

# SORTING MULTIDIMENSIONAL TYPES: THEORY AND APPLICATION

ILSE LINDENLAUB\*

JOB MARKET PAPER

19 November 2013

Please Download Latest Version at

<https://sites.google.com/site/ilselindenlaub/research/LindenlaubJMP.pdf?attredirects=0&d=1>

## Abstract

This paper studies multidimensional 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 multidimensional sorting, I develop a theoretical framework that generalizes the unidimensional notion of assortative matching. I derive the equilibrium in closed form and use this explicit solution to study *biased technological change*. The key finding is that an increase of 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 dispersion. I then estimate the model for the US and identify sizeable technology shifts: During the 90s, worker-job complementarities in cognitive inputs increased by 15% whereas complementarities in manual inputs decreased by 41%. Besides this bias in complementarities, there has also been a strong cognitive *skill*-bias in production. Counterfactual exercises suggest that these technology shifts can account for observed changes in worker-job sorting, wage polarization and a significant part of the increase in US wage dispersion.

**Keywords.** Multidimensional Heterogeneity, Assortative Matching, Closed Form, Task-Biased Technological Change.

---

\*European University Institute, Via Piazzuola 43, I-50133 Florence, email: ilse.lindenlaub@eui.eu.

I am grateful to Jan Eeckhout and Nicola Pavoni for their advice. I would also like to thank Omiros Papaspiliopoulos for his input (especially on Section 3) as well as Jérôme Adda, Vasco Carvalho, Hector Chade, David Levine, George Mailath, Kristoffer Nimark and Anja Prummer for their comments. I thank Carl Sanders for making his data on occupational skill usage available to me. 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, UPenn, Penn State, Minneapolis Fed, Washington University St. Louis. 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).

# 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.<sup>1</sup> During the 80s, a blue-collar worker in the car industry might have spent some time on cognitive tasks like reporting to his supervisor but he mainly engaged in manual labor at the assembly line. Ten years later, a newly developed machine carries out his manual task. Programming the machine requires more cognitive than manual skills, which is a different skill mix than the 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 *multidimensional assignment* problem where workers with different bundles of manual and cognitive skills sort into jobs that require different combinations of these skills.

This paper develops a general theoretical framework for multidimensional sorting, which extends the unidimensional notion of positive assortative matching. I derive the equilibrium allocation as well as the equilibrium wages in closed form. I use this explicit solution to analyze the impact on these equilibrium outcomes as cognitive (as opposed to manual) inputs become more important in production, capturing one of the main recent technological shifts. I then take this model to the data to study technological change in the US during the 90s. Using this theoretical framework of multidimensional sorting, I can infer from data on observed equilibrium outcomes how much the underlying technological determinants changed over time and study their effect on sorting and wage inequality.

A key insight from this model is that workers face a *sorting trade-off*. Whether to take a job that better fits their cognitive or their manual skills depends on worker-job *complementarities* in the cognitive versus manual task.<sup>2</sup> *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' skill demands. In light of the previous example, the blue-collar worker, who was substituted by a machine, may now be employed as a car salesman. This new job is tailored to his cognitive skills but a poor fit with his manual abilities. As a result, 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 the cognitive but compressing inequality in the manual dimension.

I estimate this model for the US and identify sizeable technological shifts: I find that during the 90s, complementarities in cognitive inputs increased by 15% whereas complementarities in manual inputs decreased by 41%, in line with cognitive task-biased technological change. Moreover, there was significant cognitive *skill-biased technological change* that affects the productivity of skills *independent* of the task, leaving worker-job complementarities unchanged.

---

<sup>1</sup>See Autor et al. [2003] for an empirical analysis of the changing skill content of tasks.

<sup>2</sup>The meaning of complementarities in this context is that workers with high cognitive skills are particularly productive in jobs that put a lot of weight on cognitive ability and similar in the manual dimension.

The key findings are that these technological shifts may account for both the observed wage polarization (i.e. stagnant lower tail but expanding upper tail wage inequality) as well as a large part of the increase in wage dispersion. More precisely, counterfactual exercises show that task-biased technological change can account for wage polarization. The reason why this technology shift differently affects upper and lower tail wage inequality is that the winners (i.e. workers with high cognitive skills) are clustered in the upper part of the wage distribution while the adversely affected workers (those 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 90s.

Biased technological change, and particularly task-biased technological 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.<sup>3</sup>

One advantage of these one-dimensional models is their tractability. However, it is important to note that collapsing the 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 such as intellectual and manual skills. Some people are strong in both skills (e.g. mechanical engineers or surgeons) and others specialize. This points to the main reason for requiring matching models with multidimensional 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 the agents are ranked and matched.<sup>4</sup>

To assess the (quantitative) importance of multidimensional matching in the labor market, one needs a tractable theoretical framework. While the literature on optimal transport has studied the existence and uniqueness of multidimensional assignments under transferable utility, it provides little insights on the characteristics of the equilibrium and comparative statics.<sup>5</sup> This paper makes a first attempt at developing a tractable framework that allows for both.

Section 2 introduces the general theoretical framework. I develop an assignment model where

---

<sup>3</sup>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 [2012]). Routine skills also capture manual skills. To fit their analysis more closely, the two skills here could be interpreted as routine and non-routine skills.

<sup>4</sup>More recently, results by Papageorgiou [2013] also favor the specialization hypothesis over a single index model.

<sup>5</sup>In non-technical terms, the optimal transport problem is to find a measure-preserving map that carries one distribution into another at minimal cost, relying on linear programming. See, for instance, Gretsky et al. [1992], Villani [2009], Chiappori et al. [2010] and Ekeland [2010].

workers and jobs match in pairs. Workers possess manual and cognitive skills. At work, each worker performs two tasks, a manual and a cognitive one. Jobs, in turn, differ in the 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 multidimensional 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 similar in the manual dimension. This captures, for instance, that the best scientists usually 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 full generality without any assumptions on the distributions or specific functional form of the production technology.

To study biased technological change, one ideally has a closed form solution that is amenable for comparative statics and estimation. Towards this goal, Section 3 specifies the environment to Gaussian distributions and linear-quadratic technology. Using this notion of assortative matching, I develop a technique to solve for the equilibrium assignment and wage function in closed form.

It is important to note that besides many parallels to the one-dimensional setting, there is also an important difference: With multidimensional heterogeneity, there is no complete order of types. As a result, there is *no unique* PAM allocation that clears the labor market. This is why, contrary to one-dimensional matching in Becker [1973], super or submodularity of the technology is not enough to pin down the output-maximizing PAM allocation. Instead, the parametric specification of the production function (i.e. the *relative level* of complementarities across tasks and not only their *sign*) is crucial to determine the unique equilibrium assignment.

This strong link between technology and assignment creates the main technical difficulty in solving the model. But it also allows for a richer analysis than one-dimensional matching and offers a natural framework to study task-biased technological change, which is all about *complementarities*. Worker-job complementarities determine the optimal PAM allocation out of all the 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 than one-dimensional PAM.

Section 4 uses this closed form to analyze task-biased technological change, which demonstrates how these matching patterns (and ultimately wages) are shaped by the technology. I also contrast these results with the more standard skill-biased technological change, which only increases the relative productivity of the workers' cognitive *skills* without affecting worker-job complementarities.

Section 5 brings this model to the data. I focus on the US economy during the 90s. I first construct the bivariate skill and skill demand distributions, combining data from the National Longitudinal Survey of Youth (NLSY) and the O\*NET. I then estimate the model by Maximum Likelihood to quantify technological change during this period and to decompose the changes in wage inequality into those driven by different technological and distributional shifts.

I also highlight in which dimensions the multidimensional model offers a richer interpretation of

the data than a similar model with one-dimensional traits. The one-dimensional model misses several important margins: first, it misses the manual-cognitive sorting trade-off and, closely related, the differential impact of biased technological change on manual and cognitive returns. Moreover, it fails to account for a sizeable group of *generalists* (holding both types of skills) whose cognitive skills allow them to buffer against the adverse shock on manual skills.

Section 6 places the main contribution of the paper into the literature. Section 7 concludes. The Appendix contains all the proofs, data details and estimation results.

## 2 Theoretical Framework for Multidimensional Sorting

Towards the goal of developing a theoretical framework for multidimensional sorting, this section outlines the general model without relying on specific assumptions on the underlying distributions or production technology. To make the results most intuitive, I will focus here on two-dimensional heterogeneity. Notice that this section fully generalizes to N-dimensional heterogeneity.<sup>6</sup>

### 2.1 Environment

*Agents:* There are two types of agents, firms and workers. All are risk-neutral. There is a continuum of each type. Every worker is endowed with a skill bundle of cognitive and manual skills,  $\mathbf{x} = (x_C, x_M) \in X \subseteq \mathbb{R}_+^2$ . Points in  $X$  represent worker types. Denote the joint c.d.f. of  $(x_C, x_M)$  by  $H(x_C, x_M)$ , which is assumed to be absolutely continuous with respect to the Lebesgue measure. In turn, each firm (which I use interchangeably with *job*) is endowed with both cognitive and manual skill demands,  $\mathbf{y} = (y_C, y_M) \in Y \subseteq \mathbb{R}_+^2$ .  $y_C$  (respectively  $y_M$ ) corresponds to the productivity or skill requirement in the 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  $G(y_C, y_M)$ , which is also assumed to be absolutely continuous. Assume that the overall masses of firms and workers coincide.<sup>7</sup>

*Production:* Every firm produces a single homogenous final good by combining all inputs. Denote the technology by  $F(x_C, x_M, y_C, y_M)$ . It is assumed that  $F$  is twice continuously differentiable.

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

### 2.2 Definitions

*Matching Function:* The sorting between workers and firms is described by a map  $\mathbf{x}^* = \nu(\mathbf{y})$ , where  $\nu(\mathbf{y})$  is the worker type that firm  $\mathbf{y}$  optimally chooses to hire ('\*' indicates an equilibrium object). The focus here is on a bijective  $C^1$  map  $\nu : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+^2$ , which can be uniquely characterized by its inverse  $\mu \equiv \nu^{-1}$ . I call  $\mu$  the *matching function*, which describes the assignment of workers to firms.

*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 monotonically increasing (decreasing) matching function, denoted as  $y = \mu(x)$ , meaning that better (worse) workers work in

---

<sup>6</sup>All proofs of this section are given for N-dimensional heterogeneity.

<sup>7</sup>Otherwise, there is equilibrium unemployment or idle firms, which unnecessarily complicates the model.

better firms. This concept captures two aspects: (a) purity of the matching (i.e.  $\mu(x)$  is one-to-one), (b) the direction of sorting. Here, I aim to define a multidimensional version of assortative matching which also incorporates these two features. Like in the one-dimensional setting, here assortativeness involves properties of the first derivative of the matching function (i.e. of its Jacobian), given by:

$$J_\mu \equiv D_x \mathbf{y}^* = \begin{bmatrix} \frac{\partial y_C^*}{\partial x_C} & \frac{\partial y_C^*}{\partial x_M} \\ \frac{\partial y_M^*}{\partial x_C} & \frac{\partial y_M^*}{\partial x_M} \end{bmatrix}$$

I define multidimensional positive and negative assortative matching as follows:

**Definition 1 (Assortative Matching with Multidimensional Types)** *The sorting pattern is PAM (NAM) if  $D_x \mathbf{y}^*$  is a P-matrix ( $P^-$ -matrix), i.e. if*

$$[i] \quad \frac{\partial y_C^*}{\partial x_C} > (<) 0 \quad [ii] \quad \frac{\partial y_M^*}{\partial x_M} > (<) 0 \quad [iii] \quad \text{Det}(J_\mu) = \frac{\partial y_C^*}{\partial x_C} \frac{\partial y_M^*}{\partial x_M} - \frac{\partial y_C^*}{\partial x_M} \frac{\partial y_M^*}{\partial x_C} > 0 \quad (1)$$

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.<sup>8</sup> In economic terms, PAM means that intellectual types work in firms where workers need to perform complex intellectual tasks (part [i]). Similarly, workers with strong manual skills work in firms that attach considerable weight to the manual task (part [ii]). Moreover, inequality [iii] says that these *within*-task matching forces dominate the *between*-task matching forces. Otherwise, scientists would work in the best garages whereas the best mechanics would work at leading universities. Hence, this definition captures the *direction of sorting*, which under PAM is a positive relation between worker and firm traits along the natural sorting dimensions.

Definition 1 also captures that the assignment is *pure*, defined as follows:

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

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.<sup>9</sup> Gale and Nikaïdo [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  $D_x \mathbf{y}^*$  is a *P-matrix* (or *P<sup>-</sup>-matrix*), then the matching function is globally one-to-one. The *P-matrix* property is also sufficient for global invertibility, justifying my approach to consider  $\mu = \nu^{-1}$  as the matching function instead of  $\nu$ .<sup>10</sup>

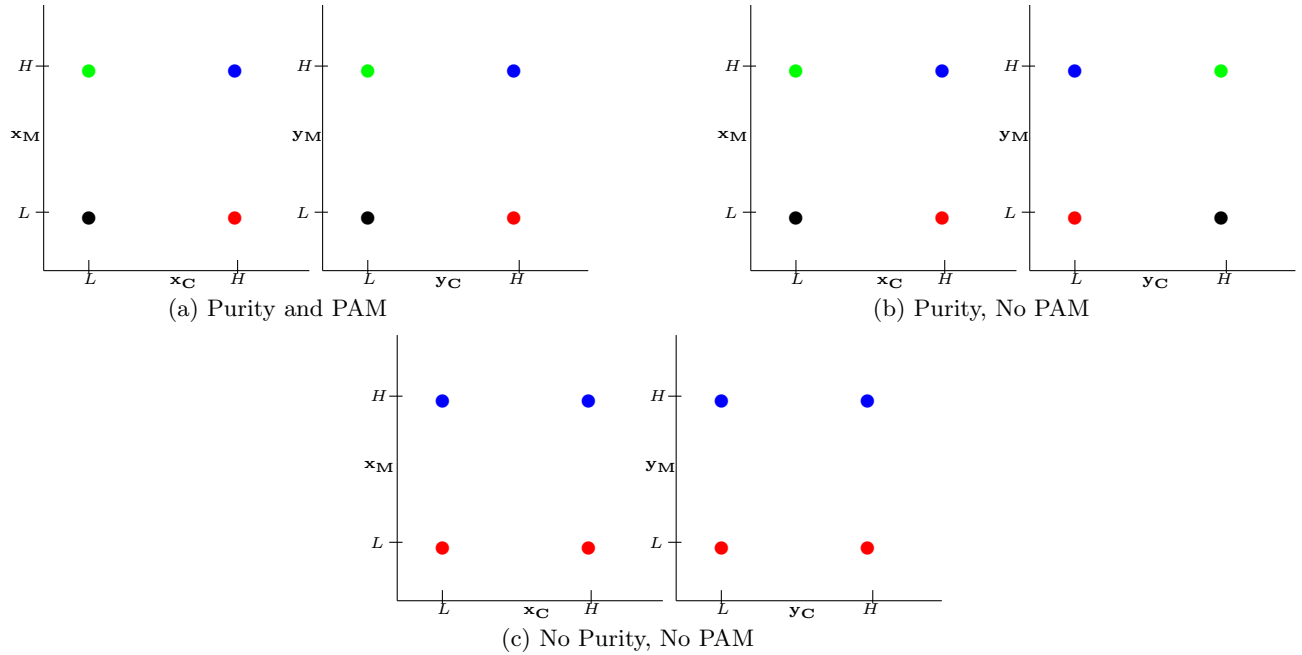
Definition 1 is a natural generalization of one-dimensional assortative matching, capturing the same two aspects: the *direction of sorting* in each task dimension (given by [i] and [ii] in (1)) and *purity* of the assignment (guaranteed by the determinant condition [iii]). In both the one-dimensional and multidimensional setting, PAM implies purity.

<sup>8</sup>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, a matrix  $M$  is *P<sup>-</sup>* if  $-M$  is *P*.

<sup>9</sup>*P-matrices* have so far not been exploited 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].

<sup>10</sup>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  $\mathbb{R}^n$ .

The figure below provides a graphical illustration of multidimensional 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 four worker and four firm types. In each subfigure, the left panel are worker types and the right panel are firm types. Dots indicate types. Assume that all dots 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 preferred type) but PAM is violated along the C dimension. In subfigure (c), matching is neither positive assortative nor pure because agents are indifferent between several matches.



### 2.3 The Firm's Problem

A firm with given productivity bundle  $(y_C, y_M)$  chooses a worker with skill bundle  $(x_C, x_M)$  in order to maximize profits. It takes the wage schedule as given, meaning that wages are not a function of productivities. In this section, I derive the firm's problem and optimality conditions heuristically, taking as given that the wage function (denoted by  $w(x_C, x_M)$ ) is twice continuously differentiable. Below, I show conditions under which  $w(x_C, x_M)$  satisfies this property. The firm's problem is given by:

$$\max_{(x_C, x_M) \in X} F(x_C, x_M, y_C, y_M) - w(x_C, x_M) \tag{2}$$

The FOCs of this maximization problem read

$$F_{x_C}(x_C, x_M, y_C, y_M) - w_{x_C}(x_C, x_M) = 0 \tag{3}$$

$$F_{x_M}(x_C, x_M, y_C, y_M) - w_{x_M}(x_C, x_M) = 0 \tag{4}$$

where subscripts denote derivatives. Equations (3) and (4) hold only at the equilibrium assignment.

