac{\partial q_1}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_1}{\partial q(X_{i+n+1}, Y_j)} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial q_{i+n}}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_{i+n}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial q_{i+1}}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+n+1}, Y_j)} \\
\frac{\partial q_{i+1}}{\partial q(X_1, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+1}, Y_j)} & \cdots & \frac{\partial q_{i+1}}{\partial q(X_{i+n+1}, Y_j)}
\end{pmatrix}

= e^{-q(X_{i+1}, Y_j)} e^{-Q(X_{i+1}, Y_j)} \cdot |B_i| > 0

<0

Next, consider the case when the most preferred type the firm attracts is $i + 2$, where $i + 2$ is odd. The numerator can be expressed as
Step 2 Assume that $|A_{i+n}| > 0$ where WLOG $n$ is some odd number such that $i + n$ is even.

Step 3 Show that $|A_{i+n+1}| < 0$.

Thus by MI we know that the numerator is positive when $I$ is even negative when $I$ is odd. Hence, $\frac{\partial g(Y_i)}{\partial g(X_{i+n+1}, Y_j)} = -\frac{|A_i|}{|B_i|} < 0 \forall I$ since $|B_i|$ is negative when $I$ is odd and positive when $I$ is even.

I then prove that $\frac{\partial g(X_{i+k})}{\partial g(X_{i+k}, Y_j)} = 0, \forall k > 1$.

Step 1 WLOG let $k$ be odd such that $k + i$ is even. Consider the case where the $X_{i+k}$ is the most preferred type that firm $Y_j$ attracts with positive probability. The numerator of $\frac{\partial g(X_{i+k})}{\partial g(X_{i+k}, Y_j)}$ can be written as
\[ |A_{i+k}| = \det \begin{pmatrix}
\frac{\partial q_1}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_1}{\partial q(X_{i-1},Y_j)} & \frac{\partial q_1}{\partial q(X_{i+k},Y_j)} & \ldots & \frac{\partial q_1}{\partial q(X_i,Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_{i+k}}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_{i+k}}{\partial q(X_{i-1},Y_j)} & \frac{\partial q_{i+k}}{\partial q(X_{i+k},Y_j)} & \ldots & \frac{\partial q_{i+k}}{\partial q(X_i,Y_j)} \\
\end{pmatrix} \]

\[ = \det \begin{pmatrix}
\frac{\partial q_1}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_1}{\partial q(X_{i-1},Y_j)} & 0 & \ldots & \frac{\partial q_1}{\partial q(X_{i+k},Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_{i+k-1}}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_{i+k-1}}{\partial q(X_{i-1},Y_j)} & 0 & \ldots & \frac{\partial q_{i+k-1}}{\partial q(X_{i+k},Y_j)} \\
\frac{\partial q_{i+k}}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_{i+k}}{\partial q(X_{i-1},Y_j)} & e^{-q(X_{i+k},Y_j)}e^{-Q(X_{i+k},Y_j)} & \ldots & \frac{\partial q_{i+k}}{\partial q(X_i,Y_j)} \\
\end{pmatrix} \]

\[ = e^{-q(X_{i+k},Y_j)}e^{-Q(X_{i+k},Y_j)} \]

Step 2 Assume that \( |A_{i+k+n}| = 0 \) where \( WLOG \) \( n \) is some even number such that \( i + k + n \) is odd.

Step 3 Show that \( |A_{i+k+n+1}| = 0 \).

\[ |A_{i+k+n+1}| = \det \begin{pmatrix}
\frac{\partial q_1}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_1}{\partial q(X_{i-1},Y_j)} & \frac{\partial q_1}{\partial q(X_{i+k+n+1},Y_j)} & \ldots & \frac{\partial q_1}{\partial q(X_i,Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_{i+k+n}}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_{i+k+n}}{\partial q(X_{i-1},Y_j)} & \frac{\partial q_{i+k+n}}{\partial q(X_{i+k+n+1},Y_j)} & \ldots & \frac{\partial q_{i+k+n}}{\partial q(X_i,Y_j)} \\
\frac{\partial q_{i+k+n+1}}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_{i+k+n+1}}{\partial q(X_{i-1},Y_j)} & \frac{\partial q_{i+k+n+1}}{\partial q(X_{i+k+n+1},Y_j)} & \ldots & \frac{\partial q_{i+k+n+1}}{\partial q(X_i,Y_j)} \\
\frac{\partial q_{i+k+n+2}}{\partial q(X_1,Y_j)} & \ldots & \frac{\partial q_{i+k+n+2}}{\partial q(X_{i-1},Y_j)} & \frac{\partial q_{i+k+n+2}}{\partial q(X_{i+k+n+1},Y_j)} & \ldots & \frac{\partial q_{i+k+n+2}}{\partial q(X_i,Y_j)} \\
\end{pmatrix} \]

\[ = -e^{-Q(X_{i+k+n+1},Y_j)}[\Phi(X_{i+k+n+1},Y_j) - \Phi(X_{i+k+n+1},Y_j)] \cdot |A_{i+k+n}| = 0 \]

