Sorting multi-dimensional Types: Theory and Application

Ilse Lindenlaub

Yale University, 28 Hillhouse Avenue, New Haven, CT 06520, USA
ilse.lindenlaub@yale.edu

November 13, 2016

Abstract

This paper studies multi-dimensional matching between workers and jobs. Workers differ in manual and cognitive skills and sort into jobs that demand different combinations of these two skills. To study this multi-dimensional sorting problem, I develop a theoretical framework that generalizes the unidimensional notion of assortative matching and sufficient conditions on the technology under which sorting obtains. I derive the equilibrium in closed form and use this explicit solution to study biased technological change. The main finding is that an increase in worker-job complementarities in cognitive relative to manual inputs leads to more pronounced sorting and wage inequality across cognitive relative to manual skills. This can trigger wage polarization and boost aggregate wage inequality. I then estimate the model for the US and identify sizable technology shifts: during the last two decades, worker-job complementarities in cognitive inputs strongly increased whereas complementarities in manual inputs decreased. In addition to this bias in complementarities, there has been a cognitive skill-bias in production. Counterfactual exercises suggest that these technology shifts (as opposed to changes in skill supply and demand) can account for observed changes in worker-job sorting, wage polarization and a significant part of the increase in US wage dispersion.

Keywords. Multi-Dimensional Heterogeneity, Assortative Matching, Closed Form, Task-Biased Technological Change.

*I am very grateful to Jan Eeckhout, Omiros Papaspiliopoulos and Nicola Pavoni for their advice. I would also like to thank Jérôme Adda, Joe Altonji, Vasco Carvalho, Hector Chade, David Levine, George Mailath, Giuseppe Moscarini, Kristoffer Nimark and Anja Prummer for their comments. I thank Carl Sanders for making his data on occupational skill usage available to me. Finally, I would like to thank Philipp Kircher (Co-Editor) and four anonymous referees for their helpful comments and suggestions. The paper has benefitted from discussions with seminar participants at the EUI, UPF, ESSET 2012 Gerzensee, XVII Workshop on Macroeconomic Dynamics Vigo, EEA 2013 Gothenburg, Matching: Theory and Estimation Paris, Barcelona GSE Winter Workshop, UPenn, Penn State, Minneapolis Fed, WUSTL, Bonn University, Stockholm University, NYU, IIES, ASU, UCLA, Edinburgh, Cambridge, Minnesota, Madison-Wisconsin, University of Toronto, ITAM, Philadelphia Fed, New York Fed, Humboldt University Berlin, Collegio Carlo Alberto, SED 2014 Meetings Toronto, NBER Summer Institute 2014, Search and Matching Workshop Philadelphia Fed, Yale University. Most of this work was carried out while visiting UPF, Barcelona. This research has been supported by a grant from the DAAD (German Academic Exchange Service). First Draft: January 2013.
1 Introduction

Technological progress has drastically changed the task composition of work and hence the structure of labor demand. Across the board, workers spend less time performing manual tasks such as assembling cars and more time performing cognitive tasks such as computer programming or selling products and services.\(^1\) During the 1980s, a blue-collar worker in the car industry might have spent some time on cognitive tasks such as reporting to his supervisor, but he mainly engaged in manual labor on the assembly line. Twenty years later, a newly-developed machine carries out his manual task. Programming the machine requires more cognitive than manual skills, and thus a different skill mix than such a worker can offer. So, who operates this machine? What is the worker’s new job? And, how does this technological shift affect wages and inequality? This is a multi-dimensional assignment problem where workers with different bundles of manual and cognitive skills sort into jobs that require different combinations of these skills.

This paper studies this problem and makes three distinct but related contributions: First, it develops a notion of multi-dimensional sorting that extends the unidimensional notion of positive and negative assortative matching (PAM and NAM). It then specifies sufficient conditions on the technology under which the multi-dimensional allocation is positively/negatively assortative in equilibrium. Second, I solve for the equilibrium allocation and wages in closed form for a class of examples, and use this explicit solution to analyze sorting and comparative statics. The proposed notions of PAM/NAM are crucial to characterize and interpret equilibrium sorting patterns and how they change in the face of technological or distributional changes. In particular, I analyze the impact on equilibrium outcomes as cognitive (as opposed to manual) inputs become more important in production, capturing one of the main recent technological shifts. Third, I illustrate the usefulness of this theory in an empirical application to technological change in the US. Using the model, I can infer from data on observed equilibrium outcomes by how much technology has changed over time, and, guided by the comparative statics predictions, I can study the effects on sorting and wage inequality.

One of the main economic insight from this multi-dimensional matching model is that workers face a sorting trade-off. The decision to take a job that better fits their cognitive or their manual skills depends on worker-job complementarities in cognitive versus manual tasks.\(^2\) Task-biased technological change, which increases the level of complementarities between cognitive skills and skill demands (relative to those in the manual dimension), puts this trade-off to work. Sorting improves along the cognitive dimension but the opposite is true in the manual dimension, where matches are characterized by a poorer fit between workers’ skills and jobs’ demands. In light of the previous example, the blue-collar worker who was replaced by a machine may now be employed as a car salesman. This new job is tailored to his cognitive skills but is a poor fit with his manual abilities. The new allocation benefits workers with high cognitive abilities but harms those with manual know-how. This makes wages more convex in cognitive but less convex in manual skills, thereby fueling wage inequality along

\(^1\)See Autor et al. \[2003\] for an empirical analysis of the changing skill content of tasks.

\(^2\)The meaning of complementarities in this context is that workers with high cognitive skills are particularly productive in jobs that put significant weight on cognitive ability, and similar in the manual dimension.
the cognitive dimension but compressing inequality in the manual dimension.

To assess whether this new mechanism of how technological change affects wages and inequality matters also quantitatively, I estimate the model for the US. I first identify sizable technological shifts: Between 1990 and 2010, complementarities in cognitive inputs strongly increased whereas complementarities in manual inputs decreased. While this cognitive *task-biased* technological change increased the productivity of workers’ cognitive skills in tasks that require a lot of cognitive skills (an example could be new computers that help programming the machine that replaced the above-mentioned production worker), there was also cognitive *skill-biased* technological change. This shift increased the productivity of cognitive skills independent of the performed task and its cognitive skill demands, leaving worker-job complementarities unchanged. Examples of skill-biased technical change are advancements in communication technology (e.g. the internet, Google) that benefit workers in cognitive tasks of any complexity.

The main finding from the structural estimation is that these technological shifts may account for observed wage polarization (i.e. declining lower tail but expanding upper tail wage inequality), increased wage skewness and much of the rise in wage dispersion. More precisely, counterfactual exercises show that task-biased technological change can account for the boost in wage skewness as well as polarization. The reason this technology shift affects upper and lower tail wage inequality differently is that winners (i.e. workers with high cognitive skills) are clustered in the upper part of the wage distribution while those adversely affected (workers with mainly manual skills) are concentrated in the lower part. In turn, cognitive skill-biased technological change, which does not affect the curvature of the wage schedule, fuels inequality across the whole distribution. It can account for a significant part of the increase in US wage dispersion over the last two decades.

Biased technological change, and particularly task-biased change, is considered an important force behind recent wage inequality trends in the developed world (Acemoglu and Autor [2011]). The idea is that technological advances like the development of computers have replaced workers in manual tasks but created stronger complementarities between skills and job attributes in cognitive tasks. However, even though two intrinsically different skills are involved (manual and cognitive), the literature has analyzed this technological change only in one-dimensional settings. In these frameworks, an adverse technology shock reduces firms’ demand for medium-skilled workers (who presumably hold manual skills). As a result, their relative wages drop and so do employment shares in medium-skilled jobs – a phenomenon that is referred to as labor market polarization.\(^3\)

One advantage of the assumption of one-dimensional heterogeneity is tractability. However, it is important to note that collapsing agents’ multiple characteristics into a single index is not innocuous. A notable study that rejects the single index model is by Willis and Rosen [1979]. They show that worker performance depends on a *bundle* of different skills including intellectual and manual skills. Some people are strong in both skills (e.g. mechanical engineers or surgeons) and others specialize.

\(^3\)See, for instance, Costinot and Vogel [2010] and Acemoglu and Autor [2011]. In the literature, task-biased technological change is often referred to as *routinization*, meaning that new machines replace those workers performing routine tasks (e.g. Autor et al. [2003], Autor et al. [2006], Autor and Dorn [2013]). Routine skills also capture manual skills. To fit their analysis more closely, the two skills here could be interpreted as routine and non-routine.
This points to the main reason for requiring matching models with multi-dimensional heterogeneity: In the data, characteristics are not perfectly correlated, which is why agents can only be partially ordered. Thus, it is problematic to aggregate different attributes into a single one-dimensional index, according to which agents are ranked and matched.

To assess the (quantitative) importance of multi-dimensional sorting in the labor market, one needs a tractable theoretical framework. While the literature on optimal transport has studied the existence, uniqueness and purity of multi-dimensional assignments under transferable utility, existing studies provide little insights into the qualitative properties of equilibrium sorting and wages or comparative statics. This paper makes a first attempt at developing a framework that allows for both.

Section 2 introduces the theoretical framework, which builds on a standard transferable utility assignment model where workers and jobs match in pairs. Workers possess manual and cognitive skills. Each worker performs two tasks, a manual and a cognitive one. Jobs, in turn, differ in productivities or skill demands for each task. Within this task-based framework, I propose a generalization of positive assortative matching (PAM) and negative assortative matching (NAM) to the multi-dimensional setting. In non-technical terms, my definition of PAM means that, ceteris paribus, workers with more cognitive skills match with jobs whose cognitive task is more demanding, and similarly in the manual dimension. This captures, for instance, that the best scientists tend to work in the best universities (universities put a lot of weight on intellectual skills but little on manual dexterity) whereas the best mechanics often work in professional motor sports (which require manual skills more than intellectual abilities). I then state conditions on the production function such that the equilibrium is assortative. Intuitively, if there are complementarities of skills and productivities within tasks but not across tasks, then the optimal assignment satisfies PAM. These properties are shown in generality and without any functional form assumptions on the distributions and production technology.

To study biased technological change, it is useful to have a closed form solution that is amenable to comparative statics and estimation. Toward this goal, Section 3 specifies the environment to Gaussian distributions and linear-quadratic technology. Using the above notion of assortative matching, I develop a technique to solve for equilibrium assignment and wage function in closed form, following similar steps as in one-dimensional decentralized assignment models. It is important to note that notwithstanding many parallels to the one-dimensional setting, there is also an important difference: with multi-dimensional heterogeneity, it is not possible to fully rank agents and, as a result, there are many PAM allocations that clear the labor market. This is why, contrary to one-dimensional matching in Becker [1973], super or submodularity of technology is not sufficient to pin down the unique optimal (i.e. output-maximizing) PAM allocation. Instead, the relative level of complementarities across tasks (and not only their signs) are crucial to determine the unique equilibrium assignment.

This stronger link between technology and assignment in multi-d settings creates the main technical difficulty in solving the model. But it also allows for a richer analysis than one-dimensional

---

4 More recently, results by Papageorgiou [2014] also favor the specialization hypothesis over a single index model.
5 The optimal transport problem involves finding a measure-preserving map that carries one distribution into another at minimal cost, relying on linear programming (Gretsky et al. [1992], Villani [2009], Chiappori et al. [2010], Ekeland [2010]). In Section 6, I will discuss the similarities and differences to optimal transport in more detail.
matching and offers a natural framework to study task-biased technological change, which focuses on complementarities. The relative level of worker-job complementarities determines the unique optimal PAM allocation from many existing ones. They range from strong assortativeness to significant mismatch between worker and job traits in one or both task(s), capturing a much richer set of assignments (including the discussed sorting trade-off) compared to one-dimensional PAM.

Section 4 derives comparative statics results in the Gaussian linear-quadratic model and uses the developed notion of sorting to interpret them. I focus on the effects of task-biased technological change, which demonstrates how sorting (and ultimately wages and inequality) are shaped by technology. Task-biased technological change triggers the discussed sorting trade-off, where workers change jobs in order to better align their cognitive skills with cognitive demands at the cost of mismatch in the manual dimension. The presented notion of PAM is instrumental for interpreting these changes in sorting. I also contrast these results with those for the more standard skill-biased technological change (the latter increases the relative productivity of workers' cognitive skills without affecting worker-job complementarities) and with those for changes in the underlying distributions.

Section 5 illustrates how this multi-dimensional model can be structurally estimated, yielding novel insights on the effects of technological change in the US, despite three major challenges: First, data on multi-dimensional skills and skill demands is not readily available. Second, the model has few parameters to fit the data. And, third, it is deterministic (i.e. there are no search frictions or unobserved heterogeneity to help explain mismatch on observable characteristics).

To overcome the first challenge, I construct bivariate skill and skill demand distributions, combining data from two cohorts of the National Longitudinal Survey of Youth (NLSY79, NLSY97) with the O*NET. I then estimate the model by Maximum Likelihood (assuming that both wages and assignment are subject to measurement error - partially addressing the third challenge) to quantify technological change between 1990 and 2010 and to decompose changes in wage inequality into those driven by technological and distributional shifts. I find that technological change (as opposed to changes in skill supply and demand distributions) has been a major force behind increased US wage inequality. This result provides an additional argument for studying technological change in a multi-dimensional framework (with its stronger link between technology and sorting, and ultimately wages) as opposed to a one-dimensional model (where distributions are the main determinant of the assignment).

I also estimate several one-dimensional models to highlight in which dimensions the multi-dimensional model offers a richer interpretation of the data: These single index models miss the manual-cognitive sorting trade-off, and closely related, the differential impact of biased technological change on manual and cognitive returns. This is why they fail to account for the observed wage polarization. Moreover, empirically observed rank switchings in the wage distribution (meaning that, e.g., cognitive specialists have overtaken manual specialists in the wage ranking) cannot be accounted for by comparable one-dimensional models but naturally occur in the multi-dimensional case as a consequence of biased technological change.

Section 6 places the main contribution of the paper into the literature. Section 7 concludes. The Appendix (and Online Appendix) contains all proofs, data details and detailed estimation results.
2 Theoretical Framework for Multi-Dimensional Sorting

This section outlines the general model. For simplicity, I will focus here on two-dimensional heterogeneity even though this section generalizes to $N$-dimensional heterogeneity (see Appendix).

2.1 Environment

Agents: There is a continuum of firms and of workers. All are risk-neutral. Every worker is endowed with a skill bundle of cognitive and manual skills, $x = (x_C, x_M) \in X = [x_C, x_M] \times [x_M, x_M]$. Points in $X$ represent worker types. Denote the joint c.d.f. of $(x_C, x_M)$ by $H$ with positive density $h$. In turn, each firm (which I use interchangeably with job) is endowed with both cognitive and manual skill demands, $y = (y_C, y_M) \in Y = [y_C, y_C] \times [y_M, y_M]$. Coordinate $y_C$ (respectively $y_M$) corresponds to the productivity or skill requirement of cognitive task $C$ (respectively manual task $M$). Points in $Y$ represent firm types. Denote the joint c.d.f. of $(y_C, y_M)$ by $G$ with positive density $g$. Assume that the overall measures of firms and workers coincide.

Production: Every firm produces a single homogenous good by combining all inputs. Denote the production function by $F: X \rightarrow \mathbb{R}_+$. I assume that $F$ is $C^4$.

Labor market: Firms and workers match pairwise. The labor market is competitive.

2.2 Definitions

Matching Function: Sorting between workers and firms is described by a function $\mu: X \rightarrow Y$, that assigns each worker with skill bundle $x$ to some job with skill requirements $y = \mu(x)$. I will call $\mu$ the matching function (and will provide more details below).

Assortative Matching: What makes assignment problems tractable in the one-dimensional world is the concept of assortative matching: There, PAM (NAM) is defined by a strictly increasing (decreasing) matching function, meaning that better (worse) workers work in better firms. This concept captures two aspects: (a) purity of matching (i.e. one-to-one matching function), and (b) sign of sorting. I aim to define a multi-dimensional version of assortative matching that also incorporates these two features. As in the one-dimensional setting, here assortativeness involves properties of the first derivative of the matching function (i.e. of its Jacobian), given by:

$$ J_\mu = D_{xy} = \begin{bmatrix} \frac{\partial y_C}{\partial x_C} & \frac{\partial y_C}{\partial x_M} \\ \frac{\partial y_M}{\partial x_C} & \frac{\partial y_M}{\partial x_M} \end{bmatrix} $$

I define multi-dimensional assortative matching as follows:

Definition 1 (Multi-Dimensional Assortative Matching) The sorting pattern is PAM (NAM) if, for all $x$, $D_{xy}$ is a $P$-matrix ($P^-$-matrix), i.e. if

$$ i. \quad \frac{\partial y_C}{\partial x_C} > (\leq) \\ ii. \quad \frac{\partial y_M}{\partial x_M} > (\leq) \\ iii. \quad \text{Det}(J_\mu) = \frac{\partial y_C}{\partial x_M} \frac{\partial y_M}{\partial x_C} - \frac{\partial y_C}{\partial x_C} \frac{\partial y_M}{\partial x_M} > 0 $$

$^6$For all but one result, $C^2$ suffices. The $C^4$ property is only used for the existence of a smooth matching function.

$^7$Note that I restrict the analysis to the search for the optimal matching function, i.e. to the search for a deterministic optimal transport, which is called the Monge problem in the optimal transport literature (e.g. Villani [2009], Ch. 1).

$^8$Throughout, I will state a function is increasing/decreasing or super/submodular if this is the case in the weak sense. Strict properties are mentioned explicitly.
First, I will give the intuition and then the technical details. To illustrate most arguments in this paper, I will focus on $P$-matrices and PAM. What underlies Definition 1 is the ability to identify the natural sorting dimensions. In my setting, the natural sorting dimensions are ‘cognitive skills-cognitive skill requirements’ on the one hand, and ‘manual skills-manual skill requirements’ on the other (if, in turn, it was more natural to sort manual skills into cognitive requirements and cognitive skills into manual requirements, I could simply relabel skills and apply the definition as it stands).

Given this classification, this notion of PAM restricts the sign of sorting in a sensible way: PAM means that, ceteris paribus, intellectual types with higher $x_C$ work in firms with higher $y_C$ where they perform complex intellectual tasks, compared to workers with less cognitive abilities (i.). Similarly, workers with strong manual skills $x_M$ work in firms with high $y_M$ that attach much weight to the manual task (ii.). Condition iii. is mainly technical (see below) but also has an intuitive interpretation: In the multi-dimensional space, sorting occurs along more than one dimension (e.g. also between cognitive skills and manual skill requirements). Inequality iii. requires that under PAM, sorting within the natural dimensions is overall more pronounced than sorting across the natural dimensions.

This definition thus captures the sign of sorting, which under PAM is a positive relation between worker and firm traits within the natural sorting dimensions (and this sorting is more pronounced than that between dimensions, i.e. $x$ and $y$ co-move more strongly within task $C$ and $M$ than across tasks).

Besides the sign of sorting, Definition 1 also captures that the assignment is pure, defined as:

**Definition 2 (Pure Matching)** Matching is pure if $\mu$ is one-to-one almost everywhere.

In economic terms, pure matching means that two firms of the same type choose the same worker. Technically, purity is closely related to the properties of the Jacobian of the matching function and particularly to the $P$-matrix property of the Jacobian (given by i-iii in Definition 1). Gale and Nikaido [1965] link the $P$-matrix property of the Jacobian of a function to the function’s injectivity, giving a sufficient condition for purity in the current setting: if $D_{xy}$ is a $P$-matrix (or $P^\prime$-matrix), then the matching function is globally one-to-one. The $P$-matrix property is also sufficient for global invertibility, which is why $\mu$ can be thought of the unique inverse of a function $\nu : Y \rightarrow X$, where $x = \nu(y)$ is worker type $x$ that firm type $y$ chooses to hire.

This discussion shows that Definition 1 is a natural generalization of one-dimensional assortative matching, capturing the same two aspects via the first derivative of the matching function: the sign of sorting in each task dimension (given by i. and ii. in (1)) and purity of the assignment (guaranteed by also considering the determinant condition iii.). With one-dimensional types the strict first derivative of $\mu$ guarantees both purity and PAM. In both settings, PAM implies purity.

The figure below provides a graphical illustration of multi-dimensional PAM, using a discrete 2x2 example: Each side of the market has two attributes that can be high (H) or low (L). Hence, there are

---

9 Generally, a matrix is a $P$-matrix if all its principal minors are positive. Hence, every positive definite matrix is a $P$-matrix but the converse statement only holds for symmetric matrices. In turn, matrix $M$ is $P^\prime$ if $-M$ is $P$.

10 $P$-matrices have so far not been used in the matching literature but have been used in other fields of economics to rule out multiple equilibria. See, for instance, Simsek et al. [2005].

11 See Theorem 1.1 in Chua and Lam [1972] and the references therein for the equivalence of the class of globally one-to-one and continuous functions from $\mathbb{R}^n$ into $\mathbb{R}^n$ and the class of globally homeomorphic functions from $\mathbb{R}^n$ to $\mathbb{R}^n$. 

---
four worker and four firm types. In each subfigure, the dots in the left panel represent worker types and those in the right panel firm types. Assume that all types carry the same mass of agents, and suppose worker and firm types of the same color match. In subfigure (a), matching is characterized by PAM (which implies purity). In subfigure (b), matching is pure (i.e. every agent matches with a single type) but PAM is violated along dimension $C$. In subfigure (c), matching is not pure (and thus not positive assortative) because two identical agents can have different matches.

2.3 The Equilibrium

A firm with given productivity bundle $y = (y_C, y_M)$ chooses a worker with skill bundle $x = (x_C, x_M)$ in order to maximize profits. It takes the wage schedule $w$ as given, meaning that wages are not a function of productivities. The firm’s problem is given by profit maximization:

$$\max_{x \in X} F(x, y) - w(x)$$

I denote $y^* = (y_C^*, y_M^*)$ the equilibrium assignment, where $y^* = \mu(x)$ denotes the firm’s productivity bundle that is optimally matched to the worker with skill bundle $x = (x_C, x_M)$. I focus on a competitive equilibrium, which is defined as follows.

**Definition 3 (Equilibrium)** An equilibrium is characterized by a matching function $\mu : X \rightarrow Y$, and a wage function $w : X \rightarrow \mathbb{R}_+$, satisfying:

(i) **Optimality**: Price-taking firms maximize profits (2) by choosing $x$ for the given $w$.

(ii) **Market Clearing**: Feasibility of $\mu$ requires that when $x \sim H$ then $y^* \sim G$.

Optimality of the firm’s choice and labor market clearing are standard requirements of equilibrium. Market Clearing requires that workers and firms are matched in a measure-preserving way for each measurable set.
I can now state the result on the existence of a unique and differentiable equilibrium. Note that a diagonal $P$-matrix is a diagonal matrix with all positive principal minors. Regarding notation, $D_x F$ denotes the gradient of $F$ with respect to the vector $x$ (and for other functions similarly) and $D^2_{xy} F$ denotes the matrix of cross-partial derivatives of $F$. Non-bold subscripts denote scalar derivatives.

**Proposition 1 (Equilibrium)** (i) If, for all $x$, $D_x F(x,\cdot)$ is injective, then there exist a unique equilibrium matching function $\mu$ (for almost every $x$), which solves the FOCs of firm problem (2),

$$D_x F(x, \mu(x)) - D_x w(x) = 0$$

and a unique equilibrium wage function $w$ (up to a constant) that decentralizes assignment $\mu$.

(ii) If, in addition, $D^2_{xy} F$ is a diagonal $P$-matrix and $\forall i \in \{C,M\}$, $F_{x_i\bar{y}_i}$ is supermodular and log-supermodular in $(x_i, \bar{y}_i)$, and if $g$ and $h$ are smooth, then $w$ and $\mu$ are smooth in the interior of $X$.

All proofs are in the Appendix. The existence of a unique deterministic equilibrium (i) follows from well-established results in the optimal transport literature (see e.g. Villani [2009], Chapter 10). The assumption that $D_x F(x,\cdot)$ is injective guarantees uniqueness and is implied by the $P$-matrix property of $D^2_{xy} F$ assumed in (ii). Part (ii), in turn, ensures that the equilibrium will be smooth (i.e. guaranteeing that $w$ is $C^2$ and $\mu$ is $C^1$) in the interior of $X$, which requires stronger assumptions on the technology.\(^{12}\) My proof is based on the Ma-Trudinger-Wang condition (MTW) from optimal transport theory (Ma et al. [2005], see Appendix), which was shown to be necessary and (nearly) sufficient for smoothness of the matching function.\(^{13}\) In its general form, MTW requires complex restrictions on third and forth order derivatives of the technology, which are difficult to satisfy and often impossible to check. My assumption that $D^2_{xy} F$ is a diagonal $P$-matrix is not necessary for MTW to hold but simplifies the condition substantially: This assumption along with $F_{x_i\bar{y}_i}$ being (log)-supermodular ensures that MTW holds. For some specific production functions, weaker assumptions than the diagonal $P$-matrix guarantee that MTW holds: For instance, for the bi-linear production function analyzed below, MTW is satisfied even if the off-diagonal elements of $D^2_{xy} F$ are non-zero.\(^{14}\)

Note that the assumptions in (ii) are purely technical and imposed here to guarantee differentiability of the equilibrium. Differentiability is crucial for applying my notion of multi-dimensional assortative matching that relies on the Jacobian of the matching function. I use this notion below for comparative statics to gain insights into the qualitative properties of the equilibrium. Finally note that part (ii) can be easily adjusted for the case of a diagonal $P^-$-matrix (see proof).

\(^{12}\)Note that smoothness of $X$ and $Y$ and uniform convexity of the sets $-D_x F(x,Y)$ and $-D_y F(X,y)$ are not needed to guarantee smoothness of $w$ and $\mu$ in the interior of the support. Instead, a weaker condition (that $-D_x F(x,Y)$ and $-D_y F(X,y)$ are convex) suffices, which I verify in the Appendix. I thank Guido de Philippis for pointing this out to me.

\(^{13}\)For necessity, see Theorems 12.7 and 12.42 in Villani [2009] or, when stated for the Euclidean space, Theorems 4.5 and 4.6 in De Philippis and Figalli [2014]. For sufficiency, see Theorem 2.1 in Ma et al. [2005] or Theorem 4.2 in De Philippis and Figalli [2014].

\(^{14}\)Also, alternative (non-MTW based) regularity results exist for bi-linear technology (Th.12.50 in Villani [2009]).
2.4 Multi-Dimensional Assortative Matching

This section relates properties of the technology to properties of the equilibrium assignment. The assignment is only optimal if the second-order conditions of the firm’s problem, evaluated at \((y^*_C, y^*_M)\), i.e. negative semi-definite Hessian, are satisfied. Using these necessary second-order conditions for optimality, I show that if the complementarities in technology follow a diagonal P-matrix, i.e.

\[
D^2_{xy} F = \begin{bmatrix}
F_{xy}^+ & 0 \\
0 & F_{xy}^-
\end{bmatrix}
\]  

then the equilibrium assignment satisfies PAM (i.e. \(D_{xy}^*\) is a P-matrix). For NAM, a similar statement holds when replacing complementarities by substitutabilities.\(^\text{15}\) Throughout this section I will maintain the Assumptions from Proposition 1.

**Proposition 2 (Assortative Matching)** If \(D^2_{xy} F\) is a diagonal P-matrix (\(P^-\)-matrix), then the equilibrium assignment \(\mu\) satisfies PAM (NAM).

To gain intuition into the assortativeness result, consider PAM. If there is complementarity between skills and productivities within both the cognitive task \((F_{xy}^C > 0)\) and the manual task \((F_{xy}^M > 0)\) and between-task complementarities are absent \((F_{xy}^C = F_{xy}^M = 0)\), then it is optimal that workers and firms match in a positive assortative way: Agents with strong intellectual skills work in firms that require these skills (and similarly for the manual dimension).\(^\text{16}\)

Proposition 2 obtains under strong restrictions on the complementarities in production. The reason for the separability assumption is that, in the multi-dimensional space, interactions between complementarities in production and properties of the type distributions can easily become a barrier to PAM – a problem that is absent under one-dimensional traits. To see this, note that sorting occurs here along all skill and productivity dimensions, i.e. also between tasks: also between manual (cognitive) productivity and cognitive (manual) skill, \(\partial y^*_M / \partial x_C \neq 0\) (\(\partial y^*_C / \partial x_M \neq 0\)). Allowing for complementarities between, say, manual skill demands and cognitive skills \((F_{xy}^C > 0)\) may render a positive relation between these attributes optimal, \(\partial y^*_M / \partial x_C > 0\). This can come at the expense of negative sorting in the manual task, \(\partial y^*_M / \partial x_M < 0\), especially when skills are negatively correlated, violating PAM.

It is important to note that the stated sufficient condition for PAM is distribution-free. If one is willing to impose restrictions on the distributions, this condition can be considerably weakened, allowing for between-task complementarities or substitutabilities \((F_{xy}^C \neq 0, F_{xy}^M \neq 0)\). In Section 8.4 of the Appendix, I show that a weaker version of (4) applies to settings where skills and productivities are either (i) uniformly distributed, (ii) identically distributed or (iii) normally distributed. For (i) and (ii), the sufficient condition for PAM is that the matrix of cross-partials of \(F\) is a symmetric \(P\)-matrix (i.e. positive definite) and for (iii) a diagonally dominant \(P\)-matrix.

\(^{15}\)The presented condition is related to the twist condition from optimal transport but is not equivalent. See Section 6.

\(^{16}\)Similarly, in the case of NAM, assortative matching within tasks dominates assortativeness across tasks, only in this case highly skilled workers are matched with low productivity firms.
Despite these complications due to multi-dimensional heterogeneity, the conditions for sorting show much similarity to those under one-dimensional traits. There, the requirement of a negative definite Hessian collapses to the requirement on the firm’s second-order condition that \(-F_{xy}\mu_x \leq 0\). If \(F_{xy}\) is strictly positive then matching is PAM, \(\mu_x > 0\). Purity is given by strict monotonicity of matching function \(\mu\) and the sorting direction by its positive slope. Similarly in this model, I impose conditions on the matrix of cross-partial \(D^2_{xy}F\) to obtain PAM. The difference is that with multiple dimensions not only the signs but also the relative magnitudes of different complementarities need to be restricted in order to ensure assortative matching.

A final insight comes from the computation of the wage schedule. Recall that Proposition 1 established the existence of a unique equilibrium wage function. Below, this equilibrium wage function needs to be computed as the solution of a system of partial differential equations (PDEs), given by the firm’s first-order conditions evaluated at the equilibrium assignment (3). As is well-known from the general theory on PDEs, to solve this (overdetermined) system, one needs to take into account an integrability condition that makes the system formally integrable.\(^{17}\) It is here given by cross-differentiating and equating both FOCs in (3), ensuring that the Hessian of the firm’s problem is symmetric, \(w_{xMxM} = w_{xMxM}\), or

\[
F_{xycy} \frac{\partial y_C}{\partial xM} + F_{xycy} \frac{\partial y_M}{\partial xM} = F_{xycy} \frac{\partial y_C}{\partial xC} + F_{xycy} \frac{\partial y_M}{\partial xC} \tag{5}
\]

Apart from being a technical condition, (5) also has implications for sorting, highlighting a crucial difference to one-dimensional settings: With multiple dimensions, there is a stronger link between technology and assignment. The equilibrium assignment (i.e. the Jacobian of the matching function) does not only depend on the signs of the cross partial derivatives, \(F_{xy}\), but also on their strength. According to (5), changing the strength (but not the signs) of \(F_{xy}\) will induce worker reallocation (i.e. a change in the Jacobian elements) without necessarily violating PAM or NAM.

In the one-dimensional setting, there is no integrability condition because the wage is the solution to a single ordinary differential equation. In such a setting, the assignment depends only on the sign of \(F_{xy}\), not on its level: supermodularity (submodularity) of the technology implies PAM (NAM). Given PAM (NAM), there exists a unique measure-preserving increasing (decreasing) map of skills to productivities, which can be pinned down by labor market clearing alone. Under PAM, this map is given by \(y = G^{-1}(H(x))\), where \(x\) and \(y\) are scalars. However, with multiple traits, there is no complete order of types. Hence, there is no unique measure-preserving positive (or negative) assortative map of skills to productivities that can be computed from labor market clearing alone. Therefore, the optimal assignment must be jointly determined by labor market clearing and the firm’s problem. This is central to the closed form derivation below.

\(^{17}\)For the linear system of first-order partial differential equations (3), there is only one integrability condition. It is given by the commutativity of mixed partial derivatives. See Frobenius’ Theorem stated in Lemma 1 (Appendix).
3 Quadratic-Gaussian Model

A main goal of this paper is to apply this multi-dimensional sorting framework to the phenomenon of biased technological change. Towards achieving this objective, this section specifies the environment to Gaussian distributions and bi-linear technology, and develops a technique to explicitly compute the assignment and wage functions following similar steps as in one-dimensional decentralized assignment models. The closed form as well as the above notion of PAM are then used to analyze the economics of multi-dimensional sorting, including equilibrium properties and comparative statics.

3.1 Environment

Let skills \( (x_C, x_M) \) and productivities \( (y_C, y_M) \) follow bivariate standard normal distributions:

\[
\begin{bmatrix}
  x_C \\
  x_M
\end{bmatrix}
\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}\right),
\]

\[
\begin{bmatrix}
  y_C \\
  y_M
\end{bmatrix}
\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_y \\ \rho_y & 1 \end{bmatrix}\right).
\]

Denote the bivariate distribution functions of skills and productivities by \( \Phi_x \) and \( \Phi_y \), respectively. Assume, \( \rho_x, \rho_y \in (-1, 1) \). The skill correlation \( \rho_x \) indicates how specialized the workforce is, with lower \( \rho_x \) pointing to more specialized workers who are either good in cognitive or in manual work; similarly for the productivity correlation \( \rho_y \). I focus on the bi-linear technology

\[
F(x_C, x_M, y_C, y_M) = \alpha x_C y_C + \beta x_M y_M = \alpha(x_C y_C + \delta x_M y_M)
\]  

(6)

where \( \alpha \) and \( \beta \) are task-weights that indicate the level of worker-job complementarities or substitutabilities. Notice that \( \delta = \frac{\beta}{\alpha} \) indicates the relative level of complementarities across tasks. Without loss of generality, set \( \alpha \geq \beta \) such that \( \delta \in [0, 1] \), meaning that worker-firm complementarities in the cognitive task are weakly stronger than in the manual task.\(^{18}\) Technology (6) captures that there is within-task complementarity but between-task complementarity is shut down, implying that \( D_x^2 F \) is a diagonal \( P \)-matrix so that the equilibrium assignment is unique and satisfies PAM.\(^{19}\) These properties will prove useful in the construction of the equilibrium. Note that not all assumptions from the general model that ensure differentiability of equilibrium in Proposition 1(ii) are satisfied here (e.g. the supports of \( X \) and \( Y \) are not bounded). This, however, will not post a problem since I construct the unique equilibrium below which is indeed differentiable.

18 Nothing hinges on this restriction but it simplifies the interpretation.

19 Analogously to the general model, the firm’s problem is given by: \( \max_{(x_C, x_M) \in X} \alpha(x_C y_C + \delta x_M y_M) - w(x_C, x_M) \).
copulas (see Online Appendix 1.1 and 1.2). However, here I focus on the simplest environment that conveys the full intuition. I solve this assignment problem in two steps. First, I construct the equilibrium assignment and then the wage schedule that supports it.