## 2.4 The Equilibrium

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  $(x_C, x_M)$  for a given  $w(x_C, x_M)$ .*
- (ii) *Market Clearing: Feasibility of  $\mu$  requires that when  $\mathbf{x} \sim H$  then  $\mathbf{y}^* \sim G$ .*

*Optimality* of the firm's choice is a standard requirement of a competitive equilibrium. *Market Clearing* requires that the amount of workers of type  $(x_C, x_M)$  demanded across all firm types cannot exceed the measure of such workers in the economy.

Existence of the equilibrium in the Walrasian assignment problem was proven in [Gretsky et al. \[1992\]](#). Since the focus of this paper is on the characterization of the equilibrium, the reader is referred to their proof for general existence (Theorem 4 of their paper).<sup>11</sup> In what follows, I focus on a *differentiable* equilibrium where  $\mu$  is  $C^1$  and  $w$  is  $C^2$ . The differentiable equilibrium can be shown to exist in various multivariate environments with absolutely continuous distributions. I will discuss some of them below, with particular emphasis on the case with Gaussian distributions.

## 2.5 The Equilibrium Assignment

This section relates properties of the production technology to the properties of the equilibrium assignment  $(y_C^*, y_M^*) = \mu(x_C, x_M)$ , which I will explicitly denote by  $y_C^* = y_C(x_C, x_M)$  and  $y_M^* = y_M(x_C, x_M)$ . This 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 technology features the following complementarities

$$D_{xy}^2 F = \begin{bmatrix} F_{x_C y_C}^+ & 0 \\ 0 & F_{x_M y_M}^+ \end{bmatrix} \quad (5)$$

then the equilibrium assignment satisfies PAM (i.e.  $D_x \mathbf{y}^*$  is a  $P$ -matrix). For NAM, a similar statement holds when replacing complementarities by substitutabilities.<sup>12</sup> Moreover, under the same condition, the assignment is a global maximum:

**Proposition 1 (Assortativeness and Global Maximum)** *If  $D_{xy}^2 F$  is a diagonal  $P$ -matrix ( $P^-$ -matrix), then the equilibrium assignment satisfies PAM (NAM), and is globally unique.*

The proof is in Appendix A.1. To gain intuition into the assortativeness result, consider PAM. If there is complementarity between skills and productivities *within* both the cognitive task ( $F_{x_C y_C} > 0$ ) and

<sup>11</sup>Closely related existence results are given in [Chiappori et al. \[2010\]](#) and [Ekeland \[2010\]](#). Their work extends Gretsky's existence result on the *endowment* economy where every seller is endowed with a given type of good to a *production* economy where sellers can choose the type of good they want to sell.

<sup>12</sup>The presented condition is related to the *twist condition* from optimal transport but is not equivalent. See Section 6.



the manual task ( $F_{x_M y_M} > 0$ ) and interfering *between*-task complementarities are absent ( $F_{x_C y_M} = F_{x_M y_C} = 0$ ), then it is optimal that workers and firms match positively assortative: Agents with strong intellectual skills work in firms that value these skills (and similarly on the manual dimension).<sup>13</sup>

This sorting result, which ensures a positive relation between skills and productivities along the *natural* dimensions (i.e. within the cognitive and within the manual task), is obtained under strong restrictions on the complementarities in production. The intuition is that, in the multidimensional world, sorting occurs along *all* skill and productivity dimensions, i.e. also *between* tasks. That is, between the manual productivity and the cognitive skill,  $\frac{\partial y_M^*}{\partial x_C} \neq 0$ , and also between the cognitive skill demands and the manual ability,  $\frac{\partial y_C^*}{\partial x_M} \neq 0$ . Allowing for complementarities between, say, manual skill demands and cognitive skills ( $F_{x_C y_M} > 0$ ) might render a positive relation between these two attributes,  $\frac{\partial y_M^*}{\partial x_C} > 0$ . This may come at the expense of negative sorting in the manual task,  $\frac{\partial y_M^*}{\partial x_M} < 0$ , 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 across-task complementarities or substitutabilities ( $F_{x_C y_M}, F_{x_M y_C} \neq 0$ ). In Section D of the Appendix, I show that a weaker version of (5) applies to settings where skills and productivities are (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*.

This section closes with a comparison to the one-dimensional setting. With one-dimensional traits, the requirement of a negative definite Hessian collapses to the requirement on the second-order condition, given by  $-F_{xy} \frac{\partial \mu(x)}{\partial x} < 0$ . If  $F_{xy}$  is positive, then matching is PAM. Purity is given by the strict monotonicity of the matching function  $\mu$  and the sorting direction by its positive slope. Similarly in this model, I impose conditions on the matrix of cross-partials  $D_{xy}^2 F$  to obtain PAM. The difference is that with multiple dimensions not only the signs but also the relative magnitude of different complementarities need to be restricted in order to ensure assortative matching.

## 2.6 The Equilibrium Wage Function

This section derives conditions for the existence of a unique wage schedule that supports the equilibrium assignment. The equilibrium wage is the solution of a *system of partial differential equations* (PDEs), which are given by the first-order conditions of the firm, (3) and (4), evaluated at the equilibrium assignment. To solve a system of PDEs, integrability conditions of the system need to be specified in order to make the system involutive (i.e. formally integrable). For the linear system of first-order PDEs given above, there is only one integrability condition. It is given by the commutativity of mixed

---

<sup>13</sup>Similarly, in the case of NAM, assortative matching within tasks dominates the assortativeness across tasks, only that in this case high productivity workers are matched with low productive firms.

partial derivatives and obtained by cross-differentiating (3) and (4), when evaluated at  $(y_C^*, y_M^*)$ :

$$\begin{aligned} w_{x_C x_M} &= w_{x_M x_C} \\ \Leftrightarrow F_{x_C y_C} \frac{\partial y_C^*}{\partial x_M} + F_{x_C y_M} \frac{\partial y_M^*}{\partial x_M} &= F_{x_M y_C} \frac{\partial y_C^*}{\partial x_C} + F_{x_M y_M} \frac{\partial y_M^*}{\partial x_C}. \end{aligned} \quad (6)$$

This condition is equivalent to the requirement that the Hessian of the firm's problem is symmetric. The next proposition states the result on existence and uniqueness of the equilibrium wage function.

**Proposition 2 (Existence and Uniqueness of the Wage Function)** *There exists a unique wage function (up to a constant) that decentralizes the equilibrium assignment if and only if the equilibrium assignment satisfies (6).*

The proof relies on *Frobenius' Theorem*. Both theorem and proof are stated in the Appendix A.2. Integrability condition (6) has both technical and economic implications. Technically, given (6), there exists a  $C^2$  wage function  $w$ , justifying the differentiation-based approach above.

Condition (6) also carries an important economic message. It highlights a crucial difference between the multidimensional and the one-dimensional setting. 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,  $F_{x_i y_j}$ ,  $i, j \in \{C, M\}$ , but also on their *strength*. Changing the strength (but not the signs) of  $F_{x_i y_j}$  will induce worker reallocation without necessarily violating PAM or NAM. Matching multidimensional types thus generates something similar to an *intensive margin* even though firms and workers match in pairs.

This is different from the one-dimensional setting, where there is no such integrability condition because the wage is the solution to a *single* ordinary differential equation. There, the assignment only depends on the sign of  $F_{x_i y_i}$  but not on its strength: 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  $y = G^{-1}(H(x))$ . 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. 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.

### 3 Quadratic-Gaussian Model

A main goal of this paper is to apply this multidimensional sorting framework to the empirically relevant phenomenon of biased technological change. This section presents an important step towards this objective. It specifies the environment to Gaussian distributions and quadratic technology and develops a technique to explicitly compute the multidimensional assignment and corresponding wage. The closed form solution then allows me to focus on the economics of multidimensional sorting: characterize the equilibrium properties and analyze comparative statics.

PAM provides the crucial link between the previous general section on multidimensional sorting, this section on the closed form and the next section on the application: First, PAM puts a *useful structure* on the equilibrium assignment that helps to solve multidimensional assignment models similarly to one-dimensional problems. Second, despite the imposed structure, PAM is *flexible* enough to allow for a wide range of assignment patterns. I will show how technology and distributions generate a rich set of positive assortative matchings, which is at the heart of the paper’s application below.

### 3.1 Environment

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

$$\begin{bmatrix} x_C \\ x_M \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}\right), \quad \begin{bmatrix} y_C \\ y_M \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_y \\ \rho_y & 1 \end{bmatrix}\right)$$

Denote the bivariate distribution functions of skills and productivities by  $\Phi_x(x_C, x_M)$  and  $\Phi_y(y_C, y_M)$ , respectively. Assume,  $\rho_x, \rho_y \in (-1, 1)$ . I focus on the bi-linear technology

$$F(x_C, x_M, y_C, y_M) = \alpha x_C y_C + \beta x_M y_M = \alpha(x_C y_C + \delta x_M y_M) \tag{7}$$

where  $\alpha$  and  $\beta$  are task-weights that indicate the level of worker-job complementarities or substitutabilities across tasks. Notice that  $\delta \equiv \frac{\beta}{\alpha}$  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.<sup>14</sup>

Technology (7) captures that there is *within*-task complementarity but *between*-task complementarity is shut down. Based on the results in Section 2, certain properties of the equilibrium assignment are already known at this point without having to check second-order conditions of the firm’s problem.<sup>15</sup> Under (7),  $D_{xy}^2 F$  is a diagonal  $P$ -matrix. Consequently, the equilibrium assignment is unique and satisfies PAM. These properties will prove useful in the construction of the equilibrium.

Notice that this model can be generalized in various ways. It can be solved in closed form under

$$F(x_C, x_M, y_C, y_M) = \alpha x_C y_M + \beta x_M y_C + \gamma x_C y_C + \delta x_M y_M \tag{8}$$

allowing for non-zero between-task complementarity (Appendix D.0.1). Moreover, I can allow for non-standard normally distributed variables or even arbitrary marginal distributions that are linked via Gaussian copulas (see Online Appendix). 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. Here, I sketch the main steps. Appendix B provides the details.

<sup>14</sup>Nothing hinges on this restriction but it simplifies the results’ interpretation and, moreover, is in line with the data.

<sup>15</sup>Analogously to the general model, the firm’s problem is given by:  $\max_{(x_C, x_M) \in X} \alpha(x_C y_C + \delta x_M y_M) - w(x_C, x_M)$ .

### 3.2 The Equilibrium Assignment Functions

The objective is to compute the equilibrium assignment functions  $y_C^* = y_C(x_C, x_M)$  and  $y_M^* = y_M(x_C, x_M)$  in closed form. They must be consistent with both labor market clearing and the firm's optimality. Due to the incomplete order of types in the multidimensional setting, there are many possibilities of how to match workers with firms in a positive assortative way. This is the main difficulty in solving for the assignment. What matters for pinning it down is not only the sign but also the relative strength of skill-productivity complementarities across tasks, captured by  $\delta$ . By temporarily converting the two-dimensional problem to two separate one-dimensional problems, I make the model tractable. I do the matching in the transformed space and then re-transform:

I first apply a measure-preserving transformation that un-correlates the Gaussian variables. In particular, let  $\mathbf{x}$  be a p-variate random vector with mean  $\mu$  and nonsingular covariance matrix  $\Sigma$ . Then,

$$\mathbf{z} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu) \quad (9)$$

has mean  $\mathbf{0}$  and covariance matrix  $\mathbf{I}_p$ . The matrix  $\Sigma^{-\frac{1}{2}}$  is the inverse of *any* square root of the covariance matrix, i.e.  $\Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T = \Sigma$ . Denote by  $\Sigma_x$  (resp.  $\Sigma_y$ ) the covariance matrix of skills (resp. productivities). Apply (9) to the standard bivariate normal skills and productivities

$$\mathbf{z}_x = \begin{bmatrix} z_{x_C} \\ z_{x_M} \end{bmatrix} = \Sigma_x^{-\frac{1}{2}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \quad \text{and} \quad \mathbf{z}_y = \begin{bmatrix} z_{y_C} \\ z_{y_M} \end{bmatrix} = \Sigma_y^{-\frac{1}{2}} \begin{bmatrix} y_C \\ y_M \end{bmatrix} \quad (10)$$

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

$$(1 - \Phi(z_{y_C}))(1 - \Phi(z_{y_M})) = (1 - \Phi(z_{x_C}))(1 - \Phi(z_{x_M})) \quad (11)$$

where  $\Phi$  again denotes the standard normal c.d.f. The interpretation of (11) is that if the firm  $(z_{y_C}, z_{y_M})$  matches with worker  $(z_{x_C}, z_{x_M})$ , then the mass of workers with better skills than  $(z_{x_C}, z_{x_M})$  must be equal to the mass of firms that are more productive than  $(z_{y_C}, z_{y_M})$  (due to PAM).<sup>16</sup>

The market clearing condition (11) implicitly defines the vector-valued matching function of transformed variables, denoted by  $\mu^z : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . The objective is to back out two real-valued assignment functions of this vector-valued matching function. To do so, set equal the quantiles of the marginal skill and productivity distributions *within* the cognitive and *within* the manual dimension

$$\Phi(z_{y_i}) = \Phi(z_{x_i}) \quad \forall \quad i \in \{C, M\} \quad (12)$$

which gives a system of two equations. In principle, there are many possible ways to match up the

<sup>16</sup>I will verify below that the market clearing in the transformed variables  $(\mathbf{z}_x, \mathbf{z}_y)$ , which is based on purity and PAM, gives rise to an assignment in  $(x, y)$  that also admits purity and PAM.

marginals in (11) but due to PAM (i.e. positive diagonal elements of  $D_x \mathbf{y}^*$ ), this is the only sensible way. System (12) can be retransformed into original variables, explicitly solving for productivities  $y_C$  and  $y_M$  as functions of skills  $x_C$  and  $x_M$ , which constitutes the candidate equilibrium assignment

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \quad (13)$$

where  $D_x \mathbf{y}^* = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}}$  is the Jacobian of the matching function. System (13) is the candidate equilibrium assignment, mapping bivariate skills into bivariate productivities. By (11), it is measure-preserving (i.e. in line with labor market clearing). Notice, however, that a covariance matrix has an *infinite number* of square roots because it is a *symmetric* positive definite matrix. Hence, there are *many* matchings that satisfy market clearing and that are potentially in line with PAM. How to pick the optimal one? I use the degree of freedom in computing the square roots to take into account the firm's optimal choice, which depends on the relative level of skill-productivity complementarities, captured by  $\delta$ .<sup>17</sup> 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)** *The equilibrium assignment  $\mu$  is given by*

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} = \underbrace{\begin{bmatrix} J_{11}(\rho_x, \rho_y, \delta) & J_{12}(\rho_x, \rho_y, \delta) \\ J_{21}(\rho_x, \rho_y, \delta) & J_{22}(\rho_x, \rho_y, \delta) \end{bmatrix}}_{D_x \mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \quad (14)$$

where  $J_{11}, J_{12}, J_{21}, J_{22}$  have closed forms. 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 two polar cases. The square roots,  $\Sigma_y^{\frac{1}{2}}, \Sigma_x^{\frac{1}{2}}$ , are obtained from a rotation of the spectral square roots. They range between the spectral square root (for  $\delta = 1$ ) and the Cholesky square root (for  $\delta = 0$ ).

See Appendix B.2 for the proof and explicit expressions. 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 completely asymmetric case ( $\delta = 0$ ), only the cognitive task matters for production. The Cholesky square root is the unique lower triangular square root, and hence *asymmetric*. It delivers an *asymmetric* assignment, which is optimal when the technology exhibits extreme asymmetries. Last, when there are some asymmetries in the production technology  $\delta \in (0, 1)$ , then the assignment is in-between these two polar cases (more intuition below).

There are two main messages from this proposition. First, the assignment can be computed in closed form and is a *linear* map from skill bundles to productivity bundles. Second, there is a much stronger link between technology and assignment compared to the one-dimensional case, where the

<sup>17</sup>This is done by taking into account the integrability condition (6), under which a wage schedule exists that induces firms to choose this assignment. With this bilinear technology, (6) collapses to  $\frac{\partial y_C^*}{\partial x_M} = \delta \frac{\partial y_M^*}{\partial x_C}$ .

matching function only depends on the underlying distributions (through skill correlation  $\rho_x$  and productivity correlation  $\rho_y$ ). Unlike in 1-d, here also the relative *level* of skill-productivity complementarities enters the assignment, which is captured by covariance square roots that adjust 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. In this quadratic-Gaussian model, the wage function admits a highly tractable closed-form solution.

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

$$w(\mathbf{x}) = \frac{1}{2}\alpha\mathbf{x}^T\tilde{J}\mathbf{x} + w_0 \tag{15}$$

where  $w_0$  is the constant of integration.

See Appendix B.3 for the proof and explicit expression.  $\tilde{J}$  is a matrix of parameters and closely related to the equilibrium assignment. It contains the assignment coefficients from the Jacobian of the matching function,  $J_\mu$ . For the special case of symmetric tasks ( $\delta = 1$ ), the two coincide,  $\tilde{J} = J_\mu$ , emphasizing the tight link between allocation and wages, which is typical for assignment models. The wage function is a quadratic form in standard normal variables, which allows 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.

### 3.4 Properties of the Equilibrium

This section discusses the equilibrium properties for the benchmark case with symmetric tasks ( $\delta = 1$ ). The next section on the application of task-biased technological change (task-biased TC hereinafter) studies in detail the case of asymmetric task weights ( $\delta \neq 1$ ). To analyze the sorting properties of this equilibrium, it is useful to first define the concepts of *perfect assortativeness* and *mismatch*.