Hence, by MI, it has been proven that \( \frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)} = 0, \forall k > 1 \).

Lastly, I show that \( |\frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)}| \geq |\frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)}| \forall i \neq i' \). Before proceeding, recall that we have shown that \( \frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)} = 0 \) for \( k > 1 \). As such, let us focus on showing that \( |\frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)}| \geq |\frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)}| \) and that \( |\frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)}| \geq |\frac{\partial q(X_i,Y_j)}{\partial q(X_{i+k},Y_j)}| \cdot \). As usual, the proof is done using MI.

Step 1 WLOG, let \( i \) be odd. Consider the case where the most preferred worker a firm type gets is \( i + 1 \). Then \( \frac{\partial q(X_i,Y_j)}{\partial q(X_{i+1},Y_j)} = -\frac{|A_{i+1}|}{|B_{i+1}|} \) and \( \frac{\partial q(X_i,Y_j)}{\partial q(X_{i+1},Y_j)} = -\frac{|A_{i+1}|}{|B_{i+1}|} \), where
Some linear algebra reveals that that \( |A^0_{i+1}| > |A^1_{i+1}| \). Therefore, \( \frac{\partial q(X_i, Y_j)}{\partial q(X_i, Y_j)} > \frac{\partial q(X_i, Y_j)}{\partial q(X_i, Y_j)} \) holds.

**Step 2** Assume that the statement holds if the best ranked worker the firm attracts with some probability is worker \( i + n \), meaning that \( |A^0_{i+n}| > |A^1_{i+n}| \), where WLOG \( n \) is even such that \( i + n \) is odd.

**Step 3** Show that the statement holds should the firm attract up to worker \( i + n + 1 \) with positive probability.

\[
|A^0_{i+n+1}| = det \begin{pmatrix} \frac{\partial q_1}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_1}{\partial q(X_i, Y_j)} & \frac{\partial q_1}{\partial q(X_i, Y_j)} & \frac{\partial q_1}{\partial q(X_i, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_i}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_i}{\partial q(X_i, Y_j)} & \frac{\partial q_i}{\partial q(X_i, Y_j)} & \frac{\partial q_i}{\partial q(X_i, Y_j)} \\
\frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} \\
\frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} \\
\end{pmatrix}
\]

\[
= -e^{Q(X_{i+n}, Y_j)} [\Phi(X_{i+n+1}, Y_j) - \Phi(X_{i+n}, Y_j)] |A^0_{i+n}|
\]

\[
|A^1_{i+n+1}| = det \begin{pmatrix} \frac{\partial q_1}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_1}{\partial q(X_i, Y_j)} & \frac{\partial q_1}{\partial q(X_i, Y_j)} & \frac{\partial q_1}{\partial q(X_i, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial q_i}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_i}{\partial q(X_i, Y_j)} & \frac{\partial q_i}{\partial q(X_i, Y_j)} & \frac{\partial q_i}{\partial q(X_i, Y_j)} \\
\frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n}}{\partial q(X_i, Y_j)} \\
\frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} & \ldots & \frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} & \frac{\partial q_{i+n+1}}{\partial q(X_i, Y_j)} \\
\end{pmatrix}
\]

\[
= e^{Q(X_{i+n}, Y_j)} [\Phi(X_{i+n+1}, Y_j) - \Phi(X_{i+n}, Y_j)] |A^1_{i+n}|
\]

Since by assumption \( |A^0_{i+n}| > |A^1_{i+n}| \), we have that \( |A^0_{i+n+1}| > |A^1_{i+n+1}| \). As such, by MI, we have shown that \( \frac{\partial q(X_i, Y_j)}{\partial q(X_i, Y_j)} > \frac{\partial q(X_i, Y_j)}{\partial q(X_i, Y_j)} \) holds regardless of the number of types the firm attracts with positive probability.