### 3.2 The Equilibrium Assignment

The objective is to compute the equilibrium assignment $\mu$ in closed form. It must be consistent with both labor market clearing and firm optimality. Due to the incomplete order of types in the multi-dimensional setting, there are many measure-preserving ways of how to match workers with firms in a positive assortative way. What matters for pinning down the unique optimal assignment is therefore not only the sign (like in 1D) but also the relative strength of skill-productivity complementarities across tasks, captured by $\delta$. This will be the main difficulty in solving for the assignment.

I solve this assignment problem in several steps. I first temporarily convert the two-dimensional problem to two separate one-dimensional problems. Second, I ‘guess’ that the matching in the transformed space is PAM. This step of matching up the marginal distributions in an increasing way to obtain PAM is familiar from solving 1D assignment problems. I then re-transform the variables. Third, I ensure that this assignment is in line with firm optimality for any level of relative complementarities across tasks $\delta$. Last, I verify that this assignment is in line with PAM. By Proposition 1, it is then the unique equilibrium assignment. In more detail (see also Appendix 8.2):

I first apply a measure-preserving transformation that un-correlates the Gaussian variables. In particular, let $x$ be a $p$-variate random vector with mean $m_x$ and nonsingular covariance matrix $\Sigma$. Then, $z = \Sigma^{-\frac{1}{2}}(x - m_x)$ (8)

has mean 0 and covariance matrix $I_p$. Matrix $\Sigma^{-\frac{1}{2}}$ is the inverse of a square root of the covariance matrix, i.e. $\Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T = \Sigma$. Denote by $\Sigma_x$ (respectively $\Sigma_y$) the covariance matrix of skills (respectively productivities). Apply (8) to the standard bivariate normal skills and productivities $z_x = [z_{xc}\ z_{xm}] = \Sigma_x^{-\frac{1}{2}}[x_c\ x_m]$ and $z_y = [z_{yc}\ z_{ym}] = \Sigma_y^{-\frac{1}{2}}[y_c\ y_m]$ (9)

where $z_x$ and $z_y$ are vectors of uncorrelated skills and productivities, respectively. The labor market clearing condition can now be specified in terms of uncorrelated variables, which is consistent with labor market clearing in $(x, y)$ because the applied transformation is measure-preserving. Since the equilibrium assignment will need to satisfy PAM, I map skills to productivities in an increasing way

$(1 - \Phi(z_{yc}))(1 - \Phi(z_{ym})) = (1 - \Phi(z_{xc}))(1 - \Phi(z_{xm}))$ (10)

where $\Phi$ denotes the standard normal c.d.f. The interpretation of (10) is that if firm $(z_{yc}, z_{ym})$ matches with worker $(z_{xc}, z_{xm})$, then the mass of workers with better skills than $(z_{xc}, z_{xm})$ must be equal to the mass of firms that are more productive than $(z_{yc}, z_{ym})$ (in line with PAM).²⁰

²⁰At this stage, it is only a guess that market clearing in transformed variables $(z_x, z_y)$, which has been specified in
The market clearing condition (10) implicitly defines the vector-valued matching function of transformed variables, denoted by \( \mu^* : \mathbb{R}^2 \to \mathbb{R}^2 \). The objective is to back out two real-valued assignment functions of (10). In principle, there are many possible ways to match up the marginals in (10) but, to obtain PAM, the natural way is to do it within the cognitive and within the manual dimension:

\[
\Phi(z_{yi}) = \Phi(z_{xi}) \quad \forall \quad i \in \{C, M\}
\]

System (11) gives \( z_y = \Phi^{-1}\Phi(z_x) \) and can be re-transformed into original variables, explicitly solving for firm attributes \( y \) as a function of worker skills \( x \),

\[
\begin{bmatrix}
y_C^* \\
y_M^*
\end{bmatrix} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}}
\begin{bmatrix}
x_C \\
x_M
\end{bmatrix}
\]

where \( D_{xy^*} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}} \) is the Jacobian of the matching function. System (12) is the candidate equilibrium assignment, mapping bivariate skills into bivariate productivities. By (10), it is measure-preserving. Notice, however, that a covariance matrix has an infinite number of square roots. Hence, there are many matchings that satisfy market clearing and are potentially in line with PAM. How to pick the optimal one? I use the degree of freedom in computing these square roots to take into account the firm’s optimal choice, which depends on the relative level of skill-productivity complementarities, captured by \( \delta \). The Appendix shows how \( \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}} \) can be parameterized by \( \delta \), such that the resulting assignment is consistent with the firm’s optimality for any level of complementarities across tasks.

**Proposition 3 (Equilibrium Assignment)** The unique equilibrium assignment \( \mu \) is given by

\[
\begin{bmatrix}
y_C^* \\
y_M^*
\end{bmatrix} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}}
\begin{bmatrix}
x_C \\
x_M
\end{bmatrix} =
\begin{bmatrix}
J_{11}(\rho_x, \rho_y, \delta) & J_{12}(\rho_x, \rho_y, \delta) \\
J_{21}(\rho_x, \rho_y, \delta) & J_{22}(\rho_x, \rho_y, \delta)
\end{bmatrix}
\begin{bmatrix}
x_C \\
x_M
\end{bmatrix}
\]

where

\[
\begin{bmatrix}
J_{11}(\rho_x, \rho_y, \delta) & J_{12}(\rho_x, \rho_y, \delta) \\
J_{21}(\rho_x, \rho_y, \delta) & J_{22}(\rho_x, \rho_y, \delta)
\end{bmatrix} =
\begin{bmatrix}
\frac{1+\delta \sqrt{1-\rho_y^2}}{\sqrt{1-\rho_y^2}} \\
\frac{\sqrt{1+2\delta(\rho_x, \rho_y, \sqrt{1-\rho_y^2})+\delta^2}}{\sqrt{1-\rho_y^2}} \\
\rho_y - \rho_x \sqrt{1-\rho_x^2} & \frac{\sqrt{1+2\delta(\rho_x, \rho_y, \sqrt{1-\rho_x^2})+\delta^2}}{\sqrt{1-\rho_x^2}} \\
\frac{1+2\delta(\rho_x, \rho_y, \sqrt{1-\rho_x^2})+\delta^2}{\sqrt{1-\rho_x^2}} & \rho_y - \rho_x \sqrt{1-\rho_x^2}
\end{bmatrix}
\]

**Remark 1** For \( \delta = 1 \): \( J_{11} = J_{22} \) and \( J_{12} = J_{21} \). For \( \delta = 0 \): \( J_{22} \neq J_{11} = 1 \) and \( J_{21} \neq J_{12} = 0 \). For \( \delta \in (0, 1) \): the assignment lies in between these cases. Square roots, \( \Sigma_y^{\frac{1}{2}}, \Sigma_x^{\frac{1}{2}} \), are obtained from a rotation of the spectral square roots, ranging between the spectral (if \( \delta = 1 \)) and the Cholesky square root (if \( \delta = 0 \)).

See Appendix 8.2.2 for the proof. With symmetric technology (\( \delta = 1 \)), the equilibrium assignment is fully symmetric across the two tasks. The spectral square root, which is the unique symmetric positive definite square root of the covariance matrix, is used to compute this assignment. In the line with PAM, gives rise to an assignment in original variables \((x, y)\) that also admits PAM. I will verify this below.

\(^{21}\text{This is because a covariance matrix is a symmetric positive-definite matrix.}\)
most asymmetric case \((\delta = 0)\), only the cognitive task matters for production. The Cholesky square root (which is the unique lower triangular square root and hence asymmetric) delivers an asymmetric assignment, which is optimal when technology exhibits extreme asymmetries. Last, when there are some asymmetries in technology \(\delta \in (0, 1)\), then the assignment is in-between these two polar cases.

This closed form assignment which is a linear map from skill to productivity bundles confirms a key result of multi-dimensional sorting that was already hinted at in Section 2: There is a much stronger link between technology and assignment compared to the one-dimensional case, where the matching function only depends on the underlying distributions. Unlike in 1-d, here not only the sign of skill-productivity complementarities matters but also the relative level of these complementarities enters the assignment, captured by a slope of the matching function that adjusts to the level of \(\delta\).

### 3.3 The Equilibrium Wage Function

I close the model by computing the wage function that supports the assignment found above.

**Proposition 4 (Equilibrium Wage Schedule)** The equilibrium wage function is given by

\[
w(x) = \frac{1}{2} \alpha x^T \tilde{J} x + w_0
\]

where \(w_0\) is the constant of integration.

Note that \(\tilde{J}\) is a matrix of parameters that closely relates to the equilibrium assignment: All entries of \(\tilde{J}\) are identical to those in \(J\) except \(\tilde{J}_{21} = \delta J_{21}\), making \(\tilde{J}\) a symmetric positive-definite matrix. For the special case of symmetric tasks \((\delta = 1)\), the two coincide, \(\tilde{J} = J\), emphasizing the tight link between allocation and wages that is typical for assignment models. The wage function is a quadratic form in standard normal variables, which allows me to compute the moments of the wage distribution in closed form. The next sections extensively discuss the properties of the wage function and how they depend on distributions and technology through the assignment.

A final remark on the equilibrium is that the one-dimensional equilibrium is subsumed by the two-dimensional model as a special case: For perfect correlations, \(\rho_x, \rho_y \to 1\), the Jacobian of the matching function \(J\) in (13) becomes the identity matrix (i.e. the assignment is independent of relative complementarities in production) and the wage collapses to a simple quadratic function in skill.

### 3.4 Properties of the Equilibrium

This section discusses equilibrium properties of the benchmark case with symmetric tasks \((\delta = 1)\). The next section on task-biased technological change (task-biased TC hereinafter) examines in detail the case of asymmetric task weights \((\delta \neq 1)\).

To analyze the sorting properties of this equilibrium as well as comparative statics in the most nuanced way, it is useful to not only be able to determine the sign of sorting (PAM versus NAM) but also its strength. To this end, I first define the concepts of perfect assortativeness and mismatch, which naturally map into properties of the Jacobian of the matching function. The Jacobian (with its \(P\)-matrix property) was already shown to be the key object for characterizing the sign of sorting.
in this multi-dimensional environment. This section shows that the Jacobian can take on different forms of $P$-matrices to reflect varying strengths of PAM.

**Definition 4 (Perfect Assortativeness and Mismatch)** An assignment in task $i \in \{C, M\}$ is perfectly assortative if $x_i - y_i^*$ for all matches (i.e. if the Jacobian of $\mu$ is the identity matrix in row $i$). In turn, an assignment exhibits mismatch if $x_i \neq y_i^*$ for almost all matches (i.e. if the Jacobian of $\mu$ deviates from the identity matrix in row $i$).

Perfect assortativeness means that workers’ skills perfectly match firms’ skill demands in a certain task $i$. If the Jacobian of the matching function is the identity matrix, then sorting in both task dimensions is perfectly positive assortative. In turn, structural mismatch is defined as deviations from perfect assortativeness. Mathematically, I define mismatch by a Jacobian that departs from the identity matrix in task $i$, with non-zero off-diagonal element and non-one diagonal element. In what follows, mismatch in task $i$ is said to be increasing the larger is the element-by-element distance between the Jacobian and the identity matrix in task (i.e. row) $i$. Therefore, an assignment is characterized by the largest degree of mismatch in task $i$ if both elements of the Jacobian in row $i$ exhibit the largest distance to the identity matrix. Note that defining mismatch via properties of the Jacobian is closely related to defining mismatch in terms of mean squared deviation between skills and job attributes, $E[(y_i^* - x_i)^2] = E[(J_{11}x_C + J_{12}x_M - x_i)^2]$. Lemma 4 (Appendix) shows that the underlying distributions that maximize mismatch in line with Definition 4 are also those that maximize mismatch according to $E[(y_i^* - x_i)^2]$. Notice that mismatch in this frictionless economy has nothing to do with inefficiencies. Instead, it refers to the misfit between workers’ and firms’ traits.

To shed light on how sorting depends on the underlying distributions and, below, how technological change interacts with the properties of the distributions, it will also prove useful to define a measure of the similarity of skill supply and demand. For illustration, I will focus on two polar cases:

**Definition 5 (Perfect and Poorest Fit of Supply and Demand)** The fit between supply and demand is perfect if $\rho_x = \rho_y$ and it is poorest at $\sup_{\rho_x, \rho_y}|\rho_x - \rho_y|$.

Under the assumption that $\rho_x, \rho_y \in (-1, 1)$, the largest discrepancy of skill supply and demand occurs at $\sup_{\rho_x, \rho_y}|\rho_x - \rho_y| = 2$, i.e. workers are completely specialized ($\rho_x \to -1$) but firms want generalists ($\rho_y \to 1$), or vice versa.

I can now state the following properties of equilibrium sorting.

**Proposition 5 (Equilibrium Sorting)** (i) The equilibrium assignment exhibits PAM. (ii) For $\delta = 1$, if there is a perfect fit of skill supply and demand, then sorting is perfectly assortative in both tasks, i.e. $y_C = x_C$ and $y_M = x_M$. In turn, if the fit of skill supply and demand is poorest, then mismatch in both tasks is largest.

Regarding the sign of sorting, the equilibrium assignment satisfies PAM (by construction), meaning that workers with more intellectual skills work in jobs that value them and similarly on the manual dimension. This is immediate from the technology that features worker-firm complementarities in each task. In turn, the strength of assortativeness depends on the underlying distributions, illustrated...
by two polar cases: First, when skill supply and demand are identical \((\rho_x = \rho_y)\), then every worker obtains the perfect job. On the other hand, if there is a large discrepancy between the skills that are needed and those that are supplied (i.e. if \(|\rho_x - \rho_y| \to 2\)), then the labor market can only clear under considerable mismatch, with every worker being in a job for which he is either under or overqualified. This result demonstrates that the proposed multi-dimensional notion of PAM captures a rich set of assignments, ranging from perfect assortativeness to significant mismatch.

Figure 1 provides a graphical interpretation, displaying contour lines of two standard normal distributions for various skill and productivity correlations. To illustrate, assume that workers are represented by blue contour lines and firms by red ones. The left panel shows a poor fit of skill supply and demand. The labor market clears under PAM but matches are characterized by a poor worker-firm fit (illustrated for a few firms – dots on the red contour lines, far out – and workers – dots on the blue curves, further in – where dots of equal color match). On the other extreme, the right panel displays a perfect fit of skill supply and demand, which would lead to perfectly assortative matches between all workers and jobs in both tasks \((y_C^* = x_C, y_M^* = x_M)\).

Figure 1: Contour Plots of Skill and Productivity Distributions

Figure 2, which plots the Jacobian elements of the matching function, offers an alternative graphical interpretation, in line with the way sorting and mismatch were defined. It plots firms’ productivity in the cognitive task (left) and the manual task (right) as a function of both skills. The constant slopes of the lines stem from the linearity of the assignment functions. The slopes of the solid lines, \(J_{11} = \partial y_C^* / \partial x_C\) and \(J_{22} = \partial y_M^* / \partial x_M\), indicate the assortativeness of the match (showing how strongly worker and firm attributes relate within tasks). The slopes of the dashed lines, \(J_{12} = \partial y_C^* / \partial x_M\) and \(J_{21} = \partial y_M^* / \partial x_C\), indicate the degree of mismatch in a pair (showing how strongly worker and firm attributes relate between tasks).

The assignment in the upper panel is perfectly assortative: the solid lines have slope one and the dashed lines slope zero, i.e. the Jacobian of \(\mu\) is the identity matrix with \(y_C^* = x_C\) and \(y_M^* = x_M\). This assignment results when underlying distributions are identical (corresponding to \(\rho_x = \rho_y\), Figure 1, right panel). The lower panel displays the other extreme. Here matches are characterized by maximum mismatch with the wrong skill dimensions contributing significantly to the match, indicated by similar slopes of solid and dashed lines, i.e. by a Jacobian that significantly differs from the identity matrix (corresponding to Figure 1, left panel, where the underlying distributions differ significantly). Notice that despite considerable mismatch, PAM is satisfied (i.e. positively sloped and steeper solid lines).
The next result summarizes selected properties of the equilibrium wage function.

**Proposition 6 (Equilibrium Wages)** (i) Wages are convex in skills. (ii) The wage distribution is positively skewed.

The central idea of assignment models is that the allocation of workers to firms shapes wages, and hence, wage inequality. Positive assortativeness (implying that $\tilde{J}$ in (14) is a symmetric P-matrix or positive definite) causes wages to be convex. Convex wages mean that workers with large (absolute) quantities of skills earn disproportionally more than workers with small (absolute) quantities of skills. Importantly, skills are not the only force behind high earnings. Due to PAM, skill differences are magnified because skilled workers are matched to more productive firms, convexifying the wage schedule.

An alternative commonly used measure of wage inequality is the skewness of the wage distribution. In line with many empirical wage distributions, the model’s wage distribution is positively skewed, indicating that a large fraction of workers earns little while a small fraction earns disproportionally much. The force behind positive skewness is again PAM.

It can also be shown that the average performance of an economy depends on the assignment of workers to firms and thus on underlying distributions. The average wage (and also output) is maximized when skill supply and demand are perfectly aligned ($\rho_x = \rho_y$). Intuitively, at that point, every worker obtains the perfect firm match in both tasks. In turn, the economy performs most poorly on average when misalignment between skills and skill requirements is largest.

This section illustrated how the strength of positive sorting depends on an economy’s skill supply and demand distributions and how this feeds into wages. The next section revisits the key message from Proposition 3 that the assignment not only depends on distributions but also on technology, i.e. $\delta$.

---

22Note that it also follows from the above-cited Theorem 10.28 in Villani [2009] that, for quadratic technology, the assignment is the gradient of a convex (wage) function.
4 Biased Technological Change

This section uses the closed form to analyze the effects of task-biased TC on sorting and wages and also contrast them to the more commonly analyzed skill-biased TC. It will also show how these effects of technological change are mitigated or reinforced by the underlying distributions.

4.1 Task-Biased Technological Change

Task-biased TC (also referred to as TBTC) is viewed as an important force behind recent wage inequality shifts in the developed world. The idea behind it is that technological advances have replaced workers in performance of manual tasks but created stronger complementarities between skills and job attributes in cognitive tasks – in line with the leading example from the introduction. The literature also refers to this technological change as routinization, where workers performing routine tasks are increasingly substituted by computers and machines.\(^{23}\) Notice that task-biased TC does not imply that the prevalence of routine tasks in the production process has diminished over time – quite the opposite (Acemoglu and Autor [2011]). What has changed is the technology to perform them.

Even though two intrinsically different skills are involved (manual and cognitive), the literature has analyzed task-biased TC only in one-dimensional settings. Contrary to these models, my model does not assume that manual skills are only used by medium-skilled workers. Instead, I make the natural assumption that both types of skills are used on every job yet in different proportions.\(^{24}\)

In the model, task-biased TC can be captured by a relative decrease in complementarities of manual inputs. Recall the technology \(F(x_c, x_M, y_C, y_M) = \alpha(x_C y_C + \delta x_M y_M)\) where \(\delta = \frac{\beta}{\alpha}\) indicates relative complementarities in the manual task. Consider a change from \(\delta\) to \(\delta'\) such that \(\delta' < \delta\). Then, \(\delta'\) is called task-biased relative to \(\delta\), with the bias favoring the cognitive task. Moreover, to obtain clean analytical results, this section focuses on cases where either \(\rho_x, \rho_y \leq 0\) or \(\rho_x, \rho_y \geq 0\).\(^{25}\)

The next result summarizes the effect of task-biased TC on the equilibrium assignment.

**Proposition 7 (TBTC and Sorting)** Suppose there is cognitive task-biased TC \((0 < \delta' < \delta \leq 1)\):

(i) Mismatch decreases (increases) in the cognitive (manual) task. (ii) As \(\delta' \to 0\), cognitive sorting becomes perfectly assortative at the expense of maximum manual mismatch. (iii) Task-biased TC has no effect on sorting if there is a perfect fit of supply and demand and it reduces cognitive mismatch the most if the fit of supply and demand is poorest.

For most underlying distributions, workers do not obtain their perfect job matches in equilibrium, simply because it is not feasible. However, in a multi-dimensional world, agents can decide in which dimension (cognitive or manual) sorting is more important. This decision depends on technology and, in particular, on relative levels of worker-firm complementarities across tasks. In the task with larger complementarities, perfect assortativeness is desired whereas in the task with weaker complementarities, mismatch is tolerated. I call this the mismatch-assortativeness trade-off across tasks.

\(^{23}\)See, e.g. Autor et al. [2003], Autor et al. [2006] and Autor and Dorn [2013]. There is a close mapping between manual and routine skills on the one hand, and between cognitive and non-routine skills on the other.

\(^{24}\)This is similar to the skill weights approach proposed by Lazear [2009].

\(^{25}\)Besides analytical convenience, the restriction captures the empirically relevant case for the US. See below.
Task-biased TC, which is defined as a change in relative complementarities, puts this trade-off to work. Consider, for instance, the development and increasing use of computers, which makes cognitive skills more productive in jobs that demand them. At the same time, new computers perform several manual tasks, replacing workers with manual know-how. As a result, sorting becomes more pronounced in the cognitive task (e.g. highly qualified computer programmers relocate to the best computer producing firms) at the expense of mismatch in the manual task (e.g. the production worker from the leading example is now employed as a car salesman which fits his cognitive but not his manual skills), part (i). The amount of worker reallocation depends on both, the size of the technology shift (ii) and the underlying distributions (iii). How the size of the shift matters is illustrated in Figure 3, which has a similar structure as Figure 2. The upper panels plot cognitive sorting, i.e. $y_C$ as a function of $x_C$ and $x_M$ before (left panel) and after (right panel) task-biased TC, as $\delta \to 0$. The lower panels plot manual sorting, i.e. $y_M$ as a function of $x_M$ and $x_C$, and have the same structure.

![Figure 3: Effects of TBTC on Sorting in Cognitive (upper panel) and Manual Dimension (lower panel)](image)

Recall that the slope of the straight lines indicates how strong sorting is within tasks. The slope of the dashed lines is an indicator of how strong sorting forces are between tasks. Due to the bi-linear technology with only within-task complementarities, the within-force is desirable whereas the between-force is not (it reflects mismatch). For illustration, before task-biased TC, cognitive and manual tasks receive identical weights in production ($\delta = 1$), hence, the left panels are identical. Going from left to right, relative complementarities in the cognitive task increase: the economy converges to the perfectly assortative allocation in the cognitive task ($y_C^*=x_C$). But this comes at the expense of significant mismatch between workers’ skills and firms’ skill needs in the manual task, with manual productivity responding even more strongly to changes in the cognitive than in the manual skill.
Besides the size of the technology shock, what also matters for the sorting response are the underlying distributions, part (iii). If skill supply and demand are perfectly aligned, task-biased TC has no effect on the assignment. This is because sorting in both tasks is perfectly assortative to start with. Thus, worker-firm assignment in the cognitive task cannot further improve as $\delta$ decreases. On the other hand, the amount of re-sorting in response to task-biased TC is largest when skill supply and demand differ considerably. In this situation, the initial assignment exhibits mismatch in both tasks. Hence, there is much to gain from improving cognitive sorting in response to task-biased TC.

It follows from this discussion that there are two sources of structural mismatch in the economy, technology and distributions. The first source stems from asymmetries in production technology ($\delta \neq 1$, Proposition 7) and affects sorting across tasks asymmetrically. The second is due to discrepancy between skill and productivity distributions or, in other words, between supply and demand ($\rho_x \neq \rho_y$, Proposition 5) and affects sorting symmetrically in both tasks. It arises because the frictionless labor market must clear no matter how different skill and productivity distributions are.

The next result summarizes how these assignment changes affect wages and inequality.

**Proposition 8 (TBTC and Wages)** Suppose there is cognitive task-biased TC ($0 < \delta' < \delta < 1$):

(i) The wage distribution becomes more positively skewed. Further, if task-biased TC is driven by an increase in cognitive complementarities (i.e. if $\delta \downarrow$ due to $\alpha \uparrow$) then wage dispersion also increases.

(ii) If $|\rho_x| \leq |\rho_y|$, wages become more convex in cognitive but less convex in manual skills. These curvature changes are smallest if there is a perfect fit of supply and demand. (iii) There are rank switchings in the wage distribution where cognitive specialists surpass manual specialists in the wage ranking.

Task-biased TC increases wage inequality if measured by the skewness of the wage distribution. In fact, the skewness is minimized absent of any technological change ($\delta = 1$). In turn, the effect on the variance is ambiguous. While the skewness responds to asymmetries in complementarities across tasks, the variance increases in the level of complementarities in both tasks. Since cognitive task-biased TC can either be driven by an increase in cognitive complementarities or by a decrease in manual ones, the overall effect depends on the relative magnitude of these two changes. However, a sufficient condition for wage dispersion to increase is that task-biased TC is driven by increased cognitive complementarities $\alpha$, part (i).

Task-biased TC also affects wage inequality by altering the curvature of the wage schedule, part (ii): Wages convexify in cognitive skills but become less convex in manual skills (for the empirically relevant case of $|\rho_y| \geq |\rho_x|$, see below). Intuitively, this technology shift favors workers with high levels of cognitive skills (who are now sorted into jobs that better fit these skills), driving up wage inequality in the cognitive dimension. On the other hand, manual workers are adversely affected by this shock. Those with strong manual skills are hit most severely (since a skilled production worker ends up in a similar job of low cognitive requirements – e.g. car salesman – as a mediocre production worker), compressing wage inequality in this dimension. The magnitude of these effects depends on the amount of worker-job re-sorting in response to task-biased TC, which in turn depends on the underlying distributions. If there is little mismatch in initial worker-job matches (which is the case when $\rho_x \approx \rho_y$), then both re-sorting and wage inequality changes are smallest. As will become clear
in the empirical part below, this (de-)convexification result is an important ingredient for generating wage polarization, which refers to expanding upper tail but compressing lower tail inequality.

Part (iii) reveals an interesting difference between this model and the one-dimensional case: Here, technological change causes workers to switch ranks in the wage distribution. Before task-biased TC, an expert in manual dexterity earned more than a mediocre worker specialized in cognitive skills. After technological change, this ranking is reversed because the manual expert now performs a job—e.g., car salesman—that values his weak (cognitive) skill more than his strong (manual) skill.

4.2 Skill-Biased Technological Change

A downside of this model is that it generates a non-monotonic wage schedule in skills, which would be difficult to reconcile with the data. To make the model more suitable for empirical analysis, I augment the production function by non-interaction skill terms and a constant, given by

\begin{equation}
F(x_C, x_M, y_C, y_M) = x_C(\alpha y_C + \lambda) + x_M(\beta y_M + \eta) + f_0 = \alpha(x_C y_C + \delta x_M y_M) + \lambda(x_C + \kappa x_M) + f_0 \tag{15}
\end{equation}

where \( \delta = \frac{\alpha}{\beta} \) is the relative manual task weight, \( \lambda, \eta \) are skill weights, \( \kappa = \frac{\eta}{\lambda} \) is the relative manual skill weight and \( f_0 \) is a constant.\(^{27}\) The assignment is unaffected by this change in technology (since it is pinned down by the interaction terms) but the wage function differs and becomes a non-homogenous quadratic form in standard normal variables,

\begin{equation}
w(x_C, x_M) = \frac{1}{2} \alpha(x - h)^T \tilde{J}(x - h) + C = \alpha \left( \frac{1}{2} J_{11} x_C^2 + J_{12} x_C x_M + \frac{1}{2} \delta J_{22} x_M^2 \right) + \lambda(x_C + \kappa x_M) + w_0. \tag{16}
\end{equation}

See Section 8.3.2 in the Appendix, for the derivation and expressions \( h, \tilde{J}, C \). Non-interaction skill terms can shift the location of the minimum wage to the left, allowing for a wage schedule that is increasing \( \forall x_C \geq \underline{x}_C \) and \( x_M \geq \underline{x}_M \), where \( \underline{x}_C, \underline{x}_M \) are, for instance, the lowest observed skills in the data.\(^{28}\) Moreover, the included constant \( f_0 \) translates into a non-zero constant in the wage function \( w_0 \), guaranteeing non-negative wages to all agents in the economy.\(^{29}\)

Technology (15) gives rise not only to a more realistic wage schedule. It also allows for a sensible definition of skill-biased technological change (skill-biased TC or SBTC hereinafter), independently of task-biased TC that works through complementarities in production. Consider a change in relative manual skill weight from \( \kappa \) to \( \kappa' \) with \( \kappa' < \kappa \). Then, \( \kappa' \) is called skill-biased relative to \( \kappa \), with the bias favoring cognitive skills. This shift increases the productivity of cognitive skills independent of a job’s cognitive skill demands. For instance, advancements in communication technology (e.g., Google) benefit both the secretary and the CEO even though their tasks require different levels of cognitive skill.

---

\(^{26}\)Under the assumed technology, the wage is folded around \((0,0)\), e.g. workers \((-1,-1)\) and \((1,1)\) earn the same.

\(^{27}\)Including additional non-interaction \( y \)-terms in the technology would not affect wages. Furthermore, I assume that \( f_0 \) equals the outside option of the workers, implying that \( w_0 = f_0 \), i.e. \( f_0 \) equals the wage of the least productive worker.

\(^{28}\)Sufficient conditions for wages to be monotone in skills above some skills \( \underline{x}_C, \underline{x}_M \) are: \( \alpha J_{11} \underline{x}_C + \alpha J_{12} \underline{x}_M + \lambda \geq 0 \), \( \beta J_{22} \underline{x}_M + \alpha J_{12} \underline{x}_C + \eta \geq 0 \).

\(^{29}\)Contrary to the baseline technology, when including non-interaction terms, wages could otherwise become negative.
Proposition 9 (SBTC) Suppose there is cognitive skill-biased TC ($\kappa' < \kappa$). (i) The assignment is unaffected. (ii) The curvature of the wage function is unaffected. (iii) The effect on the variance is ambiguous but if skill-biased TC is driven by increased cognitive productivity ($\kappa \downarrow$ due to $\lambda \uparrow$), then wage dispersion increases whenever $\lambda > -\eta \rho_x$. The effect on the skewness is generally ambiguous.

Skill-biased TC has no impact on the assignment, reiterating that what matters for the assignment is the relative level of complementarities across tasks. Moreover, from (16) it is clear that it also has no impact on the curvature of the wage function, which solely depends on task-bias parameters and the assignment. This will be important for the differential impact of task- versus skill biased technological change on wages in the data (see below). Finally, similar to task-biased TC, the effect of skill-biased TC on wage dispersion is ambiguous since the variance tends to increase in the productivity of both cognitive and manual skills ($\lambda$ and $\eta$). A sufficient condition for the variance to increase in skill-biased TC is that it is driven by increased cognitive productivity (for the empirically relevant case below with negative skill correlation, $\rho_x < 0$, and relatively large cognitive skill weight, $\lambda > \eta$). The next section brings the model to the data, in order to (a) quantify skill-biased and task-biased TC over time and (b) disentangle their roles in observed assignment and wage inequality shifts.

5 Quantitative Analysis

I first I estimate the model by Maximum Likelihood (ML), providing insights on how technology in the US has evolved over time. Then I use various counterfactual experiments to decompose wage inequality shifts into those driven by (i) task-biased technological change, (ii) skill-biased technological change and (iii) changes in the underlying distributions.

5.1 The Data

My main data source is the National Longitudinal Survey of Youth (NLSY). This data set contains information on labor market outcomes of two cohorts: The first cohort starts in 1979 (NLSY79), interviewed every year until 1994 and since then biennially. Outcomes of the second cohort are reported annually since 1997 (NLSY97). To assess the role of technological and distributional changes in US wage inequality shifts, I compare labor market outcomes across these two cohorts. Using two cohorts instead of a single one allows me to distinguish changes that are attributable to technological change from those that are due to life cycle phenomena.\textsuperscript{30}

The main reason for using the NLSY is that, contrary to other datasets like CPS, PSID or IPUMS, it contains detailed information on respondents’ training and educational degrees that I will use to construct a bivariate skill supply distribution. I also supplement the NLSY by the O*NET to learn about occupational skill requirements. This data will be crucial for constructing a skill demand distribution, where I interpret occupations as the empirical counterpart of the model’s firms.\textsuperscript{31} The analysis in this paper covers the period from 1990 to 2010. In particular, I compare labor market outcomes of employed workers aged between 27 and 29 in 1990/91 from the NLSY79 to those of employed workers aged between 27 and 29 in 1990/91 from the NLSY79 to those of employed workers aged between 27 and 29 in 1990/91 from the NLSY79 to those of employed workers aged between 27 and 29 in 1990/91 from the NLSY79 to those of employed workers aged between 27 and 29 in 1990/91 from the NLSY79 to those of employed

\textsuperscript{30}In this exercise, I rely on the assumption that individuals’ unobservable characteristics do not change across cohorts.

\textsuperscript{31}The O*NET is the US Department of Labor Occupational Characteristics Database.
workers of the same age in 2009/10 from the NLSY97. When selecting periods and age groups, three considerations played a role: First, I choose age groups in which most workers have finished their formal education and are working. Second, to enable a clean comparison between the two cohorts, their age distributions must be similar. Tables 5-7 (Appendix 8.5.1) show that this consideration is the main determinant behind the choice of age groups and years. During the selected periods, each age (i.e. 27, 28 and 29) accounts for roughly one third of the sample. Overall, I am restricted to young workers because the oldest workers in the NLSY97 are still young. Third, I pool observations from several years as well as workers from various ages to increase the number of observations in the estimation. Below, I also estimate the model on different samples to show robustness.

I restrict the sample to employed male and female workers in non-military occupations. I consider hourly wages, adjusted by the CPI. Additionally, my analysis requires measures of workers’ cognitive and manual skills \((x_C, x_M)\) as well as occupations’ cognitive and manual skill requirements \((y_C, y_M)\).

To construct these bivariate distributions, I combine information from the NLSY with O*NET data. O*NET provides detailed information on skill requirements for a large number of occupations. This information can be classified into two categories, manual and cognitive, and then aggregated to two requirements for each occupation. They indicate the level of skills needed to perform manual and cognitive tasks, which I interpret as the \((y_C, y_M)\)-bundle from my model (see Table 8 in Appendix 8.5.2 for examples).\(^{32}\) I then merge these scores into occupations of employed workers in the NLSY, which yields the bivariate skill demand distribution. Constructing the bivariate skill distribution is involved because data on manual skills are not readily available.\(^{33}\) Moreover, the literature provides little guidance on this issue.\(^{34}\) To impute agents’ manual and cognitive skills, I use responses from the NLSY observations on their college degrees, apprenticeships and vocational degrees, degrees of government programs and past on-the-job training paid by firms. I then find out for which occupations these various forms of education and degrees qualify. The final step is to combine this information with the O*NET data that indicates which cognitive and manual scores to attach to the occupations individuals qualify for and thus to the individuals’ education. From this information, I can proxy a manual and cognitive skill for each agent (see Appendix 8.5.2 for the details). After data cleaning and sample restrictions, I am left with around 3000 observations in the NLSY79 and with 4500 in the NLSY97.