**Definition 4 (Perfect Assortativeness and Mismatch)** *An assignment in task  $i \in \{C, M\}$  is perfectly assortative if  $x_i = y_i$ . An assignment is characterized by mismatch if  $|y_i - x_i| \neq 0$ . Mismatch is said to be increasing in  $|y_i - x_i|$ .*

Perfect assortativeness means that a worker’s skills perfectly match the firm’s skill requirements in a certain task. The opposite of perfect assortativeness is structural mismatch, which I define as the dissimilarity between skills and skill demands in a given match. Notice that mismatch in this frictionless economy has nothing to do with inefficiencies. Instead, it refers to the misfit between workers’ and firms’ traits. I can now state the following properties of the equilibrium assignment.

**Proposition 5 (Equilibrium Sorting)** *(i) The equilibrium assignment is characterized by PAM. (ii) For a perfect fit of skill supply and demand ( $\rho_x = \rho_y$ ), sorting is perfectly assortative in both tasks, i.e.  $y_C = x_C$  and  $y_M = x_M$ . In turn, for the poorest fit of skill supply and demand (i.e. maximal  $|\rho_x - \rho_y|$ ), mismatch along both task dimensions is maximized.*

By construction, the equilibrium assignment satisfies PAM, meaning that workers with more intellectual skills work in jobs that value them and similarly on the manual dimension. This stems from the technology that features worker-firm complementarities in each task. Interestingly, the *degree* of assortativeness depends on the underlying distributions. This is illustrated by the two polar cases: First, when skill supply and demand perfectly overlap ( $\rho_x = \rho_y$ ), then every worker matches with the firm that needs exactly his skills. On the other hand, if there is a large discrepancy between the skills needed and the skills supplied ( $|\rho_x - \rho_y| \rightarrow 2$ ), 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.

These results are illustrated in Figure 1, which displays contour plots of two standard normal distributions for various skill and productivity correlations. For the sake of illustration, assume that the workers are represented by the blue contour lines and the firms by the red ones. In the middle panel, there is a perfect fit of skill supply and demand distributions, which would lead to perfect matches between workers and jobs. The panels at the left and right show the other extreme case, where skill demand and supply are most misaligned. Focus on the left panel. In this economy, workers are *specialists* (they are either good in the manual or in the cognitive task but not in both) whereas firms want *generalists*. The labor market clears under PAM but matches are characterized by a poor fit between workers' and firms' attributes.

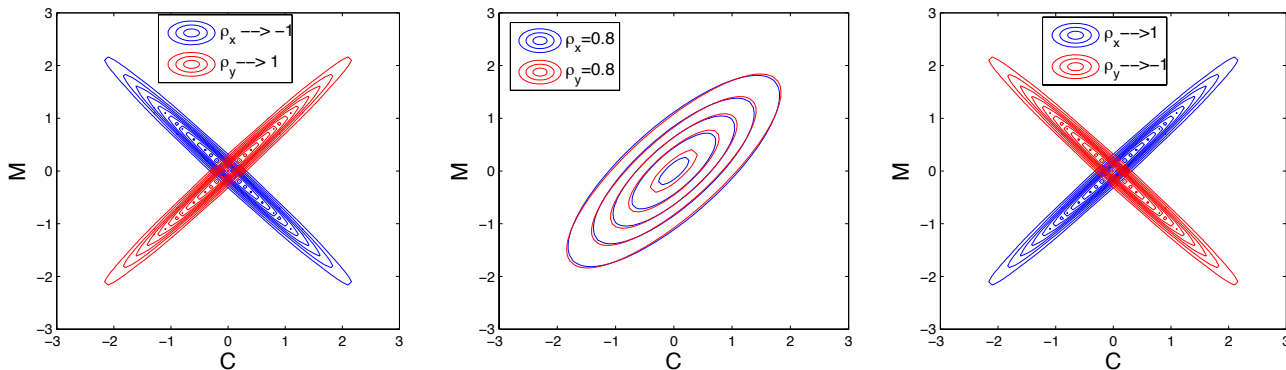


Figure 1: Contour Plots of Skill and Productivity Distributions

The Jacobian of the matching function offers an alternative graphical way of looking at the equilibrium sorting properties. In the graph below, I plot the 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 slope of solid lines resembles the assortativeness of the match (they show how strongly skills and productivities relate *within* tasks) and the slope of dotted lines indicates the degree of mismatch in a pair (they show how strongly skills and productivities relate *between* tasks). The assortativeness forces are given by the diagonal elements of the Jacobian  $J_\mu$  ( $J_{11}$  and  $J_{22}$ ), and the mismatch forces are given by the off-diagonal elements of the Jacobian, ( $J_{12}$  and  $J_{21}$ ).

The assignment in the upper panel is perfectly assortative where only the *right* skill contributes to the match: the straight lines have slope one and the dotted lines lie on the x-axis, i.e.  $y_C = x_C$  and  $y_M = x_M$ . This assignment results when the underlying distributions perfectly overlap ( $\rho_x = \rho_y$ ). The lower panel displays the other extreme. Here the matches are characterized by maximum mismatch with the *wrong* skill dimensions contributing almost as much to the match as the assortative dimensions: the slopes of straight and dotted lines are similar. Such an assignment corresponds to the left and right panel in the previous figure where the underlying distributions differ significantly. Notice that despite considerable mismatch, PAM is satisfied ((i) positively sloped straight lines, (ii) straight lines are steeper than dotted lines). One advantage of my multidimensional notion of assortative matching is that, despite the imposed structure, it is flexible enough to allow for a rich set of assignment patterns, ranging from perfect assortativeness to significant mismatch.

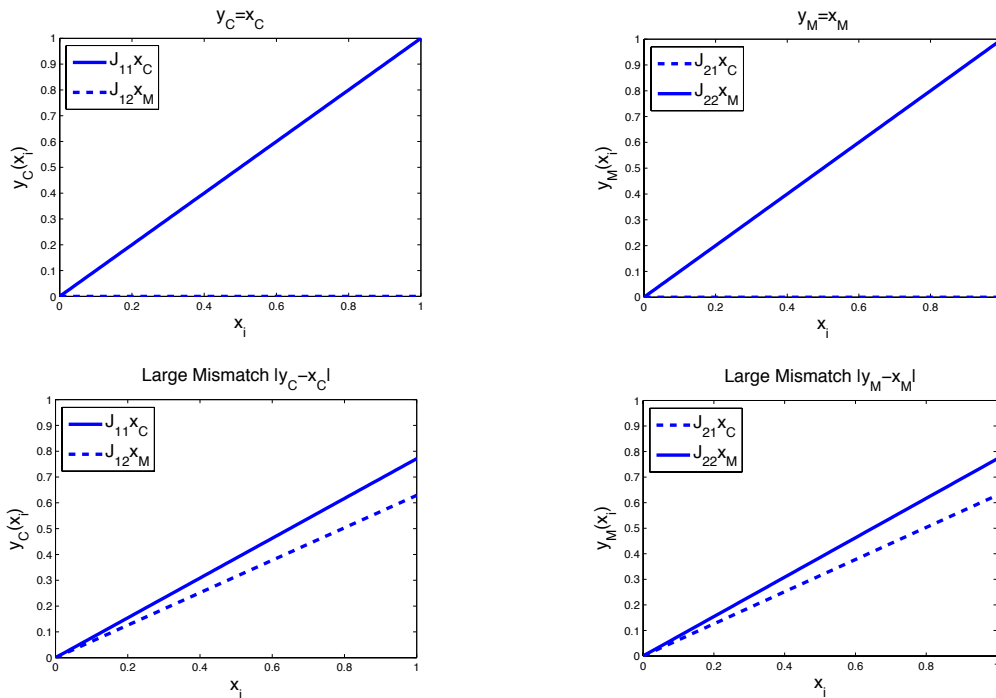


Figure 2: Perfect Assortativeness (upper panel) and Mismatch (lower panel) within PAM

The next result summarizes a selected set of properties of the 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. Since sorting is positive assortative (implying that  $\tilde{J}$  in (15) is a symmetric P - matrix or positive definite) wages are convex. Convex wages mean that workers with large (absolute) amounts of skills earn over-proportionally more than workers with small (absolute) amounts of skills. Notice that 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.



On the other hand, if sorting was negative assortative, the wage function would be concave.<sup>18</sup>

An alternative 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 over-proportionally much. The force behind positive skewness is again PAM, which is the driving factor of wage inequality in the economy.

It can also be shown that the average performance of the economy depends on the assignment of workers to firms and thus on the 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 the misalignment between skills and skill requirements is largest.

This section illustrated how sorting depends on the economy’s skill and productivity distributions and how this feeds into wages. It was shown that PAM is the major force behind wage inequality. The next section revisits the key message from Proposition 3 that the assignment not only depends on the distributions but also on the *technology* (through the relative level of firm-worker complementarities across tasks). I will use the closed form and the developed sorting framework to study the central application of this paper: How does task-biased technological change affect assignment and wages? How are these effects mitigated or reinforced by the underlying distributions?

## 4 Biased Technological Change

This section uses the closed form to study the central economic question of this paper. I analyze the effects of task-biased TC and also contrast them with skill-biased TC.

### 4.1 Task-Biased Technological Change

Task-biased TC is considered an important force behind recent wage inequality shifts in the developed world. The idea behind task-biased TC is that technological advances have replaced workers in the manual task but created stronger complementarities between skills and job attributes in the cognitive task. The literature also refers to this technological change as *routinization*, where workers performing routine tasks are increasingly substituted by computers and machines.<sup>19</sup> Notice that task-biased TC does not imply that the prevalence of those 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 more natural assumption that both types of skills are used on every job yet in different proportions.<sup>20</sup>

---

<sup>18</sup>Even though this is not the focus here, it is worth mentioning that in this model, wage data is sufficient to determine the direction of sorting. In several one-dimensional models, this is not the case (see e.g. Eeckhout and Kircher [2011]).

<sup>19</sup>See, e.g. Autor et al. [2003], Autor et al. [2006] and Autor and Dorn [2012]. There is a close mapping between manual and routine skills on the one hand, and between cognitive and non-routine skills on the other.

<sup>20</sup>This is similar to the *skill weights approach* by Lazear [2009].

In the presented model, task-biased TC can be captured by a relative decrease in skill-productivity complementarities in the manual task. 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 the relative complementarities in the manual task. Consider a change from  $\delta$  to  $\delta'$  such that  $\delta' < \delta = 1$ . Then,  $\delta'$  is called *task-biased* relative to  $\delta$ , with the bias favoring the cognitive task. Moreover, to obtain clean analytical results, I will focus on the cases where  $\rho_x, \rho_y \leq 0$  or  $\rho_x, \rho_y \geq 0$ .<sup>21</sup> The next result summarizes the effect of task-biased TC on the equilibrium assignment.

**Proposition 7 (Task-Biased TC and Sorting)** *Suppose there is cognitive task-biased TC ( $\delta' < \delta = 1$ ): (i) The equilibrium assignment still satisfies PAM. (ii) Sorting becomes more (less) pronounced in the cognitive (manual) task (i.e.  $|y_C - x_C|$  decreases,  $|y_M - x_M|$  increases). (iii) As  $\delta \rightarrow 0$ , perfect assortativeness is achieved in the cognitive task ( $y_C = x_C$ ) but manual mismatch becomes maximal. (iv) Given a perfect fit of supply and demand ( $\rho_x = \rho_y$ ), task-biased TC has no effect on the assignment. For the poorest fit (maximal  $|\rho_x - \rho_y|$ ), the assignment changes are largest.*

As long as  $\delta > 0$ , the equilibrium assignment will satisfy PAM. Hence, all matching patterns discussed in this section can be analyzed in the proposed sorting framework (part(i)).

For most underlying distributions, the equilibrium assignment will be such that workers do not obtain their perfect job matches. This is because it is simply not feasible. However, in a multidimensional world, agents can decide in which dimension (cognitive or manual) sorting is more important. This decision depends on the technology and in particular on the relative level of worker-firm complementarities across tasks. In the task with relatively large complementarities, perfect assortativeness is strongly desired whereas in the task with weaker complementarities mismatch is tolerated. This trade-off is what I call *mismatch-assortativeness trade-off* across tasks.

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. On the other hand, these new computers perform several manual tasks, substituting those workers with manual know-how. As a result, sorting becomes more pronounced in the cognitive task at the expense of mismatch in the manual task (part(ii)). The amount of worker reallocation depends on both, the size of the shock (part(iii)) and the underlying distributions (part(iv)). How the size of the shock matters is illustrated in the figure below, 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 task-biased TC (as  $\delta$  goes to zero). The lower panels plot manual sorting, i.e.  $y_M$  as a function of  $x_M$  and  $x_C$  and has the same structure.

The slope of the straight lines indicates how strong the sorting forces are *within* tasks. On the other hand, the slope of the dotted lines is an indicator of how strong sorting forces are *between* tasks. Due to the bilinear technology, the within-force is desirable whereas the between-force is not (it reflects mismatch). Before task-biased TC, cognitive and manual tasks receive identical weights in production ( $\delta = 1$ ), hence, the left panels in both figures are identical. Going from the left to the

---

<sup>21</sup>This can be easily relaxed but I would have to rely more on simulations. Moreover, the restriction captures the empirically relevant case for the US. See below.

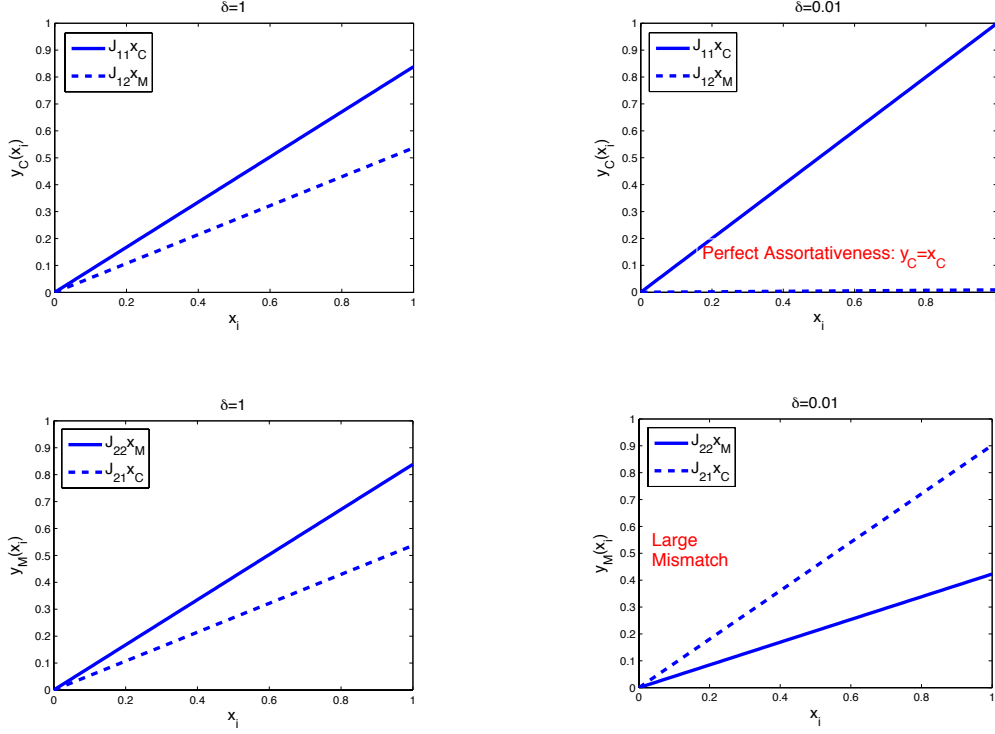


Figure 3: Effects of TBTC on Sorting in Cognitive (upper panel) and Manual Dimension (lower panel)

right panel, the relative complementarities in the cognitive task increase: The economy converges to the perfectly assortative allocation in the cognitive task. But this comes at the expense of significant misalignment between workers' skills and firms' skill needs in the manual task with the manual productivity responding even more strongly to changes in the cognitive than in the manual skill.

Besides the size of the technological shock, what matters for the sorting response to task-biased TC is the shape of the underlying distributions (part (iv)). If skill supply and demand are perfectly aligned ( $\rho_x = \rho_y$ ), task-biased TC has no effect on the assignment. This is because sorting in both tasks is perfectly assortative to start with ( $y_C = x_C$  and  $y_M = x_M$ ). Thus, the worker-firm assignment in the cognitive task cannot further improve as  $\delta$  decreases. On the other hand, the amount of resorting in response to task-biased TC is maximized when skill supply and demand differ considerably. In this situation, the initial assignment is convoluted by mismatch in *both* tasks. Hence, there is a lot to gain from improving the 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 the production technology (Proposition 7). The second source is due to the discrepancy between skill and productivity distributions or, in other words, between supply and demand ( $\rho_x \neq \rho_y$ , Proposition 5). It arises because the frictionless labor market must clear no matter how different skill and productivity distributions are.

Clearly, these assignment changes feed into wage changes, summarized by the next result.

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

(i) *Aggregate Wage Inequality: The effect on the wage variance is ambiguous.*

(ii) *Wage Curvature: Wages become more convex in cognitive but less convex in manual skills (for  $|\rho_x| < |\rho_y|$ ). For a perfect fit of supply and demand ( $\rho_x = \rho_y$ ), the changes in the curvature are smallest. In turn, for the poorest fit (i.e. maximal  $|\rho_x - \rho_y|$ ), the curvature changes are largest.*

Task-biased TC has ambiguous effects on the variance (part(i)). The wage variance is sensitive to the *level* of technology. It increases in both technology parameters  $\alpha$  and  $\beta$ . Since cognitive task-biased technological change can either be driven by an increase in the cognitive task-weight  $\alpha$  or by a decrease in the manual task-weight  $\beta$ , the overall effect depends on the relative magnitude of these two changes.