Likewise, using the same method, we can show that \( \frac{\partial q(X_i, Y_j)}{\partial q(X_i, Y_j)} \geq \frac{\partial q(X_i, Y_j)}{\partial q(X_i, Y_j)} \).
Step 1  WLOG, let $i$ be odd. Consider the case where the most preferred worker a firm type gets is $i$. Then \( \frac{\partial q_i(X_j, Y_j)}{\partial \Phi(X_j, Y_j)} = -|A_i^0| \) and \( \frac{\partial q_i(X_{i+n}, Y_j)}{\partial \Phi(X_{i+n}, Y_j)} = -|A_i^1| \), where

\[
|A_i^0| = \text{det} \begin{pmatrix}
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_1}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n}, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n}, Y_j)} \\
\end{pmatrix}
\]

and

\[
|A_i^1| = \text{det} \begin{pmatrix}
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_1}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n}, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n}, Y_j)} \\
\end{pmatrix}
\]

Some linear algebra reveals that that \( |A_i^0| = |A_i^1| = e^{-q_i(X_j, Y_j)}|B_{i-1}| \). Therefore, \( \frac{\partial q_i(X_j, Y_j)}{\partial \Phi(X_j, Y_j)} = \frac{|\partial q_i(X_j, Y_j)|}{|\partial \Phi(X_j, Y_j)|} \) holds in the first step.

Step 2  Assume that the statement holds if the best ranked worker the firm attracts with some probability is worker \( i + n \), meaning that \( |A_{i+n}^0| = |A_{i+n}^1| \), where WLOG \( n \) is odd such that \( i + n \) is even.

Step 3  Show that the statement holds should the firm attract up to worker \( i + n + 1 \) with positive probability.

\[
|A_{i+n+1}^0| = \text{det} \begin{pmatrix}
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_1}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n+1}, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n+1}, Y_j)} \\
\end{pmatrix}
\]

\[
= -e^Q(X_{i+n+1}, Y_j):[\Phi(X_{i+n+1}, Y_j) - \Phi(X_{i+n}, Y_j)]|A_{i+n}^0|
\]

\[
|A_{i+n+1}^1| = \text{det} \begin{pmatrix}
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_1}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n+1}, Y_j)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial q_1}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i-1}, Y_j)} & \frac{\partial q_i}{\partial \Phi(X_i, Y_j)} & \cdots & \frac{\partial q_i}{\partial \Phi(X_{i+n+1}, Y_j)} \\
\end{pmatrix}
\]

\[
= e^Q(X_{i+n+1}, Y_j):[\Phi(X_{i+n+1}, Y_j) - \Phi(X_{i+n}, Y_j)]|A_{i+n}^1|
\]

Since by assumption \( |A_{i+n}^0| = |A_{i+n}^1| \), we have that \( |A_{i+n+1}^0| = |A_{i+n+1}^1| \). As such, \( \frac{\partial q_i(X_j, Y_j)}{\partial \Phi(X_j, Y_j)} = \frac{|\partial q_i(X_j, Y_j)|}{|\partial \Phi(X_j, Y_j)|} \).
Therefore, since \( \partial q(X_i, Y_j) / \Phi(X_{i-1}, Y_j) = 0 \) for \( |k| > 1 \), we can thus conclude that \( |\partial q(X_i, Y_j) / \Phi(X_{i-1}, Y_j)| \geq |\partial q(X_i, Y_j) / \Phi(X_{i+1}, Y_j)| \) \( \forall k > 0 \). To conclude then, combining the previous results, \( |\partial q(X_i, Y_j) / \Phi(X_{i-1}, Y_j)| \geq |\partial q(X_i, Y_j) / \Phi(X_{i+1}, Y_j)| \) \( \forall i \neq i \).

Now assume that a given firm \( Y_j \) attracts \( N \) types of workers, which it ranks according to index \( n = 1, ..., N \), where \( n = 1 \) is the least preferred worker at firm \( Y \) and \( n = N \) is the most preferred worker. From the Theorem 2, the Jacobian \( \partial q / \Phi \) can be written as follows, with all diagonal entries positive and larger than the absolute of the sum of all off-diagonal entries.