It is important to note that, even though I factor in on-the-job training from past occupations to impute skills, my construction of the skill and skill demand distributions avoids a mechanically high correlation between contemporaneous skill supply and demand: First, and crucially, I use workers’ (university, apprenticeship and government program) degrees on top of formal occupational training to determine skills. This educational information is independent of whether the individual ever worked in a certain occupation. Second, I only use information on workers’ occupations if they have received formal training in that occupation (instead of using information on past occupations per se). Finally,

\(^{32}\)This data as well as the crosswalk linking O*NET occupational codes to NLSY occupational codes come from Sanders [2016]. Yamaguchi [2012] uses a similar approach to classify manual and cognitive occupational inputs.

\(^{33}\)To assess manual and cognitive skills, I do not want to rely on AFQT ability scores since, in my sample, they were assessed before any formal tertiary education takes place that is crucial for occupational choices.

\(^{34}\)Yamaguchi [2012] and Sanders [2016] estimate the bivariate skill distribution from their models. In turn, I aim to provide information on the skill distributions that is independent of the model.
I do not rely on occupational training in the current period but only on past training experiences.\textsuperscript{35}

The Appendix provides summary statistics of both skill supply and demand distributions. They show that changes in the marginal distributions across time are minor. In order to align the data with the model that features standard Gaussian distributions, I transform the empirical distributions into Gaussian copulas. By taking out the marginal characteristics (see Section 1.2, Online Appendix, for technical details), this transformation provides a way of studying dependency independently of what the marginals look like. To see how this transformation impacts the dependency structure, I report the correlations of transformed and original variables in Table 9 in the Appendix. Importantly, the correlations in the transformed data are similar to those in the original data. Table 9 further shows that manual and cognitive skills are negatively correlated, indicating that a worker with high cognitive skills has little manual dexterity and vice versa. In 1990/91, occupations’ skill requirements are more negatively correlated than skills. The interpretation is that US jobs demand workers with higher degrees of specialization than available workers can offer. In 2009/10, this order is reversed.

5.2 Estimation

I estimate the model by Maximum Likelihood (ML). To be able to do so, I introduce measurement error in both wage and assignment equations into this otherwise deterministic model.\textsuperscript{36}

The closed form solution is particularly useful for this estimation since it allows me to specify an exact expression for the likelihood function. Denote the parameter vector by $\theta = ((\alpha, \beta, \lambda, \eta, w_0), (s, t, u))$, which is to be estimated. The first set of parameters corresponds to technology parameters and the second set relates to measurement errors of the wage and assignment functions, respectively. The data vector is given by $z = (z_1, ..., z_n)$ where $\forall \ i = 1, ..., n$ observations, $z_i = (w_i, y_{C_i}, y_{M_i}, x_{C_i}, x_{M_i})$.

The log-likelihood function for this model is given by

$$
\ln L(\theta | z) = - \sum_{i=1}^{n} \left( w_i - \left( \frac{1}{2} \alpha J_{11} x_{C_i}^2 + \frac{1}{2} \beta J_{22} x_{M_i}^2 + \lambda x_{C_i} + \eta x_{M_i} + w_0 \right) \right)^2
$$

$$
- \sum_{i=1}^{n} \left( y_{C_i} - (J_{11} x_{C_i} + J_{12} x_{M_i}) \right)^2 - \sum_{i=1}^{n} \left( y_{M_i} - (J_{21} x_{C_i} + J_{22} x_{M_i}) \right)^2 - n \ln(stu) - \frac{3n}{2} \ln 2\pi
$$

(17)

where $J_{11}, J_{12}, J_{21}, J_{22}$ are the elements of the Jacobian of the matching function (see (13)), which are functions of skill supply and skill demand correlations $\rho_x, \rho_y$ (estimated from the empirical distributions) and the relative manual complementarity weight $\delta = \frac{\beta}{\alpha}$ (see Appendix 8.5.3 for details). Notice that an additional advantage of this model is that all parameters are identified.\textsuperscript{37}

\textsuperscript{35}Table 16 (Appendix) reports the reduced form assignment regression results and shows that the correlation between skills and skill demands is far from perfect. This shows that $x = y$ is not imposed by the way I measure skills.

\textsuperscript{36}Note that measurement error will be the only source of randomness. One can think of other ways of introducing noise, e.g. through unobserved heterogeneity or search frictions. The reason behind this choice is to keep the empirical model as much as possible in line with the theory in order to be guided by its predictions.

\textsuperscript{37}My model circumvents non-identification of similar linear-quadratic Gaussian models that arise due to collinearity (pointed out by Brown and Rosen [1982] and Ekeland et al. [2004]). There, the identification problem stems from an additional quadratic term in the production function. More generally, my model avoids such collinearity problem because the curvature of $w(x)$ in $x$ is not the same as the curvature of technology $F(x, y)$ in $x$. 

24
I estimate the model for each period separately. Table 10 (Appendix) reports the estimation results of the main specification in detail (using the sample of 27-29 year old employed workers in 1990/91 and 2009/10). It reveals that the model fit of the data is not perfect. The reasons are clear: First, the model is parsimonious with few parameters to match the data. Second, measurement error is the only reason why there is no perfect one-to-one matching, which clearly does not capture all the mismatch in the data (but has the advantage to keep the empirical application closely aligned with the theory).

5.3 Technological Change in the US

Identifying unobserved worker-firm complementarities from observed equilibrium outcomes has been of independent interest and the focus of a growing literature on the identification of sorting. Using my model as a measuring instrument, I can identify from data on wages and worker-job assignment the underlying technological determinants of the US economy and how they changed over time. Recall

\[ F(x_C, x_M, y_C, y_M) = \alpha x_C y_C + \beta x_M y_M + \lambda x_C + \eta x_M + f_0 \]

which is the specified production function, where \( \alpha, \beta \) are complementarity weights, \( \lambda, \eta \) are skill weights, and \( f_0 \) is a constant. Table 1 contains the ML-estimates of these technology parameters across cohorts. The estimation results suggest that the production technology features complementarities between worker and job attributes in both tasks and across periods (\( \alpha \) and \( \beta \) are positive). However, while the beginning of the 1990s were characterized by relatively strong complementarities in the manual task, at the end of the 2000s the opposite holds true, indicating task-biased TC in favor of cognitive tasks: Complementarities between cognitive worker and job attributes have significantly gone up whereas complementarities in manual inputs have strongly decreased. As a consequence, the relative complementarity weight \( \delta = \frac{\beta}{\alpha} \) has significantly declined.

<table>
<thead>
<tr>
<th></th>
<th>( \lambda )</th>
<th>( \eta )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( f_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990/91</td>
<td>1.686</td>
<td>-0.421</td>
<td>0.203</td>
<td>0.479</td>
<td>14.493</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.141)</td>
<td>(0.342)</td>
<td>(0.175)</td>
<td>(0.253)</td>
</tr>
<tr>
<td>2009/10</td>
<td>2.203</td>
<td>0.210</td>
<td>0.739</td>
<td>0.055</td>
<td>15.055</td>
</tr>
<tr>
<td></td>
<td>(0.152)</td>
<td>(0.152)</td>
<td>(0.198)</td>
<td>(0.154)</td>
<td>(0.169)</td>
</tr>
</tbody>
</table>

Table 1: Maximum Likelihood Estimates of Technology Parameters

Besides these shifts in relative task complementarities, there was also a change in the skill-bias of technology, indicated by changes in skill weights \( \lambda \) and \( \eta \). Between the 1990/91 and 2009/10, the US economy experienced cognitive skill-biased TC. The cognitive skill weight \( \lambda \) increased (while the manual weight \( \eta \) became statistically insignificant over this period). In sum, these estimates suggest that during the last two decades, the US faced two major technological shifts: first, a bias in favor of the cognitive task and, second, a bias favoring cognitive skills. Additionally, there was a positive trend (indicated by an increase in \( f_0 \)), which had an impact on all workers’ wages independent of their skills.

---

38See, for instance, Abowd et al. [1999], Eeckhout and Kircher [2011] and also the survey by Graham [2011].
39As before, under the assumption that \( f_0 \) equals the outside option of the workers, it holds that \( w_0 = f_0 \).
Appendix 8.5.4 shows that these estimation results on technological change are not driven by the particular sample of 27-29 year old workers but are robust to considering samples of different age groups and years (Table 11) and to weighting the observations differently (Tables 12 and 13). Moreover, the documented patterns of technological change are also present when estimating the model on data from a single cohort (NLSY79) only that in this case the technological shifts seem more pronounced, see Table 14. This suggests that using a single cohort may confound technological change with life cycle phenomena, with both forces having similar effects on labor market outcomes. Finally, notice that, due to data constraints, the second period of the estimation coincides with the Great Recession. To show that the estimated patterns of technological change are not simply a by-product of the crisis, I also re-estimate the model for the pre-crisis period 2005-2007 (Table 15). Comparing the pre-crisis period with the beginning of the 1990s from Table 10, task-biased TC is still apparent but, naturally, weaker than in the main specification because the NLSY97 workers are younger. Note that there have also been changes in the underlying distributions over time: workers have become more specialized over time (\( \rho_x \) = -0.38 in 1990/91 and \( \rho_x \) = -0.49 in 2009/10, Table 9, Appendix) but changes in skill demand were small.

5.4 The Role of Technological Change in US Sorting and Wage Inequality Shifts

Observed wage inequality shifts in the data can occur for many reasons. The advantage of estimating a structural model is that the effects of various sources can be disentangled. This section conducts counterfactual exercises to decompose the impact of task-biased TC, skill-biased TC, and changes in underlying distributions on wage inequality. For instance, to study how much of the change in wage inequality is due to task-biased TC alone, I keep both skill-bias parameters \( \lambda \) and \( \eta \) as well as distributional parameters \( \rho_x \) and \( \rho_y \) at their 1990/91-levels and only feed the estimated changes in task-bias into the model (given by \( \alpha, \beta \)); similarly, for skill-biased TC and the change in distributions.

5.4.1 Wage Polarization

A growing literature documents wage polarization in the US. This phenomenon refers to a boost in upper-tail inequality but a slow-down or even decline in lower tail wage inequality, often measured as an increase in 90/50 but a decrease in 50/10 inequality. Figure 4 (a) plots the change in US hourly wages relative to the median wage between 1990/91 and 2009/10 by wage percentile (in red), illustrating that inequality disproportionally increased in the upper part of the distribution but declined in the lower part. It also shows that the full model (blue curve), once the estimated parameters are fed in, can account for polarization, only that it under-predicts the action in the tails. What might have caused this specific change in the wage distribution?

Panels (b), (c) and (d) analyze whether wage polarization can possibly be triggered by the estimated technology or distributional changes when fed into the model (blue curves). Panel (b) shows that task-biased TC matches wage polarization fairly well, especially the decline in lower tail inequality. In turn, skill-biased TC in panel (c) boosts inequality across the board, thus failing to account for declining lower tail inequality. Finally, had only distributional changes taken place, then wage inequality would have stayed nearly constant (panel (d)).
The model offers an explanation for why only task-biased TC can account for wage polarization. Through an increase in cognitive input complementarities (\(\alpha\) goes up) and a decrease in manual input complementarities (\(\beta\) goes down), task-biased TC affects the curvature of the wage schedule. Wages become more convex in cognitive but less convex in manual skills (see Figure 8 in the Appendix).

This fuels wage inequality in the cognitive but compresses inequality in the manual dimension (in line with Proposition 8ii.). Polarization then occurs because differently skilled workers are not uniformly distributed across the wage distribution. Instead, workers with high cognitive skills are concentrated in the upper part of the wage distribution. This is why these differential wage changes lead to a disproportionate increase in upper tail inequality but a decline in lower tail wage inequality.

To see where differently skilled workers are located in the wage distribution, I separately plot the c.d.f.’s of the empirical wage distributions for low-skilled workers, manual specialists, generalists and cognitive specialists in 1990/91 and 2009/10 (Figure 5). Throughout the last two decades, the wage

---

40Low-skilled are defined as \(x_C < \mathbb{E}(x_C), x_M < \mathbb{E}(x_M)\), manual specialists as \(x_C < \mathbb{E}(x_C), x_M > \mathbb{E}(x_M)\), cognitive specialists by \(x_C > \mathbb{E}(x_C), x_M < \mathbb{E}(x_M)\), and generalists by \(x_C > \mathbb{E}(x_C), x_M > \mathbb{E}(x_M)\).
distribution of cognitive specialists first-order stochastically dominates the distribution of generalists, which in turn dominates the distributions of manual specialists and low-skilled workers. Thus, manual specialists and the low-skilled are concentrated in the lower part of the wage distribution whereas those workers with strong cognitive skills (cogn. specialists and generalists) are clustered in the upper tail.

Recall from Proposition 8 that task-biased TC affects the curvature of cognitive and manual returns (and hence polarization) through two channels. First, there is a direct effect through changes in worker-job complementarities $\alpha$ and $\beta$, and thus through $\delta$. Second, there is an indirect effect through re-sorting of workers to jobs (i.e. through a change in the Jacobian of the matching function, Proposition 7). The ML-estimates on task-biased TC (i.e. a decrease in $\delta$) imply that a mismatch-assortativeness trade-off has been at work: positive sorting along the cognitive dimension has become more pronounced at the expense of increased manual mismatch. To obtain more direct empirical support for the re-sorting of workers to jobs, I run reduced-form regressions on the matching function in both periods, to estimate the Jacobian $J_\mu$ (which fully captures the economy’s sorting patterns) directly from the data:

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} x_C \\ x_M \end{bmatrix}$$

The estimates suggest that both in 1990/91 and 2009/10, sorting satisfies PAM (i.e. $J_\mu$ is a P-matrix) with a positive relationship between skills and skill requirements in both tasks (Table 16 or Figure 7, Appendix) – in line with the ML-estimates on worker-job complementarities (i.e. $\alpha, \beta > 0$). Importantly, the estimated sorting changes over time are qualitatively consistent with the estimated task-biased TC. While in 1990/91, positive assortativeness is stronger in the manual dimension, the opposite is true for 2009/10. Notice that, in light of the model, relatively small assignment changes were expected, given that skill supply and demand (i.e. $\rho_x$ and $\rho_y$) are fairly well aligned throughout the whole period (Proposition 7iii.).
In sum, task-biased TC leads to less convex manual returns but more convex cognitive returns because of two effects, a direct one operating through the change in complementarities and an indirect one through worker-job re-sorting. Since cognitive (but not manual) workers are concentrated in the upper part of the wage distribution, these wage movements trigger wage polarization. In turn, under skill-biased TC neither of the two effects is at work (see Proposition 9). Thus, skill-biased TC has no effect on the curvature of the wage schedule but simply shifts it and, as a result, triggers an increase in wage inequality across the whole wage distribution. Finally, distributional changes over the last two decades have been relatively small and have not been driving polarization.

5.4.2 Wage Skewness and Wage Dispersion

Wage polarization is not the only recent shift in US wage inequality. Between 1990/91 and 2009/10 there have also been significant increases in skewness and variance of the wage distribution (Table 2).

The positive skewness has increased by nearly 100% over the last two decades, see Table 2. The model is able to generate a strong increase in the skewness (and slightly over-shoots), with this boost being entirely driven by task-biased TC (column 3). In contrast, skill-biased TC and distributional changes in isolation would have decreased the skewness. This may suggest a link between wage polarization and an increase in wage skewness as they both appear to be driven by task-biased TC – an issue which would require further investigation but goes beyond the scope of this paper.

<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>Task-Biased TC</th>
<th>Skill-Biased TC</th>
<th>Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>%Δ Wage Skewness</td>
<td>0.96</td>
<td>1.10</td>
<td>0.86</td>
<td>-0.24</td>
</tr>
<tr>
<td>%Δ Wage Variance</td>
<td>0.53</td>
<td>0.31</td>
<td>0.10</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Table 2: Skewness and Variance of the Wage Distribution: Data, Model and Decomposition

There also has been a substantial increase of about 50% in wage dispersion during that period. The model – despite being frictionless – can account for 58% of this increase. The decomposition exercise suggests that skill-biased TC was the most important force behind this shift while the role of task-biased TC was relatively small. The model offers an explanation for why this is the case: Over the last two decades, the increase in cognitive task-weight \( \alpha \) was accompanied by a strong decrease in manual task weight \( \beta \). The first force fuels wage dispersion while the second one compresses it, which is why the net effect of task-biased TC on wage dispersion is small. To the contrary, changes in the skill bias are dominated by the increase in cognitive skill weight \( \lambda \), which shifts the wage schedule (instead of impacting its curvature) and fuels inequality across the whole distribution (Proposition 9(iii)).

Besides technological change, there have been changes in skill supply and demand distributions during the last two decades. However, they were too small to have a large impact on wage inequality.

Both exercises show that technological change rather than changes in distributions mattered for US wage inequality shifts such as wage polarization as well as rising wage dispersion and skewness.
5.5 Performance of One-Dimensional Models

How well does a one-dimensional skill hypothesis describe the data? In what sense does the multi-dimensional model provide a richer understanding of the data than comparable one-dimensional models? To address these questions, this section compares the performance of the multi-dimensional assignment model to two types of one-dimensional models: first, to the one-dimensional assignment model which is the closest one-dimensional analogue of the presented model; second, to a single index model widely used in labor economics (Card and Lemieux [1996]; see also Altonji and Blank [1999] for a summary). I will compare their performance relative to my model in terms of various wage inequality shifts (wage polarization, change in variance and skewness of wage distribution) and regarding changes in the wage ranking over time. I will also highlight the different implications for sorting.

5.5.1 Comparable One-Dimensional Models

**One-Dimensional Assignment Model.** The one-dimensional model that compares most closely to the presented multi-dimensional assignment model is Becker [1973]. I estimate this model using data on years of education as the single index for skills (in the Appendix, I also report the estimates, using cognitive skills as single traits of workers). Education has been widely used as a proxy for skills in settings that focus on one-dimensional heterogeneity and presumably captures an overall level of both cognitive and manual skills.

To ensure comparability to my model, I assume standard normal distributions $x,y \sim N(0,1)$ and technology $F(x,y) = \alpha xy + \lambda x + f_0$. It follows that sorting is 1-d PAM and the equilibrium wage is

$$w(x) = \alpha \frac{x^2}{2} + \lambda x + w_0$$

where $w_0$ is the constant of integration. I estimate the parameters $\alpha, \lambda, w_0$ by OLS (where observations are weighted by NLSY sample weights), using (18). *Education* is given by highest grade completed, reported by individuals in the NLSY79 and NLSY97. The wage data is cleaned in the exact same way as for the estimation of the multi-dimensional assignment model.

**Single Index Model.** The second one-dimensional model that will be compared to my model is the single index model by Card and Lemieux [1996]. Variants of this model have been widely used in labor economics to explain the evolution of wages over time. Here I only sketch this model, see Card and Lemieux [1996] for the details. It is important to note that this single index model is not an assignment model. But even though its underlying structure is different, the resulting equilibrium wage function is similar to that of the one-dimensional assignment model above, which makes this model a natural candidate for comparison.

In the background, there is a competitive labor market where the wage of each skill group equals its marginal product. Worker productivity $\theta$ is a one-dimensional (unobserved) trait of individuals or, in other words a single index, that is the sole worker characteristic determining the marginal product and thus wages. Technological change may lead to an expansion or contraction of the relative productivity differences across individuals over time. A simple way of modeling this is to assume that
an individual i’s productivity (which is here interchangeably used with skill) in period s is related to his productivity in a base period t by $\theta_{is} = g(\theta_{it})$, where $g$ is a strictly increasing function. \footnote{For instance, a convex $g$ would lead to an expansion and a concave $g$ to a contraction of productivity differences.}

This single-index model can be estimated using grouped micro-data from repeated cross-sections.\footnote{The underlying assumption is that individuals in a given cell at two different points in time are exchangeable, i.e. the relation between their unobserved skills and the observed categories does not change over time.} In particular, I will group individuals from each sample, NLSY79 and NLSY97, into $J$ narrow groups according to their age, education and labor market experience. Card and Lemieux [1996] discuss in detail how this approach (along with the assumption that the single index $\theta_j$ of each group $j$ evolves according to function $g$ over time) yields the following mean log wage of group $j$ in a given period, say period 1, as a function of the mean log wage of the same group in the previous period

$$w_{j1} = g(w_{j0}) + r_j$$

(19)

where the remainder $r_j$ vanishes if the within-group variance of skills is zero or if $g$ is linear.\footnote{Note that for (19) to hold, it is also assumed that the single index $\theta$ is composed of a group-specific mean as well as a person-specific deviation from that group mean (see Card and Lemieux [1996] for more details).}

How to choose the functional form $g$? Card and Lemieux [1996] find that a quadratic $g$ describes the evolution of US wages during the 1970s and 1980s best.\footnote{Inspection of my data suggests that convex wages also yield a reasonable approximation.} Moreover, the quadratic single index model generates a wage function very similar to those from the assignment models above, which enables a clean comparison between the three models (i.e. the single index model, the one-dimensional and the multi-dimensional assignment model). I will therefore estimate the quadratic single index model, where, based on (19), the true model is given by

$$w_{j1} = aw_{j0}^2 + bw_{j0} + c.$$ \footnote{The difference is that (20) is expressed in logs whereas (18) is expressed in levels.}

(20)

with coefficients $a$ and $b$ as well as constant $c$ to be estimated. Note the similarity to (18).\footnote{Not taking labor market experience into account would yield too few categories here since I am constrained to focus on workers of age 27-29, i.e. few age groups. Also note that I am not relying on cognitive/manual skills to group workers since the NLSY lacks some training/degree measures in the 1980s that are crucial for my construction of workers’ skills.}

The mean wage of a particular group in a previous period proxies the (unobserved) single index of worker productivity which determines contemporaneous wages of that group. Therefore, the evolution of wages is given by a mapping between average mean cell wages across periods.

To estimate (20), I follow Card and Lemieux [1996] and categorize workers into narrow groups by years of education and age, and moreover by labor market experience to proxy for group-specific productivity $\theta$.\footnote{For instance, a convex $g$ would lead to an expansion and a concave $g$ to a contraction of productivity differences.} In order to estimate the wage function in 1990/91, I construct a baseline sample of workers using observations from the NLSY79 in the 1980s. In turn, observations from 1990/91 form the baseline sample for the estimation of the wage function in 2009/10. This approach yields 104 education-age-experience groups that are present in all three periods and that form the sample for estimation. In Appendix 8.5.6, I describe the data and robustness to my choices in more detail.

Since this single index model is not an assignment model, its wage equation has a different
interpretation than the wage function from the one-dimensional assignment model. It would therefore be meaningless to assess the roles of task-biased versus skill-biased technological change in this single index model. But despite different micro-foundations, both types of models (single index and the discussed assignment models) share the competitiveness of the labor market as a crucial feature, and, importantly, give rise to a similar structure of equilibrium wages, facilitating the comparison.

There are two important implications of wage equations (18) and (20) and of one-dimensional models more generally: First, workers are sufficiently characterized by a one-dimensional index, given here by education $x$ or (unobserved) worker productivity $\theta$. A key implication is that workers with a similar index should have similar wage growth over time. Second, since in both models all workers/groups are affected by the same productivity transformation, the rank ordering of wages across workers/groups remains constant over time. This is an assumption underlying any conventional single index model, which is why Card and Lemieux [1996] call it the single index assumption.

5.5.2 Results: Multi-Dimensional Assignment Model Versus One-Dimensional Models

This section compares the performance of the multi-dimensional assignment model in explaining the evolution of US wages to the performance of the discussed one-dimensional models. To save space, denote the Beckerian one-dimensional assignment model by (1dA) and the single index model by (SI).

ESTIMATES. I report the estimates of both one-dimensional models in Table 18, Appendix 8.5.8. Both (1dA) and (SI) give rise to an increasing and convex wage function, see Figure 10 in the Appendix. This implies that highly educated/productive workers earn over-proportionally more than less educated/productive workers. Across the board, the estimated coefficients are significant at conventional significance levels. For robustness, I also report the estimates of the one-dimensional assignment model using cognitive skills (as opposed to education) as the single worker characteristic as well as the estimates of the single index model for a different weighting procedure.

WAGE INEQUALITY. According to both one-dimensional models, curvature changes of the wage function over time are minor, if anything, wages have become less convex in education and in the productivity index over time. One reason may be that, through the one-dimensional aggregation of skills, differential curvature changes of the wage in the underlying attributes in the data (convexification in cognitive but de-convexification in manual skills, as shown earlier) get washed out, leading to small changes of the wage curvature in the one-dimensional attributes. These models have particular difficulty to pick up convexification at the top but de-convexification at the bottom of the distribution.

As a result, neither of these models captures the observed wage polarization. Instead both models over-predict the increase in lower tail wage inequality. This is shown in Figure 6 where, as before, the red curve indicates the change in hourly wages relative to the median between 1990/91-2009/10 in the data and the blue curve is what the one-dimensional model would predict based on the estimates. While the Beckerian assignment model approximates changes in the wage distribution poorly (left

---

47To the contrary, if multi-dimensional heterogeneity is important, the loadings on the multiple components that make up $x$ and $\theta$ are crucial. In a multi-dimensional world, these loadings may change over time and nothing guarantees that workers with a similar index face similar wage growth.

48Note that for this analysis, all log-wages from the single index model are converted into levels.
panel), the single index model generates some polarization in the upper tail but also fails to generate polarization at the bottom. Further, both models perform worse than the multi-dimensional model regarding changes in the skewness (compare Table 2 with Table 17, Appendix) – again indicating a link between wage polarization and increased skewness as also suggested by the multi-dimensional data.

Both one-dimensional models match the observed changes in wage dispersion considerably better than the changes in skewness, with (SI) outperforming the multi-dimensional model in this dimension.

Figure 6: Wage Polarization in One-Dimensional Data and Models. Left Panel: One-Dimensional Assignment Model (1dA). Right Panel: Single Index Model (SI)

Re-Sorting. Due to the absence of assignment/sorting in (SI), this model does not feature any changes in the allocation of workers to jobs in response to technological change. Further, (1dA) also fails to account for any re-sorting of workers to jobs in response to technological change unless technology shifts so drastically that negative instead of positive assortative matching becomes optimal. Changing $\alpha$ (while keeping it positive) does not produce any shift in sorting. Thus, both one-dimensional models miss the observed assortativeness-mismatch trade-off across skills, empirically documented in the previous section (for which the multi-dimensional model provides an explanation).

Rank Switchings in the Wage Distribution. Due to the lack of both re-sorting and the direct effect of decreasing manual complementarities (which is quantitatively more important than re-sorting), the one-dimensional models have difficulty to account for rank switchings in the wage distribution in response to technological change. This phenomenon refers to the following (this example is given using multi-dimensional data): At the beginning of the 1990s, a worker with some skill bundle $(\tilde{x}_C, \tilde{x}_M)$ earned a higher wage than someone with a different bundle $(\hat{x}_C, \hat{x}_M)$ whereas in the late 2000s the wage ranking between two workers with the exact same skill bundles is reversed. In terms of the leading example, pre-TC, a specialized production worker in a car factory (that demands high levels of manual skills) earned more than an office clerk whose job requires medium levels of cognitive skills and no manual skills. Due to TC the worker loses his production job and is now employed as a car salesman, under-utilizing his (strong) manual skills and mainly using his (weaker) cognitive skills. Since the office job requires and rewards more cognitive skills than the job as car salesman, the wage ranking of these two individuals is now reversed.
To document this phenomenon in the data, I compute the rank correlation of wages across periods (Table 3), measured by Kendall’s $\tau$. Higher rank correlation implies few rank switchings and Kendall’s $\tau$ equals one for perfect rank correlation, i.e. when there are no rank switchings at all.

Table 3 has three parts. Part 1 reports the rank correlation in wages in both multi-dimensional data and model where I divided the workers into narrowly defined skill types (both taking manual and cognitive skills into account) such that the number of overall observations is similar to that in the single index model. Part 2 reports the rank correlation analysis for the single index model and data, and part 3 for the one-dimensional assignment model and data.

The structure of all three parts is identical: I report the rank correlation of wages, p-value (for the test of the null hypothesis that wages across periods are uncorrelated) and the number of observations, both for data and model. I also report in each part how much of the imperfect rank correlation in the data is captured by the model under consideration, given by the statistic $(1 - \tau_{\text{Model}})/(1 - \tau_{\text{Data}})$. This statistic equals one if the model perfectly captures the observed rank correlation. The smaller is this statistic (if it is smaller than one), the worse is the explanatory power of the model, in the sense that the model over-predicts the rank correlation.

<table>
<thead>
<tr>
<th>Kendall’s $\tau$</th>
<th>P-value</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multi-Dimensional Assignment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.25</td>
<td>0</td>
</tr>
<tr>
<td>Model</td>
<td>.73</td>
<td>0</td>
</tr>
<tr>
<td>$1 - \tau_{\text{Model}}$</td>
<td>.36</td>
<td>.</td>
</tr>
<tr>
<td>$1 - \tau_{\text{Data}}$</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td><strong>Single Index</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.55</td>
<td>0</td>
</tr>
<tr>
<td>Model</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$1 - \tau_{\text{Model}}$</td>
<td>.01</td>
<td>.</td>
</tr>
<tr>
<td>$1 - \tau_{\text{Data}}$</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td><strong>One-Dimensional Assignment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.53</td>
<td>0</td>
</tr>
<tr>
<td>Model</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$1 - \tau_{\text{Model}}$</td>
<td>0</td>
<td>.</td>
</tr>
<tr>
<td>$1 - \tau_{\text{Data}}$</td>
<td></td>
<td>.</td>
</tr>
</tbody>
</table>

Table 3: Rank Correlation in the Wage Distribution 1990/91-2009/10

The results deliver the following insights: First, in both the multi-dimensional and the one-dimensional data, there is a considerable amount of rank switchings in the wage distribution over time. Second, in the multi-dimensional data, rank switchings are even more pronounced (Kendall’s $\tau$ is .25 compared to .55 and .53 in the 1d-data). Third, the multi-dimensional assignment model (as predicted by Proposition 8.iii) is able to capture a sizable share (36%) of the observed imperfect rank correlation in wages across periods, highlighting the important role of multi-dimensional heterogeneity.

49 I here report Kendall’s $\tau$ instead of Spearman’s $\rho$ since the former is more accurate in small samples. Results based on Spearman’s $\rho$ are similar.

50 I computed ten percentiles of both the marginal distribution of cognitive and of manual skills, divided workers into $10 \times 10$ cells and computed the average wage per skill cell. This is done for both periods. I then compare the wage ranking by skill cells across periods. Note that there are 99 instead of 100 observations since one skill cell was empty.
for this phenomenon. Fourth, and in contrast, both one-dimensional models fail to capture any of the observed rank switchings. They generate perfect rank correlation between wages across periods, which is one of the key predictions of any single index model that features monotone wages in skills but is clearly rejected by the US data.

Notice that rank switchings are not necessary for wage polarization to occur since polarization only requires wage growth at the bottom to dominate wage growth in the middle of the distribution without necessarily violating the monotonicity of wages in education/productivity. However, rank switchings (especially) at the bottom of the distribution may be conducive to polarization, which is another reason why the multi-dimensional model does a better job in accounting for wage polarization.

In line with the multi-dimensional assignment model and its predictions on task-biased TC, a closer look at the data suggests that rank switchings are mainly due to workers with different manual skills switching places in the wage ranking. In turn, when wages are measured by cognitive skill groups, the wage structure across time is more stable. This is documented in Table 4, where the rank correlations in wages are considerably lower when focusing on average wages by manual skill groups, compared to the case of computing average wages by cognitive skill groups. The model produces the same qualitative patterns and suggests the following explanation: It predicts that, in the face of cognitive task-biased TC, the cognitive task gains importance in production. Workers change jobs such that PAM becomes stronger along the cognitive dimension at the expense of increased manual mismatch. Moreover, workers with strong cognitive skills benefit from the direct effect of changes in the task-bias parameters. The wage ranking in terms of cognitive skills remains relatively stable but it changes drastically along the manual dimension: Workers with low manual but high cognitive skills see their wages increase while workers with high manual but low cognitive skills see their wages decrease. These workers thus switch ranks in the wage distribution, if the mass of workers with negatively correlated skills is relatively large, and if the wage changes in response to technological change are sufficiently strong (see Appendix 8.5.9 for an illustrative example).

<table>
<thead>
<tr>
<th></th>
<th>Kendall’s Tau</th>
<th>P-value</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive Skill Groups</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>.958</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>Data</td>
<td>.404</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td><strong>Manual Skill Groups</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>.595</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>Data</td>
<td>.251</td>
<td>0</td>
<td>99</td>
</tr>
</tbody>
</table>

Table 4: Rank Correlation in the Wage Distribution 1990/91-2009/10 by Skill Group

To sum, workers who mainly rely on their manual skills have lost considerable ground against other skill groups over the last two decades, leading to rank switchings in the wage distribution. While this phenomenon can be easily explained in the two-dimensional model as a consequence of cognitive task-biased TC, rank switchings cannot be rationalized in comparable one-dimensional models.

51 I first compute 100 percentiles of each marginal skill distribution. For panel 1 of the table, I divide workers into 99 groups based on their cognitive skills; in turn, panel 2 takes average wages by manual skill percentile into account.


6 Literature Review

This paper contributes to two strands of literature: multi-dimensional matching under transferable utility (including hedonic models and optimal transport); and task-biased technological change. Here I discuss those papers that are most closely related.

MULTI-DIMENSIONAL MATCHING. Variations of the quadratic-Gaussian model have been studied in several contexts: Building on Tinbergen [1956], Ekeland et al. [2004] analyze the econometric identification of hedonic models with focus on a quadratic-Gaussian setting. They discuss an identification problem which arises in their model because wage function and production technology have the same curvature in $x$. My model circumvents this collinearity problem by specifying a production technology without quadratic loss terms. Additionally, to make the Gaussian model suitable for empirical analysis I include non-interaction skill terms in the technology such that wages can be monotone over the whole observed skill support. Olkin and Pukelsheim [1982] solve a related Gaussian example, phrased as the linear algebra problem of minimizing the $L^2$ distance between two random normal vectors, but in a symmetric setting (i.e. $\delta = 1$). Bojilov and Galichon [2015] extend the quadratic-Gaussian setting to include unobserved heterogeneity. They derive closed-forms and show that the surplus function can be identified. Unlike all of these papers, I here focus on multi-dimensional sorting. I first extend the unidimensional notion of (positive) assortative matching (PAM) and provide sufficient conditions on the technology for sorting to obtain. Second, using this notion of PAM, I derive the equilibrium in closed form (for any level of $\delta$), similarly to the way one-dimensional decentralized assignment problems are solved. Third, I use the closed form to perform a series of novel comparative statics related to technological and distributional changes, where multi-dimensional PAM is instrumental for interpreting the effects on assignment and wages. Finally, this paper structurally estimates the model to study technological change in a multi-dimensional world. Both theory and empirical application highlight various – so far undocumented – differences between one- and multi-dimensional sorting.

This paper also relates to the literature on multi-dimensional matching on the marriage market. Choo and Siow [2006] propose a transferable utility model of the marriage market to estimate the marriage matching function from observed matches in the US. Their model allows for multi-dimensional (un)observed heterogeneity under the assumption that there is no interaction between unobservable characteristics of partners (separability assumption). More recently, Galichon and Salanié [2010] study optimal matching in a model with multi-dimensional (un)observed characteristics under the same separability assumption. Dupuy and Galichon [2014] extend their set-up to continuous types.