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.<sup>22</sup> Intuitively, This technology shift favors workers with high levels of cognitive skills, driving up wage inequality in the cognitive dimension. On the other hand, manual workers are adversely affected by task-biased TC. Those with lots of manual skills are hit most severely, compressing the wage inequality in this dimension. The magnitude of these effects depends on the amount of worker-job reallocation in response to task-biased TC. If there is considerable misfit in the initial worker-job matches (which is the case when  $|\rho_x - \rho_y|$  is large), then the reallocation response is strong. These allocation shifts translate into larger wage inequality movements.

To the extent that manual specialists are medium-income earners whereas cognitive specialists are high-income earners, the discussed (de)convexification fuels upper tail but compresses lower tail inequality.<sup>23</sup> This is reminiscent of *wage polarization* from the one-dimensional literature, which refers to expanding upper tail but compressing lower tail inequality, relevant for the empirics below.

Notice an important difference to the one-dimensional setting is that in my model, there exist *generalists*. Generalists have a second (i.e. cognitive) skill, which allows them to shield against the shock on manual skills. They gain over manual specialists, who additionally lose relative to low-skilled workers and cognitive specialists in the economy (see Online Appendix for a formal statement).

## 4.2 Skill-Biased Technological Change

An important advantage of the specified technology is its tractability. On the downside, it generates a non-monotonous wage schedule in skills, which would be difficult to reconcile with the data.<sup>24</sup> To make the model more suitable for empirical analysis, I augment the production technology by non-interaction skill terms and a constant, given by

$$\begin{aligned} F(x_C, x_M, y_C, y_M) &= x_C(\alpha y_C + \lambda) + x_M(\beta y_M + \eta) + f_0 \\ &= \alpha(x_C y_C + \delta x_M y_M) + \lambda(x_C + \kappa x_M) + f_0 \end{aligned} \tag{16}$$

<sup>22</sup>This is the case for  $|\rho_y| \geq |\rho_x|$ , which is the empirically relevant case for the US; see below.

<sup>23</sup>I think of manual (cognitive) specialists as workers who have manual (cognitive) but low cognitive (manual) skills.

<sup>24</sup>Under the previous technology, the wage is folded around  $(x_C, x_M)$  (0,0), e.g. workers (-1,-1) and (1,1) earn the same.

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. The assignment is unaffected by this technology but the wage becomes a *non-homogenous* quadratic form in standard normal variables,

$$w(x_C, x_M) = \frac{1}{2}\alpha(\mathbf{x}-h)'\tilde{J}(\mathbf{x}-h)+C = \alpha\left(\frac{1}{2}J_{11}x_C^2 + J_{12}x_Cx_M + \frac{1}{2}\delta J_{22}x_M^2\right) + \lambda(x_C + \kappa x_M) + w_0. \quad (17)$$

See Appendix C.2 for the derivation and the expressions  $h, \tilde{J}, C$ . The non-interaction skill terms can shift the location of the minimum wage to the left, allowing for a wage schedule that is increasing  $\forall x_C, x_M \geq \underline{x}_C, \underline{x}_M$ , where  $\underline{x}_C, \underline{x}_M$  are, for instance, the lowest observed skills in the data.<sup>25</sup> Moreover, I include a constant  $f_0$ , which then translates into a non-zero constant in the wage function  $w_0$ , guaranteeing non-negative wages to all agents in the economy.<sup>26</sup>

Technology (16) does not only give rise to a more realistic wage schedule. It also allows for a sensible definition of *skill-biased technological change* (skill-biased TC hereinafter), independently of task-biased TC that works through the complementarities in production. Consider a change in the relative manual skill weight from  $\kappa$  to  $\kappa'$  such that  $\kappa' < \kappa$ . Then,  $\kappa'$  is called *skill-biased* relative to  $\kappa$ , with the bias favoring the cognitive skill. This shift increases the productivity of the cognitive skill independent of the 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 skills. The next result highlights the effect of STBC on equilibrium outcomes.

**Proposition 9 (Skill-Biased Technological Change)** *Suppose there is cognitive skill-biased TC ( $\kappa' < \kappa$ ). Then: (i) The assignment is unaffected. (ii) The curvature of the wage function is unaffected. (iii) The effect on the wage variance is 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 (17) it is clear that it has no impact on the curvature of the wage function, which solely depends on the assignment and the task-bias parameters. Finally, similar to task-biased TC, the effect of skill-biased TC on the wage variance is ambiguous. The variance increases in both skill weights,  $\lambda$  and  $\eta$ . Depending on whether the increase in  $\lambda$  outweighs the decrease in  $\eta$ , wage dispersion increases or decreases with skill-biased TC. The next section brings the model to the data, which will allow me (a) to quantify skill-biased and task-biased TC over time and (b) to disentangle their roles in observed allocation and wage inequality shifts.

## 5 Quantitative Analysis

In this section, I first estimate the model by Maximum Likelihood (ML) and document how technology in the US has evolved over time. Then, I conduct various counterfactual experiments. The objective is 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.

<sup>25</sup>Notice that including additional non-interaction productivity terms in the technology would not affect wages.

<sup>26</sup>With the previous technology, wages were always positive (see Online Appendix). However, when including non-interaction terms, wages can become negative. Hence, the inclusion of the constant.

## 5.1 The Data

I use the National Longitudinal Survey of Youth 1979 (NLSY) as the main data source. The NLSY follows a (single) cohort since 1979 that was interviewed every year until 1994 and since then biennially. The reason for using the NLSY is that it contains detailed information on the respondents' occupations, training and degrees, which I will use to construct the skill supply distribution. I supplement the NLSY by the O\*NET data to learn about occupational skill requirements. This data will be crucial for constructing the skill demand distribution, where I interpret *occupations* as the empirical counterpart of the model's firms.<sup>27</sup> The analysis of this paper covers the period 1992-2000.<sup>28</sup> I restrict the sample to employed male and female workers in non-military occupations who work more than twenty hours per week and forty weeks per year. For the analysis, I consider hourly wages, computed as the yearly gross labor income divided by yearly hours worked and adjusted by the CPI. Additionally, my analysis requires measures of workers' cognitive and manual skills ( $x_C, x_M$ ) as well as of occupations' cognitive and manual skill requirements ( $y_C, y_M$ ).

To construct these bivariate distributions, I heavily rely on the O\*NET data, which 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 task measures for each occupation. They indicate the *level* of skills needed to perform the manual and cognitive task, which I interpret as the ( $y_C, y_M$ )-bundle from my model (see Table 3 in Appendix E for examples).<sup>29</sup> I then merge these scores into the occupations of the employed workers in the NLSY, which yields the bivariate skill demand distribution. Constructing the bivariate skill distribution is involved. Data on manual skills is not readily available. Moreover, the literature provides little guidance on this issue.<sup>30</sup> To impute the agents' manual and cognitive skills, I use information on their college degrees, apprenticeship and vocational degrees, degrees of government programs and training on-the-job paid for by the firm, provided by the NLSY.<sup>31</sup> From this information, I can proxy a manual and cognitive skill for each agent (see Appendix E.1 for details). After data cleaning and sample restrictions, I am left with around 2700 yearly observations.

The Online Appendix provides summary statistics of the bivariate skill and productivity distributions in 1992. In order to align the data with the model, I transform the empirical skill and productivity distributions into *Gaussian copulas*, which takes out the marginal characteristics (means and variances) and leaves the correlation as the only distributional parameter (see Online Appendix). The correlations between the transformed variables are plotted below (standard errors in parentheses).

---

<sup>27</sup>The O\*NET is the U.S. Department of Labor Occupational Characteristics Database.

<sup>28</sup>This period is chosen for two reasons: First, there is a consensus in the literature that task-biased TC started around the beginning of the 1990s when computers and advanced technology became widely spread. I choose the starting year 1992 because this is when I begin observing task-biased TC in the data, i.e. a technological shift away from the manual and towards the cognitive task inputs in production. Second, the years beyond 2000 are excluded because NLSY occupations are recoded in 2002, which complicates the measurement of sorting.

<sup>29</sup>This data as well as the crosswalk linking the O\*NET occupational codes to the NLSY occupational codes come from Sanders [2012]. Yamaguchi [2012] uses a similar approach to classify manual and cognitive occupational inputs.

<sup>30</sup>Yamaguchi [2012] and Sanders [2012] 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.

<sup>31</sup>I only consider training paid by the firm because it is presumably related to the occupation performed by the worker.

Manual and cognitive skills are negatively correlated, indicating that a worker with high cognitive skills has little manual dexterity and vice versa. Occupations' skill requirements are more strongly negatively correlated than skills. The interpretation is that jobs in the US demand workers with a higher degree of specialization than the available workers can offer. In light of the model, it is crucial that these empirical correlations are not equal as the model predicts a non-trivial effect of technological change on sorting and wages, which will be analyzed below.

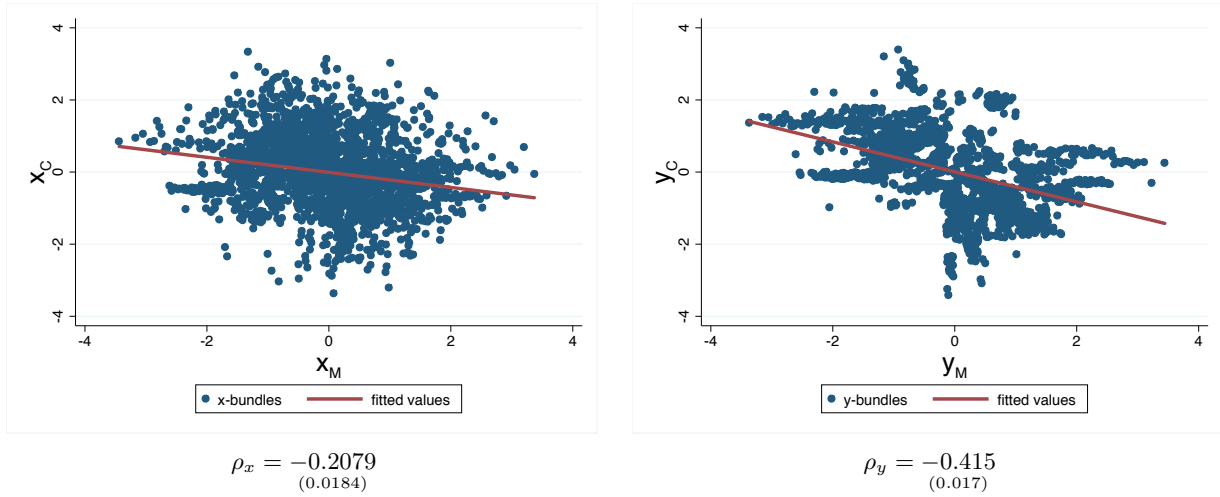


Figure 4: US Skill and Productivity Correlations in 1992

## 5.2 Estimation

I estimate the model by Maximum Likelihood (ML). The closed form solution is particularly useful for this purpose since it allows me to specify an exact expression for the likelihood function. Denote the parameter vector by  $\theta = ((J_{11}, J_{12}, J_{21}, J_{22}), (\alpha, \beta, \lambda, \eta, w_0), (s, t, u))$ , which is to be estimated. The first set of parameters corresponds to the coefficients of the assignment functions (i.e. the Jacobian of the matching function), the second set are technology parameters, the last set relates to measurement errors of the wage and assignment, respectively. The data vector is given by  $z = (z_1, \dots, z_n)$  where  $\forall i = 1, \dots, n, z_i = (w_i, y_{Ci}, y_{Mi}, x_{Ci}, x_{Mi})$ . The log-likelihood function for this model is given by:

$$\ln L(\theta|z) = - \sum_{i=1}^n \frac{(w_i - (\frac{1}{2}\alpha J_{11}x_{Ci}^2 + \alpha J_{12}x_{Ci}x_{Mi} + \frac{1}{2}\beta J_{22}x_{Mi}^2 + \lambda x_{Ci} + \eta x_{Mi} + w_0))^2}{2s^2} - \sum_{i=1}^n \frac{(y_{Ci} - (J_{11}x_{Ci} + J_{12}x_{Mi}))^2}{2t^2} - \sum_{i=1}^n \frac{(y_{Mi} - (J_{21}x_{Ci} + J_{22}x_{Mi}))^2}{2u^2} - n \ln(stu) - \frac{3n}{2} \ln 2\pi \quad (18)$$

See Appendix E.2 for details. Notice that another advantage of this model is that all parameters are identified.<sup>32</sup> I estimate the model year by year. Appendix E.2 reports the detailed estimation results.

<sup>32</sup>My model circumvents the non-identification of similar linear-quadratic Gaussian models arising due to collinearity (pointed out by [Brown and Rosen \[1982\]](#) and [Ekeland et al. \[2004\]](#)). There, the identification problem stems from an

### 5.3 Technological Change in the US

Identifying unobserved worker-job/firm complementarities from observed equilibrium outcomes has been of independent interest and the focus of a growing literature on the identification of sorting.<sup>33</sup> 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 for the years 1992 and 2000. The estimation results suggest that the production technology features complementarities between worker and job attributes in both tasks ( $\alpha$  and  $\beta$  are positive; see Appendix E.2 for the results of the remaining years). Moreover, the 90s were characterized by task-biased TC in favor of the cognitive task: Complementarities between cognitive worker and job attributes have gone up by 15% whereas complementarities in manual inputs have decreased by 41%. The relative manual complementarities,  $\delta = \frac{\beta}{\alpha}$ , dropped from 0.55 to 0.29 – a decline of 47%.

	$\alpha$	$\beta$	$\lambda$	$\eta$	$w_0$
1992	2.7291 (0.6090)	1.5009 (0.7244)	2.7962 (0.1668)	0.2079 (0.1677)	15.1680 (0.2380)
2000	3.1358 (1.0472)	0.8954 (0.8028)	4.7352 (0.2641)	-0.1773 (0.2632)	18.4752 (0.3792)

Standard errors in parenthesis.

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 the technology, indicated by the skill weights  $\eta$  and  $\lambda$ . Over the 90s, the US was characterized by a strong cognitive skill-biased TC. The cognitive skill weight  $\lambda$  increased sharply (+68%) whereas the manual weight  $\eta$  decreased (however,  $\eta$  is statistically insignificant). In sum, these estimates suggest that during the 90s, 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  $w_0$ ), which had an impact on all workers independent of their skills.

Notice that apart from technological change, there was a change in the distributions: Workers were less specialized in 2000 compared to 1992 ( $\rho_x = -0.2079$  in 1992 and  $\rho_x = -0.05$  in 2000, see Online Appendix). On the other hand, the change in the skill demand was negligible.

### 5.4 The Role of Technological Change in US Wage Inequality Shifts

Observed wage inequality shifts in the data can have many reasons. The advantage of estimating a structural model is that the effects of various sources can be disentangled. This section conducts

additional quadratic term in the production technology. More generally, my model avoids such collinearity problem because the curvature of  $w(x)$  in  $x$  is not the same as the curvature of the technology  $F(x, y)$  in  $x$ .

<sup>33</sup>See [Abowd et al. \[1999\]](#) and also [Eeckhout and Kircher \[2011\]](#).



counterfactual exercises to decompose the impact of task-biased TC, skill-biased TC and changes in the 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 the skill-bias parameters  $\lambda$  and  $\eta$  as well as distributional parameters  $\rho_x$  and  $\rho_y$  at their 1992-level and only feed the estimated changes in the 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 slow-down in lower tail wage inequality and a boost in upper-tail inequality. Figure 5 plots hourly wages by wage percentile for 1992 and 2000 (solid and dashed line, respectively), illustrating that inequality overproportionally increased in the upper part of the distribution with little action in the lower part.

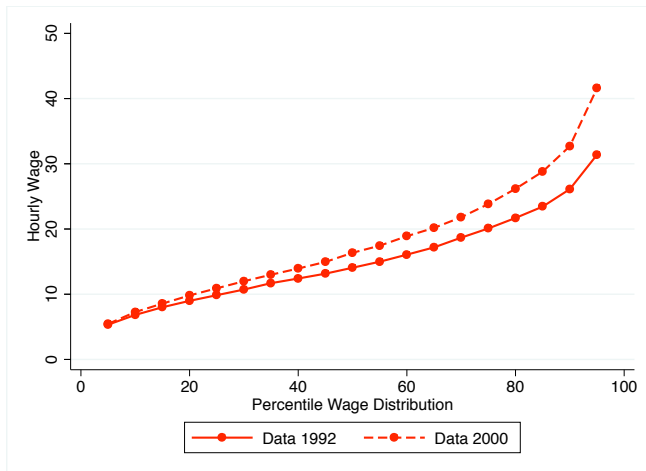


Figure 5: Wage Polarization in the US (Data)

What may have caused this specific change in the wage distribution? Panels (a) and (b) in Figure 6 analyze whether wage polarization can possibly be triggered by the estimated technology shifts in task and skill-bias. Panel (a) shows that task-biased TC matches fairly well the increase in upper tail wage inequality and exactly matches the halt in lower tail inequality. In turn, skill-biased TC can only match the expanding upper tail inequality. It fails to account for the stagnating inequality in the lower part of the distribution, overpredicting the increase in lower tail wage inequality.

The model offers an explanation 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. This fuels wage inequality in the cognitive but compresses inequality in the manual dimension. Polarization 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 an overproportionate increase in upper tail inequality.