\[
\frac{\partial q}{\partial \Phi} = \begin{pmatrix}
\frac{\partial q(X_1, Y_j)}{\Phi(X_1, Y_j)} & \frac{\partial q(X_2, Y_j)}{\Phi(X_2, Y_j)} & \cdots & \frac{\partial q(X_n, Y_j)}{\Phi(X_n, Y_j)} \\
\frac{\partial q(X_2, Y_j)}{\Phi(X_2, Y_j)} & \frac{\partial q(X_3, Y_j)}{\Phi(X_3, Y_j)} & \cdots & \frac{\partial q(X_n, Y_j)}{\Phi(X_n, Y_j)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial q(X_{n-1}, Y_j)}{\Phi(X_{n-1}, Y_j)} & \frac{\partial q(X_n, Y_j)}{\Phi(X_n, Y_j)} & \cdots & \frac{\partial q(X_n, Y_j)}{\Phi(X_n, Y_j)}
\end{pmatrix}
\]

Substituting in the 0 terms,

\[
\frac{\partial q}{\partial \Phi} = \begin{pmatrix}
\frac{\partial q(X_1, Y_j)}{\Phi(X_1, Y_j)} & \frac{\partial q(X_2, Y_j)}{\Phi(X_2, Y_j)} & 0 & 0 & \cdots & \cdots & 0 \\
\frac{\partial q(X_2, Y_j)}{\Phi(X_2, Y_j)} & \frac{\partial q(X_3, Y_j)}{\Phi(X_3, Y_j)} & 0 & \cdots & \cdots & 0 \\
0 & \frac{\partial q(X_3, Y_j)}{\Phi(X_3, Y_j)} & \frac{\partial q(X_4, Y_j)}{\Phi(X_4, Y_j)} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & \frac{\partial q(X_{n-2}, Y_j)}{\Phi(X_{n-2}, Y_j)} & \frac{\partial q(X_{n-1}, Y_j)}{\Phi(X_{n-1}, Y_j)} \\
0 & \cdots & \cdots & \cdots & 0 & \frac{\partial q(X_n, Y_j)}{\Phi(X_n, Y_j)}
\end{pmatrix}
\]

Where the diagonal terms (in bold) are strictly positive and larger in absolute terms than the sum of the off-diagonals in the first and last row, and are equal in absolute terms to the sum of the off-diagonals in the intermediate rows. The off-diagonal non-zero terms are all strictly negative.

**Proof of Corollary 1**

The probability that firm type \( Y_j \) is matched to worker type \( X_i \) is \((1 - e^{-q(X_i, Y_j)})e^{-Q(X_i, Y_j)}\).

To show that corollary 1 is true, observe that

\[
\frac{\partial(1 - e^{-q(X_i, Y_j)})e^{-Q(X_i, Y_j)}}{\partial \Phi(X_i, Y_j)} = \frac{\partial(1 - e^{-q(X_i, Y_j)})}{\partial \Phi(X_i, Y_j)} \frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)} e^{-Q(X_i, Y_j)} + (1 - e^{-q(X_i, Y_j)}) \frac{\partial e^{-Q(X_i, Y_j)}}{\partial \Phi(X_i, Y_j)} \frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)}
\]

\[
= e^{-q(X_i, Y_j)} \frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)} e^{-Q(X_i, Y_j)} - (1 - e^{-q(X_i, Y_j)}) e^{-Q(X_i, Y_j)} \frac{\partial q(X_i, Y_j)}{\partial \Phi(X_i, Y_j)}
\]

\[
> 0
\]

70
Next, consider any worker $X_{i'}$ who is ranked better than worker $X_i$ at firm $Y_j$.

\[
\frac{\partial}{\partial \Phi(X_i, Y_j)} \left( 1 - e^{-q(X_{i'}, Y_j)} \right) e^{-Q(X_{i'}, Y_j)} = \frac{\partial}{\partial q(X_{i'}, Y_j)} q(X_{i'}, Y_j) \frac{\partial}{\partial \Phi(X_i, Y_j)} e^{-Q(X_{i'}, Y_j)}
\]

Using the result of Theorem 2, one can note that the above derivative is 0 when for workers ranked more than one rank above $X_i$ and is only strictly negative for the worker ranked just above $X_i$ at the firm $Y_j$.