These studies differ from this paper in terms of objective and modeling choices. Choo and Siow [2006] estimate the gains from marriage, i.e. their focus is empirical. In contrast, Galichon and Salanié [2010] and Dupuy and Galichon [2014] develop techniques to estimate complementarities in the surplus function from observed matches. Conversely, this paper aims at developing a multi-

---

52 Decker et al. [2013] analyze the existence and uniqueness of equilibrium, provide a closed form as well as comparative statics of the Choo-Siow model. Chiappori et al. [2016] also provide a closed form of a multi-dimensional matching model and then test predictions of how spouses trade off education and non-smoking. Their assumptions are as follows: (i) Smoking status (binary) and education (continuously uniform) are independent. (ii) In the surplus, the disutility of smoking is proportional to the surplus generated by the spouses’ skills.
dimensional sorting framework that allows for comparative statics and interpretability of the results. Notice, however, that there is an important conclusion common to the papers by Galichon and Salanié [2010], Dupuy and Galichon [2014], and my own: With multi-dimensional matching, there is a trade-off between matching along different characteristics that depends on complementarity weights in the surplus function.

The proposed sorting framework also sheds further light on several results in the literature where either matching is based on multiple variables (in settings other than the quadratic-Gaussian one) or where uni-dimensional firms have more than one choice variable. McCann et al. [2015] analyze a marriage market where agents match based on cognitive and social skills. Under the assumption of identical distributions of men and women and their specified technology they prove that matching is positively assortative in both dimensions – a result that can be captured by my sorting framework (see Proposition 11 (b) in Appendix 8.4.2). My multi-dimensional sorting conditions also offer a re-interpretation of the sorting conditions in Eeckhout and Kircher [2016] where matching is one-dimensional but where firms (that differ in a single attribute: quality) have two choices: worker type (quality) and firm size (quantity). Similar to this paper where sorting occurs in worker-firm attributes both within and across tasks, in their setting, sorting can occur within quality (i.e. within worker and firm quality) and across quality-quantity (i.e. across firm quality and the hired labor quantity). Like here, the relative complementarities in production resolve this trade-off. In fact, my sorting condition that the matrix of cross-partial derivatives is a diagonal $P$-matrix, which guarantees within-task sorting, is sufficient for PAM within quality variables in their model.

Notice that my notion of multi-d PAM is different from the notion proposed by Topkis [1998]. His results on increasing solutions can be applied to multi-dimensional types to deliver a notion of multi-d PAM, which implies that better workers are matched to better firms in the vector sense. (For instance, if worker A dominates worker B in both skills, then worker A will be matched with job A that dominates job B in both productivities). Similarly, extending the PAM-definition of Shimer and Smith [2000] to multiple dimensions would imply (for a frictionless model) that better workers are matched to better firms in the vector sense, i.e. along both dimensions. Neither of these notions can capture the sorting trade-off present in my model, according to which a better worker (in the vector sense) can be matched to a worse firm in the dimension that features relatively weak worker-firm complementarities. Moreover, and contrary to my definition, with these notions one can only analyze sorting of those workers and firms that can be ranked in the vector sense. They cannot be used, for instance, to analyze sorting when some workers are specialized in cognitive and others in manual skills. In this sense, my notion is more broadly applicable.

Finally, this paper relates to the literature on optimal transport. In non-technical terms, the optimal transport problem involves finding a measure-preserving map that carries one distribution

53In their papers, modeling devices are (un)observed heterogeneity and extreme value distributions of unobserved traits. I rely on observed heterogeneity and Gaussian copulas.

54The conditions in Eeckhout and Kircher [2016] for PAM under any distributions are that both (i) the determinant of the matrix of cross-partialis and (ii) the cross-partial between labor quantity and some fixed resource quantity are positive. Naturally, my conditions are stronger because of the fully multi-dimensional setting (for a distribution-free result, they would also require a diagonal $P$-matrix as well as a positive cross-partial between worker and firm quality).
into another at minimal cost, using linear programming. A tight link has been established between the following two formulations of the assignment problem: a hedonic pricing problem with transferable utility (like the one in this paper) and an optimal transport problem. Shapley and Shubik [1971] show this equivalence in a discrete setting and Gretsky et al. [1992] for the continuous case. I make use of this link and apply several results from that literature (particularly from Villani [2009] and Ma et al. [2005]) to establish existence and uniqueness of the (differentiable) equilibrium (Proposition 1).

Apart from providing existence and uniqueness results, this literature has also established purity of the assignment (see, e.g. Chiappori et al. [2010] and Ekeland [2010]): If technology satisfies the twist condition, which states that $D_x F$ is injective with respect to $y$, then the assignment is one-to-one, i.e. pure. Notice that the $P$-matrix property of $D^2_{xy} F$ from this paper is stronger than the twist condition (more precisely, it is sufficient for twist since $D^2_{xy} F$ is the Jacobian of $D_x F$, see Gale and Nikaido [1965]). It is assumed here to guarantee not only purity but especially positive sorting (Proposition 2). In sum, while the optimal transport literature has developed powerful general tools to study multi-dimensional matching problems, it has not analyzed multi-dimensional sorting (Definition 1 and Proposition 2), which is the main theoretical contribution of this paper.

Task-Biased Technological Change. Costinot and Vogel [2010] and Acemoglu and Autor [2011] use one-dimensional assignment models to analyze (amongst other issues) task-biased TC. In these frameworks, an adverse technology shock reduces firms’ demand for medium-skilled workers and hence their relative wages. This fuels upper-tail but compresses lower tail wage inequality – a phenomenon referred to as wage polarization.

Instead of implicitly assuming that manual skills are only used by medium-skilled workers, I make the assumption that every worker has both skills, yet in different proportions. This assumption of multi-dimensional skills and skill demands yields several new theoretical insights: First, by including a second dimension, I can analyze the differential effect of task-biased TC on sorting and wage inequality in manual and cognitive skills. I identify a new mechanism of how this technology shift affects wage inequality and polarization (i.e. the re-sorting in line with an assortativeness-mismatch trade-off across tasks, rendering cognitive (manual) returns more (less) convex).

Second, due to multiple skills I can distinguish between generalists and different types of specialists, thereby capturing that generalists can shield themselves against adverse shocks to manual inputs. Closely related, my model predicts rank switchings in the wage distribution in response to technological change – in line with the data – which one-dimensional models of technological change fail to capture. Finally, this paper adds to this literature a unified framework of task-biased and the more standard skill-biased TC, that trigger different responses of equilibrium outcomes (and particularly so in the multi-dimensional setting). My empirical application of the multi-dimensional model highlights the

\[^{55}\text{Optimal transport has a long tradition in mathematical theory. See Villani [2009] for a recent reference book.}\]

\[^{56}\text{See also Dizdar and Moldovanu [2012] for recent work on the intersection of multi-dimensional matching and mechanism design that makes use of the twist condition as well.}\]

\[^{57}\text{In their frameworks, task-biased TC also leads to employment polarization, which is beyond the scope of my model since jobs and workers match one-to-one in this frictionless and competitive labor market.}\]

\[^{58}\text{It is noteworthy that this assortativeness-mismatch trade-off across tasks takes place here despite pairwise matching (there is no intensive margin) and without switching from PAM to NAM.}\]
relevance of these novel predictions. Also note that although there is abundant empirical evidence on labor market polarization in developed countries, there is little structural analysis into the causes of this phenomenon – something that this paper seeks to address.⁵⁹

7 Conclusion

Technological change has drastically changed the structure of production in favor of cognitive relative to manual inputs in the developed world. How does this shift affect worker-job assignments, wages and inequality? This is a multi-dimensional assignment problem where workers with different bundles of manual and cognitive skills sort into jobs that require different combinations of these skills. To study this problem, this paper develops a theoretical framework for multi-dimensional sorting that extends the unidimensional notion of assortative matching. I derive the equilibrium allocation as well as equilibrium wages in closed form. I then analyze the impact on these equilibrium outcomes as cognitive (as opposed to manual) inputs become more prevalent in production, capturing one of the main recent technological shifts. Finally, I take this model to the data to study technological change in the US between 1990 and 2010. The empirical analysis reveals that technological change was strongly biased toward cognitive inputs. Counterfactual exercises suggest that this technology shift (as opposed to changes in skill and skill demand distributions) can account for observed changes in worker-job sorting, wage polarization and wage dispersion.

It is worth pointing out that the theoretical framework developed here is of independent interest and can be used beyond this paper’s application to technological change. It could be applied to a variety of matching problems that involve multi-dimensional heterogeneity, not only in the labor but also in the marriage or education markets. To broaden the applicability of this theory even further, it will be important to extend this analysis to settings with search frictions (see Section 1.3, Online Appendix, for the extension of the proposed framework to search frictions and directed search, and Lindenlaub and Postel-Vinay [2016] for a first analysis of multi-dimensional sorting under random search) and to settings where the two sides of the market have different numbers of characteristics, preventing pure matching. These are challenging problems for future research.

⁵⁹ An exception is Boehm [2013] who studies wage polarization in an empirical Roy model where a variety of abilities determine three occupation-specific skills (for low, medium and high-skilled occupations).
8 Appendix

8.1 Proofs of General Model (Section 2)

Note that the proofs for this Section will be given for N-dimensional heterogeneity where firms are characterized by \( y = (y_1, ..., y_N) \in Y \subset \mathbb{R}^N_+ \) (with c.d.f. \( G \) and density \( g \)) and workers are characterized by \( x = (x_1, ..., x_N) \in X \subset \mathbb{R}^N_+ \) (with c.d.f. \( H \) and density \( h \)) while the text focuses on two-dimensional heterogeneity for illustration.

8.1.1 Proof of Proposition 1

Throughout, I will make the following assumption.

**Assumption 1** \( D^2_{xy} F \) is a diagonal \( P \)-matrix.

**Proof.** Assumption 1 implies that \( D_y F \) is injective w.r.t. \( x \) and \( D_x F \) is injective w.r.t. \( y \).

(i). Under the stated assumptions, the existence of a unique, deterministic optimal assignment \( \mu \) follows from Theorem 10.28 in Villani [2009] or, when directly applied to \( \mathbb{R}^n \), from Theorem 3.6 in De Philippis and Figalli [2014]. The existence of a unique wage function (up to an additive constant) follows from Theorem 10.28 and Remark 10.30 in Villani [2009].

(ii). Smoothness of the equilibrium matching function \( \mu \) (and wage function \( w \)) in the interior of \( X \) follows from a minor modification of Theorem 4.2 in De Philippis and Figalli [2014] which is based on Ma et al. [2005].

Note that their assumptions (a)-(c), stating that \( X \) and \( Y \) are smooth and that 
\[
\|D_x F(x, y)\| \geq \|D_y F(x, y)\|, \quad \text{and for all } x \in X, \text{ and for all } y \in Y,
\]
are not needed for interior regularity. Instead, it suffices that the sets \( -D_x F(x, y) \) and \( -D_y F(x, y) \) are convex, which follows here from the assumptions of rectangular supports of \( X \) and \( Y \), additive separability of the technology, \( F_{x,y} = 0 \) for \( i \neq j \), and continuous differentiability of \( F \). To see this note that the image of an interval under a continuous function is also an interval and thus a convex set. Therefore, \( -D_x F(x, y) = - \times_{i=1}^N F_{x,i}(x, [y_i, \bar{y}_i]) \) is a product of intervals and thus a convex set – and similarly for \( -D_y F(x, y) \). What remains to be verified from Theorem 4.2 in De Philippis and Figalli [2014] is their assumption “MTW(0)”, which states that the Ma-Trudinger-Wang tensor, given by

\[
\mathcal{S}_{x,y}(\xi, \eta) := \sum_{i,j,k,p,q,r,s} (\tilde{F}_{p,q} \tilde{F}_{ij,p} \tilde{F}_{q,rs} - \tilde{F}_{ij,rs}) \tilde{F}_{r,k} \tilde{F}_{s,l} \xi_i \xi_j \eta_k \eta_l, \quad \xi, \eta \in \mathbb{R}^N
\]

satisfies, what will be called, the Ma-Trudinger-Wang condition (in short by MTW):

\[
\mathcal{S}_{x,y}(\xi, \eta) \geq 0 \quad \text{whenever } \xi \perp \eta
\]

where the notation is as follows: \( \tilde{F} = -F \) (note that the original MTW condition is stated for a cost-minimization problem, where their cost function, \( c \), and the production function from this paper, \( F \)),
are linked by \( c = -F \), \( F_j = \frac{\partial F(x,y)}{\partial x_j} \), \( F_{jk} = \frac{\partial^2 F(x,y)}{\partial x_j \partial x_k} \), \( F_{i,j} = \frac{\partial F(x,y)}{\partial x_i \partial y_j} \), \( F^{i,j} = (F_{i,j})^{-1} \) (i.e. \( F^{i,j} \) denotes the \( i,j \) entry of the inverse matrix of \( F_{i,j} \)), and where all indices run from 1, ..., \( N \).

In order for their theorem to hold in my context, it thus remains to show that MTW is satisfied under the assumptions in part (ii) of the Proposition. (Note that to ensure that MTW holds here I add two assumptions compared to De Philippis and Figalli [2014]: 1. \( D^2_{xy} F \) is a diagonal \( P \)-matrix and 2. \( F_{x_i y_i} \) is (log)-supermodular in \( (x_i, y_i) \). In contrast, De Philippis and Figalli [2014] do not specify any assumptions that guarantee that MTW holds.)

First, notice that the inverse matrix of \( F_{i,j} \) is given by:

\[
(D^2_{xy} F(x,y))^{-1} = \begin{bmatrix}
1/F_{x_1 y_1} & 0 & \cdots & \cdots & 0 \\
0 & 1/F_{x_2 y_2} & \ddots & & \\
\vdots & \ddots & \ddots & \ddots & \\
\vdots & & \ddots & \ddots & 0 \\
0 & \cdots & 0 & 1/F_{x_N y_N}
\end{bmatrix}
\]  

(23)

The assumption that \( D^2_{xy} F \) is a diagonal \( P \)-matrix together with (23) implies that \((\tilde{F}^{p,q} \tilde{F}_{ij,p} \tilde{F}_{q,rs} \tilde{F}_{r,k} \tilde{F}_{s,l}) = \left((-F^{p,q})(-F_{ij,p})(-F_{q,rs}) - (-F_{ij,rs})\right)(-F^{r,k})(-F^{s,l})\) in (21) can only be non-zero if \( i = j = s = l = r = k \). In this case, \((-F^{r,k})(-F^{s,l})\) is positive (due to the \( P \)-matrix property) and expression \([(-F^{p,q})(-F_{ij,p})(-F_{q,rs}) - (-F_{ij,rs})\) takes one of the following two values, \( \forall i \):

\[
\frac{F_{x_i x_i y_i} F_{x_i y_i y_i}}{-F_{x_i y_i}} - \frac{(-F_{x_i x_i y_i})}{-F_{x_i y_i}}.
\]

The first expression is non-negative by the assumption of log-supermodularity of \( F_{x_i y_i} \) and the second expression is non-negative by the assumption of supermodularity of \( F_{x_i y_i} \). Also notice, that whenever \((\tilde{F}^{p,q} \tilde{F}_{ij,p} \tilde{F}_{q,rs} \tilde{F}_{r,k} \tilde{F}_{s,l})\) is non-zero (i.e. if \( i = j = s = l = r = k \)), then \( \xi_i \xi_j = (\xi_i)^2 \geq 0, \eta_k \eta_l = (\eta_k)^2 \geq 0 \). Therefore, under the stated assumptions, all summands in (21) are non-negative, implying (22) (and thus assumption MTW(0) of Theorem 4.2 in De Philippis and Figalli [2014]) is satisfied.

**Remark:** If \( D^2_{xy} F \) is a diagonal \( P^- \)-matrix the argument is similar. Again, \((\tilde{F}^{p,q} \tilde{F}_{ij,p} \tilde{F}_{q,rs} \tilde{F}_{r,k} \tilde{F}_{s,l})\) is non-zero in two cases (see above). Further, \( F_{x_i x_i y_i} \) is non-negative by the assumption of \( F_{x_i y_i} \) being supermodular; \( F_{x_i x_i y_i} F_{x_i y_i y_i}/(-F_{x_i y_i}) - (-F_{x_i x_i y_i}) \) is non-negative if \( F_{x_i y_i} \) is log-submodular (recall that here \( F_{x_i y_i} < 0 \)). Under these assumptions, all the summands in (21) are non-negative and (22) holds.

### 8.1.2 Proof of Proposition 2

**Claim:** If \( D^2_{xy} F \) is a diagonal \( P^- \)-matrix (\( P^- \)-matrix), then \( J_\mu \equiv D_{xy}^* \) is a \( P \)-matrix (\( P^- \)-matrix).

**Suppose** Assumption 1 holds.

**Proof.** It will be shown that under Assumption 1, optimality of the firm’s choice requires that the
Jacobian of the matching function, $D_xy$, is a $P$-matrix. The proof for the case when $D^2_{xy}F$ is a $P^-$-matrix is analogous and therefore omitted. I proceed in several steps.

1. The Hessian of the firms’ problem evaluated at the equilibrium assignment, given by $\mathcal{H} = D^2_{xx}F(x, y^*) - D^2_{xx}w(x)$, is negative-definite. By the necessary second order conditions for optimality, $\mathcal{H}$ is negative semi-definite. It follows that $-\mathcal{H}$ must be positive semi-definite.

To show that $\text{Det}(-\mathcal{H})$ is strictly positive, differentiate the first order conditions evaluated at the optimal assignment $y^*$, (3), with respect to skill vector $x$, which gives

$$\mathcal{H}^* = D^2_{xx}F(x, y^*) - D^2_{xx}w(x) = -(D^2_{xy}F(x, y^*)) (D_xy^*)$$

where $D_xy^*$ is the Jacobian of the matching function and $D^2_{xx}F(x, y^*)$ is defined as

$$D^2_{xx}F(x, y^*) = \begin{bmatrix}
    F_{x_1x_1}(x, y^*) & \cdots & F_{x_1x_N}(x, y^*) \\
    \vdots & \ddots & \vdots \\
    F_{x_Nx_1}(x, y^*) & \cdots & F_{x_Nx_N}(x, y^*)
\end{bmatrix}$$

and where $D^2_{xx}w(x)$ is defined similarly. Since $D^2_{xy}F$ is a $P$-matrix everywhere (and, hence, also along the equilibrium allocation $y^*$), it follows that $\text{Det}(D^2_{xy}F) > 0$. Moreover, by Proposition 1, $\mu$ is measure-preserving and differentiable, and under Assumption 1, also injective (see Theorem 3.6 in De Philippis and Figalli [2014]). It then follows from Theorem 11.1 in Villani [2009] which establishes the Jacobian equation (and which can be obtained in $\mathbb{R}^n$ by combining Lemma 5.5.3 in Ambrosio et al. [2004] with Theorem 3.83 in Ambrosio et al. [2000]) that $\text{Det}(D_xy^*) > 0$. Then from (24), $\text{Det}(-\mathcal{H}) = \text{Det}(D^2_{xy}F)\text{Det}(D_xy^*)$, and therefore $\text{Det}(-\mathcal{H}) > 0$.

2. If $D_xy^*$ is sign-symmetric then it is a $P$-matrix. Suppose that $D_xy^*$ is sign symmetric, i.e. $\frac{\partial y_i}{\partial x_j} = \frac{\partial y_j}{\partial x_i} > 0, \forall i, j \in \{1, 2, ..., N\}, i \neq j$. For sign-symmetric matrices, positivity of principal minors and stability are equivalent (see Theorem 2.6. in Hershkowitz and Keller [2005]). In the following, I show that $D_xy^*$ has positive eigenvalues, i.e. is stable. From (24), $\mathcal{H}^* = (D^2_{xy}F(x, y^*)) (D_xy^*)$, where $-\mathcal{H}^*$ has all positive eigenvalues (Step 1). Denote $M = D^2_{xy}F(x, y^*)$, $J = D_xy^*$. Denote the eigenvalues of $-\mathcal{H}^*$ by $\lambda^H$. They must obey the characteristic equation $det(MJ - \lambda^H I) = 0$. Since $M$ is a P-matrix (Assumption 1), it is invertible and the characteristic equation can be reformulated as $det(J - \lambda^H M^{-1}) = 0$, where $\lambda^H$ is the generalized eigenvalue of the square matrices $(J, M^{-1})$. Given $(J, M^{-1})$, the generalized Schur decomposition factorizes both matrices $J = QSZ^*$ and $M^{-1} = QTZ^*$, where $(Q, Z)$ are orthogonal matrices and $(S, T)$ are upper triangular matrices with the eigenvalues of $(J, M^{-1})$ on their diagonals. The (real) generalized eigenvalues can be computed as $\lambda^H_i = \frac{S_{ii}}{T_{ii}}$. Notice that $T_{ii} > 0 \ \forall i$ because $M$ is a diagonal P-matrix, which implies stability (i.e. positive real part of eigenvalues) and $\lambda^M_{ii} = 1$. For $\lambda^H_i > 0$, it must be that $S_{ii} > 0$, i.e. $J = D_xy^*$ has positive eigenvalues, i.e. is stable.

\textsuperscript{60}If $J$ has complex eigenvalues, $S$ is quasi-upper triangular.
$D_{xy}^*$ is sign-symmetric. To see this, notice that by symmetry of the Hessian and Assumption 1 (i.e. $F_{x_iy_j} = 0, i, j \in \{1, 2, ..., N\}, i \neq j$),

$$H_{ij}^* = H_{ji}^* \quad \iff \quad F_{x_iy_j} \frac{\partial y_i^*}{\partial x_j} = F_{x_jy_j} \frac{\partial y_j^*}{\partial x_i} \quad \forall \quad i, j \in \{1, 2, ..., N\}, i \neq j,$$

and hence $D_{xy}^*$ is sign-symmetric, i.e. $\frac{\partial y_i^*}{\partial x_j} \frac{\partial y_j^*}{\partial x_i} > 0 \quad \forall i, j \in \{1, 2, ..., N\}, i \neq j$. Moreover, $D_{xy}^*$ is stable (see Step 2). A sign-symmetric and stable matrix is a $P$-matrix (Theorem 2.6. in Hershkowitz and Keller [2005]), which proves the result.

8.2 Proofs of Quadratic-Gaussian Model (Section 3)

8.2.1 Labor Market Clearing under PAM (or NAM) for Independent Variables

Consider some independently distributed skills $z_{xC}, z_{xM}$ and independently distributed productivities $z_{yC}, z_{yM}$. The labor market clearing of independent variables under PAM reads

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\hat{z}_{yC}, \hat{z}_{yM}) d\hat{z}_{yM} d\hat{z}_{yC} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\hat{z}_{xC}, \hat{z}_{xM}) d\hat{z}_{xM} d\hat{z}_{xC}$$

where $h$ and $g$ denote the p.d.f.’s of the skills and productivities, respectively. Equation (10) follows immediately, taking into account that the $z’s$ are independent and standard normally distributed.

Similarly, under NAM, the market clearing would read

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\hat{z}_{yC}, \hat{z}_{yM}) d\hat{z}_{yM} d\hat{z}_{yC} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\hat{z}_{xC}, \hat{z}_{xM}) d\hat{z}_{xM} d\hat{z}_{xC}.$$

8.2.2 The Equilibrium Assignment

The following three lemmas are building blocks for the proof of Proposition 3.

Lemma 1 (Integrability Condition) Given a continuously differentiable assignment $y^*$, the following integrability condition

$$F_{xCyC} \frac{\partial y_C^*}{\partial x_M} + F_{xCyM} \frac{\partial y_M^*}{\partial x_C} = F_{xMyC} \frac{\partial y_C^*}{\partial x_C} + F_{xMyM} \frac{\partial y_M^*}{\partial x_M} \quad \iff \quad \frac{\partial y_C^*}{\partial x_M} = \delta \frac{\partial y_M^*}{\partial x_C}$$

is necessary and sufficient for the existence of a unique solution to system (3), given by $w(x)$, such that $w(x) = w_0$.

Proof of Lemma 1. The proof is based on Frobenius Theorem. Consider a system of linear

---

$w_0$ is the reservation wage of the least productive worker $\bar{x}$, set s.t. he is indifferent between working and not working.
first-order partial differential differential equations

\[ \frac{\partial u^\rho}{\partial x^i} = \psi_i^\rho(x, u) \quad i = 1, ..., N; \rho = 1, ..., n \]  

(27)

where \( u : \mathbb{R}^N \rightarrow \mathbb{R}^n \). Consider the following theorem.

**Frobenius Theorem.** The necessary and sufficient conditions for the unique solution \( u^\alpha = u^\alpha(x) \) to system (27) to exist, such that \( u(x_0) = u_0 \) for any initial data \( (x_0, u_0) \in \mathbb{R}^{N+n} \), is that the relations

\[ \frac{\partial \psi_i^\alpha}{\partial x^j} - \frac{\partial \psi_j^\alpha}{\partial x^i} + \sum_\beta \left( \frac{\partial \psi_i^\alpha}{\partial u^\beta} \psi_j^\beta \right) = 0 \quad \forall i, j = 1, ..., N, \quad \alpha, \beta = 1, ..., n. \]  

(28)

hold where \( \psi_j^\alpha = \frac{\partial u^\beta}{\partial x^j}, \psi_i^\beta = \frac{\partial u^\beta}{\partial x^i} \).

Applying Frobenius’ Theorem to my model implies: \( u = w, x = (x_1, x_2, ..., x_N) \) and \( \psi_i(x, u) = F_{x_i}(x, y(x)) \). Notice that \( n = 1 \) because \( w \) is a real-valued function. Then, (28) reduces to

\[ \frac{\partial \psi_i}{\partial x^j} - \frac{\partial \psi_j}{\partial x^i} = 0 \]

which in the presented 2-dimensional model is given by

\[ F_{x_{1}x_{1}} + F_{x_{1}y_{1}} \frac{\partial y_{1}^2}{\partial x_{1}} + F_{x_{2}y_{1}} \frac{\partial y_{1}^2}{\partial x_{2}} - F_{x_{1}x_{2}} \frac{\partial y_{1}^2}{\partial x_{1}} - F_{x_{1}y_{2}} \frac{\partial y_{1}^2}{\partial x_{2}} = 0. \]  

(29)

(29) coincides with condition (26) under the assumption of bi-linear technology. Hence, given (26), the integrability condition from Frobenius Theorem is satisfied.

**Lemma 2 (Continuum of Square Roots)** (i) There exists a continuum of square roots of the covariance matrix \( \Sigma \), denoted by \( S \). Denote its elements by \( \Sigma^\frac{1}{2} \in S \), where \( \Sigma^\frac{1}{2}(\Sigma^\frac{1}{2})^T = \Sigma \).

(ii) The elements of \( S \) can be computed by applying an orthonormal transformation to any given square root. In particular, let \( R \) be an orthogonal matrix, i.e. its columns are mutually orthogonal unit vectors. Hence, \( R^{-1} = R^T \). Then, \( \Sigma^\frac{1}{2} R(\Sigma^\frac{1}{2} R)^T = \Sigma^\frac{1}{2} RR^T(\Sigma^\frac{1}{2})^T = \Sigma^\frac{1}{2}(\Sigma^\frac{1}{2})^T = \Sigma \).

**Proof.** (i) The existence of an infinite number of square roots of the covariance matrix follows from its symmetry. The following non-linear system

\[ \Sigma^\frac{1}{2}(\Sigma^\frac{1}{2})^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \Sigma \]  

(30)

or,

\[ a^2 + b^2 = 1 \]
\[ c^2 + d^2 = 1 \]
\[ ac + bd = \rho \]
is underdetermined. Thus, it either has none or an infinite number of solutions. Since \( \Sigma \) is positive-definite, one square root can be computed using the spectral square root decomposition

\[
\Sigma = CDC'^t
\]

\[
\Leftrightarrow \Sigma^\frac{1}{2} = CD^\frac{1}{2}C'^t
\]  

(31)

where \( D \) is a diagonal matrix with the eigenvalues of \( \Sigma \) as diagonal entries and \( C \) is a matrix of orthonormal eigenvectors of \( \Sigma \). Since the spectral square root is one solution to (30), it follows that the system has an infinite number of solutions.

(ii) follows directly from orthonormality of \( R \), as stated in the Lemma. ■

I now state how the orthogonal transformation matrices \( R_i, i \in \{x, y\} \) can be parameterized by \( \delta \).

**Lemma 3 (Orthogonal Transformation Matrices)** The system of equations to be solved is given by:

\[
\alpha_x^2 + \beta_x^2 = 1 \quad \alpha_y^2 + \beta_y^2 = 1
\]

\[
\frac{\partial y^*_C}{\partial x_M} = \delta \frac{\partial y^*_M}{\partial x_C}
\]

(32)

(33)

(34)

where \( \alpha_x, \beta_x, \alpha_y, \beta_y \) are the elements of the orthogonal transformation matrices:

\[
R_x = \begin{bmatrix} \alpha_x & -\beta_x \\ \beta_x & \alpha_x \end{bmatrix}, \quad R_y = \begin{bmatrix} \alpha_y & -\beta_y \\ \beta_y & \alpha_y \end{bmatrix}
\]

and where (34) is the integrability condition from Lemma 1.

(i) For all \( \delta \in [0, 1] \), the solution to system (32)-(34) is given by \( \alpha_x = \pm 1, \beta_x = 0 \) and

\[
\alpha_y = \pm \frac{(1 + \delta \left( \sqrt{\frac{1+\rho_y}{1+\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}} \right))}{\sqrt{(1-\delta)^2 \left( \sqrt{\frac{1+\rho_y}{1+\rho_x}} - \sqrt{\frac{1-\rho_y}{1+\rho_x}} \right)^2 + (1 + \delta)^2 \left( \sqrt{\frac{1+\rho_y}{1+\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}} \right)^2}}
\]

\[
\beta_y = \sqrt{1 - \alpha_y^2}.
\]

(35)

(36)

(ii) For \( \rho_x \leq \rho_y \), set \( \alpha_i > 0 \). For \( \rho_x > \rho_y \), set \( \alpha_i < 0 \), where \( i \in \{x, y\} \).

**Proof.** To solve (32)-(34), first express the off-diagonal elements of \( D_x \mathbf{y}^*, \frac{\partial y^*_C}{\partial x_M} \) and \( \frac{\partial y^*_M}{\partial x_C} \), as functions of the unknowns. To this end, I compute a candidate equilibrium assignment from (12) where I use rotations of the spectral square root (given by (31)) to uncorrelate skills and productivities, given by:

\[
\Sigma^\frac{1}{2} = \begin{bmatrix} \frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) \\ \frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i}) \end{bmatrix} \begin{bmatrix} \alpha_i & -\beta_i \\ \beta_i & \alpha_i \end{bmatrix}, \forall \ i \in \{x, y\}.
\]  

(37)
Using (37), the candidate equilibrium assignment can be computed from (12) as:

\[
\begin{bmatrix}
y_C \\
y_M
\end{bmatrix} = \frac{1}{2} \left[ \begin{array}{cc}
(\alpha_{y|x} + \delta_{y|x}) \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right) + (\beta_{y|x} - \alpha_{y|x}) \left( \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} \right) \\
(\alpha_{y|x} + \delta_{y|x}) \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right) - (\beta_{y|x} - \alpha_{y|x}) \left( \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} \right)
\end{array} \right] \begin{bmatrix} x_C \\ x_M \end{bmatrix}
\]

(i) The underdetermined system (32)-(34) has one degree of freedom. I exploit it by setting \(\beta_x = 0\), which immediately gives \(\alpha_x = \pm 1\) from equation (32). It remains to determine two unknowns, \(\alpha_y, \beta_y\), from two equations (33) and (34). From (33), \(\beta_y = \pm \sqrt{1 - \alpha_y^2}\). Using this relation along with \(\alpha_x = \pm 1, \beta_x = 0\) and (38), integrability condition (34) reads:

\[
(\alpha_y \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} - \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right) - \sqrt{1 - \alpha_y^2} \left( \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right)) = \\
\delta \left( \alpha_y \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} - \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right) + \sqrt{1 - \alpha_y^2} \left( \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right) \right)
\]

Reorganizing terms and solving for \(\alpha_y\) yields:

\[
\alpha_y = \pm \frac{(1 + \delta) \left( \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right)}{\sqrt{(1 - \delta)^2 \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} - \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right)^2 + (1 + \delta)^2 \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right)^2}}
\]

Using (40), \(\beta_y\) can be backed out from (33)\(^6\)

\[
\beta_y = \sqrt{1 - \alpha_y^2}
\]

(ii) Rearranging (39) yields:

\[
\alpha_y(1 - \delta) \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} - \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right) = \sqrt{1 - \alpha_y^2} (\delta + 1) \left( \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right)
\]

While RHS \(\geq 0, \forall \rho_x, \rho_y\), LHS \(\leq 0\) iff \(\rho_y \leq \rho_x\) (if \(\alpha_y > 0\)). It follows that \(\alpha_y \geq 0\) if \(\rho_x \leq \rho_y\) and \(\alpha_y < 0\) if \(\rho_x > \rho_y\). ■

**Proof of Proposition 3.**

**Computing the Assignment:** I will compute the assignment for all \(\delta \in [0, 1]\), verifying the three cases described in Remark 1.