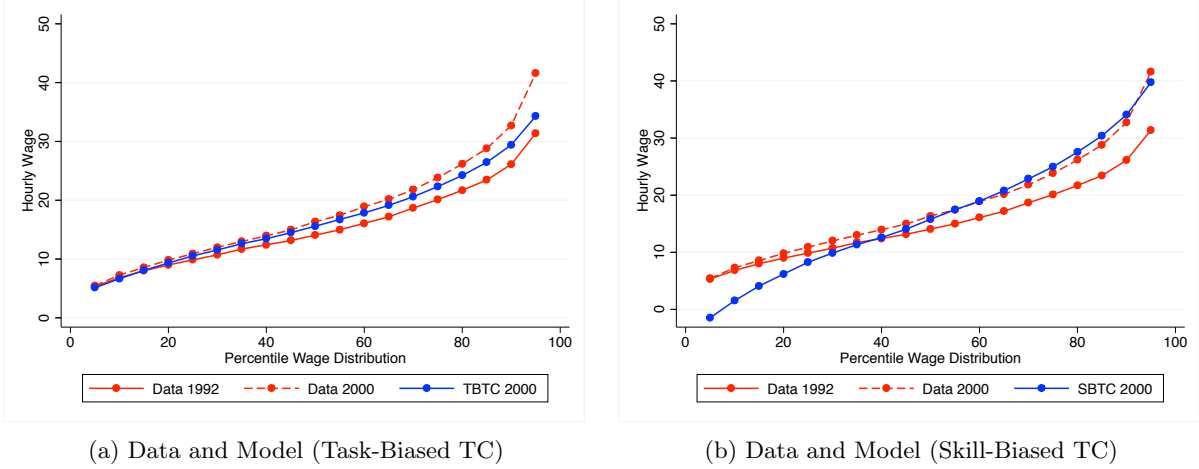


Figure 6: Wage Polarization 1992-2000: Data, Task-Biased TC and Skill-Biased TC

To see this, I plot the c.d.f.'s of the empirical wage distributions for low-skilled workers, manual specialists, generalists and cognitive specialists in 1992 and 2000 (Figure 7).<sup>34</sup> Cognitive specialists and generalists form the group of high-income earners in the US economy. In 1992, the wage 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. Strikingly, in the course of the 90s, there is no increase in lower tail inequality (the difference between low-skilled and manual wages remains nearly unchanged). In turn, the first order stochastic dominance of the wage distributions of cognitive specialists and generalists over the distributions of low-skilled and manual specialists has become more pronounced over time (compare panels (a) and (b)). This implies that generalists and cognitive specialists gain significantly relative to low and medium income earners, fueling upper tail wage inequality.<sup>35</sup>

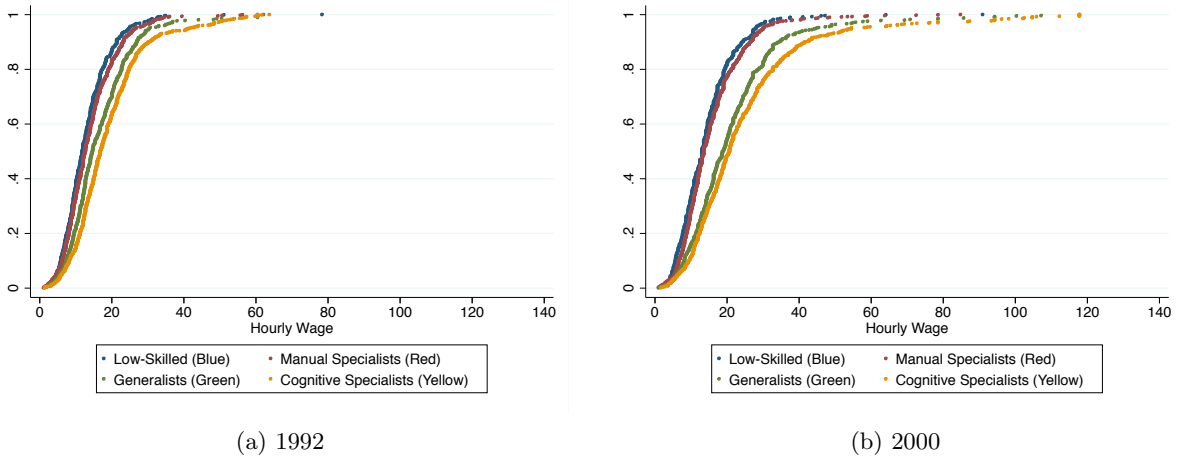


Figure 7: Wage Distributions by Worker Group (Data)

<sup>34</sup>Low-skilled are defined as  $x_C < E(x_C)$ ,  $x_M < E(x_M)$ , manual specialists as  $x_C < E(x_C)$ ,  $x_M > E(x_M)$  etc.

<sup>35</sup>Figure 10 in Appendix E.2 makes the same point with wage densities.

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. Second, there is an indirect effect through re-sorting of workers to jobs. Due to task-biased TC, sorting along the cognitive dimension should improve whereas sorting along the manual dimension is expected to deteriorate. To evaluate these predictions, recall that the sorting patterns are fully captured by the assignment coefficients of the matching function, given by:

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

$J_\mu$  is estimated via ML. If it is a P-matrix (i.e. with positive diagonal elements and positive determinant) then sorting satisfies PAM. In Figure 8, I plot the assignment estimates for the year 1992 in blue (left panels). For the year 2000, I only plot the estimates whose change was statistically significant compared to 1992 (in red, right panels). See Appendix E.2 for the estimates. The structure is as in Figure 3: The slope of the solid lines indicates assortativeness (diagonal elements of  $J_\mu$ ,  $J_{11}$  and  $J_{22}$ ) whereas the steepness of the dashed lines indicates mismatch (off-diagonal elements of  $J_\mu$ ,  $J_{12}$  and  $J_{21}$ ).

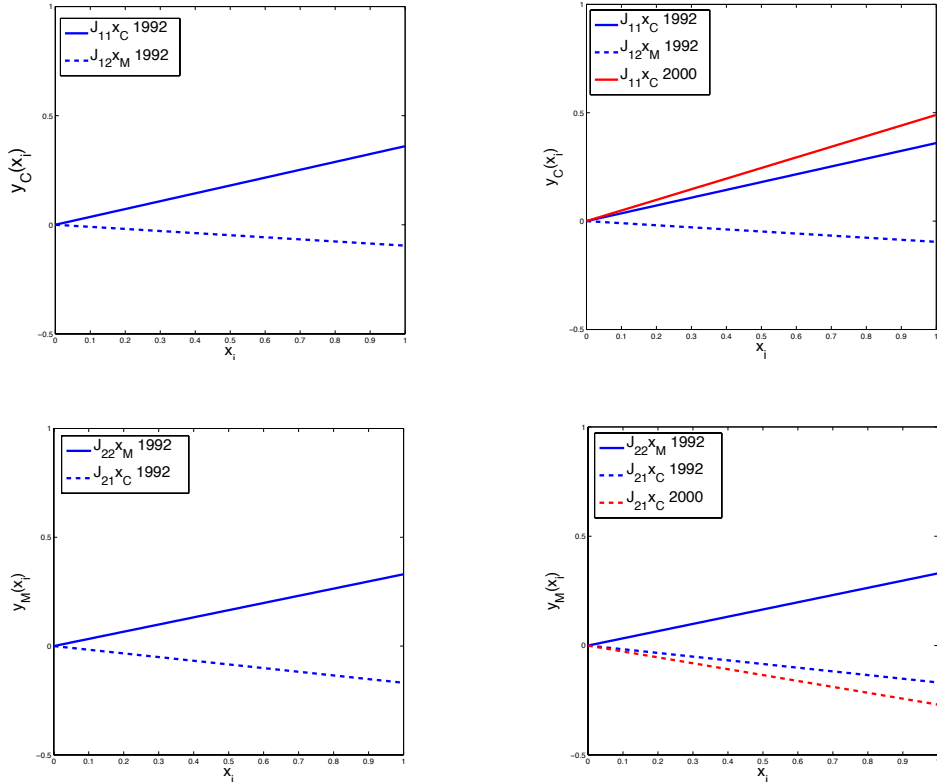


Figure 8: Maximum Likelihood Assignment Estimates 1992 and 2000: Cognitive Dimension (upper panel), Manual Dimension (lower panel)

Both in 1992 and 2000, sorting satisfies PAM with a positive relation between skills and skill requirements in both tasks (given by positively sloped solid lines). This is in line with the estimated worker-job *complementarities* (i.e.  $\alpha, \beta > 0$ ).<sup>36</sup> Moreover, the sorting *changes* over time are consistent with task-biased TC: assortativeness in the cognitive task significantly increased between 1992 and 2000, indicated by a steeper red solid line in the upper right panel. In turn, there is a statistically significant deterioration in the manual fit, indicated by a steeper red dashed line (lower right panel). Quantitatively small assignment changes were expected: Given that skill supply and demand (i.e.  $\rho_x$  and  $\rho_y$ ) are fairly well aligned in 1992, my model predicts minor effects of task-biased TC on sorting.

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 reallocation. Since cognitive (but not manual) workers are concentrated in the upper part of the wage distribution, these wage movements trigger wage polarization. In turn, for skill-biased TC neither of the two effects is at work. Thus, skill-biased TC has no effect on the curvature of the wage schedule but simply shifts it. As a result, skill-biased TC triggers an increase in wage inequality across the *whole* wage distribution. For more analysis, see Figure 11 in Appendix E.2.

#### 5.4.2 Wage Dispersion

The proposed model – despite being frictionless – does a decent job in matching the increase in US *wage dispersion* during the 90s. It generates a substantial increase of 126%, compared to an increase of 145% in the data. Moreover, the model matches well the *shape* of this shift. Only in terms of the *level* of the variance, the model is off (the data variance in 1992 is 6.7 times higher than the model variance). A possible reason is that the model is frictionless whereas search frictions are believed to play an important role in wage dispersion (e.g. Uren and Virag [2011]). Moreover, in the data not only skills but many other factors impact wage dispersion, which are not part of my model. To be able to better compare the variance change in data and model, I normalize the model variance so that it has same level as the data variance in 1992 (Figure 9).

Notice that the driving force in the wage inequality increase is the *within*-variance of the manual workers,  $E[Var(w|x_B)]$ , as opposed to their *between*-variance,  $Var(E[w|x_B])$ . This multidimensional model offers a natural way to think about these concepts since, for instance, every group of workers with similar manual skills has a whole distribution of wages due to differences in their cognitive skills. Technological change in favor of the cognitive inputs exacerbates this within-group wage dispersion. In contrast, between-wage dispersion of manual types contributes little to the overall variance and its shift (indicating that the driving force of wage inequality is the cognitive and not the manual skill).<sup>37</sup>

Table 2 reports a decomposition for the shift in the wage variance. The results suggest that skill-biased TC was the driving force behind the increase in wage dispersion, generating a boost of 119%. Compared to skill-biased TC, the role of task-biased TC is moderate, only achieving an increase

<sup>36</sup>Moreover, the solid lines are steeper than the dashed lines, fulfilling the requirement on the determinant of  $J_\mu$ .

<sup>37</sup>The within and between variance predictions of the model can be computed since closed forms are available. Regarding the data, I first categorize workers into thirty bins depending on their manual skill, and then compute the wage variance within and across bins. The results are robust to using more or less than thirty bins.

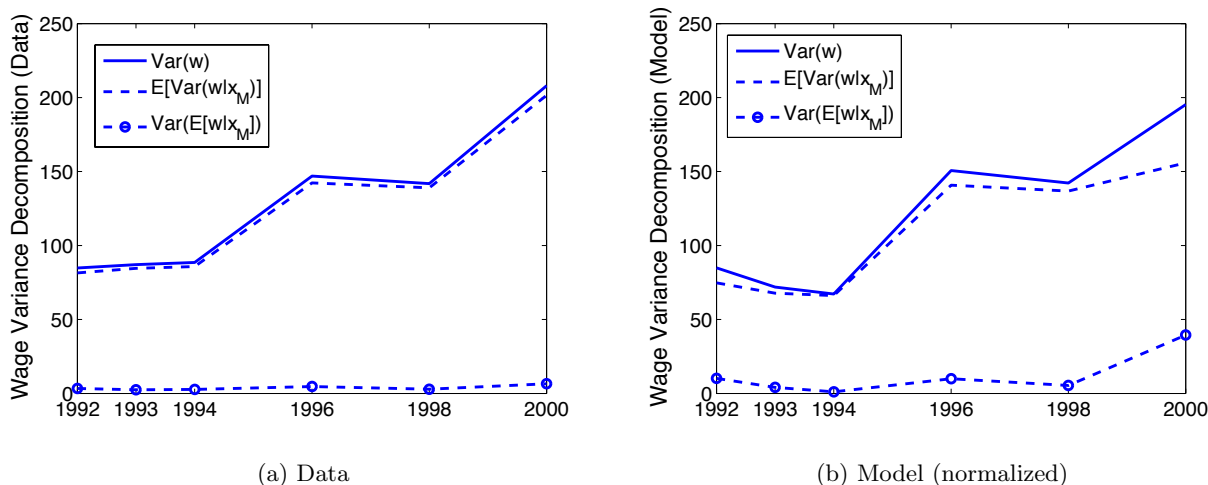


Figure 9: US Wage Dispersion over 1992-2000 (Data and Model)

of 7% over this period. The model offers two reasons why task-biased TC played a minor role for wage dispersion: First, over the 90s, the increase in cognitive task-weight  $\alpha$  is accompanied but a strong decrease in manual task weight  $\beta$ . Since what matters for the wage variance is the *level* of the technology parameters, the net effect of task-biased TC on wage dispersion is small. To the contrary, the strong increase in cognitive skill bias ( $\lambda$ ) is not counteracted by a drop in  $\eta$  because the manual skill weight was low to start with. The second reason for the minor impact of task-biased TC is that sorting shifts are quantitatively small (see above). Stronger re-sorting would have fueled more inequality.

	Data	Model	Task-Bias ( $\alpha, \beta$ )	Skill-Bias ( $\eta, \lambda$ )	Distributions ( $\rho_x, \rho_y$ )	Trend ( $w_0$ )
$\Delta Var(w)$	+ 145%	+126%	+7%	+119%	-4%	-

Table 2: Change in Wage Variance over 1992-2000 (Data versus Model)

Besides technological progress, distributions also changed during the 90s. There was a shift in skill supply, with workers becoming less specialized, but skill demand remained constant (see Online Appendix). At odds with the observed increase in inequality, the change in the skill distribution had a negative effect on the wage variance. Finally, the TFP shifter does not affect wage dispersion because it enters as a constant in the wage function.

In sum, this exercise shows that technological change rather than changes in distributions mattered for US wage inequality shifts.<sup>38</sup> Skill-biased TC accounts for a significant part of the increase in wage dispersion. On the other hand, task-biased TC played a critical role for wage polarization, being particularly important for stagnating lower tail wage inequality.

<sup>38</sup>Notice that similar to the variance exercise, the change in distributions had little effect on the change in the curvature of the wage function, which is why it is not included here.

## 5.5 Comparison to the One-Dimensional Assignment Model

In what sense does the multidimensional model provide a richer understanding of the data than a comparable one-dimensional model? In order to address this question, I specify the one-dimensional analogue of my model and estimate it using cognitive skills and skill requirements only. I interpret cognitive skills as a proxy for years of schooling, which is often used as the single worker characteristics in one-dimensional settings.<sup>39</sup>

To ensure comparability of the two models, I assume standard normal distributions  $x_C, y_C \sim N(0, 1)$  and the technology  $F(x_C, y_C) = \alpha x_C y_C + \lambda x_C + f_0$ . It is immediate that the wage is given by

$$w(x_C) = \alpha \frac{x_C^2}{2} + \lambda x_C + w_0 \quad (19)$$

where  $w_0$  is the constant of integration. I estimate the parameters  $\alpha, \lambda, w_0$  by OLS, using (19): During the 90s,  $\alpha$  increased by 88% and  $\lambda$  by 74% (see the Online Appendix for details), suggesting that technological change favored those workers with high cognitive skills.

The one-dimensional model captures well the *convexification* in cognitive returns, indicated by an increase in  $\alpha$ . However, it misses that manual returns have become concave (see previous section). As a result, this model overpredicts the change in wage dispersion during 1992-2000. It predicts an increase of 220%, compared to an observed increase of 145%.

Moreover, the one-dimensional model misses that not all workers with manual skills suffer from cognitive-biased technological change. Looking at the data through the two-dimensional lense suggests that *generalists* (who hold above average skills in both dimensions) experienced a substantial real wage increase of 27% over the 90s. Their second skill allows them to buffer against the shock on manual skills. Notice that generalists form a sizeable group, forming almost one fifth of the US workforce in 1992 (see Online Appendix). This suggests that the distinction between generalists and specialists is important. Yet, it falls short of the one-dimensional model.

Finally, this model cannot account for reallocation of workers to jobs in response to technological change unless the technology shifts so drastically that negative instead of positive assortative matching becomes optimal.<sup>40</sup> For changing (but still positive)  $\alpha$ , the model predicts no shift in assignment. Moreover, since there is only one skill, the one-dimensional model entirely misses the assortativeness-mismatch trade-off across skills, which was present in the US during the 90s.

## 6 Literature Review

This work primarily relates to two strands of literature: As far as the theory is concerned, to multidimensional matching under transferable utility (including hedonic models and optimal transport); as far as the application is concerned, to the literature on task-biased technological change. I will discuss those papers that are most relevant to my research.<sup>41</sup>

---

<sup>39</sup>This is justified since I construct skills from educational attainment (i.e. degrees) and training data.