Lastly, consider any worker $X_{i''}$ who is ranked worse than worker $X_i$ at firm $Y_j$.

\[
\frac{\partial}{\partial \Phi(X_i, Y_j)} \left( 1 - e^{-q(X_{i''}, Y_j)} \right) e^{-Q(X_{i''}, Y_j)} = \frac{\partial}{\partial q(X_{i''}, Y_j)} q(X_{i''}, Y_j) \frac{\partial}{\partial \Phi(X_i, Y_j)} e^{-Q(X_{i''}, Y_j)}
\]

\[
\leq 0
\]

**Proof of Theorem 3**

$\frac{\partial \mu(X, Y)}{\partial \Phi(X, Y)} > 0$ and $\frac{\partial \mu(X, Y)}{\partial \Phi(X, Y)} \leq 0$ have already been proven above. Simple algebra reveals that $\frac{\partial \mu(X, Y)}{\partial \Phi(X, Y)} > 0$. The Jacobian matrix $\frac{\partial \mu}{\partial \Phi}$ for a firm $Y_j$ who has ranked potential workers in descending order of preference using the index $n = 1, \ldots, N$ can be written as

\[
\frac{\partial \mu}{\partial \Phi} = \begin{pmatrix}
\frac{\partial \mu(X_1, Y_j)}{\partial \Phi(X_1, Y_j)} & \frac{\partial \mu(X_2, Y_j)}{\partial \Phi(X_1, Y_j)} & \cdots & \frac{\partial \mu(X_N, Y_j)}{\partial \Phi(X_1, Y_j)} \\
\frac{\partial \mu(X_1, Y_j)}{\partial \Phi(X_2, Y_j)} & \frac{\partial \mu(X_2, Y_j)}{\partial \Phi(X_2, Y_j)} & \cdots & \frac{\partial \mu(X_N, Y_j)}{\partial \Phi(X_2, Y_j)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial \mu(X_1, Y_j)}{\partial \Phi(X_N, Y_j)} & \frac{\partial \mu(X_2, Y_j)}{\partial \Phi(X_N, Y_j)} & \cdots & \frac{\partial \mu(X_N, Y_j)}{\partial \Phi(X_N, Y_j)}
\end{pmatrix}
\]
Hence the probability of a firm being unmatched is equivalent to the probability that no worker type applies. The Jacobian except for a firm is left vacant and letting workers be ranked in index $j$, match probabilities must sum to 1. Otherwise put, denoting $\mu(0, Y_j)$ as the probability that a firm is left vacant and letting workers be ranked in index $n = 1, ..., N$, the following must hold:

$$\sum_{n=1}^{N} \mu(X_n, Y_j) + \mu(0, Y_j) = 1$$

$$\leftrightarrow \mu(X_n, Y_j) = 1 - \mu(0, Y_j) - \sum_{n' \neq n} \mu(X_{n'}, Y_j)$$

$$\frac{\partial \mu(X_n, Y_j)}{\partial \Phi(X_n, Y_j)} = \frac{\partial \mu(0, Y_j)}{\partial \Phi(X_n, Y_j)} - \sum_{n' \neq n} \frac{\partial \mu(X_{n'}, Y_j)}{\partial \Phi(X_n, Y_j)}$$

Since we have already established that $\frac{\partial \mu(X_{n'}, Y_j)}{\partial \Phi(X_n, Y_j)} < 0 \forall n' \neq n$, $\frac{\partial \mu(X_{n}, Y_j)}{\partial \Phi(X_n, Y_j)} > \frac{\partial \mu(X_{n'}, Y_j)}{\partial \Phi(X_n, Y_j)}$ only if $\frac{\partial \mu(0, Y_j)}{\partial \Phi(X_n, Y_j)} \leq 0$. Denote $Q(0, Y_j)$ as the total expected queue length of all applicants that $Y_j$ attracts with non-zero probability. I.e. $Q(0, Y_j) = q(X_1, Y_j) + ... + q(X_n, Y_j) + ... + q(X_N, Y_j)$. The probability of a firm being unmatched is equivalent to the probability that no worker type applies. Hence $\mu(0, Y_j) = e^{-Q(0, Y_j)}$. Taking the derivative with respect to $\Phi(X_n, Y_j)$ and using the results from Theorem 2 gives

$$\frac{\partial \mu(0, Y_j)}{\partial \Phi(X_n, Y_j)} = \frac{\partial e^{-Q(0, Y_j)}}{\partial \Phi(X_n, Y_j)}$$

$$= \frac{\partial e^{-Q(0, Y_j)}}{\partial \Phi(X_n, Y_j)} e^{-\sum_{n' \neq n} q(X_{n'}, Y_j)} + \frac{\partial e^{-Q(0, Y_j)}}{\partial \Phi(X_{n-1}, Y_j)} e^{-\sum_{n' \neq n-1} q(X_{n'}, Y_j)} + \frac{\partial e^{-Q(0, Y_j)}}{\partial \Phi(X_{n+1}, Y_j)} e^{-\sum_{n' \neq n+1} q(X_{n'}, Y_j)}$$

$$= -e^{-Q(0, Y_j)} \left( \frac{\partial q(X_n, Y_j)}{\partial \Phi(X_n, Y_j)} + \frac{\partial q(X_{n-1}, Y_j)}{\partial \Phi(X_{n-1}, Y_j)} + \frac{\partial q(X_{n+1}, Y_j)}{\partial \Phi(X_{n+1}, Y_j)} \right)$$