(1) For \(\delta = 1\), it follows from (35) in Lemma 3

\[
\alpha_y = \pm 1.
\]

\(^6\)Notice that \(\beta_y = -\sqrt{1 - \alpha_y^2}\) is also possible but does not affect the result, which is why I focus on the positive square root.
The orthogonal transformation (37) then gives:

\[
\begin{bmatrix}
\frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) \\
\frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i})
\end{bmatrix}
\begin{bmatrix}
\pm 1 & 0 \\
0 & \pm 1
\end{bmatrix}
= \pm \begin{bmatrix}
\frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) \\
\frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i})
\end{bmatrix}
\text{ for } i \in \{x, y\}.
\]

(44)

To see that these are the spectral square roots of the covariance matrix (or minus one times them), I derive them below using the spectral square root decomposition,

\[
\Sigma = CD C'
\]

\[
\iff
\Sigma^{\frac{1}{2}} = CD^{\frac{1}{2}} C'
\]  

(45)

where \( D \) is a diagonal matrix with the eigenvalues of \( \Sigma \) as diagonal entries and \( C \) is a matrix of orthonormal eigenvectors of \( \Sigma \). Matrix \( \Sigma^{\frac{1}{2}} \) in (45) is called the spectral square root of \( \Sigma \). Notice that for \( \Sigma^{\frac{1}{2}} \) to be positive-definite, the positive square roots of the diagonal entries of \( D \) are used. From (45) it follows that \( \Sigma_y^{\frac{1}{2}} \) is given by:

\[
\Sigma_y^{\frac{1}{2}} = \begin{bmatrix}
\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\
\sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}}
\end{bmatrix}
\begin{bmatrix}
\sqrt{1+\rho_y} & 0 \\
0 & \sqrt{1-\rho_y}
\end{bmatrix}
\begin{bmatrix}
\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\
\sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}}
\end{bmatrix} = \begin{bmatrix}
\frac{1}{2}(\sqrt{1+\rho_y} + \sqrt{1-\rho_y}) & \frac{1}{2}(\sqrt{1+\rho_y} - \sqrt{1-\rho_y}) \\
\frac{1}{2}(\sqrt{1+\rho_y} - \sqrt{1-\rho_y}) & \frac{1}{2}(\sqrt{1+\rho_y} + \sqrt{1-\rho_y})
\end{bmatrix}
\]

which coincides with (44) (or with minus one times the spectral sq. root). Moreover, since

\[
\Sigma_y^{-\frac{1}{2}} = CD^{-\frac{1}{2}} C' = C
\begin{bmatrix}
\frac{1}{\sqrt{\lambda_1}} & 0 \\
0 & \frac{1}{\sqrt{\lambda_2}}
\end{bmatrix}
C'.
\]

(47)

where \( \lambda_1, \lambda_2 \) are the eigenvalues of \( \Sigma \) and where \( C \) is a matrix of the corresponding orthonormal eigenvectors, the matrix \( \Sigma_x^{-\frac{1}{2}} \) is given by

\[
\Sigma_x^{-\frac{1}{2}} = \begin{bmatrix}
\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\
\sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}}
\end{bmatrix}
\begin{bmatrix}
\frac{1}{\sqrt{1+\rho_x}} & 0 \\
0 & \frac{1}{\sqrt{1-\rho_x}}
\end{bmatrix}
\begin{bmatrix}
\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\
\sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}}
\end{bmatrix} = \begin{bmatrix}
\frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}}\right) \\
\frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}}\right)
\end{bmatrix}
\]

(48)

It follows that for \( \delta = 1 \), the Jacobian of the matching function in (12) is given by:

\[
D_{xy}^* = (\Sigma_y^{\frac{1}{2}} R_y)(\Sigma_x^{\frac{1}{2}} R_x)^{-1} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}} = \begin{bmatrix}
\frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_y}} \sqrt{1+\rho_x} + \frac{1}{\sqrt{1-\rho_y}} \frac{1}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_y}} \frac{1}{\sqrt{1-\rho_x}} - \frac{1}{\sqrt{1-\rho_y}} \sqrt{1+\rho_x}\right) \\
\frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_y}} \frac{1}{\sqrt{1-\rho_x}} - \frac{1}{\sqrt{1-\rho_y}} \sqrt{1+\rho_x}\right) & \frac{1}{2}
\left(\frac{1}{\sqrt{1+\rho_y}} \sqrt{1+\rho_x} + \frac{1}{\sqrt{1-\rho_y}} \frac{1}{\sqrt{1-\rho_x}}\right)
\end{bmatrix}
\]

(49)

(2) For \( \delta = 0 \), it follows from Lemma 3 that \( R_y \) and \( R_x \) are respectively given by

\[
R_y = \begin{bmatrix}
\frac{\alpha_y}{\sqrt{1-\alpha_y^2}} & -\sqrt{1-\alpha_y^2} \\
\frac{\alpha_y}{\sqrt{1-\alpha_y^2}} & \alpha_y
\end{bmatrix}
= \begin{bmatrix}
\pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}\right) & \pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} - \sqrt{(1-\rho_y)(1-\rho_x)}\right) \\
\pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} - \sqrt{(1-\rho_y)(1-\rho_x)}\right) & \pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}\right)
\end{bmatrix}
\]

\[
R_x = \begin{bmatrix}
\frac{\alpha_y}{\sqrt{1-\alpha_y^2}} & -\sqrt{1-\alpha_y^2} \\
\frac{\alpha_y}{\sqrt{1-\alpha_y^2}} & \alpha_y
\end{bmatrix}
= \begin{bmatrix}
\pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}\right) & \pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} - \sqrt{(1-\rho_y)(1-\rho_x)}\right) \\
\pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} - \sqrt{(1-\rho_y)(1-\rho_x)}\right) & \pm \frac{1}{2} \left(\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}\right)
\end{bmatrix}
\]

(50)
\[
R_x = \begin{bmatrix}
\alpha_x & -\sqrt{1 - \alpha_x^2} \\
\sqrt{1 - \alpha_x^2} & \alpha_x \\
\end{bmatrix} = \begin{bmatrix}
\pm 1 & 0 \\
0 & \pm 1 \\
\end{bmatrix}.
\] (51)

Let \(\Sigma_y^2\) and \(\Sigma_x^2\) be the spectral square roots of skill and productivity covariance matrices, given by (46) and by the inverse of (48), respectively. Then,

\[
\Sigma_y^2 R_y = \left[ \frac{1}{2} \left( \sqrt{1 + \rho_y} + \sqrt{1 - \rho_x} \right) \right] \frac{1}{2} \left( \sqrt{1 + \rho_x} - \sqrt{1 - \rho_x} \right)
\]

\[
= \left[ \frac{1}{2} \left( \frac{1}{\sqrt{1 + \rho_x}} - \frac{1}{\sqrt{1 - \rho_x}} \right) \right] \frac{1}{2} \left( \frac{1}{\sqrt{1 + \rho_x}} + \frac{1}{\sqrt{1 - \rho_x}} \right)
\]

\[
\Sigma_x^2 R_x^{-1} = \left[ \frac{1}{2} \left( \sqrt{1 + \rho_x} + \sqrt{1 - \rho_x} \right) \right] \frac{1}{2} \left( \sqrt{1 + \rho_x} - \sqrt{1 - \rho_x} \right)
\]

It follows that the Jacobian is given by:

\[
D_x y^* = \Sigma_y^2 R_y (\Sigma_x^2 R_x)^{-1} = \begin{bmatrix}
1 & 0 \\
\rho_y - \rho_x \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}} & \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}}
\end{bmatrix}
\] (53)

In the following, it is shown that (53) is equivalent to \(L_y(L_x)^{-1}\) where \(L_i, i \in \{x, y\}\), is the Cholesky square root of skill and productivity covariance matrices, which is the unique lower triangular matrix \(L_i\) such that \(L_i(L_i)^T = \Sigma_i\). By definition, \(L_i\) is a square root of \(\Sigma_i\). Under the assumption of standard normality, \(L_i\) is given by:

\[
L_i = \begin{bmatrix}
1 \\
\rho_i \\
0 \\
\sqrt{1 - \rho_i^2}
\end{bmatrix} \forall i \in \{x, y\}
\] (54)

Hence,

\[
L_y(L_x)^{-1} = \begin{bmatrix}
1 & 0 \\
\rho_y \sqrt{1 - \rho_y^2} \\
\rho_x \sqrt{1 - \rho_x^2} \\
\rho_y - \rho_x \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}}
\end{bmatrix}^{-1} = \begin{bmatrix}
1 & 0 \\
\rho_y - \rho_x \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}} & \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}}
\end{bmatrix}
\] (55)

which coincides with (53). The equilibrium assignment is then given by (12):

\[
\begin{bmatrix}
y_C^* \\
y_M^*
\end{bmatrix} = (\Sigma_y^2 R_y)(\Sigma_x^2 R_x)^{-1} \begin{bmatrix}
x_C \\
x_M
\end{bmatrix} = L_y(L_x)^{-1} \begin{bmatrix}
x_C \\
x_M
\end{bmatrix}
\]

(3) For each \(\delta \in (0, 1)\), the equilibrium assignment is given by (12) where the Jacobian is computed using the rotation matrices \(R_y\) and \(R_x\) from Lemma 3 that correspond to the specific \(\delta\). Note that \(\Sigma_x^{1/2}\) and \(\Sigma_y^{1/2}\) in \((\Sigma_y^2 R_y)(\Sigma_x^2 R_x)^{-1}\) are the spectral square roots of the covariance matrices.

Verify that the Obtained Assignment Function is the Equilibrium Assignment. Three properties have to be verified, (a) Consistency with market clearing; (b) the assignment satisfies PAM; (c) the inte-
grability condition is satisfied: (a) Market clearing is satisfied by (10) and because the transformation (8) is measure-preserving. (b) Verifying the PAM-property amounts to checking that $D_{x'y'}$ is a P-matrix. Using Lemmas 2 and 3, equilibrium assignment (38) can be simplified by substituting in the expressions for $\alpha$’s and $\beta$’s from the rotation matrices:

$$
\begin{bmatrix}
y_C \\
y_M
\end{bmatrix}
= D_{x'y'}
\begin{bmatrix}
x_C \\
x_M
\end{bmatrix}
= 
\frac{\partial}{\partial x_C}
\begin{bmatrix}
y_C \\
y_M
\end{bmatrix}
\begin{bmatrix}
x_C \\
x_M
\end{bmatrix}$$

Using Lemmas 2 and 3, equilibrium assignment (38) can be simplified by substituting in the expressions for $\alpha$’s and $\beta$’s from the rotation matrices:

$$
\begin{bmatrix}
y_C \\
y_M
\end{bmatrix}
= 
\begin{bmatrix}
J_{11}(p_r, \rho_r, \delta)
& J_{12}(p_r, \rho_r, \delta)

J_{21}(p_r, \rho_r, \delta)
& J_{22}(p_r, \rho_r, \delta)
\end{bmatrix}
\begin{bmatrix}
x_C \\
x_M
\end{bmatrix}
= 
\frac{1+\sqrt{1-\rho_r}}{\sqrt{1-\rho_r}}
\frac{\delta(p_r-\rho_r, \sqrt{1-\rho_r})}{\sqrt{1+2\delta(p_r-\rho_r, \sqrt{1-\rho_r})+4\delta}}
\begin{bmatrix}
x_C \\
x_M
\end{bmatrix}
$$

Taking derivatives yields:

$$
\frac{\partial y^*_C}{\partial x_C} > 0 \quad (57)
$$

$$
\frac{\partial y^*_M}{\partial x_M} > 0 \quad (58)
$$

where (57) and (58) follow immediately from (56). Hence, $D_{x'y'}$ is a P-matrix. (c) The assignment was derived under the integrability condition (34), i.e. it is satisfied.

8.2.3 The Equilibrium Wage Function

**Proof of Proposition 4.** I proceed by guess and verify. The guess is that equilibrium wage function is given by the sum of marginal products integrated along the assignment paths

$$
w(x_C, x_M) = \alpha \left( \int_0^{x_C} J_{11} \hat{x}_C d\hat{x}_C + \frac{1}{2} \int_0^{x_M} J_{12} \delta \hat{x}_C + \frac{1}{2} \int_0^{x_M} J_{21} \delta \hat{x}_M + \delta \int_0^{x_M} J_{22} \hat{x}_M d\hat{x}_M \right)
= \alpha \left( \frac{1}{2} J_{11} x_C^2 + J_{12} x_M x_C + \frac{1}{2} \delta J_{22} x_M^2 \right) + w_0
$$

where $w_0$ is the constant of integration, which I set to zero. The elements $J_{11}, J_{12}, J_{21}, J_{22}$ are the entries of the matching function’s Jacobian, given by (56). Hence,

$$
w(x_C, x_M) = \frac{1}{2} \alpha [x_C \times x_M]
$$

where $(x_C, x_M) = (0, 0)$ is the least productive worker. He produces zero output. $w_0$ is his reservation wage, making the least productive worker, $(x_C, x_M) = (0, 0)$, indifferent between working and not working.
which is equivalent to (14) in Proposition 4, taking into account \( w_0 = 0 \). The guess needs to be verified. Given (60), the partial derivatives of the wage with respect to skills \((x_C, x_M)\) are given by

\[
\frac{\partial w(x_C, x_M)}{\partial x_C} = \alpha (J_{11} x_C + J_{12} x_M) \tag{61}
\]

\[
\frac{\partial w(x_C, x_M)}{\partial x_M} = \alpha \delta (J_{22} x_M + J_{21} x_C) \tag{62}
\]

which coincide with the first-order conditions of the firm,

\[
\frac{\partial w(x_C, x_M)}{\partial x_C} = \alpha y_C^* \tag{63}
\]

\[
\frac{\partial w(x_C, x_M)}{\partial x_M} = \alpha \delta y_M^* \tag{64}
\]

evaluated at the equilibrium assignment (56). Moreover, the integrability condition

\[
\frac{\partial^2 w(x_C, x_M)}{\partial x_C \partial x_M} = \frac{\partial^2 w(x_C, x_M)}{\partial x_M \partial x_C} \iff J_{12} = \delta J_{21} \tag{65}
\]

is satisfied by construction of the equilibrium assignment, ensuring the Hessian of the wage function is symmetric. ■

**Proof of Proposition 5.** (i) See proof of Proposition 3 (last part, under *Verify that the Obtained Assignment Function is the Equilibrium Assignment.*).

(ii) To show the first claim, note that for \( \rho_x = \rho_y \), (35) yields

\[\alpha_y = \pm 1\]

and hence, \( \beta_y = 0 \). Substituting this (along with \( \alpha_x = \pm 1, \beta_x = 0 \)) into equilibrium assignment (38) yields \( y_C^* = x_C \) and \( y_M^* = x_M \) and, hence, \( J_{11} = \frac{\partial y_C^*}{\partial x_C} = 1, J_{22} = \frac{\partial y_M^*}{\partial x_M} = 1 \) and \( J_{12} = \frac{\partial y_C^*}{\partial x_M} = 0, J_{21} = \frac{\partial y_M^*}{\partial x_C} = 0 \). There is perfect assortativeness according to Definition 4.

The second claim is that mismatch in both dimensions is maximized at sup \( \rho_x, \rho_y \mid |\rho_x - \rho_y| = 2 \), where maximum mismatch is defined by the largest deviations of \( J \mu \) from the identity matrix, i.e. by the largest deviation of \( J_{11} \) and \( J_{22} \) from one and of \( J_{12} \) and \( J_{21} \) from zero. By assumption, \( \rho_x, \rho_y \) are bounded away from \( \pm 1 \), which is why I here restrict attention to the compact sets \( \rho_x, \rho_y \in [-1+\varepsilon, 1-\varepsilon] \) for \( \varepsilon > 0 \) and small. Moreover, since for \( \delta = 1 \), \( J_{22} = J_{11} \) and \( J_{12} = J_{21} \), it suffices to show the claim for \( J_{22}(\rho_x, \rho_y) \equiv \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}} \right) \) and \( J_{21}(\rho_x, \rho_y) \equiv \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}} \right) \). I will show that \( |J_{22}(\cdot, \cdot) - 1| \) and \( |J_{21}(\cdot, \cdot)| \) are maximized at two points: \( (\rho_x^*, \rho_y^*) = (1-\varepsilon, 1+\varepsilon) \) and \( (\rho_x^*, \rho_y^*) = (-1+\varepsilon, 1-\varepsilon) \).

Note that by PAM, \( J_{22} > 0 \). For tractability, I will maximize \( J_{22}(\rho_x, \rho_y) \) instead of \( |J_{22}(\rho_x, \rho_y) - 1| \) and will show that \( \max(J_{22}(\rho_x, \rho_y)) > 2 \) and therefore \( \max(J_{22}(\rho_x, \rho_y) - 1) > 1 \) where 1 is an upper
bound for the absolute value $|\min(J_{22}(\rho_x, \rho_y)) - 1|$. Hence, maximizing $J_{22}(\rho_x, \rho_y)$ is equivalent to maximizing $|J_{22}(\rho_x, \rho_y) - 1|$. I proceed in several steps to show the claim.

**Step 1:** For any $\rho_y$, $J_{22}(\cdot, \rho_y)$ is maximized either at $\rho_x^* = 1 - \varepsilon$ or at $\rho_x^* = -1 + \varepsilon$.

Suppose contrary to the claim that the maximum of $J_{22}$ is achieved at an interior point $\rho_x^*, \rho_y^* \in (-1 + \varepsilon, 1 - \varepsilon)$, where $\rho_x^*, \rho_y^*$ denote the maximizers. Then, at this interior solution it must hold that for any other $\rho_x, \rho_y$,

$$J_{22}(\rho_x^*, \rho_y^*) \geq \frac{1}{2} \left( \frac{1 + \rho_y}{\sqrt{1 + \rho_x}} + \frac{1 - \rho_y}{\sqrt{1 - \rho_x}} \right).$$

Towards a contradiction, set $\rho_y^* = 1 - \varepsilon$ for $\rho_y^* \in (-1 + \varepsilon, 1 - \varepsilon)$. Then, $\exists \hat{\varepsilon}$ s.t. for $\varepsilon < \hat{\varepsilon}$, (66) is violated; similarly, for setting $\rho_x^* = 1 + \varepsilon$. This rules out an optimum where $\rho_x^*$ is interior when $\rho_y^*$ is interior. Moreover, it is easy to verify that when $\rho_y^*$ is at a corner, then $J_{22}$ is maximized when $\rho_x^*$ is at a corner and $J_{22}(1 - \varepsilon, -1 + \varepsilon) = J_{22}(-1 + \varepsilon, 1 - \varepsilon) = J_{22}(1 - \varepsilon, 1 - \varepsilon) = J_{22}(1 + \varepsilon, 1 + \varepsilon) = 1$ when $\varepsilon$ is small. Thus, $J_{22}(\cdot, \rho_y)$ is maximized at $\rho_x^* = 1 - \varepsilon$ for $\rho_y \in (-1 + \varepsilon, 1 - \varepsilon)$, or $\rho_x^* = 1 + \varepsilon$ for $\rho_y \in [-1 + \varepsilon, 1 - \varepsilon]$.

**Step 2:** There are two points at which $J_{22}(\cdot, \cdot)$ is maximized: at $(\rho_x^*, \rho_y^*) = (1 - \varepsilon, 1 - \varepsilon)$ and at $(\rho_x^*, \rho_y^*) = (-1 + \varepsilon, -1 + \varepsilon)$. Since $J_{22}$ is a continuous function over a compact support, there exists a solution to the problems $\max_{\rho_y \in [-1 + \varepsilon, 1 - \varepsilon]} J_{22}(1 - \varepsilon, \rho_y)$ and $\max_{\rho_y \in [-1 + \varepsilon, 1 - \varepsilon]} J_{22}(-1 + \varepsilon, \rho_y)$ and they are picked by the Kuhn-Tucker conditions. Focus on $\max_{\rho_y \in [-1 + \varepsilon, 1 - \varepsilon]} J_{22}(1 - \varepsilon, \rho_y)$ (the argument for $\max_{\rho_y \in [-1 + \varepsilon, 1 - \varepsilon]} J_{22}(-1 + \varepsilon, \rho_y)$ is analogous). Since Step 1 rules out that $\rho_y^* = 1 - \varepsilon$, one constraint can be dropped and the following Kuhn-Tucker conditions must hold at an optimum:

$$\begin{align*}
(1) \quad \frac{\partial J_{22}(1 - \varepsilon, \rho_y)}{\partial \rho_y} + \lambda &= 0 \\
(2) \quad \lambda &\geq 0 \\
(3) \quad \rho_y &\geq -1 + \varepsilon \\
(4) \quad \lambda(\rho_y + 1 - \varepsilon) &= 0
\end{align*}$$

where $\lambda$ is the multiplier on constraint $\rho_y \geq -1 + \varepsilon$. First, suppose that $\rho_y > -1 + \varepsilon$ (and $\lambda = 0$). For (1) to hold it must be that $\frac{\partial J_{22}}{\partial \rho_y} = 0$, but

$$\frac{\partial J_{22}(1 - \varepsilon, \rho_y)}{\partial \rho_y} = \frac{1}{4} \left( \frac{1}{\sqrt{2 - \varepsilon} \sqrt{1 + \rho_y}} - \frac{1}{\sqrt{\varepsilon} \sqrt{1 - \rho_y}} \right) < 0$$

is negative $\forall \rho_y > -1 + \varepsilon$. Hence, there is no interior critical point. Second, suppose that $\rho_y = -1 + \varepsilon$. Then, $\frac{\partial J_{22}(1 - \varepsilon, -1 + \varepsilon)}{\partial \rho_y} = \lambda \Leftrightarrow \lambda = 0$, implying that (1)-(4) are satisfied. A result, the only critical point of $J_{22}(1 - \varepsilon, \rho_y)$ is at $\rho_y^* = -1 + \varepsilon$, which is the solution to the max-problem of Step 2.

Therefore, Steps 1 and 2 show that $J_{22}(\rho_x, \rho_y)$ has a maximum at $(\rho_x^*, \rho_y^*) = (1 - \varepsilon, -1 + \varepsilon)$. Similarly, there exists a second maximum at $(\rho_x^*, \rho_y^*) = (-1 + \varepsilon, 1 - \varepsilon)$ (i.e. the solution to $\max_{\rho_y \in [-1 + \varepsilon, 1 - \varepsilon]} J_{22}(-1 + \varepsilon, \rho_y)$). Note that $J_{22}(1 - \varepsilon, -1 + \varepsilon) = J_{22}(-1 + \varepsilon, 1 - \varepsilon) = \frac{1}{2} \left( \frac{\sqrt{2 - \varepsilon}}{\sqrt{\varepsilon}} + \frac{\sqrt{\varepsilon}}{\sqrt{2 - \varepsilon}} \right)$ and thus for small $\varepsilon$: $|J_{22}(\rho_x^*, \rho_y^*) - 1| > |\min(J_{22}(\rho_x, \rho_y)) - 1|$ (where $\min(J_{22}(\rho_x, \rho_y)) \approx 0.5$ is the lowest possible value of $J_{22}$).

**Step 3:** There are two points at which $|J_{21}(\cdot, \cdot)|$ is maximized: at $(\rho_x^*, \rho_y^*) = (-1 + \varepsilon, 1 - \varepsilon)$ and
at \((\rho_x^*, \rho_y^*) = (1 - \varepsilon, -1 + \varepsilon)\). First consider the case when \(\rho_x < \rho_y\), i.e., when \(J_{21}(\rho_x, \rho_y) > 0\). The objective function that is to be maximized is given by:

\[
L(\rho_x, \rho_y) = J_{21}(\rho_x, \rho_y) + \lambda_1(\rho_x + 1 - \varepsilon) + \lambda_2(-\rho_x + 1 - \varepsilon) + \lambda_y1(\rho_y + 1 - \varepsilon) + \lambda_y2(-\rho_y + 1 - \varepsilon)
\]

where I neglect the constraint \(\rho_x < \rho_y\) for now; it will be verified ex-post. Since \(J_{22}\) is a continuous function over a compact support, there exists a solution to this problem and any solution must satisfy the Kuhn-Tucker conditions:

\[
(1) \frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_x} + \lambda_1 - \lambda_2 = 0 \quad \frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_y} + \lambda_1 - \lambda_2 = 0 \quad (2) \lambda_1, \lambda_2 \geq 0, \lambda_y1, \lambda_y2 \geq 0
\]

\[
(3) \rho_y + 1 - \varepsilon \geq 0 \quad -\rho_y + 1 - \varepsilon \geq 0 \quad \rho_x + 1 - \varepsilon \geq 0 \quad -\rho_x + 1 - \varepsilon \geq 0
\]

\[
(4) \lambda_y1(\rho_y + 1 - \varepsilon) = 0 \quad \lambda_y2(-\rho_y + 1 - \varepsilon) = 0 \quad \lambda_x1(\rho_x + 1 - \varepsilon) = 0 \quad \lambda_x2(-\rho_x + 1 - \varepsilon) = 0
\]

where

\[
\frac{\partial J_{21}(\rho_x, \rho_y)}{\partial \rho_y} = \frac{1}{4} \left( \frac{1}{\sqrt{1 + \rho_y \sqrt{1 + \rho_x}}} + \frac{1}{\sqrt{1 - \rho_y \sqrt{1 - \rho_x}}} \right)
\]

\[
\frac{\partial J_{21}(\rho_x, \rho_y)}{\partial \rho_x} = \frac{1}{4} \left( -\frac{\sqrt{1 + \rho_y}}{(1 + \rho_x)^{3/2}} - \frac{\sqrt{1 - \rho_y}}{(1 - \rho_x)^{3/2}} \right)
\]

(67) (68)

I will show that only the solution, \((\rho_x^*) = -1 + \varepsilon, (\rho_y^*) = -1 - \varepsilon\), satisfies the KT-conditions (1)-(4). To see this, consider all cases:

(i) Interior Solution: \(-1 + \varepsilon < \rho_x < 1 - \varepsilon, -1 + \varepsilon < \rho_y < 1 - \varepsilon\) and \(\lambda_x1 = \lambda_x2 = \lambda_y1 = \lambda_y2 = 0\). Then, by (1), it must be that \(\frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_x} = \frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_y} = 0\). But from (67) and (68), \(\frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_x} < 0\), \(\frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_y} > 0\), violating (1).

(ii) \(\rho_x = 1 - \varepsilon, -1 + \varepsilon < \rho_y < 1 - \varepsilon\) and \(\lambda_x1 = 0, \lambda_y1 = \lambda_y2 = 0\). Again, (1) is violated.

(iii) \(\rho_x = 1 + \varepsilon, -1 + \varepsilon < \rho_y < 1 - \varepsilon\) and \(\lambda_x2 = 0, \lambda_y1 = \lambda_y2 = 0\). Similar to (ii), (1) is violated.

(iv) \(-1 + \varepsilon < \rho_x < 1 - \varepsilon, \rho_y = 1 - \varepsilon\) and \(\lambda_x1 = 0, \lambda_x2 = 0, \lambda_y1 = 0\). Then (1) is violated since \(\frac{\partial J_{12}(\rho_x, 1-\varepsilon)}{\partial \rho_x} < 0\) all interior \(\rho_x\).

(v) \(-1 + \varepsilon < \rho_x < 1 - \varepsilon, \rho_y = -1 + \varepsilon\) and \(\lambda_x1 = 0, \lambda_x2 = 0, \lambda_y1 = 0\). As in (iv), (1) is violated.

(vi) Corner Solution: \(\rho_x = 1 - \varepsilon, \rho_y = 1 - \varepsilon\) and \(\lambda_x1 = 0, \lambda_y1 = 0\). From (1), one obtains \(\frac{\partial J_{12}(1-\varepsilon, 1-\varepsilon)}{\partial \rho_x} = \lambda_x2\) but, from (68), \(\frac{\partial J_{12}(1-\varepsilon, 1-\varepsilon)}{\partial \rho_y} < 0\), violating (2).

(vii) Corner Solution: \(\rho_x = 1 - \varepsilon, \rho_y = -1 + \varepsilon\) and \(\lambda_x1 = 0, \lambda_y2 = 0\). From (1), one obtains \(\frac{\partial J_{12}(1-\varepsilon, -1+\varepsilon)}{\partial \rho_x} = \lambda_x2\) but, from (68), \(\frac{\partial J_{12}(1-\varepsilon, -1+\varepsilon)}{\partial \rho_y} < 0\), violating (2).

(viii) Corner Solution: \(\rho_x = -1 + \varepsilon, \rho_y = -1 + \varepsilon\) and \(\lambda_x2 = 0, \lambda_y2 = 0\). From (1), one obtains \(\frac{\partial J_{12}(-1+\varepsilon, -1+\varepsilon)}{\partial \rho_x} = \lambda_y1\) but, from (67), \(\frac{\partial J_{12}(-1+\varepsilon, -1+\varepsilon)}{\partial \rho_y} > 0\), and thus \(\lambda_y1 < 0\), violating (2).

(ix) Corner Solution: \(\rho_x = -1 + \varepsilon, \rho_y = 1 - \varepsilon\) and \(\lambda_x2 = 0, \lambda_y1 = 0\). From (1), \(\frac{\partial J_{12}(-1+\varepsilon, 1-\varepsilon)}{\partial \rho_x} = \lambda_x1 > 0\) and \(\frac{\partial J_{12}(-1+\varepsilon, 1-\varepsilon)}{\partial \rho_y} = \lambda_y2 > 0\), satisfying (1)-(4). This is the unique critical point/solution.

Note that at this solution, \(\rho_y > \rho_x\), verifying this constraint ex-post. Since the objective is to maximize the absolute value of \(J_{21}\), one also has to consider the case of \(\rho_y < \rho_x\) and thus \(J_{21} < 0\).
i.e. the problem of minimizing $J_{21}$. An analogous argument to the one above establishes that $J_{21}$ is minimized at $\rho_x^* = 1 - \varepsilon, \rho_y^* = -1 + \varepsilon$. Note that $|J_{12}(1 - \varepsilon, -1 + \varepsilon)| = J_{12}(-1 + \varepsilon, 1 - \varepsilon) = \frac{1}{2} \left( \frac{\sqrt{2+\varepsilon}}{-\varepsilon} - \frac{\sqrt{2}}{-2+\varepsilon} \right) > 0$ (for $\varepsilon$ small), which proves the claim of Step 3.

Step 1-3 imply that mismatch in the task $M$, defined as the deviation of $J_{22}$ from 1 and of $J_{21}$ from zero, is largest for $\sup_{\rho_x,\rho_y}|\rho_x - \rho_y| = 2$. The results for task $C$ follow from symmetry ($\delta = 1$).

Note that the maximum mismatch, as defined by the maximum distance between the Jacobian of the matching function $J_{\mu}$ and the identity matrix (which is achieved at $\sup_{\rho_x,\rho_y}|\rho_x - \rho_y| = 2$), coincides with the maximum mean squared deviation of skills and job attributes in the equilibrium assignment, which gives an alternative intuitive way to capture mismatch and shows that these two measures of mismatch are closely related. Consider the following result:

**Lemma 4** (Maximum Mean Squared Deviation of Skills and Skill Demands) $\mathbb{E}[(y_i^* - x_i)^2]$ is maximized at two points, at $\rho_x^* = -1 + \varepsilon, \rho_y^* = 1 - \varepsilon$ and at $\rho_x^* = 1 - \varepsilon, \rho_y^* = -1 + \varepsilon$, which coincide with the two points at which $J_{\mu}$ deviates most from the identity matrix (by Proposition 5).

**Proof.** First note that for $i = \{C, M\}$:

$$\mathbb{E}[(y_i^* - x_i)^2] = \mathbb{E}[(J_{11}x_C + J_{12}x_M - x_i)^2]$$

Focus on $i = C$ and thus $\mathbb{E}[(y_C^* - x_C)^2] = (J_{11} - 1)^2 + 2J_{12}(J_{11} - 1)\rho_x + J_{12}^2$. The objective is to show that $\mathbb{E}[(y_C^* - x_C)^2]$ is maximized at two points: at $\rho_x^* = 1 - \varepsilon, \rho_y^* = 1 - \varepsilon$ and at $\rho_x^* = 1 - \varepsilon, \rho_y^* = -1 + \varepsilon$. (The case of maximizing $\mathbb{E}[(y_M^* - x_M)^2]$ is identical.) The objective function of problem $\max_{\rho_x,\rho_y\in[-1+\varepsilon,1-\varepsilon]} \mathbb{E}[(y_C^* - x_C)^2]$ is given by

$$L(\rho_x, \rho_y) = \left( (J_{11}(\rho_x, \rho_y) - 1)^2 + 2J_{12}(\rho_x, \rho_y)(J_{11}(\rho_x, \rho_y) - 1)\rho_x + J_{12}^2 \right)$$

where, as before, $J_{11}(\rho_x, \rho_y) \equiv \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right)$ and $J_{12}(\rho_x, \rho_y) = \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right)$. The Kuhn-Tucker conditions are given by

1. $\frac{\partial L(\rho_x, \rho_y)}{\partial \rho_x} + \lambda_{x1} - \lambda_{x2} = 0 \quad \frac{\partial L(\rho_x, \rho_y)}{\partial \rho_y} + \lambda_{y1} - \lambda_{y2} = 0$
2. $\lambda_{x1}, \lambda_{x2} \geq 0, \lambda_{y1}, \lambda_{y2} \geq 0$
3. $\rho_x + 1 - \varepsilon \geq 0 \quad -\rho_y + 1 - \varepsilon \geq 0 \quad \rho_x + 1 - \varepsilon \geq 0 \quad -\rho_x + 1 - \varepsilon \geq 0$
4. $\lambda_{y1}(\rho_y + 1 - \varepsilon) = 0 \quad \lambda_{y2}(-\rho_y + 1 - \varepsilon) = 0 \quad \lambda_{x1}(\rho_x + 1 - \varepsilon) = 0 \quad \lambda_{x2}(-\rho_x + 1 - \varepsilon) = 0$

where

$$\frac{\partial L(\rho_x, \rho_y)}{\partial \rho_x} = 2(J_{11}(\rho_x, \rho_y) - 1)\frac{\partial J_{11}(\rho_x, \rho_y)}{\partial \rho_x} + 2 \left( \frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_x}(J_{11}(\rho_x, \rho_y) - 1) + J_{12}(\rho_x, \rho_y)\frac{\partial J_{11}(\rho_x, \rho_y)}{\partial \rho_x} \right) \rho_x$$

$$+ 2J_{12}(\rho_x, \rho_y)(J_{11}(\rho_x, \rho_y) - 1) + 2J_{12}(\rho_x, \rho_y)\frac{\partial J_{12}(\rho_x, \rho_y)}{\partial \rho_x}$$
To find the solution(s) to this maximization problem, consider all cases:

(i) Interior Solution: \(-1 + \varepsilon < \rho_x, \rho_y < 1 - \varepsilon\) and \(\lambda_{x1} = \lambda_{x2} = \lambda_{y1} = \lambda_{y2} = 0\). Then, by (1), it must be that \(\frac{\partial L(\rho_x, \rho_y)}{\partial \rho_x} = \frac{\partial L(\rho_x, \rho_y)}{\partial \rho_y} = 0\), which holds if \(\rho_x = \rho_y\), which is thus a critical point.

(ii) \(\rho_x = 1 - \varepsilon, -1 + \varepsilon < \rho_y < 1 - \varepsilon\) and \(\lambda_{x1} = \lambda_{y1} = \lambda_{y2} = 0\). (1) is violated since \(\frac{\partial L(1-\varepsilon, \rho_y)}{\partial \rho_y} < 0\) for \(\rho_x > \rho_y\).

(iii) \(\rho_x = -1 + \varepsilon, -1 + \varepsilon < \rho_y < 1 - \varepsilon\) and \(\lambda_{x2} = \lambda_{y1} = \lambda_{y2} = 0\). (1) is violated since \(\frac{\partial L(-1+\varepsilon, \rho_y)}{\partial \rho_y} > 0\) for \(\rho_x < \rho_y\).

(iv) \(-1 + \varepsilon < \rho_x < 1 - \varepsilon, \rho_y = 1 - \varepsilon\) and \(\lambda_{x1} = 0, \lambda_{x2} = 0, \lambda_{y1} = 0\). (1) is violated since \(\frac{\partial L(1-\varepsilon, \rho_y)}{\partial \rho_y} < 0\) for \(\rho_x < \rho_y\).

(v) \(-1 + \varepsilon < \rho_x < 1 - \varepsilon, \rho_y = -1 + \varepsilon\) and \(\lambda_{x1} = \lambda_{y2} = 0\). As in (iv), (1) is violated.

(vi) Corner Solution: \(\rho_x = 1 - \varepsilon, \rho_y = 1 - \varepsilon\) and \(\lambda_{x1} = 0, \lambda_{y1} = 0\). From (1), one obtains

\[
\frac{\partial L(1-\varepsilon, 1-\varepsilon)}{\partial \rho_x} = \lambda_{x2} = 0 \quad \text{and} \quad \frac{\partial L(1-\varepsilon, 1-\varepsilon)}{\partial \rho_y} = \lambda_{y2} = 0
\]

for \(\rho_x = \rho_y\). Hence, this is a critical point.