<sup>40</sup>This would be the case if  $\alpha$  switches from positive to negative, which is according to the estimates not the case.

<sup>41</sup>I will not discuss papers with non-transferable utility because there is little relation.

MULTIDIMENSIONAL 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 that model because the wage function and production technology have the same curvature in  $x$ . To address this collinearity issue, the authors propose a change of the environment, for instance, by considering Gaussian mixtures. In turn, my model circumvents this problem by specifying a production technology without quadratic loss terms. Additionally, to make my model suitable for empirical analysis I include non-interaction skill terms in the technology such that marginal wages can be positive over the whole observed skill support. [Olkin and Pukelsheim \[1982\]](#) solve a related Gaussian example but in a symmetric setting (i.e.  $\delta = 1$ ). [Bojilov and Galichon \[2013\]](#) extend the quadratic-Gaussian setting to include unobserved heterogeneity.

My contribution to this literature is as follows: First, I develop a framework for multidimensional sorting that extends the unidimensional notion of (positive) assortative matching (PAM). Second, using this notion of PAM, I develop a technique for deriving the equilibrium in closed form, which can be used not only for the quadratic-Gaussian model (as I illustrate) but also in other settings (Appendix D.0.2). Third, I use PAM to characterize equilibrium sorting. Moreover, I study a new application (i.e. technological change) in this setting. Last, I make the model amenable to empirical analysis and bring it to the data.

This paper also relates to the literature on multidimensional 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 multidimensional (un)observed heterogeneity under the assumption that there is no interaction between unobservable characteristics of partners (separability assumption).<sup>42</sup> More recently, [Galichon and Salanié \[2010\]](#) study optimal matching in a model with multidimensional (un)observed characteristics. Under the same separability assumption, the authors show that optimal matching on observable characteristics is *non-pure*. In related work, [Dupuy and Galichon \[2012\]](#) extend their set-up to continuous types.

These studies differ from my research in terms of objective and modeling choices. [Choo and Siow \[2006\]](#) estimate the gains from marriage, i.e. their focus is empirical. In turn, [Galichon and Salanié \[2010\]](#) and [Dupuy and Galichon \[2012\]](#) develop techniques to estimate the complementarities in the surplus function from the observed matches. They pursue this objective without providing a closed form. Conversely, my paper aims at developing a multidimensional sorting framework that allows for closed form characterization and comparative statics. In the above-mentioned papers, modeling devices are (un)observed heterogeneity and extreme value distributions of the unobserved traits. I rely on observed heterogeneity and Gaussian copulas. Notice, however, that there is an important conclusion common to the papers by [Galichon and Salanié \[2010\]](#), [Dupuy and Galichon \[2012\]](#) and my own: With multidimensional matching, there is a trade-off between matching along different

---

<sup>42</sup>[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. \[2012\]](#) also provide a closed form of a multidimensional 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.

characteristics which depends on the complementarity weights in the surplus function.

McCann et al. [2012] develop a model of marriage, educational and occupational choices when agents have both cognitive and social skills. This discussion focusses on their marriage market.<sup>43</sup> Under the assumption of complete overlap in distributions (i.e. equal male-female sex ratio by type) and their specified technology they prove that matching is positively assortative in both dimensions. When looking at this result through the lens of my model, it can be shown that it is captured by Proposition 11 (b) in Appendix D.0.2. Similarly, this model would capture sorting results from the environment specified in Eeckhout and Kircher [2012], if they extended their model to a fully bi-dimensional setting where firms and workers not only sort on the quality but also on the quantity dimension (Proposition 11 (a), Appendix D.0.2). This suggests that the developed sorting framework is useful for deriving closed forms beyond the quadratic-Gaussian case.

Finally, this paper is related to the literature on optimal transport. In non-technical terms, the optimal transport problem is to find a measure-preserving map that carries one distribution into another at minimal cost, using linear programming.<sup>44</sup> A tight link has been established between the following two formulations of the assignment problem: a hedonic pricing problem with transferable utility (like the problem in this paper) and an optimal transport problem. Shapley and Shubik [1971] show this equivalence in a discrete and Gretsky et al. [1992] in a continuous setting.<sup>45</sup>

Different from Gretsky et al. [1992], in the multidimensional assignment problems of Chiappori et al. [2010] and Ekeland [2010], sellers can also *choose* the characteristics of the good they sell.<sup>46</sup> Apart from providing existence and uniqueness results, both papers establish *purity* of the assignment: Their sufficient condition for purity is the *twist condition*, which states that  $D_x F(\mathbf{x}, \mathbf{y})$  is injective with respect to  $\mathbf{y}$ . Notice that the  $P$ -matrix property of  $D_{xy}^2 F(\mathbf{x}, \mathbf{y})$  from my paper is sufficient for the twist condition to hold. Since  $D_{xy}^2 F(\mathbf{x}, \mathbf{y})$  is the Jacobian of  $D_x F(\mathbf{x}, \mathbf{y})$ , the  $P$ -matrix property ensures that  $D_x F(\mathbf{x}, \mathbf{y})$  is injective (by Gale and Nikaido [1965]). While this literature has developed powerful general tools to study multidimensional matching problems, it provides little guidance on how to solve them explicitly. This is what my paper seeks to address.

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*.<sup>47</sup>

Instead of implicitly assuming that manual skills are only used by medium-skilled workers, I make

---

<sup>43</sup>In their paper, the marriage market is the only one, in which choices are based on two characteristics on *both* sides of the market and, hence, where a comparison to my set-up makes sense.

<sup>44</sup>Optimal transport has a long tradition in mathematical theory. See Villani [2009] for a recent reference book.

<sup>45</sup>Additionally, both show the equivalence to a third formulation, namely the market game. Notice that the provided examples in Gretsky et al. [1992] are restricted to one-dimensional types. See also, Dizdar and Moldovanu [2012] for a recent paper on an application of optimal transport techniques to multidimensional matching.

<sup>46</sup>In that setting, Chiappori et al. [2010] establish a similar equivalence, namely between the hedonic pricing, the stable matching and the optimal transport problem.

<sup>47</sup>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 a frictionless and competitive labor market.



the assumption that every worker has both skills, yet in different proportions. This makes it possible to distinguish between generalists and different types of specialists, thereby capturing that generalists can shield against the adverse shock on manual inputs. Moreover, 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 channel of how this technology shift affects wage inequality and polarization: Task-biased TC endogenously changes the allocation of workers to jobs, improving *the fit* of worker-firm pairs along the cognitive task relative to the manual task. It is noteworthy that this *assortativeness-mismatch trade-off* across tasks takes place despite pairwise matching and without violating positive assortative matching.<sup>48</sup> Finally, what my paper adds to this literature is a *unified* framework of task-biased TC and the more standard skill-biased TC, which allows me to distinguish between their effects both theoretically and empirically.

There is plenty of empirical evidence on labor market polarization in developed countries but little structural analysis into the causes of this phenomenon. 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). Also using NLSY data, he finds that the US has been characterized by a relative decrease of the medium-skill price and wage polarization.

## 7 Concluding Remarks

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 the worker-job assignment, wages and inequality? This is a *multidimensional assignment* problem where workers with different bundles of manual and cognitive skills sort into jobs that require different combinations of these skills. To make it tractable, this paper develops a theoretical framework for multidimensional sorting that extends the unidimensional notion of assortative matching. I derive the equilibrium allocation as well as the equilibrium wages in closed form. I then analyze the impact on these equilibrium outcomes as cognitive (as opposed to manual) inputs become more important 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 during the 90s. The empirical analysis reveals that technological change was strongly biased towards cognitive inputs. Counterfactual exercises suggest that this technology shift 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 multidimensional heterogeneity, not only on the labor but also on the marriage or education market. To broaden its applicability even further, it would be important to extend this framework to settings with search frictions (see the Online Appendix for first insights on sorting conditions in such an environment) and to settings where the two sides of the market have a different number of characteristics, preventing pure matching. These are challenging problems that are part of my ongoing research agenda.

---

<sup>48</sup>Contrary to [Costinot and Vogel \[2010\]](#) and [Acemoglu and Autor \[2011\]](#), there is no intensive margin here.

## A Proofs General Model (Section 2)

### A.1 The Equilibrium Assignment

In order to prove Proposition 1, the following Lemma and Corollary are useful. Notice that the proofs will be given for N-dimensional heterogeneity where firms are characterized by  $\mathbf{y} = (y_1, \dots, y_N) \in Y \subset \mathbb{R}_+^N$  (with abs. continuous c.d.f.  $G(\mathbf{y})$ ) and workers are characterized by  $\mathbf{x} = (x_1, \dots, x_N) \in X \subset \mathbb{R}_+^N$  (with abs. continuous c.d.f.  $H(\mathbf{x})$ ).

**Lemma 1 (P-Matrix Property)** *If  $D_{xy}^2 F(\mathbf{x}, \mathbf{y})$  is a diagonal P-matrix ( $P^-$ -matrix), then  $J_\mu(\mathbf{x}) \equiv D_x \mathbf{y}^*$  is a P-matrix ( $P^-$ -matrix).*

Throughout the proof I will make the following assumption:

**Assumption 1**  *$D_{xy}^2 F$  is a diagonal P-matrix.*

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

1. *The Hessian evaluated at the equilibrium assignment, given by  $H^* = D_{xx}^2 F(\mathbf{x}, \mathbf{y}^*) - D_{xx}^2 w(x)$ , is negative semi-definite.* These are the necessary second order conditions for optimality.

2.  *$\text{Det}(H^*) > 0$ .* Differentiate the first order conditions, evaluated at the optimal assignment  $\mathbf{y}^* = \mu(\mathbf{x})$ , with respect to the skill vector  $\mathbf{x}$ , which gives

$$H^* = D_{xx}^2 F(\mathbf{x}, \mathbf{y}^*) - D_{xx}^2 w(\mathbf{x}) = -(D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*))(D_x \mathbf{y}^*) \quad (20)$$

where  $D_x \mathbf{y}^*$  is the Jacobian of the matching function. Since  $D_{xy}^2 F$  is a P-matrix everywhere (and, hence, also along the equilibrium allocation  $\mathbf{y}^*$ ), it is non-singular and hence the inverse  $(D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*))^{-1}$  exists. From (20), it is given by

$$(D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*))^{-1} = -(D_x \mathbf{y}^*)(D_{xx}^2 F(\mathbf{x}, \mathbf{y}^*) - D_{xx}^2 w(x))^{-1}. \quad (21)$$

It follows that  $(D_{xx}^2 F(\mathbf{x}, \mathbf{y}^*) - D_{xx}^2 w(x))^{-1}$  exists, and thus  $\text{Det}(H^*) \neq 0$ . Then, by Step 1, it must be  $\text{Det}(H^*) > 0$ . Hence,  $-H^*$  is a P-matrix.

3. *If  $D_x \mathbf{y}^*$  is sign-symmetric then it is a P-matrix.* Suppose that  $D_x \mathbf{y}^*$  is sign symmetric, i.e.  $\frac{\partial y_i^*}{\partial x_j} \frac{\partial y_j^*}{\partial x_i} > 0, \forall i, j \in \{1, 2, \dots, N\}, i \neq j$ . 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  $D_x \mathbf{y}^*$  has positive eigenvalues, i.e. is stable. From (20)  $-H^* = (D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*))(D_x \mathbf{y}^*)$ , which has all positive eigenvalues (Step 2). Denote  $M = D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*)$ ,  $J = D_x \mathbf{y}^*$ . Denote the eigenvalues of  $-H^*$  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  $\det(R - \lambda^H M^{-1}) = 0$ , 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.<sup>49</sup> The (real) generalized eigenvalues can be computed as  $\lambda_i^H = \frac{S_{ii}}{T_{ii}}$ . Notice that  $T_{ii} > 0 \quad \forall i$  because  $M$  is a diagonal P-matrix, which implies stability (i.e. positive real part of eigenvalues) and  $\lambda^{M^{-1}} = \frac{1}{\lambda^M}$ . For  $\lambda_i^H > 0$ , it must be that  $S_{ii} > 0$ , i.e.  $J = D_x \mathbf{y}^*$  has positive eigenvalues. It follows that  $D_x \mathbf{y}^*$  has positive principal minors and, thus, is a P-matrix.

4.  $D_x \mathbf{y}^*$  is sign-symmetric. To see this, notice that by symmetry of the Hessian and  $F_{x_i y_j} = 0, i, j \in \{1, 2, \dots, N\}, i \neq j$ ,

$$\begin{aligned} H_{ij}^* &= H_{ji}^* \\ \Leftrightarrow F_{x_i y_i} \frac{\partial y_i^*}{\partial x_j} &= F_{x_j y_j} \frac{\partial y_j^*}{\partial x_i} \quad \forall i, j \in \{1, 2, \dots, N\}, i \neq j, \end{aligned} \quad (22)$$

and hence  $D_x \mathbf{y}^*$  is sign-symmetric, i.e.  $\frac{\partial y_i^*}{\partial x_j} \frac{\partial y_j^*}{\partial x_i} > 0 \quad \forall i, j \in \{1, 2, \dots, N\}, i \neq j$ . Moreover,  $D_x \mathbf{y}^*$  is stable (see Step 3). A sign-symmetric and stable matrix is a P-matrix (Theorem 2.6. in [Hershkowitz and Keller \[2005\]](#)), which proves the result. ■

### Corollary 1 (Assortativeness and Local Maximum) .

If  $D_{xy}^2 F(\mathbf{x}, \mathbf{y})$  is a diagonal P-matrix ( $P^{(-)}$ -matrix), then (i) the assignment  $\mu$  satisfies PAM (NAM) and (ii) it is a local maximum.

#### Proof.

- (i) Assortativeness: Follows from the definition of assortativeness (Definition 1) and Lemma 1.
- (ii) Local Maximum: If the Jacobian of a function is a P-matrix (or a  $P^{(-)}$ -matrix), then the function is injective (one-to-one) on any rectangular region of  $\mathbb{R}^n$  ([Gale and Nikaido \[1965\]](#), Theorem 4). It follows from Lemma 1 and the Gale-Nikaido theorem that the solution to the firm's problem a *local* maximum. ■

#### Proof of Proposition 1.

- (i) Assortativeness: Follows directly from Corollary 1.
- (ii) Global Maximum: It will be shown that the solution to the firm's problem is a *global* maximum. I proceed by contradiction. Consider a firm  $\mathbf{y}$  which optimally chooses worker  $\mathbf{x}$ , i.e.  $\mathbf{y} = \mu(\mathbf{x})$ .<sup>50</sup> Consider another firm  $\mathbf{y}'$ ,  $\mathbf{y}' \neq \mathbf{y}$ , for which worker  $\mathbf{x}'$ ,  $\mathbf{x} \neq \mathbf{x}'$ , is an optimal choice, and hence  $\mathbf{y}' = \mu(\mathbf{x}')$ . Let  $\mathbf{y} = \mu(\mathbf{x})$  and  $\mathbf{y}' = \mu(\mathbf{x}')$  be the local optima from Corollary 1. Now suppose that worker  $\mathbf{x}'$  is also an optimal choice for firm  $\mathbf{y}$ , that is  $\mathbf{x}'$  satisfies the optimality (first-order) conditions

<sup>49</sup>If  $J$  has complex eigenvalue,  $S$  is quasi-upper triangular.

<sup>50</sup>More precisely, this is  $\mathbf{x} = \nu(\mathbf{y})$ . But recall that  $\nu^{-1} = \mu$  is the unique inverse and hence the assignment can be completely characterized by the inverse  $\mu$ .

of both firms:

$$F_x(\mathbf{x}', \mathbf{y}) = w_x(\mathbf{x}') \quad (23)$$

$$F_x(\mathbf{x}', \mathbf{y}') = w_x(\mathbf{x}'). \quad (24)$$

I will show that, under Assumption 1, (23) and (24) cannot hold simultaneously. It suffices to show that the function  $F_x = (F_{x_C}, F_{x_M})$  is one-to-one, i.e.  $F_x(\mathbf{x}, \mathbf{y}) = F_x(\mathbf{x}, \mathbf{y}')$  implies  $\mathbf{y} = \mathbf{y}'$ . By Assumption 1,  $D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*)$  is a P-matrix. Moreover,  $F_x$  is defined over a rectangular region on  $\mathbb{R}^4$ . It follows from the Gale-Nikaido Theorem ([Gale and Nikaido \[1965\]](#)) that  $F_x$  is injective with respect to  $\mathbf{y}$ . Thus, (23) and (24) cannot hold simultaneously because

$$F_x(\mathbf{x}', \mathbf{y}) = F_x(\mathbf{x}', \mathbf{y}') \quad (25)$$

only if  $\mathbf{y} = \mathbf{y}'$ , contradicting the assumption that  $\mathbf{y} \neq \mathbf{y}'$ . It follows that the singleton solution to the firm's problem found in Corollary 1 is not only a local but also a global maximum. ■

## A.2 The Wage Function

In technical terms, Proposition 2 states: Given a continuously differentiable assignment  $\mathbf{y}^* = \mu(\mathbf{x})$ , condition (6) is necessary and sufficient for the existence of a unique solution to the system (3) and (4), given by  $w(\mathbf{x})$ , such that  $w(\underline{\mathbf{x}}) = w_0$ .<sup>51</sup>

### Proof of Proposition 2.

The proof is based on *Frobenius Theorem*. Consider a system of linear first-order partial differential equations

$$\frac{\partial u^\rho}{\partial x^i} = \psi_i^\rho(x, u) \quad i = 1, \dots, N; \rho = 1, \dots, n \quad (26)$$

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

**Theorem 1 (Frobenius Theorem)** *The necessary and sufficient conditions for the unique solution  $u^\alpha = u^\alpha(x)$  to the system (26) such that  $u(x_0) = u_0$  to exist for any initial data  $(x_0, u_0) \in \mathbb{R}^{N+n}$  is that the relations*

$$\frac{\partial \psi_i^\alpha}{\partial x^j} - \frac{\partial \psi_j^\alpha}{\partial x^i} + \sum_{\beta} \left( \frac{\partial \psi_i^\alpha}{\partial u^\beta} \psi_j^\beta - \frac{\partial \psi_j^\alpha}{\partial u^\beta} \psi_i^\beta \right) = 0 \quad \forall i, j = 1, \dots, N, \quad \alpha, \beta = 1, \dots, n. \quad (27)$$

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

Applying Frobenius' Theorem to this model implies:  $u = w$ ,  $x = (x_1, x_2, \dots, x_N)$  and

---

<sup>51</sup> $w_0$  is the reservation wage of the least productive worker  $\underline{\mathbf{x}}$ , set s.t. he is indifferent between working and not working.