From Theorem 2, we know that the term in the large curve brackets is equal to 0 for all $n$ except for $n = 1$ and $n = N$, when it is strictly positive. As such, $\frac{\partial \mu(0, Y_j)}{\partial \Phi(X_n, Y_j)} \leq 0$, $\frac{\partial \mu(X_n, Y_j)}{\partial \Phi(X_n, Y_j)} \geq \left| \frac{\partial \sum_{n' \neq n} \mu(X_{n'}, Y_j)}{\partial \Phi(X_n, Y_j)} \right|$ for $2 \geq n \leq N - 1$ and $\frac{\partial \mu(X_n, Y_j)}{\partial \Phi(X_n, Y_j)} \geq \left| \frac{\partial \sum_{n' \neq n} \mu(X_{n'}, Y_j)}{\partial \Phi(X_n, Y_j)} \right|$ for $n = 1$ and $n = N$. Thus, the Jacobian $\frac{\partial \mu}{\partial \Phi}$ is an irreducible diagonally dominant matrix.

72
Proof of Theorem 4

For ease of notation, re-express $\Phi(x, y)$ as $\Phi_{xy}$ and $\mu(x, y)$ as $\mu_{xy}$. Assume that $\Phi(x, y)$ is supermodular, such that $\Phi_{xy} + \Phi_{x'y} > \Phi_{xy'} + \Phi_{x'y'}$. For log supermodularity of the matching function to hold, we require $\frac{\partial \mu_{xy}}{\Phi_{xy} + \Phi_{x'y}} > 0$. In words, as the difference in productivity of $x$ at $y$ increases relative to that of $x'$ at $y$, the ratio of assignment probability of $x$ to $y$ relative to that of $x'$ to $y$ should increase.

Let $s = \Phi_{xy} - \Phi_{x'y}$ and let $t = \Phi_{xy} + \Phi_{x'y}$. As such $\Phi_{xy} = (s + t)/2$ and $\Phi_{x'y} = m(s, t) = (t - 2)/2$. Let $F(\mu_{xy}(\Phi_{xy}, \Phi_{x'y}), \mu_{x'y}(\Phi_{xy}, \Phi_{x'y})) = \frac{\mu_{xy}}{\mu_{x'y}}$. Observe that

$$F(\mu_{xy}(\Phi_{xy}, \Phi_{x'y}), \mu_{x'y}(\Phi_{xy}, \Phi_{x'y})) = F(\mu_{xy}(h(s, t), m(s, t)), \mu_{x'y}(h(s, t), m(s, t)))$$

Next, via the chain rule

$$\frac{\partial F}{\partial s} = \frac{\partial F}{\partial \mu_{xy}} \frac{\partial \mu_{xy}}{\partial h} \frac{\partial h}{\partial s} + \frac{\partial F}{\partial \mu_{x'y}} \frac{\partial \mu_{x'y}}{\partial h} \frac{\partial h}{\partial s} + \frac{\partial F}{\partial m} \frac{\partial m}{\partial \mu_{xy}} \frac{\partial \mu_{xy}}{\partial h} \frac{\partial h}{\partial s} + \frac{\partial F}{\partial m} \frac{\partial m}{\partial \mu_{x'y}} \frac{\partial \mu_{x'y}}{\partial h} \frac{\partial h}{\partial s}$$

$$= \frac{1}{\mu_{x'y}} \frac{\partial \mu_{xy}}{\partial h} \frac{1}{2} - \frac{1}{\mu_{x'y}} \frac{\partial \mu_{xy}}{\partial m} \frac{v}{2u^2} \left(\frac{-1}{2}\right)$$

$$- \frac{\mu_{xy}}{\mu_{x'y}} \frac{\partial \mu_{x'y}}{\partial h} \frac{1}{2} + \frac{\mu_{xy}}{\mu_{x'y}} \frac{\partial \mu_{x'y}}{\partial m} \frac{v}{2u^2} \left(\frac{-1}{2}\right)$$

$$> 0 \quad \text{Using Corollary 1}$$

Thus, having just shown that $\frac{\partial \mu_{xy}}{\Phi_{xy} - \Phi_{x'y}} > 0$ and given that $\Phi_{xy} - \Phi_{x'y} > \Phi_{xy'} - \Phi_{x'y'}$, $\frac{\mu_{xy}}{\mu_{x'y}} > \frac{\mu_{xy'}}{\mu_{x'y'}}$ holds.