(vii) Corner Solution: \(\rho_x = 1 - \varepsilon, \rho_y = 1 - \varepsilon\) and \(\lambda_{x1} = 0, \lambda_{y2} = 0\). From (1), one obtains

\[
\frac{\partial L(1-\varepsilon, 1-\varepsilon)}{\partial \rho_x} = \lambda_{x2} > 0 \quad \text{and} \quad \frac{\partial L(1-\varepsilon, 1-\varepsilon)}{\partial \rho_y} = \lambda_{y1} > 0,
\]

satisfying (1)-(4). Thus, this is a critical point.

(viii) Corner Solution: \(\rho_x = -1 + \varepsilon, \rho_y = -1 + \varepsilon\) and \(\lambda_{x2} = 0, \lambda_{y2} = 0\). From (1), one obtains

\[
\frac{\partial L(-1+\varepsilon, -1+\varepsilon)}{\partial \rho_x} = \lambda_{x1} = 0 \quad \text{and} \quad \frac{\partial L(-1+\varepsilon, -1+\varepsilon)}{\partial \rho_y} = \lambda_{y1} = 0
\]

for \(\rho_x = \rho_y\). Hence, this is a critical point.

(ix) Corner Solution: \(\rho_x = -1 + \varepsilon, \rho_y = 1 - \varepsilon\) and \(\lambda_{x2} = \lambda_{y1} = 0\). From (1), one obtains

\[
\frac{\partial L(-1+\varepsilon, 1-\varepsilon)}{\partial \rho_x} = \lambda_{x1} > 0 \quad \text{and} \quad \frac{\partial L(-1+\varepsilon, 1-\varepsilon)}{\partial \rho_y} = \lambda_{y2} > 0,
\]

satisfying (1)-(4). Thus, this is a critical point.

Thus, the KT conditions pick three critical points: (a) any \(\rho_x, \rho_y\) with \(\rho_x = \rho_y\), at which \(J_{11}(\rho_x, \rho_x) = 1\) and \(J_{12}(\rho_x, \rho_x) = 0\), and thus the objective function evaluated at such points has zero value: \(J_{11}(\rho_x, \rho_x)(1-\varepsilon)+2J_{12}(\rho_x, \rho_x)(J_{11}(\rho_x, \rho_x)-1)\rho_x = 0\); (b) \(\rho_x = -1 + \varepsilon, \rho_y = 1 - \varepsilon\) and (c) \(\rho_x = 1 - \varepsilon, \rho_y = -1 + \varepsilon\). Evaluating the objective function at point (b) or (c) yields identical and strictly positive values, \((1-\varepsilon)(1-\varepsilon)(1-\varepsilon)(1-\varepsilon)+J_{12}(1-\varepsilon, 1-\varepsilon) + J_{12}(1-\varepsilon, 1-\varepsilon) + J_{12}(1-\varepsilon, 1-\varepsilon) > 0\) (for small and positive \(\varepsilon\)). Hence, (b) and (c) are solutions to the maximization problem and they coincide with the points that maximize the distance between \(J_{\mu}\) and the identity matrix, see proof of Proposition 5ii., which proves the claim. ■
**Proof of Proposition 6.** (i) Wages are convex in skills: Recall, the wage function is given by

\[ w(x_C, x_M) = \frac{1}{2} \alpha [x_C, x_M] J [\begin{array}{ccc} J_{11} & J_{12} \\ \delta J_{21} & \delta J_{22} \end{array}] [\begin{array}{c} x_C \\ x_M \end{array}] \]

where \( J_{11}, J_{12}, J_{21}, J_{22} \) are the elements of the matching function’s Jacobian given in (13). The Hessian of the wage function is given by

\[ H(w) = \frac{1}{2} \alpha J \]

Since the Jacobian of the matching function is a P-matrix, i.e. \( J_{11}, J_{22} > 0, J_{11}J_{22} - J_{12}J_{21} > 0 \) and by assumption, \( \delta > 0, \alpha > 0 \), it follows that \( \alpha J_{11} > 0, \alpha \delta J_{22} > 0 \) and \( \text{Det}(H(w)) = \frac{1}{2} \alpha \delta (J_{11}J_{22} - J_{12}J_{21}) > 0 \). Hence, \( H(w) \) is positive-definite, i.e. the wage function is convex.

(ii) The moments of the wage distribution (mean, variance and skewness) are given by

\[ E(w(x_C, x_M)) = \alpha \text{tr}(\bar{J}\Sigma_x) = \alpha \left[ \frac{1}{2} J_{11} + J_{12}\rho_x + \frac{1}{2} \delta J_{22} \right] \]

\[ \text{Var}(w(x_C, x_M)) = \alpha^2 2\text{tr}(\bar{J}\Sigma_x\bar{J}\Sigma_x) = \alpha^2 \left[ \frac{1}{2} ((J_{11} + J_{12}\rho_x)^2 + 2(J_{11}\rho_x + J_{12})(J_{12} + \delta J_{22}\rho_x) + (J_{12}\rho_x + \delta J_{22})^2) \right] \]

\[ E\left[ \left( \frac{w - E(w)}{\sqrt{\text{Var}(w)}} \right)^3 \right] = \frac{E(w^3) - 3E(w)\text{Var}(w) - E(w)^3}{\text{Var}(w)^{\frac{3}{2}}} = \frac{8\text{tr}(\bar{J}\Sigma_x\bar{J}\Sigma_x\bar{J}\Sigma_x)}{(2\text{tr}(\bar{J}\Sigma_x\bar{J}\Sigma_x))^\frac{3}{2}} \]

where \( J_{11}, J_{12}, J_{21}, J_{22} \) are defined in (13), \( \bar{J} \) denotes the Hessian of the wage function, and where

\[ E(w^3) = \alpha^3 [\text{tr}(J_{\mu} \Sigma_x)^3 + 6\text{tr}(J_{\mu} \Sigma_x)\text{tr}(J_{\mu} \Sigma_x J_{\mu} \Sigma_x) + 8\text{tr}(J_{\mu} \Sigma_x J_{\mu} \Sigma_x J_{\mu} \Sigma_x)] \]

\[ \text{tr}(\bar{J}\Sigma_x\bar{J}\Sigma_x\bar{J}\Sigma_x) = (J_{11} + J_{12}\rho_x)((J_{11} + J_{12}\rho_x)^2 + (J_{11}\rho_x + J_{12})(J_{12} + \delta J_{22}\rho_x)) + 2(J_{12} + \delta J_{22}\rho_x)(J_{11}\rho_x + J_{12})(J_{11} + J_{12}\rho_x + \delta J_{22}) + (J_{12}\rho_x + \delta J_{22})((J_{12} + \delta J_{22}\rho_x)(J_{11}\rho_x + J_{12}) + (J_{12}\rho_x + \delta J_{22})^2) \]

See e.g. Magnus [1978] for the derivation of moments of quadratic forms in normal variables.

Skewness: \( J_{\mu} \Sigma_x J_{\mu} \) and \( \Sigma_x J_{\mu} \Sigma_x \) are positive definite matrices. Since the trace of the product of two positive-definite matrices is positive and since the variance is positive, the result follows from (71).

**8.3 Proofs on Technological Change (Section 4)**

**8.3.1 Task-Biased Technological Change**

**Proof of Proposition 7.** An underlying assumption here is that either \( \rho_x, \rho_y < 0 \) or \( \rho_x, \rho_y > 0 \).

(i) **Mismatch-Assortativeness Trade-Off Across Tasks:** The claim is that under TBTC \( \delta' \) (with \( 0 < \)
\( \delta' < \delta \), \( J_{11} \) approaches 1 and \( J_{12} \) zero (while \( J_{22} \) diverges from 1 and \( J_{21} \) from zero). Recall \( J_\mu \):

\[
J_\mu = \begin{bmatrix}
J_{11}(\rho_x, \rho_y, \delta) & J_{12}(\rho_x, \rho_y, \delta) \\
J_{21}(\rho_x, \rho_y, \delta) & J_{22}(\rho_x, \rho_y, \delta)
\end{bmatrix}
= \begin{bmatrix}
\frac{1+\delta \sqrt{1-\rho_x^2}}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}} & \frac{\delta(\rho_y - \rho_x \sqrt{1-\rho_y^2})}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}} \\
\frac{\rho_x - \rho_y \sqrt{1-\rho_x^2}}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}} & \frac{1-\rho_x^2 \sqrt{1-\rho_y^2}}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}}
\end{bmatrix}
\tag{72}
\]

I will show this claim in two steps:

**Step 1:** I will show that

\[
\begin{align*}
J_{11} &\geq 1 \quad \text{iff} \quad |\rho_x| \geq |\rho_y| \\
J_{12} &\geq 0 \quad \text{iff} \quad \rho_x \leq \rho_y \\
J_{21} &\geq 0 \quad \text{iff} \quad \rho_x \leq \rho_y \\
J_{22} &\geq 1 \quad \text{iff} \quad |\rho_x| \geq |\rho_y|.
\end{align*}
\]

This follows from (72), since

\[
J_{11} = \frac{1+\delta \sqrt{1-\rho_x^2}}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}} \geq 1
\]

\[
\Leftrightarrow 2\delta \sqrt{1-\rho_x^2} - \rho_x \rho_y \sqrt{1-\rho_x^2} + \delta^2 \left( \frac{1-\rho_y^2}{1-\rho_x^2} - 1 \right) \geq 0 \quad \text{iff} \quad |\rho_x| \geq |\rho_y|
\]

\[
J_{12} = \frac{\delta(\rho_y - \rho_x \sqrt{1-\rho_y^2})}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}} \geq 0
\]

\[
\Leftrightarrow \rho_y \sqrt{1-\rho_x^2} - \rho_x \sqrt{1-\rho_y^2} \geq 0 \quad \text{iff} \quad \rho_x \leq \rho_y.
\]

\[
J_{21} = \frac{\rho_x - \rho_y \sqrt{1-\rho_x^2}}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}} \geq 0
\]

\[
\Leftrightarrow \rho_y \sqrt{1-\rho_x^2} - \rho_x \sqrt{1-\rho_y^2} \geq 0 \quad \text{iff} \quad \rho_x \leq \rho_y.
\]

\[
J_{22} = \frac{\delta + \sqrt{1-\rho_y^2}}{\sqrt{1+2\delta(\rho_x \rho_y + \sqrt{1-\rho_x^2} \sqrt{1-\rho_y^2})+\delta^2}} \geq 1
\]

\[
\Leftrightarrow 2\delta \sqrt{1-\rho_y^2} - \rho_y \rho_x \sqrt{1-\rho_y^2} + \left( \frac{1-\rho_y^2}{1-\rho_x^2} - 1 \right) \geq 0 \quad \text{iff} \quad |\rho_x| \geq |\rho_y|
\]
Step 2: I will show that
\[
\frac{\partial J_{11}}{\partial \delta} \geq 0 \quad \text{iff} \quad |\rho_x| \geq |\rho_y|
\]
\[
\frac{\partial J_{12}}{\partial \delta} \geq 0 \quad \text{iff} \quad \rho_x \leq \rho_y
\]
\[
\frac{\partial J_{21}}{\partial \delta} \geq 0 \quad \text{iff} \quad \rho_x \geq \rho_y
\]
\[
\frac{\partial J_{22}}{\partial \delta} \geq 0 \quad \text{iff} \quad |\rho_x| \leq |\rho_y|
\]
where
\[
\frac{\partial J_{11}}{\partial \delta} = \frac{\sqrt{1 - \rho_y^2} - \delta + \left( \delta \frac{1 - \rho_y^2}{\sqrt{1 - \rho_z^2}} - 1 \right) (\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2})}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2}) + \delta^2)^{\frac{3}{2}}}
\]
\[
\frac{\partial J_{12}}{\partial \delta} = \frac{\left( \rho_y - \rho_x \frac{1 - \rho_y^2}{\sqrt{1 - \rho_z^2}} \right) (1 + \delta(\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2}))}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2}) + \delta^2)^{\frac{3}{2}}} \geq 0 \quad \text{iff} \quad \rho_x \leq \rho_y
\]
\[
\frac{\partial J_{21}}{\partial \delta} = \frac{- \left( \rho_y - \rho_x \frac{1 - \rho_y^2}{\sqrt{1 - \rho_z^2}} \right) (\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2} + \delta)}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2}) + \delta^2)^{\frac{3}{2}}} \geq 0 \quad \text{iff} \quad \rho_x \geq \rho_y
\]
\[
\frac{\partial J_{22}}{\partial \delta} = \frac{1 - \delta \frac{1 - \rho_y^2}{\sqrt{1 - \rho_z^2}} + \left( \delta - \frac{1 - \rho_y^2}{\sqrt{1 - \rho_z^2}} \right) (\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2})}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2}) + \delta^2)^{\frac{3}{2}}}
\]

The signs of (74) and (75) follow immediately from the expressions.

Next consider (73), which will be shown to be nonnegative iff $|\rho_x| \geq |\rho_y|$. The denominator is positive. Focus on the numerator, which is nonnegative for both $\delta = 1$ and $\delta = 0$:
\[
\left. \frac{\partial J_{11}}{\partial \delta} \right|_{\delta=1} = \left( \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_z^2}} - 1 \right) (1 + \rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2}) \geq 0 \quad \text{iff} \quad |\rho_x| \geq |\rho_y|
\]
\[
\left. \frac{\partial J_{11}}{\partial \delta} \right|_{\delta=0} = \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_z^2}} - (\rho_x\rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_z^2}) \geq 0
\]
\[
\Leftrightarrow \frac{\sqrt{1 - \rho_y^2} - \sqrt{1 - \rho_y^2} \rho_x \rho_y}{\sqrt{1 - \rho_z^2}} \geq 0 \quad \text{iff} \quad |\rho_x| \geq |\rho_y|
\]

Moreover the numerator of (73) is affine and thus monotone in $\delta$, proving that (73) is nonnegative for all $\delta \in (0, 1]$ iff $|\rho_x| \geq |\rho_y|$. 
Next, consider (76). I proceed similarly. The numerator is nonnegative for both $\delta = 1$ and $\delta = 0$:

$$\frac{\partial J_{22}}{\partial \delta} \bigg|_{\delta=1} = \frac{1 \mp \sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_y^2}} \left( 1 + \rho_x \rho_y + \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}} \right) \geq 0 \quad \text{iff} \quad |\rho_x| \leq |\rho_y|$$

$$\frac{\partial J_{22}}{\partial \delta} \bigg|_{\delta=0} = 1 - \frac{1 - \rho_y^2}{\sqrt{1 - \rho_x^2}} \left( \rho_x \rho_y + \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}} \right) \geq 0 \quad \text{iff} \quad |\rho_x| \leq |\rho_y|$$

$$\iff \quad \frac{\sqrt{1 - \rho_x^2} \rho_y^2 - \sqrt{1 - \rho_y^2} \rho_y \rho_x}{\sqrt{1 - \rho_x^2}} \geq 0 \quad \text{iff} \quad |\rho_x| \leq |\rho_y|$$

Moreover, the numerator of (76) is affine and thus monotone in $\delta$, proving that (76) is nonnegative for all $\delta \in (0,1]$ if $|\rho_x| \leq |\rho_y|$.

Hence, Steps 1 and 2 prove the claim: For any $\rho_x, \rho_y$ that satisfy either $\rho_x < 0, \rho_y < 0$ or $\rho_x \geq 0, \rho_y \geq 0$, as $\delta$ decreases, $J_{11}$ approaches 1 and $J_{12}$ approaches 0, while $J_{22}$ diverges from 1 and $J_{21}$ diverges from 0. Hence, in dimension $C$, the distance between the Jacobian and the identity matrix shrinks, while for dimension $M$, the opposite is true, meaning that mismatch decreases in the cognitive and increases in the manual dimension.

(ii) Sorting when $\delta = 0$: In the cognitive task, $y_C = x_C$. This follows immediately from (72). In turn, in the manual task, mismatch is largest. To see this, notice that

$$J_{22} \bigg|_{\delta=0} = \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}} \geq \frac{\delta + \sqrt{1 - \rho_y^2}}{\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}}) + \delta^2}} = J_{22} \bigg|_{\delta \neq 0} \quad \text{iff} \quad |\rho_x| \geq |\rho_y|$$

since iff $|\rho_x| \geq |\rho_y|$:}

$$(1 - \rho_y^2)(1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}}) + \delta^2) \geq (1 - \rho_y^2)\delta^2 + 2\delta \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}} + (1 - \rho_y^2)$$

$$\iff \quad (1 - \rho_y^2)(2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}}) + \delta^2) \geq (1 - \rho_y^2)\delta^2 + 2\delta \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}}$$

$$\iff \quad 2(\rho_x \rho_y(1 - \rho_y^2) - \rho_y^2 \sqrt{1 - \rho_x^2} \sqrt{1 - \rho_y^2} + \delta^2(\rho_x^2 - \rho_y^2)) \geq 0.$$}

Since $J_{22} \geq 1$ iff $|\rho_x| \geq |\rho_y|$, it then follows that $J_{22} \big|_{\delta=0} \geq J_{22} \big|_{\delta \neq 0} \geq 1$ iff $|\rho_x| \geq |\rho_y|$. Moreover,

$$J_{21} \bigg|_{\delta=0} = \rho_y - \rho_x \sqrt{1 - \rho_y^2} \geq \frac{\delta \left( \rho_y - \rho_x \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}} \right)}{\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}}) + \delta^2}} = J_{21} \bigg|_{\delta \neq 0} \quad \text{iff} \quad \rho_x \leq \rho_y$$

since iff $\rho_x \leq \rho_y$, the following inequality holds:

$$\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2 \sqrt{1 - \rho_x^2}}) + \delta^2} \geq \delta$$

Since $J_{21} \geq 0$ iff $\rho_x \leq \rho_y$, it then follows that $J_{21} \big|_{\delta=0} \geq J_{21} \big|_{\delta \neq 0} \geq 0$ iff $\rho_x \leq \rho_y$. 

This completes the proof that for \( \delta = 0 \), there is the largest distance between \( J_\mu \) and the identity matrix in the manual task dimension, and thus the largest mismatch in that dimension.

(iii) Claim 1: Task-biased TC has no effect on sorting for \( \rho_x = \rho_y \). This follows from Proposition 5ii.

Claim 2: Task-biased TC has the maximal mismatch-reducing effect on cognitive sorting (i.e. \( |\hat{J}_{\hat{\delta}}|, |\hat{J}_{\hat{\delta}2}|, \) are maximized) at \( \sup_{\rho_x, \rho_y} |\rho_x - \rho_y| = 1 \). Recall that Section 4 assumes that \( \rho_x, \rho_y \) are either both nonnegative (henceforth, Case 1) or both nonpositive (Case 2). Moreover, by assumption, \( \rho_x, \rho_y \) are bounded away from \( \pm 1 \), which is why I here restrict attention to the compact sets \( \rho_x, \rho_y \in [-1 + \varepsilon, 0] \) and \( \rho_x, \rho_y \in [0, 1 + \varepsilon] \) for \( \varepsilon > 0 \) and small.

I start with analyzing \( |\frac{\partial J_{\hat{\delta}}}{\partial \hat{\delta}}| \), given by (73). I denote \( A(\rho_x, \rho_y) = \frac{\partial J_{\hat{\delta}}}{\partial \hat{\delta}} \). I prove the claim in two steps.

Step 1: For all \( \rho_y \), \( |A(\cdot, \rho_y)| \) is maximized either at \( \rho_x^* = 1 - \varepsilon \) (Case 1) or at \( \rho_x^* = -1 + \varepsilon \) (Case 2). Consider the case where \( |\rho_x| \geq |\rho_y| \), implying that \( A(\rho_x, \rho_y) \geq 0 \) (by Step 1 in Proposition 7, (i)). The constraint \( |\rho_x| \geq |\rho_y| \) will be verified below. First, focus on Case 1 where \( \rho_x, \rho_y \) are non-negative. Suppose now that the maximum of \( A(\cdot, \cdot) \) is achieved at an interior point \( \rho_x^*, \rho_y^* \in (0, 1 - \varepsilon) \) (where \( \rho_x^*, \rho_y^* \) denote the maximizers), at which the following must hold:

\[
A(\rho_x^*, \rho_y^*) \geq \frac{\sqrt{1-\rho_y^2} - (\delta + 1)}{(1 + \delta)^3}
\]  

(77)

The RHS is a lower bound of (73). (I chose \( \rho_x, \rho_y \) in \( \rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2} \) in numerator and denominator of (73) s.t. (73) becomes least positive\(^{64}\), where \( \rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2} \in (0, 1) \) for \( \rho_x, \rho_y \geq 0 \). That is, in the denominator of (73) set \( \rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2} = 1 \); in the numerator, if \( \rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2} \) is multiplied by something positive (negative) set it to 0 (1).)

Towards a contradiction, set \( \rho_x^* = 1 - \varepsilon \) for \( \rho_y^* \in (0, 1 - \varepsilon) \). Then, \( \exists \varepsilon \) s.t. for \( \varepsilon < \hat{\varepsilon} \), (77) is violated. This rules out an optimum where both \( \rho_x^*, \rho_y^* \) are interior. If \( \rho_y^* \) is interior, it must be that \( \rho_x^* = 1 - \varepsilon \). Regarding solutions where both \( \rho_x^*, \rho_y^* \) are at a corner, it is easy to verify that \( A(1 - \varepsilon, 0) \) is larger than \( A(1 - \varepsilon, 1 - \varepsilon), A(0, 1 - \varepsilon) \) and \( A(0, 0) \). Finally, there is no solution where \( \rho_y^* \) is at a corner and \( \rho_x^* \) is interior since for every \( \rho_x \in (0, 1 - \varepsilon) \), \( A(1 - \varepsilon, 0) > A(\rho_x, 0) \) and \( A(1 - \varepsilon, 0) > A(\rho_x, 1 - \varepsilon) \). Thus, \( A(\rho_x, \rho_y) \) is maximized at \( \rho_x^* = 1 - \varepsilon \) and \( \rho_y^* \in [0, 1 - \varepsilon) \). Note that, since \( \rho_x^* \) is at the boundary, it follows that \( \rho_x \geq \rho_y \), and hence \( |\rho_x| \geq |\rho_y| \), verifying the imposed constraint as well as \( |A(\rho_x, \rho_y)| = A(\rho_x, \rho_y) \).

(Note that for Case 2, an analogous argument shows that \( \rho_x^* = -1 + \varepsilon \) and \( \rho_y^* \in (-1 + \varepsilon, 0] \).)

Step 2: \( A(1 - \varepsilon, \cdot) \) is maximized at \( \rho_y^* = 0 \) (Case 1). (Case 2: \( A(-1 + \varepsilon, \cdot) \) is maximized at \( \rho_y^* = 0 \).) Since \( A \) is a continuous function over a compact support, there exists a solution to the problem \( \max_{\rho_y \in [0, 1 - \varepsilon]} A(1 - \varepsilon, \rho_y) \) and it is picked by the Kuhn-Tucker conditions. Since Step 1 rules out that

---

\(^{64}\)Note that the RHS of (77) could even be negative here, in which case the bound is irrelevant since (77) is automatically satisfied due to \( A(\rho_x, \rho_y) \geq 0 \) for any \( |\rho_x| \geq |\rho_y| \) that are consistently chosen in numerator and denominator.
\( \rho_y^* = 1 - \varepsilon \), the following Kuhn-Tucker conditions must hold at an optimum:

\[
\begin{align*}
(1) & \quad \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} \leq 0 & (2) & \quad \rho_y \geq 0 & (3) & \quad \rho_y \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} = 0
\end{align*}
\]

I will first show that for \( \rho_y = 0 \) (corner solution), all conditions (1)-(3) are satisfied. Second, I will show that for \( \rho_y > 0 \) (interior solution), \( \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} = 0 \) fails to hold, i.e. condition (3) fails to hold.

To see this, note that:

\[
\frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} = \frac{-3\delta \left( 1 - \varepsilon \right) - \frac{\sqrt{1 - (1 - \varepsilon)^2} \rho_y}{\sqrt{1 - \rho_y^2}} \left( \sqrt{\frac{1 - \rho_y^2}{1 - (1 - \varepsilon)^2}} - \delta \right)}{\left( \delta^2 + 2\delta \left( \sqrt{1 - (1 - \varepsilon)^2} \sqrt{1 - \rho_y^2} + (1 - \varepsilon) \rho_y \right) + 1 \right)^{5/2}}
\]

\[
+ \frac{1}{\sqrt{1 - \rho_y^2}} \left( 1 - \varepsilon \right) \left( \sqrt{\frac{1 - \rho_y^2}{1 - (1 - \varepsilon)^2}} - 1 \right) \left( \delta - \sqrt{1 - (1 - \varepsilon)^2} \sqrt{1 - \rho_y^2} - (1 - \varepsilon) \rho_y \right) + 1 \right) \left( \delta^2 + 2\delta \left( \sqrt{1 - (1 - \varepsilon)^2} \sqrt{1 - \rho_y^2} + (1 - \varepsilon) \rho_y \right) + 1 \right)^{5/2}
\]

The denominator is positive \( \forall \rho_y \). Setting \( \rho_y = 0 \) in the numerator gives

\[
(1 - \varepsilon) \left( \sqrt{\frac{1 - \rho_y^2}{1 - (1 - \varepsilon)^2}} - 1 \right) \left( \delta - \sqrt{1 - (1 - \varepsilon)^2} + 1 \right) - 3\delta(1 - \varepsilon) \left( \sqrt{\frac{1}{1 - (1 - \varepsilon)^2}} - \delta \right) < 0
\]

which is negative for \( \varepsilon > 0 \) and small. Hence, \( \rho_y = 0 \) satisfies conditions (1)-(3) above.

Next, set \( \rho_y > 0 \). Set \( \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} = 0 \) and evaluate it at \( \varepsilon \to 0 \):

\[
- \frac{\delta \left( \delta^2 (2\rho_y^2 - 1) + \delta \rho_y (\rho_y^2 + 2) + \rho_y^2 + 2 \right) + \rho_y}{\sqrt{1 - \rho_y^2}} = 0
\]

which cannot hold (this expression is strictly negative \( \forall \rho_y \in (0, 1 - \varepsilon) \) and \( \delta \in [0, 1] \)). By continuity of \( \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} \) in \( \varepsilon \), this also holds for \( \varepsilon > 0 \) but small. Hence, condition (3) is violated, there cannot be an interior critical point. As a result, the only critical point of \( A(1 - \varepsilon, \cdot) \) is \( \rho_y^* = 0 \), which is the solution to the maximization problem of Step 2.

Therefore, Steps 1 and 2 show that \( \rho_x = 1 - \varepsilon, \rho_y = 0 \) solves \( \max_{\rho_x, \rho_y \in [0, 1 - \varepsilon]} A(\rho_x, \rho_y) \) in Case 1.

(Case 2: When restricting attention to \( \rho_x, \rho_y \in [-1 + \varepsilon, 0] \), an analogous argument shows that \( A(\cdot, \cdot) \) is maximized at \( \rho_x^* = -1 + \varepsilon, \rho_y = 0 \).

The proof for \( \left| \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} \right| \), where \( \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} \) is given by (74), is similar. I denote \( B(\rho_x, \rho_y) \equiv \frac{\partial A(1 - \varepsilon, \rho_y)}{\partial \rho_y} \).

Step 1: For all \( \rho_y \), \( B(\cdot, \rho_y) \) is maximized either at \( \rho_x^* = 1 - \varepsilon \) (Case 1) or at \( \rho_x^* = -1 + \varepsilon \) (Case 2). First, consider Case 1 and focus on \( \rho_x \geq \rho_y \), implying that \( B(\rho_x, \rho_y) \leq 0 \) (by Step 1 in Proposition 7, (i)), where \( B(\rho_x, \rho_y) \) is to be minimized in order for \( B(\rho_x, \rho_y) \) to be maximized. The constraint \( \rho_x \geq \rho_y \) will be verified below. Suppose that the minimum is achieved at an interior point \( \rho_x^*, \rho_y^* \in \)
Finally, there is no solution where \( B(x, \rho) \) is unbounded, which rules out an optimum where both \( \rho^*_x, \rho^*_y \) are interior. Regarding solutions where both \( \rho^*_x, \rho^*_y \) are at a corner, it is easy to verify that \( B(1, \rho) \) is unbounded for \( \rho \geq 0 \). For instance, in the denominator of (74) set \( \rho_x, \rho_y \), and \( \rho^*_x, \rho^*_y \) are interior since for every \( \rho_x \in (0, 1 - \epsilon), B(1, \rho) \) is unbounded for \( \rho \geq 0 \). Towards a contradiction, set \( \rho^*_x = 1 - \epsilon \) and \( \rho^*_y \in (0, 1 - \epsilon) \). Then, \( \exists \epsilon \) s.t. for \( \epsilon < \epsilon, (78) \) is violated. This rules out an optimum where both \( \rho^*_x, \rho^*_y \) are interior. Regarding solutions where both \( \rho^*_x, \rho^*_y \) are at a corner, it is easy to verify that \( B(1, \rho) \) is unbounded for \( \rho \geq 0 \). For instance, in the denominator of (74) set \( \rho_x, \rho_y \), and \( \rho^*_x, \rho^*_y \) are interior since for every \( \rho_x \in (0, 1 - \epsilon), B(1, \rho) \) is unbounded for \( \rho \geq 0 \). Thus, \( B(\rho_x, \rho_y) \) is unbounded for \( \rho \geq 0 \). Note that, since \( \rho^*_y \) is at the boundary, it follows that \( \rho_x \geq \rho_y \), implying that \( B(\rho_x, \rho_y) \leq 0 \), verifying the imposed constraint.

(For Case 2: An analogous argument establishes that for the case of \( \rho_x < \rho_y \) (under which \( \rho^*_x = 1 - \epsilon, \rho^*_y = 0 \)), \( B(\cdot, \cdot) \) is maximized at \( (\rho^*_x, \rho^*_y) = (1 - \epsilon, 0) \).

Step 2: \( B(1 - \epsilon, \cdot) \) is minimized at \( \rho^*_y = 0 \) (Case 1). (Case 2: \( B(-1 + \epsilon, \cdot) \) is maximized at \( \rho^*_y = 0 \).

Since \( B \) is a continuous function over a compact support, there exists a solution to the problem \( \min_{\rho_y \in [0, 1 - \epsilon]} B(1 - \epsilon, \rho_y) \) and it is picked by the Kuhn-Tucker conditions. Since Step 1 rules out that \( \rho^*_y = 1 - \epsilon, \) the following Kuhn-Tucker conditions must hold at an optimum:

\[
\begin{align*}
(1) & \quad \frac{\partial B(1 - \epsilon, \rho_y)}{\partial \rho_y} \geq 0 \\
(2) & \quad \rho_y \geq 0 \\
(3) & \quad \rho_y \frac{\partial B(1 - \epsilon, \rho_y)}{\partial \rho_y} = 0
\end{align*}
\]

Note that

\[
\frac{\partial B(1 - \epsilon, \rho_y)}{\partial \rho_y} = \frac{1}{\rho_y} \left( 1 + \frac{1 - \rho_y^2}{\sqrt{1 - (1 - \epsilon)^2} \sqrt{1 - \rho_y^2}} \right) \left( \delta \sqrt{1 - (1 - \epsilon)^2} \sqrt{1 - \rho_y^2} + (1 - \epsilon) \rho_y + 1 \right) \left( \delta^2 + 2 \delta \sqrt{1 - (1 - \epsilon)^2} \sqrt{1 - \rho_y^2} + (1 - \epsilon) \rho_y + 1 \right)^{\delta^2}
\]

\[
+ \frac{\delta}{\sqrt{1 - \rho_y^2}} \left( 1 - \epsilon \right) \sqrt{\frac{1 - \rho_y^2}{1 - (1 - \epsilon)^2} \sqrt{1 - \rho_y^2} - \rho_y} \left( \delta^2 + 2 \delta \sqrt{1 - (1 - \epsilon)^2} \sqrt{1 - \rho_y^2} + (1 - \epsilon) \rho_y + 1 \right)^{\delta^2}
\]

The denominator is positive. Setting \( \rho_y = 0 \) in the numerator gives

\[
\left( \delta \sqrt{1 - (1 - \epsilon)^2} + 1 \right) \left( \delta^2 + 2 \delta \sqrt{1 - (1 - \epsilon)^2} + 1 \right) = (1 - \epsilon) \left( \delta \sqrt{1 - (1 - \epsilon)^2} + 2 - \delta^2 \right) > 0
\]

which is positive for \( \epsilon > 0 \) and small. Hence, \( \rho_y = 0 \) satisfies conditions (1)-(3) above.

65 Note that the RHS of (78) could even be positive here, in which case the bound is irrelevant since (78) is automatically satisfied due to \( B(\rho_x, \rho_y) \leq 0 \) for any \( \rho_x \geq \rho_y \) that are consistently chosen in numerator and denominator.
Next, set \( \frac{\partial B(1-\varepsilon, \rho_y)}{\partial \rho_y} = 0 \) and evaluate it at \( \varepsilon \to 0 \), which gives
\[
\rho_y (\delta \rho_y + 1) \left( \delta^2 + 2 \delta \rho_y + 1 \right) + \delta \left( \delta^2 \left( -\sqrt{1 - \rho_y^2} \right) + \delta \rho_y \sqrt{1 - \rho_y^2} + 2 \sqrt{1 - \rho_y^2} \right) = 0
\]
which cannot hold (the expression is strictly positive \( \forall \rho_y \in (0, 1 - \varepsilon) \) and \( \delta \in [0, 1] \)). By continuity of \( \frac{\partial B(1-\varepsilon, \rho_y)}{\partial \rho_y} \) in \( \varepsilon \), this also holds for \( \varepsilon > 0 \) but small. Hence, condition (3) is violated, there cannot be an interior critical point. As a result, the only critical point of \( B(1-\varepsilon, \cdot) \) is \( \rho_y^* = 0 \), which is the solution to the minimization problem of Step 2.

Therefore, Steps 1 and 2 show that \( \rho_x^* = 1 - \varepsilon, \rho_y^* = 0 \) solves \( \min_{\rho_x, \rho_y \in [0, 1-\varepsilon]} B(\rho_x, \rho_y) \) in Case 1. (Case 2: When restricting \( \rho_x, \rho_y \in [-1 + \varepsilon, 0] \), an analogous argument shows that \( B(\rho_x, \rho_y) > 0 \) on this entire support, where \( B(\cdot, \cdot) \) is maximized at \( \rho_y^* = 1 - \varepsilon, \rho_y^* = 0 \).)

**Proof of Proposition 8.**

(i) **Skewness:** Recall the skewness of the wage distribution (71). Denote \( \text{Skew}(\delta) \equiv E \left[ \left( \frac{w - E(w)}{\sqrt{\text{Var}(w)}} \right)^3 \right] \).

It follows from (71) (after simplifying) that
\[
\frac{\partial \text{Skew}(\delta)}{\partial \delta} = \frac{-6 \sqrt{2} \delta (1 - \delta^2)(1 - \rho_x^2)(1 - \rho_y^2)}{(1 + \delta^2 + 2 \delta \rho_x \rho_y)(1 + \delta (\delta + 2 \rho_y \rho_y + 2 \sqrt{1 - \rho_x^2} \sqrt{1 - \rho_y^2}))^{3/2}} \leq 0
\]
for \( \delta \in [0, 1] \), proving the claim that the skewness is increasing as \( \delta \) decreases.