$\psi_i(x, u) = F_{x_i}(\mathbf{x}, \mathbf{y}(\mathbf{x}))$ . Notice that  $n = 1$  because  $w$  is a real-valued function. Then, (27) 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

$$F_{x_C x_M} + F_{x_C y_C} \frac{\partial y_C^*}{\partial x_M} + F_{x_C y_M} \frac{\partial y_M^*}{\partial x_M} - F_{x_M x_C} - F_{x_M y_C} \frac{\partial y_C^*}{\partial x_C} + F_{x_M y_M} \frac{\partial y_M^*}{\partial x_C} = 0. \quad (28)$$

(28) coincides with condition (6) from the main text since  $F_{x_C x_M} = F_{x_M x_C}$ . Hence, given (6), the involutivity condition from Frobenius theorem is satisfied. A unique (local) solution to the system of linear partial differential equations (3) and (4) exists. ■

## B Proofs of Quadratic-Gaussian Model (Section 3)

### B.1 Labor Market Clearing under PAM (or NAM)

Having applied the measure-preserving transformation (9) to skills and productivities, the labor market clearing of transformed variables under PAM reads

$$\int_{z_{y_C}}^{\infty} \int_{z_{y_M}}^{\infty} g(\hat{z}_{y_C}, \hat{z}_{y_M}) d\hat{z}_{y_M} d\hat{z}_{y_C} = \int_{z_{x_C}}^{\infty} \int_{z_{x_M}}^{\infty} h(\hat{z}_{x_C}, \hat{z}_{x_M}) d\hat{z}_{x_M} d\hat{z}_{x_C} \quad (29)$$

where  $h$  and  $g$  denote the standard normal p.d.f.'s of the uncorrelated skills and productivities, respectively. Equation (11) follows immediately, taking into account that the  $z$ 's are independent and standard normally distributed. Similarly, under NAM, the market clearing would read

$$\int_{z_{y_C}}^{\infty} \int_{z_{y_M}}^{\infty} g(\hat{z}_{y_C}, \hat{z}_{y_M}) d\hat{z}_{y_M} d\hat{z}_{y_C} = \int_{-\infty}^{z_{x_C}} \int_{-\infty}^{z_{x_M}} h(\hat{z}_{x_C}, \hat{z}_{x_M}) d\hat{z}_{x_M} d\hat{z}_{x_C}.$$

### B.2 The Equilibrium Assignment

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

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

(ii) *The elements of  $\mathcal{S}$  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,  $\Sigma^{\frac{1}{2}} R(\Sigma^{\frac{1}{2}} R)^T = \Sigma^{\frac{1}{2}} R R^T (\Sigma^{\frac{1}{2}})^T = \Sigma^{\frac{1}{2}} (\Sigma^{\frac{1}{2}})^T = \Sigma$ .*

**Proof.**

(i) The existence of an infinite number of square roots of the covariance matrix follows from its

symmetry. The following non-linear system

$$\Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \Sigma \quad (30)$$

or,

$$\begin{aligned} a^2 + b^2 &= 1 \\ c^2 + d^2 &= 1 \\ ac + bd &= \rho \end{aligned}$$

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

$$\begin{aligned} \Sigma &= CDC' \\ \Leftrightarrow \Sigma^{\frac{1}{2}} &= CD^{\frac{1}{2}}C' \end{aligned} \quad (31)$$

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. ■

The next lemma states 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:*

$$\alpha_x^2 + \beta_x^2 = 1 \quad (32)$$

$$\alpha_y^2 + \beta_y^2 = 1 \quad (33)$$

$$\frac{\partial y_C^*}{\partial x_M} = \delta \frac{\partial y_M^*}{\partial x_C} \quad (34)$$

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

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

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

$$\alpha_y = \pm \frac{(1 + \delta) \left( \sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}} \right)}{\sqrt{(1 - \delta)^2 \left( \sqrt{\frac{1+\rho_y}{1+\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_x}} \right)^2 + (1 + \delta)^2 \left( \sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}} \right)^2}} \quad (35)$$

$$\beta_y = \sqrt{1 - \alpha_y^2}. \quad (36)$$

(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 (13) where I use *rotations* of the spectral square root (given by (31)) to uncorrelate skills and productivities. They are given by:

$$\Sigma_i^{\frac{1}{2}} = \begin{bmatrix} \frac{1}{2}(\sqrt{1 + \rho_i} + \sqrt{1 - \rho_i}) & \frac{1}{2}(\sqrt{1 + \rho_i} - \sqrt{1 - \rho_i}) \\ \frac{1}{2}(\sqrt{1 + \rho_i} - \sqrt{1 - \rho_i}) & \frac{1}{2}(\sqrt{1 + \rho_i} + \sqrt{1 - \rho_i}) \end{bmatrix} \begin{bmatrix} \alpha_i & -\beta_i \\ \beta_i & \alpha_i \end{bmatrix}, \forall i \in \{x, y\}. \quad (37)$$

Using (37), the candidate equilibrium assignment can be computed from (13) as:

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{2} \left( (\alpha_y \alpha_x + \beta_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) + (\beta_y \alpha_x - \alpha_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}} \right) \right) & \frac{1}{2} \left( (\alpha_y \alpha_x + \beta_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) - (\beta_y \alpha_x - \alpha_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}} \right) \right) \\ \frac{1}{2} \left( (\alpha_y \alpha_x + \beta_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) + (\beta_y \alpha_x - \alpha_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}} \right) \right) & \frac{1}{2} \left( (\alpha_y \alpha_x + \beta_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) - (\beta_y \alpha_x - \alpha_y \beta_x) \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}} \right) \right) \end{bmatrix} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \quad (38)$$

(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 candidate assignment (38), integrability condition (34) reads:

$$\begin{aligned} & \left( \alpha_y \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} - \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right) - \sqrt{1 - \alpha_y^2} \left( \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right) \right) = \\ & \delta \left( \alpha_y \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} - \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right) + \sqrt{1 - \alpha_y^2} \left( \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right) \right) \end{aligned} \quad (39)$$

Reorganizing terms and solving for  $\alpha_y$  yields:

$$\alpha_y = \pm \frac{(1 + \delta) \left( \sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}} \right)}{\sqrt{(1 - \delta)^2 \left( \sqrt{\frac{1+\rho_y}{1+\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_x}} \right)^2 + (1 + \delta)^2 \left( \sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}} \right)^2}} \quad (40)$$

Using (40),  $\beta_y$  can be backed out from (33)<sup>52</sup>

$$\beta_y = \sqrt{1 - \alpha_y^2}. \quad (41)$$

(ii) Rearranging (39) yields:

$$\alpha_y(1 - \delta) \left( \sqrt{\frac{1 + \rho_y}{1 + \rho_x}} - \sqrt{\frac{1 - \rho_y}{1 - \rho_x}} \right) = \sqrt{1 - \alpha_y^2}(\delta + 1) \left( \sqrt{\frac{1 + \rho_y}{1 - \rho_x}} + \sqrt{\frac{1 - \rho_y}{1 + \rho_x}} \right) \quad (42)$$

While  $RHS \geq 0, \forall \rho_x, \rho_y$ ,  $LHS \lesseqgtr 0$  for  $\rho_y \lesseqgtr \rho_x$ . It follows that  $\alpha_y \geq 0$  for  $\rho_x \leq \rho_y$  and  $\alpha_y < 0$  for  $\rho_x > \rho_y$ . ■

### Proof of Proposition 3.

*Computing the Assignment:*

(i) For  $\delta = 1$ , from (35)

$$\alpha_y = \pm 1. \quad (43)$$

The orthogonal transformation (37) delivers:

$$\begin{bmatrix} \frac{1}{2}(\sqrt{1 + \rho_i} + \sqrt{1 - \rho_i}) & \frac{1}{2}(\sqrt{1 + \rho_i} - \sqrt{1 - \rho_i}) \\ \frac{1}{2}(\sqrt{1 + \rho_i} - \sqrt{1 - \rho_i}) & \frac{1}{2}(\sqrt{1 + \rho_i} + \sqrt{1 - \rho_i}) \end{bmatrix} \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix} = \pm \begin{bmatrix} \frac{1}{2}(\sqrt{1 + \rho_i} + \sqrt{1 - \rho_i}) & \frac{1}{2}(\sqrt{1 + \rho_i} - \sqrt{1 - \rho_i}) \\ \frac{1}{2}(\sqrt{1 + \rho_i} - \sqrt{1 - \rho_i}) & \frac{1}{2}(\sqrt{1 + \rho_i} + \sqrt{1 - \rho_i}) \end{bmatrix} \quad \text{for } i \in \{x, y\}. \quad (44)$$

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, which is given by

$$\begin{aligned} \Sigma &= CDC' \\ \Leftrightarrow \Sigma^{\frac{1}{2}} &= CD^{\frac{1}{2}}C' \end{aligned} \quad (45)$$

where  $D$  is a diagonal matrix with the eigenvalues of  $\Sigma$  as diagonal entries and  $C$  is a matrix of orthonormal eigenvectors of  $\Sigma$ . The 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:

$$\Sigma_y^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{1 + \rho_y} & 0 \\ 0 & \sqrt{1 - \rho_y} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(\sqrt{1 + \rho_y} + \sqrt{1 - \rho_y}) & \frac{1}{2}(\sqrt{1 + \rho_y} - \sqrt{1 - \rho_y}) \\ \frac{1}{2}(\sqrt{1 + \rho_y} - \sqrt{1 - \rho_y}) & \frac{1}{2}(\sqrt{1 + \rho_y} + \sqrt{1 - \rho_y}) \end{bmatrix} \quad (46)$$

---

<sup>52</sup>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.



Moreover, since

$$\Sigma^{-\frac{1}{2}} = CD^{-\frac{1}{2}}C' = C \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{bmatrix} C'. \quad (47)$$

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

$$\Sigma_x^{-\frac{1}{2}} = \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{1+\rho_x}} & 0 \\ 0 & \frac{1}{\sqrt{1-\rho_x}} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}} \right) & \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}} \right) \\ \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}} \right) & \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}} \right) \end{bmatrix} \quad (48)$$

It follows that the Jacobian of the matching function is given by:

$$D_x \mathbf{y}^* = (\Sigma_y^{\frac{1}{2}} R_y) (\Sigma_x^{\frac{1}{2}} R_x)^{-1} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) & \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) \\ \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) & \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) \end{bmatrix} \quad (49)$$

The assignment is then computed using (13).

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

$$R_y = \begin{bmatrix} \alpha_y & -\sqrt{1-\alpha_y^2} \\ \sqrt{1-\alpha_y^2} & \alpha_y \end{bmatrix} = \begin{bmatrix} \pm \frac{1}{2} (\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}) & -\frac{1}{2} (\sqrt{(1+\rho_y)(1-\rho_x)} - \sqrt{(1-\rho_y)(1+\rho_x)}) \\ \frac{1}{2} (\sqrt{(1+\rho_y)(1-\rho_x)} - \sqrt{(1-\rho_y)(1+\rho_x)}) & \pm \frac{1}{2} (\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}) \end{bmatrix} \quad (50)$$

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

Let  $\Sigma_y^{\frac{1}{2}}$  and  $\Sigma_x^{\frac{1}{2}}$  be the spectral square roots of skill and productivity covariance matrices, given by (46) and by the inverse of (48), respectively. Then,

$$\Sigma_y^{\frac{1}{2}} R_y = \begin{bmatrix} \frac{1}{2} (\sqrt{1+\rho_y} + \sqrt{1-\rho_x}) & \frac{1}{2} (\sqrt{1+\rho_y} - \sqrt{1-\rho_x}) \\ \frac{1}{2} (\sqrt{1-\rho_x}(\rho_y + \sqrt{1-\rho_y^2}) + \sqrt{1+\rho_x}(\rho_y - \sqrt{1-\rho_y^2})) & \frac{1}{2} (\sqrt{1-\rho_x}(-\rho_y + \sqrt{1-\rho_y^2}) + \sqrt{1+\rho_x}(\rho_y + \sqrt{1-\rho_y^2})) \end{bmatrix}$$

$$(\Sigma_x^{\frac{1}{2}} R_x)^{-1} = \begin{bmatrix} \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}} \right) & \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}} \right) \\ \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}} \right) & \frac{1}{2} \left( \frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}} \right) \end{bmatrix}. \quad (52)$$

It can be shown that the Jacobian is then given by:

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

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  $L_i(L_i)^T = \Sigma_i, i \in \{x, y\}$ . By definition,  $L_i$  is a square root of  $\Sigma_i$ . Under the assumption of standard normality,  $L_i$  is given by:

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

Hence,

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

which coincides with (53). The equilibrium assignment is then given by (13)

$$\begin{aligned} \begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} &= (\Sigma_y^{\frac{1}{2}} R_y)(\Sigma_x^{\frac{1}{2}} R_x)^{-1} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \\ &= (L_y)(L_x)^{-1} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ \rho_y & \sqrt{1-\rho_y^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho_x & \sqrt{1-\rho_x^2} \end{bmatrix}^{-1} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} & \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \end{bmatrix} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \end{aligned} \quad (56)$$

(iii) The equilibrium assignment is obtained by using the suitable rotation matrices from Lemma 3 for each value of  $\delta \in (0, 1)$  together with the candidate equilibrium assignment (13).

*Consistency of the Assignment Functions from (i)-(iii) with the Equilibrium.* Three properties have to be verified: (a) Consistency with market clearing; (b) the assignment satisfies PAM; (c) the integrability condition is satisfied. (a) Market clearing is satisfied by (11) and because the transformation (9) is measure-preserving. (b) Verifying the PAM-property amounts to checking that  $D_x \mathbf{y}^*$  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:

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \underbrace{\begin{bmatrix} J_{11}(\rho_x, \rho_y, \delta) & J_{12}(\rho_x, \rho_y, \delta) \\ J_{21}(\rho_x, \rho_y, \delta) & J_{22}(\rho_x, \rho_y, \delta) \end{bmatrix}}_{D_x \mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1+\delta\sqrt{1-\rho_y^2}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\rho_y-\rho_x\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \\ \frac{\rho_y-\rho_x\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta+\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \end{bmatrix}}_{D_x \mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \quad (57)$$

Taking derivatives yields:

$$\frac{\partial y_C^*}{\partial x_C} > 0 \quad (58)$$

$$\frac{\partial y_M^*}{\partial x_M} > 0 \quad (59)$$

$$\frac{\partial y_C^*}{\partial x_C} \frac{\partial y_M^*}{\partial x_M} - \frac{\partial y_C^*}{\partial x_M} \frac{\partial y_M^*}{\partial x_C} = \sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}} > 0 \quad \text{for } \rho_y \ll 1 \quad (60)$$

where (58) and (59) follow immediately from (57). Hence,  $D_x \mathbf{y}^*$  is a P-matrix. (c) The assignment was derived under the integrability condition (34), i.e. it is satisfied. ■

### B.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

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

where  $w_0$  is the constant of integration, which I set to zero.<sup>53</sup>  $J_{11}, J_{12}, J_{21}, J_{22}$  are the elements of the matching function's Jacobian, which are given by (57). Using the assignment (57) explicitly, this is

$$w(x_C, x_M) = \frac{1}{2} \alpha [x_C \quad x_M] \underbrace{\begin{bmatrix} \frac{1+\delta\sqrt{1-\rho_y^2}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\rho_y-\rho_x\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \\ \frac{\delta\left(\rho_y-\rho_x\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta+\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \end{bmatrix}}_{\equiv \tilde{J}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} = \frac{1}{2} \alpha [x_C \quad x_M] \underbrace{\begin{bmatrix} J_{11} & J_{12} \\ \delta J_{21} & \delta J_{22} \end{bmatrix}}_{\equiv \tilde{J}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \quad (61)$$

<sup>53</sup>Type  $(x_C, x_M) = (0, 0)$  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 expression (15) in Proposition 4, taking into account  $w_0 = 0$ . The guess needs to be verified. Given (61), the partial derivatives of the wage with respect to skills  $x_C, x_M$  are given by

$$\frac{\partial w(x_C, x_M)}{\partial x_C} = \alpha(J_{11}x_C + J_{12}x_M) \quad (62)$$

$$\frac{\partial w(x_C, x_M)}{\partial x_M} = \alpha\delta(J_{22}x_M + J_{21}x_C) \quad (63)$$

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

$$\frac{\partial w(x_C, x_M)}{\partial x_C} = \alpha y_C^* \quad (64)$$

$$\frac{\partial w(x_C, x_M)}{\partial x_M} = \alpha \delta y_M^* \quad (65)$$

evaluated at the equilibrium assignment (57). Moreover, the integrability condition

$$\begin{aligned} \frac{\partial^2 w(x_C, x_M)}{\partial x_C \partial x_M} &= \frac{\partial^2 w(x_C, x_M)}{\partial x_M \partial x_C} \\ \Leftrightarrow J_{12} &= \delta J_{21} \end{aligned} \quad (66)$$

is satisfied by construction of the equilibrium assignment, which gives uniqueness of (61) by Proposition 2. ■

### Proof of Proposition 5. .

(i) See Proof of Proposition 3 (last part, under *Consistency of the Assignment Functions from (i)-(iii) with the Equilibrium.*).