Proof of Theorem 5

The proof is similar to that for the unidimensional case. First consider dimension 1. Let $X$ and $Y$ denote the vector of characteristics of worker and firm respectively. For notational simplicity, let $X = (x_1, x_2)$ and $X' = (x_1', x_2)$, where $x_1 > x_1'$. Let $Y = (y_1, y_2)$. From Corollary 1, we know that $\frac{\partial \mu(X, Y)}{\Phi(X, Y)} > 0$ and that $\frac{\partial \mu(X, Y)}{\Phi(X', Y)} \leq 0$ for $X' \neq X$. For log supermodularity in the first dimension to hold, we must first have $\frac{\partial \mu(X, Y)}{\Phi(X, Y) - \Phi(X', Y)} > 0$. Otherwise put, the ratio of assignment probabilities must increase in the difference in productivities between $X$ and $X'$ at some firm $Y$.

Let $s = \Phi(X, Y) - \Phi(X', Y)$ and let $t = \Phi(X, Y) + \Phi(X', Y)$. As such, $\Phi(X, Y) = h(s, t) = \frac{t + s}{2}$ and $\Phi(X', Y) = m(s, t) = \frac{t - s}{2}$. For ease of notation, re-express $\Phi(X, Y)$ as $\Phi_{XY}$ and $\mu(X, Y)$ as $\mu_{XY}$. Let $F(\mu_{XY}(\Phi_{XY}, \Phi_{X'Y}), \mu_{X'Y}(\Phi_{XY}, \Phi_{X'Y})) = \frac{\mu_{XY}(\Phi_{XY}, \Phi_{X'Y})}{\mu_{X'Y}(\Phi_{XY}, \Phi_{X'Y})}$. Note that

$$F(\mu_{XY}(\Phi_{XY}, \Phi_{X'Y}), \mu_{X'Y}(\Phi_{XY}, \Phi_{X'Y})) = F(\mu_{XY}(h(s, t), m(s, t)), \mu_{X'Y}(h(s, t), m(s, t)))$$
Then, using the chain rule,

\[
\frac{\partial F}{\partial s} = \frac{\partial F}{\partial \mu_{XY}} \frac{\partial \mu_{XY}}{\partial h} \frac{\partial h}{\partial s} + \frac{\partial F}{\partial \mu_{X'Y}} \frac{\partial \mu_{X'Y}}{\partial m} \frac{\partial m}{\partial s} + \frac{\partial F}{\partial \mu_{X'Y}} \frac{\partial \mu_{X'Y}}{\partial m} \frac{\partial m}{\partial s} \]

\[
= \frac{1}{2} \frac{1}{\mu_{X'Y}} \frac{\partial \mu_{XY}}{\partial h} + (-\frac{1}{2}) \mu_{X'Y} \frac{\partial \mu_{XY}}{\partial m} + \frac{1}{2} \left(-\frac{\mu_{XY}}{\mu_{X'Y}}\right) \frac{\partial \mu_{X'Y}}{\partial h} + \left(-\frac{1}{2}\right) \left(-\frac{\mu_{XY}}{\mu_{X'Y}}\right) \frac{\partial \mu_{X'Y}}{\partial m} > 0 \quad \text{Using Corollary 1}
\]

Since we have assumed that the difference in productivities between \(X\) and \(X'\) at \(Y\) and \(Y'\) satisfy \(\Phi(X, Y) - \Phi(X', Y') > \Phi(X, Y') - \Phi(X', Y)\), and have established that \(\frac{\partial \mu(X, Y)}{\partial \Phi(X, Y)} > 0\), we can thus conclude that \(\mu(X, Y), \mu(X', Y') > \mu(X, Y'), \mu(X', Y)\) holds. Using the same method, we can show that \(\mu(X, Y), \mu(X'', Y') > \mu(X, Y''), \mu(X'', Y'')\) holds so long as \(\Phi(X, Y) - \Phi(X'', Y') > \Phi(X, Y'') - \Phi(X'', Y'')\) is satisfied, where \(X, X'', Y\) and \(Y''\) are defined as before.