Moreover, the skewness is positive for all \( \delta \in [0, 1] \). To see this, evaluate (71) at \( \delta = 1 \):
\[
\text{Skew}(1) = \left( \sqrt{(1 + \rho_y)(1 + \rho_x)} + \sqrt{(1 - \rho_y)(1 - \rho_x)} \right) \frac{2 + 2 \rho_y \rho_y - \sqrt{(1 - \rho_y^2)(1 - \rho_x^2)}}{(1 + \rho_x \rho_y)^{3/2}}.
\]

It follows that \( \text{Skew}(0) \geq \text{Skew}(1) > 0 \) for all \( \rho_x, \rho_y \in (-1, 0] \) and \( \rho_x, \rho_y \in [0, 1) \).

(ii) **Variance:** Recall the variance of the wage distribution (70). It follows that
\[
\frac{\partial \text{Var}(w)}{\partial \alpha} = \alpha (1 + \delta \rho_x \rho_y)
\]
\[
\frac{\partial \text{Var}(w)}{\partial \beta} = \alpha (\delta + \rho_x \rho_y)
\]
which are both positive on \( \rho_x, \rho_y \in (-1, 0] \) and on \( \rho_x, \rho_y \in [0, 1) \). It follows that a decrease in \( \delta \) has ambiguous effects on the variance since it can stem from an increase in \( \alpha \) or a decrease in \( \beta \). However, a sufficient condition for TBTC to increase the variance is that it is triggered by an increase in \( \alpha \).
(ii) Wage Curvature. Wages are convex in $x_C$ and $x_M$ since

$$\frac{\partial^2 w(x_C, x_M)}{\partial x_C^2} = \alpha J_{11} > 0$$
$$\frac{\partial^2 w(x_C, x_M)}{\partial x_M^2} = \beta J_{22} > 0.$$  

Consider task-biased TC ($\delta$ decreases), triggered by an increase in $\alpha$ (one could additionally assume that $\beta$ decreases). For $|\rho_x| \leq |\rho_y|$, $J_{11} \leq 1$ and $\frac{\partial J_{11}}{\partial \delta} \leq 0$ as well as $J_{22} \leq 1$ and $\frac{\partial J_{22}}{\partial \delta} \geq 0$ (see Proposition 7i), where $J_{11}, J_{22}$ are defined in (72). Recall that $\delta = \frac{\beta}{\alpha}$. It follows that for $|\rho_x| \leq |\rho_y|$,  

$$\frac{\partial}{\partial \alpha} \left( \frac{\partial^2 w(x_C, x_M)}{\partial x_C^2} \right) = J_{11} + \alpha \frac{\partial J_{11}}{\partial \delta} \frac{\partial \delta}{\partial \alpha} > 0$$
$$\frac{\partial}{\partial \alpha} \left( \frac{\partial^2 w(x_C, x_M)}{\partial x_M^2} \right) = \beta \frac{\partial J_{22}}{\partial \delta} \frac{\partial \delta}{\partial \alpha} < 0.$$  

Hence, TBTC makes wages more convex in $x_C$ but less convex in $x_M$. Notice that additionally decreasing $\beta$ reinforces these effects. Finally, the result that the curvature changes are smallest when $|\rho_x - \rho_y| \to 0$ follows from $\frac{\partial J_{11}}{\partial \delta} = \frac{\partial J_{22}}{\partial \delta} = 0$ if $\rho_x = \rho_y$ (see Proposition 7iii).

(iii) Rank Switchings in the Wage Distribution: Let a worker with $(x_C, x_M) = (|x|, 0), |x| < \infty$ be a specialist in task $C$ and the worker $(x_C, x_M) = (0, |x| + \varepsilon), |x| < \infty, \varepsilon > 0$ and small, be a specialist in task $M$. Notice that their relative wage is given by:

$$\frac{w(|x|, 0)}{w(0, |x| + \varepsilon)} = \frac{|x| \left( 1 + \delta \sqrt{1 - \rho_y^2} \right)}{(|x| + \varepsilon) \left( \delta \left( \delta + \sqrt{1 - \rho_y^2} \right) \right)}$$

which is decreasing on $\delta \in (0, 1)$:

$$\frac{\partial}{\partial \delta} \frac{w(|x|, 0)}{w(0, |x| + \varepsilon)} = \frac{|x| \left( -\sqrt{1 - \rho_y^2} \delta^2 + 1 \right)}{(|x| + \varepsilon)^2 \left( \delta \left( \delta + \sqrt{1 - \rho_y^2} \right) \right)^2} < 0.$$  

Notice that $\lim_{\delta \to 0} \frac{w(|x|, 0)}{w(0, |x| + \varepsilon)} = \infty$ and thus $\frac{w(|x|, 0)}{w(0, |x| + \varepsilon)} > 1$ for $\delta$ close to zero. Also, $\frac{w(|x|, 0)}{w(0, |x| + \varepsilon)} < 1$ for $\delta = 1$. Hence, $\exists 0 < \delta^* < 1$ for which the wage ratio switches from $< 1$ to $> 1$, meaning that the wage ranking of these workers changes. Moreover, I now show that $\delta^*$ spans the whole support of $\delta$, implying that for any task-biased TC $\delta'$ with $\delta' < \delta \leq 1$, there exists some worker types $(x_C, x_M) = (|x|, 0)$ and $(x_C, x_M) = (0, |x| + \varepsilon)$ who switch ranks in the wage distribution. To see this,
set (79) equal to 1, which implicitly gives $\delta^*$. Then solve for $\delta^*$:

$$
\delta^* = -\frac{\varepsilon}{2(|x| + \varepsilon)} \sqrt{1 - \rho_y^2} + \frac{2(|x| + \varepsilon)}{\sqrt{1 - \rho_y^2}} + \frac{|x|}{2(|x| + \varepsilon)}
$$

Hence, as $|x| \to 0$, one obtains $\delta^* \to 0$ and as $|x| \to \infty$, one obtains $\delta^* \to 1$. Since $\delta^*$ is continuous in $|x|$, it follows that $\delta^*$ spans $(0, 1)$. Note that since the skill density is continuous, the same argument holds if I compare workers in a small neighborhood centered around $(x_C, x_M) = (|x|, 0)$ with those in a small neighborhood around $(x_C, x_M) = (0, |x| + \varepsilon)$. ■

### 8.3.2 Skill-Biased Technological Change

The wage function under augmented technology

$$
F(x_C, x_M, y_C, y_M) = \alpha x_CY_C + \beta x_My_M + \lambda x_C + \eta x_M + f_0
$$

is given by

$$
w(x_C, x_M) = \alpha \left( \frac{1}{2} \mathbf{x}' \tilde{J} \mathbf{x} + \theta' \mathbf{x} \right) + w_0 = \frac{1}{2} \alpha (\mathbf{x} - \mathbf{h})' \tilde{J} (\mathbf{x} - \mathbf{h}) + C
$$

where

$$
\tilde{J} = \begin{bmatrix} J_{11}(\rho_x, \rho_y, \delta) & J_{12}(\rho_x, \rho_y, \delta) \\ J_{21}(\rho_x, \rho_y, \delta) & J_{22}(\rho_x, \rho_y, \delta) \end{bmatrix}, \quad \theta = \begin{bmatrix} \lambda \\ \eta \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_C \\ x_M \end{bmatrix}, \quad \mathbf{h} = -\tilde{J}^{-1} \theta, \quad C = w_0 - \frac{1}{2} \alpha \theta' \tilde{J}^{-1} \theta
$$

**Proof of Proposition 9.** (i) Assignment (56) satisfies the first-order conditions of the firm under technology (80) and is independent of $\lambda, \kappa$. (ii) (81) satisfied integrability condition (5), i.e. is the unique wage schedule supporting the equilibrium assignment (56) (by Lemma 1). From (81), skill-biased TC parameters $\lambda$ and $\eta$ do not affect the curvature of the wage function.

(iii) Variance: Under wage function (81), the variance of the wage distribution is given by:

$$
Var(w) = \alpha^2 (1 - \rho_x^2) \left( (J_{12} + J_{11}\rho_x)^2 + \frac{J_{11}^2}{2} (1 - \rho_x^2) \right) + (1 - \rho_x^2)\lambda^2 + \alpha^2 \left( \frac{J_{11}}{2} \rho_x^2 + \rho_x J_{12} \right)^2 + (\rho_x \lambda + \eta)^2
$$

$$
= \alpha^2 (1 - \rho_x^2) \left( (J_{12} + J_{11}\rho_x)^2 + \frac{J_{11}^2}{2} (1 - \rho_x^2) \right) + \alpha^2 \left( \frac{J_{11}}{2} \rho_x^2 + \rho_x J_{12} \right)^2 + \lambda^2 \left( 1 + 2\rho_x \kappa + \kappa^2 \right)
$$

(82)

where $\kappa = \frac{\eta}{\lambda}$. It follows that the effect of a decrease in $\kappa$ (i.e. cognitive skill-biased TC) has an ambiguous effect on $Var(w)$. (82) is increasing in $\lambda$ if $\lambda + \rho_x \eta > 0$ and increasing in $\eta$ if $\eta + \rho_x \lambda > 0$:

$$
\frac{\partial Var(w)}{\partial \lambda} = 2(\lambda + \rho_x \eta)
$$

$$
\frac{\partial Var(w)}{\partial \eta} = 2(\eta + \rho_x \lambda)
$$
and a decrease in $\kappa$ can either be due to an increase in $\lambda$ or a decrease in $\eta$. A sufficient condition for the variance to be increasing in TBTC is that it is driven by an increase in $\lambda$ ($\kappa \downarrow$ due to $\lambda \uparrow$) and, moreover, $\lambda + \rho_x \eta > 0$ holds.

To derive (82), notice that

$$E(w|x_M) = \alpha \left( \frac{J_{11}}{2} E(x_C^2|x_M) + J_{12} x_M E(x_C|x_M) + \frac{\delta J_{22}}{2} x_M^2 \right) + \eta x_M + \lambda E(x_C|x_M)$$

$$= \alpha \left( \frac{J_{11}}{2} (\rho_x^2 x_M + (1 - \rho_x^2)) + \frac{\delta J_{22}}{2} x_M^2 \right) + \rho_x x_M (\alpha J_{12} x_M + \lambda) + \eta x_M$$

$$Var(E(w|x_M)) = \alpha^2 Var(x_M^2) \left( \frac{J_{11}}{2} \rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2} \right)^2 + Var(x_M) (\rho_x \lambda + \eta)^2$$

$$= \alpha^2 \left( \frac{J_{11}}{2} \rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2} \right)^2 + (\rho_x \lambda + \eta)^2$$

(83)

since $cov(x_M^2, x_M) = E(x_M^3) - E(x_M)E(x_M^2) = 0$. Moreover,

$$Var(w|x_M) = Var(x_C|x_M) (\alpha^2 J_{12} x_M^2 + \lambda^2) + Var(x_C^2|x_M) (\alpha^2 J_{11} x_M x_C + \lambda^2)$$

$$+ cov(x_C^2, x_C|x_M) (\alpha^2 J_{11} - \lambda^2)$$

where

$$Var(x_C|x_M) = E(x_C^4|x_M) - (E(x_C^2|x_M))^2 = 4 \rho_x^2 x_M^2 (1 - \rho_x^2) + 2 (1 - \rho_x^2)^2$$

$$cov(x_C^2, x_C|x_M) = E(x_C^3|x_M) - E(x_C^2|x_M) E(x_C|x_M) = 2 \rho_x x_M (1 - \rho_x^2).$$

Hence,

$$Var(w|x_M) = \alpha^2 (1 - \rho_x^2) (J_{12} + \rho_x J_{11})^2 + \frac{J_{11}^2}{2} (1 - \rho_x^2) (1 - \rho_x^2) \rho_x J_{11} x_M + \lambda^2$$

$$E(Var(w|x_M)) = \alpha^2 (1 - \rho_x^2) ((J_{12} + \rho_x J_{11})^2 + \frac{J_{11}^2}{2} (1 - \rho_x^2)) + (1 - \rho_x^2) \lambda^2.$$ (84)

(82) follows from adding (83) and (84), i.e. $Var(w) = E(Var(w|x_M)) + Var(E(w|x_M)).$

Skewness: Given wage function (81), the skewness of the wage distribution has a closed form

$$E \left[ \frac{(w - E(w))^3}{\sqrt{Var(w)}} \right] = \frac{E(w^3) - 3E(w)Var(w) - E(w)^3}{Var(w)^{3/2}} = \frac{8 \mu' \tilde{J} \Sigma_x \tilde{J} \Sigma_x \tilde{J} \Sigma_x \tilde{J} \mu}{(2 \mu' \tilde{J} \Sigma_x \tilde{J} \Sigma_x \tilde{J} \mu)^{3/2}}$$

(85)

where $\tilde{J}$ is the Hessian of the wage function, $\Sigma_x$ is the covariance matrix of workers’ skills, and

$$\mu = 0.5 \alpha \lambda \begin{bmatrix} 0.5 J_{11} & 0.5 J_{12} \\ 0.5 \delta J_{21} & 0.5 \delta J_{22} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \kappa \end{bmatrix}.$
Expression (85) is involved. Providing analytical sufficient conditions for monotone comparative statics is possible but they do not have a clean interpretation. They are available upon request.

8.4 Relaxed Sufficient Conditions for PAM

Section 2 provides a distribution-free sufficient condition for assortative matching, under which between-task complementarities are shut down. In contrast, this section makes assumptions on the skill and productivity distributions, under which the sufficient conditions for PAM/NAM can be relaxed, allowing for non-zero between-task complementarities. The first subsection deals with Gaussian distributions. The subsequent one deals with independent uniform distributions or arbitrary (but identical) skill and productivity distributions.

8.4.1 Gaussian Distributions

Assume bivariate Gaussian skill and productivity distributions and the following technology:

$$ F(x_C, x_M, y_C, y_M) = \gamma(x_C y_C + \alpha x_C y_M + \beta x_M y_C + \delta x_M y_M) \quad (86) $$

With non-zero between-task complementarities, the sufficient condition for PAM/NAM is stated as:

**Proposition 10 (Sufficient Condition for PAM in Gaussian-Quadratic Setting)** Suppose that $(x_C, x_M)$ and $(y_C, y_M)$ follow bivariate Gaussian distributions and the technology is given by (86). If

$$ D_{xy}^2 F(x, y) = \begin{bmatrix} F_{x_C y_C} & F_{x_C y_M} \\ F_{x_M y_C} & F_{x_M y_M} \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ \beta & \delta \end{bmatrix} $$

is a strictly diagonal dominant $P$-matrix ($P^-$-matrix) by row and column, then the equilibrium assignment satisfies PAM (NAM).

**Proof.** A matrix $M$ is strictly column diagonally dominant if $|m_{ii}| > \sum_{j \neq i} |m_{ij}|, i = 1, 2, ..., n$ and row diagonally dominant if $|m_{ii}| > \sum_{j \neq i} |m_{ji}|, i = 1, 2, ..., n$. Assume w.l.o.g. that $\delta \in [0, 1]$. Then, $D_{xy}^2 F$ is strictly diagonally dominant if $\delta > |\alpha|$ and $\delta > |\beta|$, which is assumed to hold. The proof will be given for PAM and standard Gaussian distributions. The proof for NAM is equivalent (just match up the marginal c.d.f.’s in a decreasing instead of increasing way). The extension to non-standard Gaussian variables is given in Section 1.1, Online Appendix.

Under (86), integrability condition (5), is given by:

$$ w_{x_C x_M} = w_{x_M x_C} $$

$$ \Leftrightarrow \quad J_{12} + \alpha J_{22} = \beta J_{11} + \delta J_{21} \quad (87) $$

where $J_{11} = \frac{\partial y_C}{\partial x_C}, J_{12} = \frac{\partial y_C}{\partial x_M}, J_{21} = \frac{\partial y_M}{\partial x_C}, J_{22} = \frac{\partial y_M}{\partial x_M}$ denote the elements of the matching function’s
Jacobian. Using (87), the equilibrium assignment can be computed as in Proposition 3. It is given by:

\[
J_{11} = \frac{4}{2Z\sqrt{1-\rho_x^2}}((1+\alpha \rho_y)\sqrt{1-\rho_z^2} + (\delta + \alpha \rho_x)\sqrt{1-\rho_y^2})
\]

\[
J_{12} = \frac{4}{2Z\sqrt{1-\rho_z^2}}((\beta + \delta \rho_y)\sqrt{1-\rho_x^2} - (\delta \rho_x + \alpha)\sqrt{1-\rho_y^2})
\]

\[
J_{21} = \frac{4}{2Z\sqrt{1-\rho_x^2}}((\alpha + \rho_y)\sqrt{1-\rho_z^2} - (\rho_x + \beta)\sqrt{1-\rho_y^2})
\]

\[
J_{22} = \frac{4}{2Z\sqrt{1-\rho_z^2}}((\delta + \beta \rho_y)\sqrt{1-\rho_x^2} + (1+\beta \rho_x)\sqrt{1-\rho_y^2})
\]

where

\[
Z = \sqrt{((1+\delta)\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1-\rho_z}}\right) + (\alpha + \beta)\left(\sqrt{\frac{1+\rho_z}{1-\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_z}}\right))^2 + ((1-\delta)\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_z}}\right) + (\alpha - \beta)\left(\sqrt{\frac{1+\rho_z}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1-\rho_z}}\right))^2}.
\]

PAM holds since \(\forall \rho_x, \rho_y,\)

\[
J_{11} > 0 \quad \text{if} \quad \delta > |\alpha| \\
J_{22} > 0 \quad \text{if} \quad \delta > |\beta|
\]

\[\text{Det}(J_\mu) = J_{11}J_{22} - J_{12}J_{21} > 0 \quad \text{if} \quad \delta > |\alpha| \quad \text{and} \quad \delta > |\beta|,
\]

where \(\text{Det}(J_\mu)\) reads

\[
J_{11}J_{22} - J_{12}J_{21} = \frac{16}{2Z\sqrt{1-\rho_x^2}}((1-\rho_z^2)((1+\alpha \rho_y)(\delta + \rho_y \beta) - (\rho_y + \alpha)(\rho_y \delta + \beta))
\]

\[
+ \sqrt{1-\rho_z^2}\sqrt{1-\rho_y^2}\left((1+\alpha \rho_y)(1+\rho_x \beta) + (\delta + \alpha \rho_x)(\rho_y \beta) + (\rho_y + \alpha)(\rho_x + \alpha) + (\rho_x + \beta)(\rho_y \delta + \beta)\right)
\]

\[
+ (1-\rho_y^2)((\delta + \alpha \rho_x)(1+\rho_x \beta) - (\rho_x + \beta)(\rho_y \delta + \alpha)]]
\]

where \(X_1\) and \(X_3\) are positive under diagonal dominance and where \(X_2\) can be expressed as:

\[
X_2 = 1 + \alpha^2 + \beta^2 + \delta^2 + 2\beta \rho_x + 2\delta(\beta + \rho_x)\rho_y + 2\alpha(\delta \rho_x + \rho_y + \beta \rho_x \rho_y)
\]

It remains to show that \(X_2\) is positive. Notice that \(X_2\) is linear in \(\rho_x\) and \(\rho_y\). Hence, the infimum of \(X_2\) must be in a corner. If \(X_2\) is positive in all corners, then \(\text{Det}(J_\mu) > 0\). To simplify this argument, I evaluate \(X_2\) at \(\rho_x = \pm 1\) and \(\rho_y = \pm 1\) (since \(X_2\) is continuous and is positive at the corners it is also positive arbitrarily close to the corners)

\[
X_2|_{\rho_x = \rho_y = 1} = ((1 + \delta) + (\alpha + \beta))^2 > 0
\]

\[
X_2|_{\rho_x = \rho_y = -1} = ((1 + \delta) - (\alpha + \beta))^2 > 0
\]

\[
X_2|_{\rho_x = 1, \rho_y = -1} = ((1 - \delta) - (\alpha - \beta))^2 > 0
\]

\[
X_2|_{\rho_x = -1, \rho_y = 1} = ((1 - \delta) - (\alpha - \beta))^2 > 0
\]

which proves the result. \(\blacksquare\)
8.4.2 Non-Gaussian Distributions

Specify the distributions to be uniform (Case (a)) or arbitrary (but identical) (Case (b)).

**Proposition 11** For (a) independent uniform skills $x \sim U([\underline{x}, \bar{x}]^N)$ and productivities $y \sim U([\underline{y}, \bar{y}]^N)$ or (b) for (arbitrary) $G=H$, if $D_{xy}^2 F(x, y)$ is a P-matrix everywhere and, moreover, symmetric positive definite along the equilibrium path, then there exists a unique equilibrium where the assignment satisfies PAM.

**Proof.** I prove this result in four steps:

**Step 1:** There exists a feasible PAM allocation. Consider (a). Denote by $H_{x_C}, H_{x_M}, G_{y_C}, G_{y_M}$ the marginal cdf’s of $x_C, x_M, y_C$ and $y_M$, respectively. Due to independence, the market clearing condition can be specified in line with PAM as

$$(1 - H_{x_C})(1 - H_{x_M}) = (1 - G_{y_C})(1 - G_{y_M}).$$

(92)

Because of PAM, I match up the marginals within each dimension (which is consistent with (92))

$$H_{x_C} = G_{y_C}$$
$$H_{x_M} = G_{y_M}$$

which gives the assignment functions:

$$y_C = \frac{\bar{y} - y}{\bar{x} - \underline{x}} x_C - \frac{\bar{y} - y}{\bar{x} - \underline{x}} + \frac{\underline{y}}{\underline{x}}$$
$$y_M = \frac{\bar{y} - y}{\bar{x} - \underline{x}} x_M - \frac{\bar{y} - y}{\bar{x} - \underline{x}} + \frac{\underline{y}}{\underline{x}}$$

(93)
(94)

Both (93) and (94) are in line with PAM since $\frac{\partial y_C}{\partial x_C} > 0$ and $\frac{\partial y_M}{\partial x_M} > 0$ as well as $\frac{\partial^2 y_C}{\partial x_C \partial x_M} - \frac{\partial^2 y_M}{\partial x_M \partial x_C} > 0$.

Consider (b). A feasible PAM allocation is given by $y_C = x_C$ and $y_M = x_M$.

**Step 2:** The PAM allocation from Step 1 satisfies the firms’ necessary second-order conditions for optimality under the P-matrix property of $D_{xy}^2 F$. To show this, recall from the proof of Proposition 2 that the Hessian of the firm’s problem is given by:

$$H^* = D_{xx}^2 F(x, y^*) - D_{xx}^2 w(x) = -D_{xy}^2 F(x, y^*) D_{xy} y^*$$

(95)

In both PAM allocations from Step 1, $D_{xy} y^*$ is a diagonal matrix. Since $D_{xy}^2 F$ is a P-matrix, the matrix product $D_{xy}^2 F(x, y^*) D_{xy} y^*$ is positive-definite and, hence, the Hessian in (95) is negative-definite.

**Step 3:** The PAM allocation from Step 1 satisfies the integrability condition (5). To see this, first
focus on (a). Since $D_{xy}^* is diagonal, (5) collapses to
\[ F_{xyc} \frac{\partial y^*_M}{\partial x} = F_{xyc} \frac{\partial y^*_C}{\partial x} \]  
which must hold along the equilibrium path. Using (93) and (94), condition (96) simplifies to
\[ F_{xyc} \frac{\gamma - y}{\overline{\gamma} - \bar{y}} = F_{xyc} \frac{\gamma - y}{\overline{\gamma} - \bar{y}} \]
which holds under the assumption of symmetric positive-definiteness of $D_{xy}^2 F (F_{xyc} = F_{xyc})$ along the equilibrium path.
Consider (b). Under the assignment $y_C = x_C$ and $y_M = x_M$, (5) collapses to
\[ F_{xyc} = F_{xyc} \]
which again holds under the assumption of positive-definiteness of $D_{xy}^2 F$. Hence, for (a) and (b), there exists a unique wage schedule, that induces firms to optimally choose the PAM allocation from Step 1.

**Step 4.** The constructed equilibrium is unique. Since $D_{xy}^2 F$ is a P-matrix everywhere, the assumptions from Proposition 1i. are satisfied, i.e. the equilibrium is unique. ■

### 8.5 Quantitative Analysis

#### 8.5.1 The Data

I drop observations with missing wage data or those with hourly wages smaller than one dollar or larger than hundred dollar. Also, I drop those workers who have worked zero/negative hours or weeks since last interview and also those who have zero sample weight. In both cohorts, I focus on the core sample of the NLSY, meaning that I drop the military samples from the NLSY79 and exclude oversamples of special demographic/racial groups from both the NLSY79 and NLSY97. My preferred specification focuses on 27-29 year old workers in both cohorts, comparing years 1990/91 with 2009/10, which ensures that age distributions are similar across cohorts. Tables 5-7 display the age distribution by year for both cohorts, with my preferred sample being highlighted in red.

**Table 5: NLSY79: Age by Year**

<table>
<thead>
<tr>
<th>Year</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>107</td>
<td>513</td>
<td>504</td>
<td>502</td>
<td>473</td>
<td>453</td>
<td>337</td>
<td>350</td>
<td>213</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3,452</td>
</tr>
<tr>
<td>1991</td>
<td>0</td>
<td>148</td>
<td>484</td>
<td>517</td>
<td>507</td>
<td>452</td>
<td>404</td>
<td>340</td>
<td>186</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3,390</td>
</tr>
<tr>
<td>1992</td>
<td>0</td>
<td>0</td>
<td>143</td>
<td>489</td>
<td>508</td>
<td>496</td>
<td>477</td>
<td>414</td>
<td>350</td>
<td>346</td>
<td>199</td>
<td>0</td>
<td>0</td>
<td>3,422</td>
</tr>
<tr>
<td>1993</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>165</td>
<td>499</td>
<td>506</td>
<td>517</td>
<td>483</td>
<td>404</td>
<td>360</td>
<td>352</td>
<td>167</td>
<td>0</td>
<td>3,453</td>
</tr>
<tr>
<td>1994</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>102</td>
<td>381</td>
<td>392</td>
<td>411</td>
<td>407</td>
<td>350</td>
<td>289</td>
<td>277</td>
<td>154</td>
<td>2,763</td>
</tr>
<tr>
<td>1995</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>125</td>
<td>313</td>
<td>321</td>
<td>347</td>
<td>329</td>
<td>280</td>
<td>247</td>
<td>1,955</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>171</td>
<td>375</td>
<td>387</td>
<td>397</td>
<td>363</td>
<td>1,693</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>130</td>
<td>440</td>
<td>459</td>
<td>1,029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>141</td>
<td>141</td>
<td>141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>107</td>
<td>661</td>
<td>1,131</td>
<td>1,673</td>
<td>2,089</td>
<td>2,288</td>
<td>2,252</td>
<td>2,323</td>
<td>2,206</td>
<td>1,964</td>
<td>1,677</td>
<td>1,561</td>
<td>1,364</td>
<td>21,296</td>
</tr>
</tbody>
</table>
Table 6: NLSY79: Age by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Age 38</th>
<th>Age 39</th>
<th>Age 40</th>
<th>Age 41</th>
<th>Age 42</th>
<th>Age 43</th>
<th>Age 44</th>
<th>Age 45</th>
<th>Age 46</th>
<th>Age 47</th>
<th>Age 48</th>
<th>Age 49</th>
<th>Age 50</th>
<th>Age 51</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>260</td>
<td>102</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>362</td>
</tr>
<tr>
<td>1998</td>
<td>318</td>
<td>263</td>
<td>255</td>
<td>122</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>988</td>
</tr>
<tr>
<td>2000</td>
<td>444</td>
<td>449</td>
<td>395</td>
<td>325</td>
<td>195</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2,152</td>
</tr>
<tr>
<td>2002</td>
<td>445</td>
<td>454</td>
<td>486</td>
<td>439</td>
<td>393</td>
<td>320</td>
<td>321</td>
<td>160</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3,018</td>
</tr>
<tr>
<td>2004</td>
<td>0</td>
<td>58</td>
<td>358</td>
<td>416</td>
<td>427</td>
<td>354</td>
<td>294</td>
<td>292</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2,866</td>
</tr>
<tr>
<td>2006</td>
<td>0</td>
<td>0</td>
<td>193</td>
<td>393</td>
<td>432</td>
<td>412</td>
<td>386</td>
<td>337</td>
<td>293</td>
<td>280</td>
<td>76</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2,802</td>
</tr>
<tr>
<td>2008</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>232</td>
<td>416</td>
<td>430</td>
<td>428</td>
<td>395</td>
<td>352</td>
<td>297</td>
<td>284</td>
<td>71</td>
<td>0</td>
<td>0</td>
<td>2,905</td>
</tr>
<tr>
<td></td>
<td>1,467</td>
<td>1,346</td>
<td>1,504</td>
<td>1,495</td>
<td>1,552</td>
<td>1,606</td>
<td>1,503</td>
<td>1,270</td>
<td>1,057</td>
<td>930</td>
<td>635</td>
<td>373</td>
<td>284</td>
<td>71</td>
<td>15,093</td>
</tr>
</tbody>
</table>

Table 7: NLSY97: Age by Year

| Year | Age 16 | Age 17 | Age 18 | Age 19 | Age 20 | Age 21 | Age 22 | Age 23 | Age 24 | Age 25 | Age 26 | Age 27 | Age 28 | Age 29 | Age 30 | Age 31 | Age 32 | Total |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|
| 2000 | 2      | 19     | 485    | 792    | 746    | 62     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0     | 2,106 |
| 2001 | 1      | 25     | 504    | 790    | 854    | 709    | 57     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0     | 2,940 |
| 2002 | 0      | 0      | 517    | 788    | 849    | 843    | 723    | 50     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0     | 3,770 |
| 2003 | 0      | 0      | 64     | 742    | 852    | 810    | 829    | 609    | 34     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0     | 4,000 |
| 2004 | 0      | 0      | 0      | 40     | 765    | 819    | 811    | 803    | 679    | 45     | 0      | 0      | 0      | 0      | 0      | 0      | 0     | 3,962 |
| 2005 | 0      | 0      | 0      | 0      | 50     | 771    | 803    | 780    | 790    | 650    | 34     | 0      | 0      | 0      | 0      | 0      | 0     | 3,878 |
| 2006 | 0      | 0      | 0      | 0      | 0      | 79     | 797    | 854    | 817    | 840    | 651    | 23     | 0      | 0      | 0      | 0      | 0     | 4,061 |
| 2007 | 0      | 0      | 0      | 0      | 0      | 106    | 791    | 848    | 815    | 807    | 625    | 11     | 0      | 0      | 0      | 0      | 0     | 4,003 |
| 2008 | 0      | 0      | 0      | 0      | 0      | 0      | 102    | 746    | 832    | 803    | 822    | 633    | 19     | 0      | 0      | 0      | 0     | 3,957 |
| 2009 | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 145    | 776    | 800    | 771    | 815    | 593    | 7      | 0      | 0      | 0     | 3,907 |
| 2010 | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 74     | 721    | 754    | 800    | 767    | 632    | 13     | 0      | 0     | 3,761 |
| 2011 | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 145    | 756    | 770    | 756    | 759    | 552    | 7      | 0     | 3,745 |
|      | 3      | 44     | 1,570  | 3,152  | 4,116  | 4,093  | 4,126  | 4,049  | 4,059  | 4,032  | 3,961  | 3,751  | 3,029  | 2,135  | 1,398  | 565   | 7     | 44,090 |

8.5.2 Construction of Skill and Skill Demand Distributions

**Skill Demand Distribution:** To construct the bivariate skill demand distribution (i.e. the distribution of $y$'s) is concerned, I use a dataset based on O*NET, constructed by Sanders (2012), who classifies occupational skill requirements into two categories, manual and cognitive. He then aggregates this information, using Principal Component Analysis, to get two task scores for each occupation (i.e. $y_C$ and $y_M$). Using this procedure, manual and cognitive task scores are obtained for over 400 occupations. The scores have an ordinal interpretation and allow to rank occupations according to their manual and cognitive skill requirements. I interpret these occupational task scores as the $p_{y_C}, y_M$-bundle from my model. I drop the observations whose $p_{y_C}, y_M$-bundles are missing. Table 8 provides some examples of occupations and their manual and cognitive skill requirements, starting with low-skilled jobs (requiring low amounts of both skills), followed by manual jobs, generalist jobs (requiring a fair amount of both skills) and purely cognitive jobs.  

**Skill Supply Distribution:** To construct the skill distribution (i.e. distribution of the $x$'s), I proceed as follows. I use the responses from the individuals in the NLSY on college, apprenticeships and government training degrees as well as occupational training (see the attached document on ‘Data Description’ in the Supplementary Data Appendix for the exact variables I used). I then make use of the fact that (a) college, apprenticeships and government training programs as well as (b) occupational training qualify workers for particular occupations. Regarding (a), I match the data on college degrees, apprenticeships and government programs to occupations. For instance, if an individual holds a college degree, apprenticeship, or government program, I assign them to the occupational task score associated with that occupation. Apprenticeships and government programs map into occupations in a straightforward way. This is more complicated with college degrees. While the college degrees reported in the NLSY79 are very detailed, this is not the case for those reported in the NLSY97. To ensure comparability, I aggregate the NLSY79 college degrees such that they resemble the NLSY97 college degrees and construct a crosswalk that maps from the NLSY97 college degrees to occupations.
college degree in ‘economics’, I assume he is qualified for the occupation ‘economist’. For the NLSY97, I use 2002 Census 4-digit Occupation Codes throughout. In the NLSY79, occupational coding has changed over time: I use 1970 3-digit Census Occupation Codes until year 2000, 2000 Census 3-digit Occupation Codes for year 2002, and 2002 Census 4-digit Occupation Codes since 2004.

Regarding (b), I proceed as follows. In the NLSY97, I first take training spells into account that led to a certificate or license in a certain profession. I created a crosswalk between NLSY97-license/certificate codes and occupational codes to map certificates to occupations. Second, if training spells did not lead to a certificate or license I still take them into account if respondents indicate that they have completed the training and if the employer helped to pay for the training (indicating that the training is related to the performed occupation). Likewise in the NLSY79, I take all past formal training spells into account that can be linked to a specific occupation. Depending on the period, there are different variables in the data set that indicate occupation-specific training: First, in 1979-1980, respondents reported which training certificates they obtained, where these certificates can be linked to occupations. Second, during 1982-1984, I take those training spells at the date of last interview into account that were encouraged by the employer (and thus related to the performed occupation). Third, during 1989-2010, I use the training data if the respondent indicated that he/she used the acquired skills in his/her occupation. Fourth, during 1991-2008, the NLSY79 contains information on who paid for the training, where I consider the training if it was sponsored by the employer.