**Discussion of O*NET job task measures**

From the O*NET, job tasks are constructed by conducting Principal Components Analysis (PCA) for 3 loosely categorized groups of job characteristics provided by O*NET. Some precise characteristics in each of these three broad categories are listed below.

1. \(y_c\): Written Comprehension, Written Expression, Deductive Reasoning, Inductive Reasoning, Information Ordering, Mathematical Reasoning, Number Facility, Memorization, Reading Comprehension, Writing Mathematics, Science, Critical Thinking, Complex Problem Solving.

2. \(y_i\): Social Perceptiveness, Service Orientation, Interpersonal Relationships, Assisting and Caring for Others, Resolve Conflict, Contact With Others, Deal With External Customers, Frequent Conflict, Angry People, Aggressive People.


What information do O*NET job descriptors provide exactly? The information on the O*NET descriptors for the version 15 of the O*NET was obtained from job analysts, occupation experts as well as incumbents. Respondents were essentially asked to rate on different scales, the level of each job descriptor involved in each occupation. This paper interprets these as tasks in a job. However, these descriptors could certainly be interpreted differently. In particular, these could be read as the average level of skill that workers have in each job, the minimum level of skill required, or the optimal level of skill required. These possible alternatives are discussed below, along with their shortcomings.

1. Average: If job descriptors in O*NET were actually the average level of skills of workers in each job, one would expect the average level of skills of workers from the NLSY in each occupation to correlate strongly to the to job descriptors in O*NET. As this is not the case, it seems unlikely that O*NET describes the average skill levels in each job.
2. Minimum: Assuming that the O*NET really describes the minimum skill requirement strictly required for each job, then one should not observe a substantial number of workers below the requisite skill requirement employed in each occupation. Since worker skills in the NLSY were measured in a different way to the O*NET, one first has to suppose that skills from both datasets not only cohere in terms of relative ranks (which has been thus far assumed) but also that the cohere in terms of absolute values. Accepting this assumption, one can invalidate the minimum interpretation by checking if there is a non-negligible number of workers with less than the requisite skills for their jobs, which is indeed the case. If one disagrees that the NLSY and O*NET measures are comparable in absolute terms, then it is unclear how to proceed with verifying this interpretation of O*NET tasks.

3. Optimum: If the O*NET really described the optimum level of skills for each job, then one should observe a production penalty for workers who have more or less than the O*NET-indicated level of skill. Yet, it is not even conceptually clear why workers with extra skill than the ‘optimum’ should produce less strictly than workers with fewer skill, since this would call into question the notion of skill entirely. Moreover, this interpretation, like the previous one, also necessitates the assumption that the O*NET and NLSY scales are comparable in absolute terms.

This paper hence advocates for the interpretation of the O*NET job descriptors as job tasks for the following reasons: Firstly, it does not require that the $x_k$ and $y_k$ be compared in absolute terms, which is not possible anyway from the data. Instead, $x_k$ and $y_k$ need only be ordered separately for workers and firms. Secondly, a task interpretation is the more natural one when studying worker-firm complementarities, since it implies that jobs by themselves have some production-relevant content. If O*NET job descriptors were taken to be minimum, average or optimum skill requirements, then the idea of production complementarities would seem somewhat redundant.
Local identification of parameter estimates

Figure 18: Local identification for production parameters

Figure shows how the sum of the distance between the model and data moments is minimized at the estimates of the parameters for the 1979 cohort
Figure 19: Local identification for $\gamma(Y_j)$s

Figure shows how the sum of the distance between the model and data moments is minimized at the estimates of the parameters for the 1979 cohort.
Figure 20: Local identification for $\gamma(Y_j)_s$

Figure shows how the sum of the distance between the model and data moments is minimized at the estimates of the parameters for the 1979 cohort.
Figure 21: Local identification for production parameters

Figure shows how the sum of the distance between the model and data moments is minimized at the estimates of the parameters for the 1997 cohort.
Figure 22: Local identification for $\gamma(Y_j)s$

Figure shows how the sum of the distance between the model and data moments is minimized at the estimates of the parameters for the 1997 cohort.
Figure 23: Local identification for $\gamma(Y_j)_s$

Figure shows how the sum of the distance between the model and data moments is minimized at the estimates of the parameters for the 1997 cohort.