Once I matched data on education and training to occupations, I use the constructed \((y_C, y_M)\)-bundles from the O*NET data to learn about the skills required for these occupations. I assume that if a worker is trained in a particular occupation (through college, an apprenticeship or other training), then he also holds the skills required for that occupation. For instance, if a worker holds a degree in economics and, based on the O*NET data, the occupation economist has skill requirements \((y_C = 1.34, y_M = -1.58)\), then the worker holds the skill bundle \((x_C = 1.34, x_M = -1.58)\). Importantly, in this example of a college degree, the worker’s skills are independent of whether he has actually ever

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Cognitive score ((y_C))</th>
<th>Manual score ((y_M))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payroll Clerks</td>
<td>.3801</td>
<td>.0292</td>
</tr>
<tr>
<td>Telephone Operators</td>
<td>.2994</td>
<td>.1383</td>
</tr>
<tr>
<td>File Clerks</td>
<td>.3190</td>
<td>.3099</td>
</tr>
<tr>
<td>Legal Secretaries</td>
<td>.3796</td>
<td>.0731</td>
</tr>
<tr>
<td>Brickmasons, Blockmasons, Stonemasons</td>
<td>.1706</td>
<td>.8360</td>
</tr>
<tr>
<td>Helpers–Pipelayers, Plumbers, Pipefitters, Steamfitters</td>
<td>.1759</td>
<td>.6792</td>
</tr>
<tr>
<td>Helpers–Carpenters</td>
<td>.1984</td>
<td>.7187</td>
</tr>
<tr>
<td>Dancers</td>
<td>.3374</td>
<td>1</td>
</tr>
<tr>
<td>Radiologic Technicians</td>
<td>.4280</td>
<td>.6469</td>
</tr>
<tr>
<td>Machinists</td>
<td>.4303</td>
<td>.7152</td>
</tr>
<tr>
<td>Physical Therapist Assistants</td>
<td>.4758</td>
<td>.5494</td>
</tr>
<tr>
<td>Electricians</td>
<td>.4879</td>
<td>.8146</td>
</tr>
<tr>
<td>Economists</td>
<td>.6149</td>
<td>.0340</td>
</tr>
<tr>
<td>Public Relations and Fundraising Managers</td>
<td>.6199</td>
<td>.0587</td>
</tr>
<tr>
<td>Judges, Magistrate Judges, and Magistrates</td>
<td>.6752</td>
<td>.0517</td>
</tr>
<tr>
<td>Physicists</td>
<td>1</td>
<td>.1113</td>
</tr>
</tbody>
</table>

Table 8: Examples of Occupations’ Manual and Cognitive Skill Requirements
worked in the occupation ‘economist’ and thus do not depend in his match history.

If a worker has had several educational experiences, I proceed as follows: I use the worker’s formal occupational training experiences up to the previous year as well as his educational history, giving a vector of manual and cognitive skills for every agent in every year of the survey. To obtain a single cognitive and a single manual skill from these skill vectors, I take the mean skills from each vector. This is done for all workers in the sample, which gives the entire skill distribution. Note that, as a robustness check to this construction of the skill distribution, I alternatively took the maximum of each skill vector (instead of the average), which did not have a big impact on the constructed skill distribution, nor on the estimates of the model. The robustness checks are available upon request.

In the NSLY, there are observations to whom I cannot assign any skills because there is no information on training or secondary education. These agents might be educated and the information is simply missing. Or they are low-skilled and do not have any degrees.

If the information on training or higher education is missing but if the worker has received some education (at least a high school degree), I assign him cognitive and manual skills through a random draw from the skill distribution of similarly educated people. For instance, if the skill data of a worker with high school degree is missing, he gets a random draw of cognitive and manual skills from the skill distribution of other high school graduates of similar age for whom I was able to construct their skills.68 This way, the skill sets of those with missing skill data reflect the skill bundles of those workers who have a similar level of formal education. In contrast, in the case of high-school dropouts with missing skill data, it is most likely that these workers have not acquired any skills. This is supported by the fact that there are very few observations of high-school dropouts with acquired skills in both cohorts. Therefore, I drop high-school dropouts with missing skills from my sample. Notice that the problem of missing skills can only affect those workers with high-school degree or less.69

Finally, to reduce the discreteness of the skill distributions and to better align them with the continuous distributions of the model, I add random noise to each skill observation, which is in size 5% of the variance of the corresponding skills. I proceed similarly with the skill demand distribution.

It is important to note that even though I make use of the occupational training for the construction of both skill supply and skill demand distributions, there is no mechanically high correlation between $x$ and $y$ for the following reasons: First, and crucially, I heavily rely on workers’ responses from the NLSY on university, apprenticeship and government program degrees. The O*NET scores assigned to the associated occupations (which then become the workers’ skill bundles) are independent of whether the workers have actually ever worked in these occupations. This implies that the $x$-bundles can be completely unrelated to the $y$-bundles of the jobs these individuals work in. I only complement these educational experiences with occupational training when determining skills. Second, I only use information on workers’ occupations if they have received formal training in that occupation, which was completed (instead of using information on any past occupation). Finally, I

---

68I might over-estimate the skills of those workers with missing data if their skills are missing because they have not acquired any skills after high school. For this reason and for robustness, I also drew the missing skills from the lower part of the skill distribution of other high school graduates, which leads to similar estimation results.

69Whenever a worker reports a degree $>$ than high-school, I can link this degree to an occupation and impute his skill bundle.
do not rely on occupational training in the current period but only on past training experiences.

Table 9, left panel, provides summary statistics of the untransformed skill and productivity distributions of 27-29 year old individuals across cohorts. I indicate untransformed variables by hat. The table shows that jobs in the US require on average a higher level of cognitive than manual skills in both periods. Similarly, workers hold on average more cognitive than manual skills. Over time, both skill supply and skill demand distributions are relatively stable.

<table>
<thead>
<tr>
<th></th>
<th>( \hat{x}_C )</th>
<th>( \hat{x}_M )</th>
<th>( \hat{y}_C )</th>
<th>( \hat{y}_M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean 1990/91</td>
<td>0.3596</td>
<td>-0.2912</td>
<td>0.0135</td>
<td>-0.1189</td>
</tr>
<tr>
<td>standard deviation 1990/91</td>
<td>0.7423</td>
<td>0.9923</td>
<td>0.8490</td>
<td>1.024</td>
</tr>
<tr>
<td>Min 1990/91</td>
<td>-2.0595</td>
<td>-1.7004</td>
<td>-2.0622</td>
<td>-1.6949</td>
</tr>
<tr>
<td>Max 1990/91</td>
<td>2.1649</td>
<td>2.1855</td>
<td>2.0925</td>
<td>2.1895</td>
</tr>
<tr>
<td>mean 2009/10</td>
<td>0.5667</td>
<td>-0.6601</td>
<td>0.0468</td>
<td>-0.2509</td>
</tr>
<tr>
<td>standard deviation 2009/10</td>
<td>0.7556</td>
<td>0.8358</td>
<td>0.9280</td>
<td>0.9656</td>
</tr>
<tr>
<td>Min 2009/10</td>
<td>-2.3019</td>
<td>-1.8116</td>
<td>-2.5200</td>
<td>-1.6597</td>
</tr>
<tr>
<td>Max 2009/10</td>
<td>1.9160</td>
<td>2.1838</td>
<td>3.0504</td>
<td>2.1351</td>
</tr>
</tbody>
</table>

\( \rho_y \) -0.4877 -0.4877
\( \rho_x \) -0.4031 -0.5221
\( \rho_y \) -0.4226 -0.4501
\( \rho_x \) -0.3756 -0.4916

Table 9: Summary Statistics of Skills and Skill Demand Distributions (Untransformed Data), left table. Skill Supply and Demand Correlations in Untransformed Data \((\rho_x, \rho_y)\) and Transformed Data \((\rho_x, \rho_y)\), right table.

The skill correlation indicates how specialized the workforce is, with a more negative correlation pointing to more specialized workers who are either good in the cognitive or in the manual task. Similarly, a strongly negative productivity correlation indicates that most jobs require either manual dexterity or cognitive ability and few jobs require a balanced skill set. Table 9, right panel, shows that in the US at the beginning of the 1990s most jobs are specialized, indicated by a negative productivity correlation of around -5. Notice that the demand for skills is more (less) specialized than skill supply in 1990/91 (2009/10). The correlations of the transformed data (i.e. after transforming the skill and skill demand distributions to Gaussian copulas; see Section 1.2 in the Online Appendix), are given in the lower part of the table and show that this transformation did not change the dependency structure of the data in important ways. Using Gaussian Copulas is a compromise between the empirical application and my theory, which allows for a closed form of the equilibrium allocation and wage function for Gaussian distributions. Gaussian copulas provide considerably more flexibility than Gaussian distributions by allowing for arbitrary marginal distributions (while maintaining a parametric assumption about the dependence between the marginals).

8.5.3 Maximum Likelihood Estimation

The system of equations used for the ML-estimation is given by

\[
\begin{align*}
\hat{w} &= \frac{1}{2} J_{11} \hat{x}_C^2 + \alpha J_{12} \hat{x}_C \hat{x}_M + \frac{1}{2} \beta J_{22} \hat{x}_M^2 + \lambda \hat{x}_C + \eta \hat{x}_M + w_0 + \varepsilon_w \\
\hat{y}_C &= J_{11} \hat{x}_C + J_{12} \hat{x}_M + \varepsilon_C \\
\hat{y}_M &= J_{21} \hat{x}_C + J_{22} \hat{x}_M + \varepsilon_M
\end{align*}
\]
where I assume measurement errors \( \varepsilon_w, \varepsilon_C, \varepsilon_M \) with \( \varepsilon_w \sim N(0, s^2), \varepsilon_C \sim N(0, t^2), \varepsilon_M \sim N(0, u^2) \). Then,

\[
\begin{align*}
&w | x_C, x_M \sim N\left( \frac{1}{2} \alpha J_{11} x_C + \frac{1}{2} \beta J_{12} x_C x_M + \frac{1}{2} \beta J_{22} x_M^2 + \lambda x_C + \eta x_M + w_0, s^2 \right) \\
y_C | x_C, x_M &\sim N(J_{11} x_C + J_{12} x_M, t^2) \\
y_M | x_C, x_M &\sim N(J_{21} x_C + J_{22} x_M, u^2).
\end{align*}
\]  

(97)  
(98)  
(99)

Denote the parameter vector by \( \theta = (\alpha, \beta, \lambda, \eta, w_0, s, t, u) \) and the data vector by \( z = (z_1, \ldots, z_n) \) where \( \forall i = 1, \ldots, n, z_i = (w_i, y_{Ci}, y_{Mi}, x_{Ci}, x_{Mi}) \) and where \( J_{11}, J_{12}, J_{21}, J_{22} \) are functions of \( \delta = \frac{\beta}{\alpha} \); \( n \) is the number of observations. Denote the joint density of \( (w, y_C, y_M) \) by \( p \) and the marginal densities by \( p_w, p_{y_C}, p_{y_M} \). Due to conditional independence of \( w, y_C, y_M \) given \( x_{Ci}, x_{Mi}, \alpha, \beta, \lambda, \eta, w_0 \), the likelihood function is given by

\[
L(\theta | z) = \prod_{i=1}^{n} p(w_i | x_{Ci}, x_{Mi}, \alpha, \beta, \lambda, \eta, w_0) \\
\times \prod_{i=1}^{n} p_{y_C}(y_{Ci} | x_{Ci}, x_{Mi}, \alpha, \beta) \\
\times \prod_{i=1}^{n} p_{y_M}(y_{Mi} | x_{Ci}, x_{Mi}, \alpha, \beta)
\]  

(100)

From (100), one obtains (17) when using (97)-(99) and taking logs. The parameter estimates are obtained by maximizing (17) with respect to \( \theta \) (or rather by minimizing \( \text{minus} \ (17) \)). The optimization algorithm \textit{fminunc} reports the Hessian evaluated at the MLE. The standard errors of the MLE are then computed as the square roots of the diagonal elements of the inverse of the Hessian.

Since there is no theoretical argument that the \textit{fminunc} algorithm achieves numerical convergence, I estimated this model using multiple starts, i.e. taking different starting values for the parameters into account. The estimation results for 1990/91 appear to be more sensitive to variations in starting values than those for 2009/10.

Finally, since I pool observations across years, there may be a concern of serial correlation in the errors. To address this issue, I also compute cluster-robust standard errors (clustered by individual) for my main specification. The results of the main specification are in Table 10. Tables 11-15 contain estimates of other specifications (robustness checks).

\textbf{Main Specification}

\begin{center}
Table 10: Maximum Likelihood Estimates on Sample of 27-29 Year Old Workers Across Cohorts
\end{center}

\begin{center}
\begin{tabular}{cccccccc}
\hline
 & \( \lambda \) & \( \eta \) & \( \alpha \) & \( \beta \) & \( f_0 \) & \( s \) & \( t \) & \( u \) \\
\hline
\textbf{1990/91} & 1.686 & -0.421 & 0.203 & 0.479 & 14.493 & 7.083 & 1.080 & 1.097 \\
 & (0.143) & (0.141) & (0.342) & (0.175) & (0.253) & (0.092) & (0.014) & (0.014) \\
 & (0.171) & (0.169) & (0.831) & (0.275) & (0.529) & (0.183) & (0.018) & (0.019) \\
\textbf{2009/10} & 2.203 & 0.210 & 0.739 & 0.055 & 15.055 & 8.823 & 1.119 & 1.222 \\
 & (0.152) & (0.152) & (0.198) & (0.154) & (0.169) & (0.093) & (0.012) & (0.013) \\
 & (0.180) & (0.156) & (0.252) & (0.162) & (0.187) & (0.275) & (0.015) & (0.016) \\
\hline
\end{tabular}
\end{center}

Standard errors in parentheses (first row gives \textit{standard} standard errors and second row gives \textit{clustered} standard errors by individual). Observations are weighed by their sampling weights.
8.5.4 Robustness

Different Age Groups and Different Years

Table 11: Maximum Likelihood Estimates Across Cohorts for Different Samples

<table>
<thead>
<tr>
<th></th>
<th>λ</th>
<th>η</th>
<th>α</th>
<th>β</th>
<th>f₀</th>
<th>s</th>
<th>t</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990-92 Age 27-29</td>
<td>1.724</td>
<td>-0.403</td>
<td>0.218</td>
<td>0.485</td>
<td>14.452</td>
<td>7.139</td>
<td>1.091</td>
<td>1.094</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.122)</td>
<td>(0.231)</td>
<td>(0.144)</td>
<td>(0.185)</td>
<td>(0.079)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>1991/92 Age 27-29</td>
<td>1.673</td>
<td>-0.391</td>
<td>0.563</td>
<td>0.590</td>
<td>14.134</td>
<td>7.237</td>
<td>1.091</td>
<td>1.103</td>
</tr>
<tr>
<td></td>
<td>(0.157)</td>
<td>(0.154)</td>
<td>(0.223)</td>
<td>(0.184)</td>
<td>(0.199)</td>
<td>(0.100)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>1990-92 Age 27-30</td>
<td>1.799</td>
<td>-0.411</td>
<td>0.350</td>
<td>0.447</td>
<td>14.581</td>
<td>7.431</td>
<td>1.112</td>
<td>1.090</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
<td>(0.110)</td>
<td>(0.184)</td>
<td>(0.123)</td>
<td>(0.150)</td>
<td>(0.071)</td>
<td>(0.011)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>2009-11 Age 27-29</td>
<td>2.255</td>
<td>0.178</td>
<td>0.703</td>
<td>0.130</td>
<td>14.846</td>
<td>8.618</td>
<td>1.095</td>
<td>1.200</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.121)</td>
<td>(0.159)</td>
<td>(0.114)</td>
<td>(0.136)</td>
<td>(0.074)</td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>2010/11 Age 27-29</td>
<td>2.292</td>
<td>-0.009</td>
<td>0.587</td>
<td>0.051</td>
<td>14.762</td>
<td>8.571</td>
<td>1.043</td>
<td>1.172</td>
</tr>
<tr>
<td></td>
<td>(0.147)</td>
<td>(0.146)</td>
<td>(0.195)</td>
<td>(0.116)</td>
<td>(0.161)</td>
<td>(0.089)</td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>2010/11 Age 27-30</td>
<td>2.501</td>
<td>0.013</td>
<td>0.587</td>
<td>0.069</td>
<td>15.168</td>
<td>8.846</td>
<td>1.036</td>
<td>1.159</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.132)</td>
<td>(0.174)</td>
<td>(0.116)</td>
<td>(0.147)</td>
<td>(0.081)</td>
<td>(0.010)</td>
<td>(0.011)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses. Observations are weighed by their sampling weights.

Use Different Types of Weights for Observations

Table 12: Maximum Likelihood Estimates on Sample of 27-29 Year Old Workers Across Cohorts

<table>
<thead>
<tr>
<th></th>
<th>λ</th>
<th>η</th>
<th>α</th>
<th>β</th>
<th>f₀</th>
<th>s</th>
<th>t</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990/91</td>
<td>1.592</td>
<td>-0.595</td>
<td>0.118</td>
<td>0.538</td>
<td>14.662</td>
<td>7.032</td>
<td>1.066</td>
<td>1.093</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.140)</td>
<td>(0.313)</td>
<td>(0.188)</td>
<td>(0.248)</td>
<td>(0.091)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>2009/10</td>
<td>2.126</td>
<td>0.201</td>
<td>0.606</td>
<td>0.095</td>
<td>15.411</td>
<td>8.370</td>
<td>1.098</td>
<td>1.206</td>
</tr>
<tr>
<td></td>
<td>(0.145)</td>
<td>(0.146)</td>
<td>(0.196)</td>
<td>(0.137)</td>
<td>(0.161)</td>
<td>(0.088)</td>
<td>(0.012)</td>
<td>(0.013)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses. Observations weighed by usual hours worked per week, interacted with sampling weights.

Table 13: Maximum Likelihood Estimates on Sample of 27-29 Year Old Workers Across Cohorts

<table>
<thead>
<tr>
<th></th>
<th>λ</th>
<th>η</th>
<th>α</th>
<th>β</th>
<th>f₀</th>
<th>s</th>
<th>t</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990/91</td>
<td>1.734</td>
<td>-0.368</td>
<td>0.146</td>
<td>0.514</td>
<td>14.524</td>
<td>7.045</td>
<td>1.081</td>
<td>1.094</td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.141)</td>
<td>(0.374)</td>
<td>(0.200)</td>
<td>(0.280)</td>
<td>(0.092)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>2009/10</td>
<td>2.079</td>
<td>0.107</td>
<td>0.676</td>
<td>0.051</td>
<td>14.854</td>
<td>8.755</td>
<td>1.119</td>
<td>1.218</td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td>(0.151)</td>
<td>(0.202)</td>
<td>(0.145)</td>
<td>(0.165)</td>
<td>(0.092)</td>
<td>(0.012)</td>
<td>(0.013)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses. Observations are not weighed.
**Single Cohort (NLSY79)**

Table 14: Maximum Likelihood Estimates on Single Cohort, No Age Restriction (NLSY79)

<table>
<thead>
<tr>
<th></th>
<th>λ</th>
<th>η</th>
<th>α</th>
<th>β</th>
<th>f₀</th>
<th>s</th>
<th>t</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991/92</td>
<td>2.121</td>
<td>-0.056</td>
<td>0.464</td>
<td>0.271</td>
<td>15.013</td>
<td>7.965</td>
<td>1.134</td>
<td>1.100</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.105)</td>
<td>(0.167)</td>
<td>(0.104)</td>
<td>(0.136)</td>
<td>(0.068)</td>
<td>(0.010)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>2004-06</td>
<td>5.086</td>
<td>-0.221</td>
<td>2.316</td>
<td>-0.692</td>
<td>19.889</td>
<td>12.633</td>
<td>1.056</td>
<td>1.058</td>
</tr>
<tr>
<td></td>
<td>(0.184)</td>
<td>(0.184)</td>
<td>(0.242)</td>
<td>(0.236)</td>
<td>(0.232)</td>
<td>(0.119)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>2006-08</td>
<td>5.037</td>
<td>-0.195</td>
<td>1.975</td>
<td>-0.694</td>
<td>19.933</td>
<td>12.609</td>
<td>1.052</td>
<td>1.071</td>
</tr>
<tr>
<td></td>
<td>(0.183)</td>
<td>(0.183)</td>
<td>(0.244)</td>
<td>(0.236)</td>
<td>(0.236)</td>
<td>(0.118)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses. Observations are weighed by their sampling weights.

**Pre Great Recession (NLSY97)**

Table 15: Maximum Likelihood Estimates on NLSY97 Sample Pre Recession

<table>
<thead>
<tr>
<th></th>
<th>λ</th>
<th>η</th>
<th>α</th>
<th>β</th>
<th>f₀</th>
<th>s</th>
<th>t</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/06 Age 23-25</td>
<td>0.763</td>
<td>0.102</td>
<td>0.242</td>
<td>0.071</td>
<td>12.544</td>
<td>7.601</td>
<td>1.256</td>
<td>1.284</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.119)</td>
<td>(0.148)</td>
<td>(0.090)</td>
<td>(0.151)</td>
<td>(0.078)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>2005/06 Age 23-26</td>
<td>0.917</td>
<td>0.077</td>
<td>0.337</td>
<td>0.228</td>
<td>12.645</td>
<td>7.870</td>
<td>1.254</td>
<td>1.279</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.115)</td>
<td>(0.131)</td>
<td>(0.128)</td>
<td>(0.149)</td>
<td>(0.076)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>2005-07 Age 24-26</td>
<td>1.074</td>
<td>0.047</td>
<td>0.335</td>
<td>0.127</td>
<td>13.235</td>
<td>8.059</td>
<td>1.215</td>
<td>1.262</td>
</tr>
<tr>
<td></td>
<td>(0.113)</td>
<td>(0.113)</td>
<td>(0.134)</td>
<td>(0.142)</td>
<td>(0.152)</td>
<td>(0.072)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>2006/07 Age 23-26</td>
<td>1.081</td>
<td>0.100</td>
<td>0.295</td>
<td>-0.055</td>
<td>13.006</td>
<td>7.904</td>
<td>1.195</td>
<td>1.251</td>
</tr>
<tr>
<td></td>
<td>(0.108)</td>
<td>(0.107)</td>
<td>(0.152)</td>
<td>(0.042)</td>
<td>(0.120)</td>
<td>(0.070)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>2006/07 Age 24-27</td>
<td>1.309</td>
<td>0.049</td>
<td>0.389</td>
<td>0.152</td>
<td>13.523</td>
<td>8.408</td>
<td>1.186</td>
<td>1.238</td>
</tr>
<tr>
<td></td>
<td>(0.126)</td>
<td>(0.127)</td>
<td>(0.156)</td>
<td>(0.122)</td>
<td>(0.155)</td>
<td>(0.081)</td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses. Observations are weighed by their sampling weights.
### 8.5.5 Mechanism Behind Wage Polarization

Table 16: Reduced Form Assignment Regressions in 1990/91 and 2009/10

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_C$</td>
<td>0.324***</td>
<td>-0.095***</td>
<td>0.337***</td>
<td>-0.088***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.016)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>$y_M$</td>
<td>-0.172***</td>
<td>0.373***</td>
<td>-0.045***</td>
<td>0.217***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.016)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>constant</td>
<td>0.039**</td>
<td>-0.021</td>
<td>0.011</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.017)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>$N$</td>
<td>2984</td>
<td>2984</td>
<td>4495</td>
<td>4495</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.177</td>
<td>0.166</td>
<td>0.133</td>
<td>0.073</td>
</tr>
</tbody>
</table>

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

![Figure 7: Assignment Estimates: Cognitive (upper panel) and Manual Dimension (lower panel). The figure plots the assignment estimates for 1990/91 in blue and, on top of that, for 2009/10 in red (right panels) from Table 16. Remarks: Assortativeness in the cognitive task has increased over time, indicated by a steeper red solid line (i.e. a slope closer to 1) and a flatter red dashed line (i.e. a slope closer to 0), compared to the blue lines in the upper right panel. But there has been a deterioration in the manual fit, indicated by a flatter red solid line (lower right panel).]
Figure 8: (De)Convexification of Wages 1990/91-2009/10. This figure displays the wage as a function of cognitive skills (upper panels of each subplot (a)-(e)) and manual skills (lower panels of each subplot), comparing data and full model (subplots (a) and (b)) and conducting counterfactual exercises (subplots (c)-(e)). To construct these figures, I use residual wages. For instance, the upper panels in each subplot plot the fitted residual wage (obtained from a regression of hourly wages on manual skills) against cognitive skills. Remarks: In both, data and full model, cognitive returns become more convex but manual returns become less convex over time. The counterfactual exercises show that only task-biased TC can account for the shifts in the wage curvature, making cognitive (but not manual) returns more convex.
8.5.6 One-Dimensional Models - Data

**One-Dimensional Assignment Model.** I estimate two versions of this model: One (my main specification) using education as the workers' single characteristic (measured by highest grade completed); one using the workers' cognitive skills as their single characteristic (measured as described in Section 8.5.2, Appendix). On the job-side this corresponds to requirements in terms of education or cognitive skills. Analogous to the estimation of the multi-dimensional model, I transform the agents’ empirical distribution of attributes into Gaussian distributions. Also, like in the multi-dimensional exercise, I weigh all observations by their sample weight. Regarding data cleaning, I follow the same steps as for the multi-dimensional data, see Section 8.5.2.

**Single Index Model.** Define Categories and Data Cleaning: I proxy the unobserved single index by workers’ age, education and labor market experience.\(^70\) In particular, there are three age groups (age 27, 28, 29), eleven education groups (< 8, 9, ..., 17, > 18 years of education), and seven labor market experience groups (< 2, 2 – 4, 4 – 6, ..., > 12 years of experience), measured by accumulated weeks working.\(^71\) This potentially yields \(3 \times 11 \times 7 = 231\) different educ-age-experience cells. However, not all cells have observations. Further, in order to ensure comparability of the wage structure across periods, I only keep those groups that have observations in all periods under consideration. Finally, I drop cells that have fewer than 5 observations, to not rely on too few observations (potentially outliers) that could have a big impact on the estimation that relies on group wages. (The estimates are very similar if I drop cells with less than 2, 3, 4 or 10 observations.) This leaves me with 104 categories that I use for the estimation.\(^72\) Finally - in line with Card and Lemieux [1996] - in my main specification the mean cell wages in both periods are weighted by the number of observations in each cell in 1990/91 (and, in turn, the observations in each cell are weighted by the NLSY79 sample weights), to address the concern about the exchangeability of individuals in each cell across periods. Below, for robustness, I also report the estimates for different weighting schemes. As far as other data cleaning is concerned, I apply the same data cleaning as for my multi-dimensional model in order to ensure comparability (see Section 8.5.1 for the details).

**Baseline Period:** The choice of the baseline period is not straightforward due to data limitations that stem from the panel structure of the NLSY and, in particular, from the lack of observations of age 27-29 in the late 1970s or early 1980s. The baseline period of my most preferred specification uses NLSY79 observations from years 1984-1988. I also experimented with different years for this baseline period. On the one hand, I cannot include years prior to 1984 since no 29 year old is present then. On the other hand, the baseline period should not be too close to the estimation period, which is why I leave out 1989. I conducted robustness, including 1989 as well or dropping both 1988 and 1989 (which comes at the cost of losing many observations), without significantly affecting the results.

---

\(^{70}\)Notice that I did not use manual and cognitive skills to classify the workers because several training variables that I use to determine an individual’s manual and cognitive skills are not available for the baseline sample in the 1980s.

\(^{71}\)I also did the analysis using potential labor market experience, which is often used in the literature as a proxy for actual experience. There are no significant changes. Moreover, I am not using finer experience cells because many cells would have only one or two observations which is problematic for the estimation that uses average wages by group.

\(^{72}\)In comparison, Card and Lemieux [1996] have 227 cells due to many more age groups.
an additional robustness check, I also used CPS data from (i) the beginning of the 1980s or (ii) the beginning of the 1970s as a baseline period (where I cleaned the data in the exact same way as I cleaned the NLSY data) to make sure that the results do not depend in important ways on the choice of the baseline period (available upon request). This approach gives similar estimates but reduces the number of cells considerably, which is why I prefer using the baseline sample from the NLSY.

8.5.7 One-Dimensional Models - Estimation

ONE-DIMENSIONAL ASSIGNMENT MODEL. In the one-dimensional assignment model, my main specification uses ‘education’ as the single worker characteristic, which is presumably a better summary measure for overall skills than cognitive skills. I estimate equation (18) by weighted least squares. Note that the sample consists of a pooled panel which is why there may be serial correlation in the error terms of the estimation. This does not impact the estimates but can bias the standard errors. To address this point, I also estimated the one-dimensional assignment model with clustered standard errors (by individual identifier), which naturally increases the standard errors but it does not change the statistical significance of the estimates of interest in important ways. See Table 18.

SINGLE INDEX MODEL. In my main specification of the single index model, I estimate equation (20) by weighted least squares where the weights of the observations are given by the weighted (by sample weight) number of cell observations in the base year (as in Card and Lemieux [1996]). Note that the estimated \( \hat{w}_j^2 \) is not an unbiased estimate of \( w_j^2 \), which is why I use \( \hat{w}_j^2 - \hat{s}_j^2 / N_j \) as an unbiased estimator of squared cell mean wages, where \( N_j \) is the number of observations in cell/group \( j \) and where \( \hat{s}_j \) represents the estimated standard deviation of wages in cell/group \( j \). All results are in Table 18.

Note that for the single index model of grouped data, and with one observation per group in each period, the weighted least squares estimator readily produces robust standard errors.

Finally, note that in the single index model, wages (which are used also as independent variables here) may be measured with error, which would cause the estimates to be biased. To address this, I also ran the regressions using the error-in-variable estimator (not reported here).
8.5.8 One-Dimensional Models - Results

![Graphs showing wages in One-Dimensional Models. Left Panel: One-Dimensional Assignment Model (1dA). Right Panel: Single Index Model (SI).](image)

Figure 10: Wages in One-Dimensional Models. Left Panel: One-Dimensional Assignment Model (1dA). Right Panel: Single Index Model (SI).

<table>
<thead>
<tr>
<th>Data Model</th>
<th>Wages 1990/91 (Model)</th>
<th>Wages 2009/10 (Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-Dimensional Assignment Model: Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%Δ Wage Skewness</td>
<td>0.33</td>
<td>-0.65</td>
</tr>
<tr>
<td>%Δ Wage Variance</td>
<td>0.30</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Single Index Model (Weights 1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>%Δ Wage Skewness</td>
<td>2.56</td>
<td>-0.50</td>
</tr>
<tr>
<td>%Δ Wage Variance</td>
<td>-0.19</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

Weights 1: All observations weighted by number of cell observations in 1990/91 where cell observations in turn are weighted by sample weights.

Table 17: Skewness and Variance of the Wage Distribution: One-Dimensional Data and Models
<table>
<thead>
<tr>
<th></th>
<th>α</th>
<th>λ</th>
<th>f₀</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-Dimensional Assignment Model (Education - Main Specification 1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990/91</td>
<td>0.884</td>
<td>2.540</td>
<td>14.530</td>
<td>3654</td>
</tr>
<tr>
<td></td>
<td>(0.190)</td>
<td>(0.125)</td>
<td>(0.158)</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>(0.267)</td>
<td>(0.159)</td>
<td>(0.188)</td>
<td>.</td>
</tr>
<tr>
<td>2009/10</td>
<td>0.201</td>
<td>2.756</td>
<td>15.053</td>
<td>4989</td>
</tr>
<tr>
<td></td>
<td>(0.185)</td>
<td>(0.125)</td>
<td>(0.154)</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>(0.243)</td>
<td>(0.154)</td>
<td>(0.172)</td>
<td>.</td>
</tr>
<tr>
<td><strong>One-Dimensional Assignment Model (Cognitive Skills)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990/91</td>
<td>0.482</td>
<td>1.868</td>
<td>14.590</td>
<td>2984</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.133)</td>
<td>(0.159)</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>(0.214)</td>
<td>(0.142)</td>
<td>(0.158)</td>
<td>.</td>
</tr>
<tr>
<td>2009/10</td>
<td>0.769</td>
<td>2.099</td>
<td>15.071</td>
<td>4495</td>
</tr>
<tr>
<td></td>
<td>(0.186)</td>
<td>(0.131)</td>
<td>(0.162)</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>(0.207)</td>
<td>(0.146)</td>
<td>(0.156)</td>
<td>.</td>
</tr>
<tr>
<td><strong>Single Index Model (Weights 1 - Main Specification 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990/91</td>
<td>0.710</td>
<td>-3.043</td>
<td>5.641</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(0.853)</td>
<td>(1.060)</td>
<td>.</td>
</tr>
<tr>
<td>2009/10</td>
<td>0.437</td>
<td>-1.521</td>
<td>3.521</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>(0.140)</td>
<td>(0.701)</td>
<td>(0.879)</td>
<td>.</td>
</tr>
<tr>
<td><strong>Single Index Model (Weights 2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990/91</td>
<td>0.710</td>
<td>-3.043</td>
<td>5.641</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(0.853)</td>
<td>(1.060)</td>
<td>.</td>
</tr>
<tr>
<td>2009/10</td>
<td>0.298</td>
<td>-0.876</td>
<td>2.801</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.665)</td>
<td>(0.850)</td>
<td>.</td>
</tr>
</tbody>
</table>

**One-Dimensional Assignment Model**: Standard errors in parentheses (with clustered standard errors in second row.) Observations weighted by sample weights.

**Single Index Model**: Standard errors in parentheses.

Weights 1: Observations weighted by number of cell observations in 1990/91 (where cell observations are weighted by sample weights), as in Card and Lemieux [1996].

Weights 2: Observations weighted by number of cell observations in 1990/91 and 2009/10 where cell observations in turn are weighted by sample weights.
8.5.9 Illustrative Example of Rank Switchings in the Wage Distribution

This section provides an illustrative example for when rank switchings in the wage distribution happen among workers who are heterogeneous in manual skills. Consider 9 worker types whose skills in each dimension can be low (L), medium (M) or high (H). Suppose that, for illustration, pre-TC complementarities across tasks are symmetric ($\delta = \delta'$). Post-TC, I assume that $\delta' < \delta$.

First, consider wages conditional on cognitive skills before and after TC. Suppose that before TC, average wages by cognitive skills satisfy the following: $E[w|x_C = H] > E[w|x_C = M] > E[w|x_C = L]$. After TC, more weight is put on the cognitive skills and less on the manual, which is why these inequalities are even more pronounced. No change in the wage ranking occurs.

Second, consider wages conditional on manual skills before and after TC. Before TC, suppose that average wages by manual skills satisfy the following: $E[w|x_M = H] > E[w|x_M = M] > E[w|x_M = L]$, where

$$E[w|x_M = H] = p_{lh}w(L, H) + p_{mh}w(M, H) + p_{hh}w(H, H)$$
$$E[w|x_M = M] = p_{mn}w(L, M) + p_{mm}w(M, M) + p_{mh}w(H, M)$$
$$E[w|x_M = L] = p_{ml}w(L, L) + p_{ml}w(M, L) + p_{hl}w(H, L)$$

where $p_{ij}, i, j \in \{l, m, h\}$ are the fractions of workers in the corresponding skill group.

Now, due to TC which is biased towards cognitive inputs, $w(L, H)$ and $w(M, H)$ in $E[w|x_M = H]$ decrease, and $w(M, L)$ and $w(H, L)$ in $E[w|x_M = L]$ increase. If technological change is sufficiently strong and the mass of workers with negatively correlated skills (captured by $p_{lh}, p_{mh}, p_{ml}, p_{hl}$) is sufficiently large, then a drop in $E[w|x_M = H]$ and an increase in $E[w|x_M = L]$ can occur, which – if sufficiently large – causes switches in the wage ranking compared to pre-TC.

References