(ii) 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 the candidate assignment given by (13) yields  $y_C^* = x_C$  and  $y_M^* = x_M$  and, hence,  $\frac{\partial y_C^*}{\partial x_C} = \frac{\partial y_M^*}{\partial x_M} = 1$  and  $\frac{\partial y_C^*}{\partial x_M} = \frac{\partial y_M^*}{\partial x_C} = 0$ . Hence, there is perfect assortativeness according to Definition 4.

The second result is that the maximum mismatch is achieved at  $|\rho_x - \rho_y| \rightarrow 2$ , i.e.  $J_{22} \equiv \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right)$  and  $abs(J_{21}) \equiv abs \left( \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right) \right)$  have suprema at  $\rho_x = 1, \rho_y = -1$  as well as  $\rho_x = -1, \rho_y = 1$ . I show this in two steps.

1. Step: No interior maximum:

$$\max_{\rho_x, \rho_y} \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right)$$

$$\rho_y : \left( \frac{1}{\sqrt{1+\rho_y}\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_y}\sqrt{1-\rho_x}} \right) = 0 \quad (67)$$

$$\rho_x : \frac{1}{2} \left( -\frac{\sqrt{1+\rho_y}}{(1+\rho_x)^{\frac{3}{2}}} + \frac{\sqrt{1-\rho_y}}{(1-\rho_x)^{\frac{3}{2}}} \right) = 0 \quad (68)$$

Solving the system of first order conditions (67) and (68) (necessary for an interior solution) yields  $\rho_x^2 + 1 = 0$ . A contradiction. There is no interior maximum. Likewise, maximize  $J_{21}$  with respect to  $\rho_x$  and  $\rho_y$ :

$$\max_{\rho_x, \rho_y} \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}} \right)$$

$$\rho_y : \left( \frac{1}{\sqrt{1+\rho_y}\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_y}\sqrt{1-\rho_x}} \right) = 0 \quad (69)$$

$$\rho_x : \frac{1}{2} \left( -\frac{\sqrt{1+\rho_y}}{(1+\rho_x)^{\frac{3}{2}}} - \frac{\sqrt{1-\rho_y}}{(1-\rho_x)^{\frac{3}{2}}} \right) = 0 \quad (70)$$

where it is immediate that (69) does not omit a solution for  $-1 < \rho_y < 1, -1 < \rho_x < 1$ .

2. Step: Suprema at the boundary. First, focus on  $J_{22}$ . Fix  $\rho_x = 1 - \epsilon, \epsilon > 0$ . Maximize  $J_{22}$  with respect to  $\rho_y$ .

$$\max_{\rho_y} \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{2-\epsilon} + \frac{\sqrt{1-\rho_y}}{\epsilon} \right)$$

$$\left( \frac{1}{(2-\epsilon)\sqrt{1+\rho_y}} - \frac{1}{\epsilon\sqrt{1-\rho_y}} \right) = 0$$

$$\Leftrightarrow \rho_y = -\frac{(2-\epsilon)^2 - \epsilon^2}{(2-\epsilon)^2 + \epsilon^2} \approx -1$$

Now fix  $\rho_x = -1 + \epsilon, \epsilon > 0$ .

$$\max_{\rho_y} \frac{1}{2} \left( \frac{\sqrt{1+\rho_y}}{\epsilon} + \frac{\sqrt{1-\rho_y}}{2-\epsilon} \right)$$

$$\left( \frac{1}{\epsilon\sqrt{1+\rho_y}} - \frac{1}{(2-\epsilon)\sqrt{1-\rho_y}} \right) = 0$$

$$\Leftrightarrow \rho_y = \frac{(2-\epsilon)^2 - \epsilon^2}{(2-\epsilon)^2 + \epsilon^2} \approx 1$$

Moreover, the same value of  $J_{22}$  is achieved at  $(\rho_x, \rho_y) = (1 - \epsilon, -\frac{(2-\epsilon)^2 - \epsilon^2}{(2-\epsilon)^2 + \epsilon^2})$  and  $(\rho_x, \rho_y) = (-1 + \epsilon, \frac{(2-\epsilon)^2 - \epsilon^2}{(2-\epsilon)^2 + \epsilon^2})$ .

Then, focus on  $J_{21}$ . In a similar way, fixing  $\rho_x$  to (i)  $\rho_x = 1 - \epsilon$  and (ii)  $\rho_x = -1 + \epsilon$ , and maximizing with respect to  $\rho_y$  gives (i)  $\rho_y = -\frac{(-2+\epsilon)^2 - \epsilon^2}{(-2+\epsilon)^2 + \epsilon^2} \approx -1$  and (ii)  $\rho_y = \frac{(-2+\epsilon)^2 - \epsilon^2}{(-2+\epsilon)^2 + \epsilon^2} \approx 1$ . Both points give the same absolute value of  $J_{21}$ . Both steps together imply that mismatch in the manual task, defined as  $|y_M - x_M| = |y_M - cx_C - dx_M|$  is maximal for  $|\rho_x - \rho_y| \rightarrow 2$ . ■

### Proof of Proposition 6. .

(i) Wages are convex in skills: Recall, the wage function is given by

$$w(x_C, x_M) = \frac{1}{2}\alpha[x_C \quad x_M] \underbrace{\begin{bmatrix} J_{11} & J_{12} \\ \delta J_{21} & \delta J_{22} \end{bmatrix}}_{\equiv \tilde{J}} \begin{bmatrix} x_C \\ x_M \end{bmatrix}$$

where  $J_{11}, J_{12}, J_{21}, J_{22}$  are the elements of the matching function's Jacobian given in (57). The Hessian of the wage function is given by

$$H(w) = \frac{1}{2}\alpha\tilde{J}$$

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

(ii) The moments of the wage distribution are given by

$$E(w(x_C, x_M)) = tr(\tilde{J}\Sigma_x) = \frac{1}{2}J_{11} + J_{12}\rho_x + \frac{1}{2}\delta J_{22} \quad (71)$$

$$Var(w(x_C, x_M)) = 2tr(\tilde{J}\Sigma_x\tilde{J}\Sigma_x) = \frac{1}{2}((J_{11} + J_{12}\rho_x)^2 + 2(J_{11}\rho_x + J_{12})(J_{12} + \delta J_{22}\rho_x) + (J_{12}\rho_x + \delta J_{22})^2) \quad (72)$$

$$E\left[\left(\frac{w - E(w)}{\sqrt{Var(w)}}\right)^3\right] = \frac{E(w^3) - 3E(w)Var(w) - E(w)^3}{Var(w)^{\frac{3}{2}}} = \frac{8tr(\tilde{J}\Sigma_x\tilde{J}\Sigma_x\tilde{J}\Sigma_x)}{(2tr(\tilde{J}\Sigma_x\tilde{J}\Sigma_x))^{\frac{3}{2}}} \quad (73)$$

where  $J_{11}, J_{12}, J_{21}, J_{22}$  are defined in (57),  $\tilde{J}$  denotes the Hessian of the wage function and where

$$\begin{aligned} tr(\tilde{J}\Sigma_x\tilde{J}\Sigma_x\tilde{J}\Sigma_x) &= (J_{11} + J_{12}\rho_x)((J_{11} + J_{12}\rho_x)^2 + (J_{11}\rho_x + J_{12})(J_{12} + \delta J_{22}\rho_x)) \\ &\quad + 2(J_{12} + \delta J_{22}\rho_x)(J_{11}\rho_x + J_{12})(J_{11} + 2J_{12}\rho_x + \delta J_{22}) \\ &\quad + (J_{12}\rho_x + \delta J_{22})((J_{12} + \delta J_{22}\rho_x)(J_{11}\rho_x + J_{12}) + (J_{12}\rho_x + \delta J_{22})^2) \end{aligned}$$

See e.g. [Magnus \[1978\]](#) for the derivation of moments of quadratic forms in normal variables. For the symmetric case,  $\delta = 1$ , these expressions simplify to

$$\begin{aligned}
E(w(x_C, x_M)) &= \text{tr}(J_\mu \Sigma_x) = \frac{1}{2} \left( \sqrt{(1 + \rho_x)(1 + \rho_y)} + \sqrt{(1 - \rho_x)(1 - \rho_y)} \right) \\
\text{Var}(w(x_C, x_M)) &= 2\text{tr}(J_\mu \Sigma_x J_\mu \Sigma_x) = 1 + \rho_x \rho_y \\
E(w^3) &= \text{tr}(J_\mu \Sigma_x)^3 + 6\text{tr}(J_\mu \Sigma_x) \text{tr}(J_\mu \Sigma_x J_\mu \Sigma_x) + 8\text{tr}(J_\mu \Sigma_x J_\mu \Sigma_x J_\mu \Sigma_x) \\
E \left[ \left( \frac{w - E(w)}{\sqrt{\text{Var}(w)}} \right)^3 \right] &= \frac{E(w^3) - 3E(w)\text{Var}(w) - E(w)^3}{\text{Var}(w)^{\frac{3}{2}}} = \frac{8\text{tr}(J_\mu \Sigma_x J_\mu \Sigma_x J_\mu \Sigma_x)}{(2\text{tr}(J_\mu \Sigma_x J_\mu \Sigma_x))^{\frac{3}{2}}} \\
&= \frac{(\sqrt{(1 + \rho_y)(1 + \rho_x)} + \sqrt{(1 - \rho_y)(1 - \rho_x)}) \left( 2 + 2\rho_x \rho_y - \sqrt{(1 - \rho_y^2)(1 - \rho_x^2)} \right)}{(1 + \rho_x \rho_y)^{\frac{3}{2}}}.
\end{aligned}$$

where, as before,  $J_\mu$  denotes the Jacobian of the matching function.

Skewness:  $J_\mu \Sigma_x J_\mu$  and  $\Sigma_x J_\mu \Sigma_x$  are two 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.

Finally, a result mentioned in the main text is that the average wage is maximized at  $\rho_x = \rho_y$  and has two infima at  $\rho_x = -1, \rho_y = 1$  and  $\rho_y = -1, \rho_x = 1$ . To see this, maximize  $E(w(x_C, x_M))$  with respect to  $\rho_y$  and  $\rho_x$  yields:

$$\begin{aligned}
\max_{\rho_x, \rho_y} \frac{1}{2} \left( \sqrt{(1 + \rho_x)(1 + \rho_y)} + \sqrt{(1 - \rho_x)(1 - \rho_y)} \right) \\
\rho_x : \quad \rho_x = \rho_y \\
\rho_y : \quad \rho_x = \rho_y
\end{aligned}$$

The Hessian of this maximization problem is given by:

$$H(E(w)) = \begin{bmatrix} -\frac{1}{8} \left( \frac{\sqrt{1 - \rho_y}}{(1 - \rho_x)^{\frac{3}{2}}} + \frac{\sqrt{1 + \rho_y}}{(1 + \rho_x)^{\frac{3}{2}}} \right) & \frac{1}{8} \left( \frac{1}{\sqrt{1 - \rho_x} \sqrt{1 - \rho_y}} + \frac{1}{\sqrt{1 + \rho_x} \sqrt{1 + \rho_y}} \right) \\ \frac{1}{8} \left( \frac{1}{\sqrt{1 - \rho_x} \sqrt{1 - \rho_y}} + \frac{1}{\sqrt{1 + \rho_x} \sqrt{1 + \rho_y}} \right) & -\frac{1}{8} \left( \frac{\sqrt{1 - \rho_x}}{(1 - \rho_y)^{\frac{3}{2}}} + \frac{\sqrt{1 + \rho_x}}{(1 + \rho_y)^{\frac{3}{2}}} \right) \end{bmatrix}$$

Notice that  $H_{11}(E(w)) < 0$  and  $\text{Det}(H(E(w)))$  simplifies to

$$\text{Det}(H(E(w))) = ((1 - \rho_y)(1 + \rho_x) - (1 - \rho_x)(1 + \rho_y))^2 \geq 0$$

Hence, the Hessian is negative semi-definite  $\forall \rho_x, \rho_y$ . Hence,  $E(w(x_C, x_M))$  is concave. At  $\rho_x = \rho_y$ ,  $\text{Det}(H(E(w))) = 0$ , indicating that  $\rho_x = \rho_y$  are (degenerate) maxima. Due to concavity of  $E(w)$ , its infimum is achieved at the boundary of the domain  $-1 < \rho_x < 1, -1 < \rho_y < 1$ .  $E(w)$  has two infima at  $\rho_x = -1, \rho_y = 1$  and  $\rho_y = -1, \rho_x = 1$ . ■

## C Proofs on Technological Change

### C.1 Task-Biased Technological Change (Section 4.1)

#### Proof of Proposition 7. .

Notice that I focus on the cases for which  $\rho_x, \rho_y < 0$  or  $\rho_x, \rho_y > 0$ . I will make explicit where this assumption is needed to arrive at *clean* analytical expressions.

(i) Follows from proof of Proposition 5 part (i). Notice that for this result to hold, it is not necessary to impose the restriction  $\rho_x, \rho_y < 0$  or  $\rho_x, \rho_y > 0$ .

(ii) Mismatch-Assortativeness Trade-Off Across Tasks. It needs to be shown that as  $\delta' < \delta = 1$ ,  $|y_C - x_C|$  decreases and  $|y_M - x_M|$  increases. Recall that the Jacobian of the matching function is given as in (57):

$$\underbrace{\begin{bmatrix} J_{11}(\rho_x, \rho_y, \delta) & J_{12}(\rho_x, \rho_y, \delta) \\ J_{21}(\rho_x, \rho_y, \delta) & J_{22}(\rho_x, \rho_y, \delta) \end{bmatrix}}_{D_x \mathbf{y}^*} = \underbrace{\begin{bmatrix} \frac{1 + \delta \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}}}{\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2}) + \delta^2}} & \frac{\delta \left( \rho_y - \rho_x \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}} \right)}{\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2}) + \delta^2}} \\ \frac{\rho_y - \rho_x \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}}}{\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2}) + \delta^2}} & \frac{\delta + \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}}}{\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2}) + \delta^2}} \end{bmatrix}}_{D_x \mathbf{y}^*} \quad (74)$$

I will show this claim in two steps:

1. Step:

$$\begin{aligned} J_{11} &\geq 1 && \text{for } |\rho_x| \geq |\rho_y| \\ J_{12} &\geq 0 && \text{for } \rho_x \leq \rho_y \\ J_{21} &\geq 0 && \text{for } \rho_x \leq \rho_y \\ J_{22} &\geq 1 && \text{for } |\rho_x| \geq |\rho_y| \end{aligned}$$

This follows from (74), since

$$\begin{aligned} J_{11} &= \frac{1 + \delta \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}}}{\sqrt{1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2}) + \delta^2}} \geq 1 \\ &\Leftrightarrow \left( 1 + \delta \frac{\sqrt{1 - \rho_y^2}}{\sqrt{1 - \rho_x^2}} \right)^2 \geq 1 + 2\delta(\rho_x \rho_y + \sqrt{1 - \rho_y^2} \sqrt{1 - \rho_x^2}) + \delta^2 \\ &\Leftrightarrow 2\delta \frac{\sqrt{1 - \rho_y^2} \rho_x^2 - \rho_x \rho_y \sqrt{1 - \rho_x^2}}{\sqrt{1 - \rho_x^2}} + \delta^2 \left( \frac{1 - \rho_y^2}{1 - \rho_x^2} - 1 \right) \geq 0 \quad \text{for } |\rho_x| \geq |\rho_y| \end{aligned}$$



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

$$\Leftrightarrow \rho_y\sqrt{1-\rho_x^2} - \rho_x\sqrt{1-\rho_y^2} \geq 0 \quad \text{for } \rho_x \leq \rho_y.$$

Notice that for  $\delta = 1$ ,  $J_{11} = J_{22}$  and  $J_{12} = J_{21}$ .

2. Step: At  $\delta = 1$ ,

$$\begin{aligned} \frac{\partial J_{11}}{\partial \delta} &\geq 0 \quad \text{for } |\rho_x| \geq |\rho_y| \\ \frac{\partial J_{12}}{\partial \delta} &\geq 0 \quad \text{for } \rho_x \leq \rho_y \\ \frac{\partial J_{21}}{\partial \delta} &\geq 0 \quad \text{for } \rho_x \geq \rho_y \\ \frac{\partial J_{22}}{\partial \delta} &\geq 0 \quad \text{for } |\rho_x| \leq |\rho_y| \end{aligned}$$

where

$$\begin{aligned} \frac{\partial J_{11}}{\partial \delta} &= \frac{\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - \delta + \left( \delta \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1 \right) (\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2)^{\frac{3}{2}}} \\ \frac{\partial J_{11}}{\partial \delta} \Big|_{\delta=1} &= \frac{\left( \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1 \right) \left( 1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2} \right)}{(2 + 2(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}))^{\frac{3}{2}}} \geq 0 \quad \text{for } |\rho_x| \geq |\rho_y| \\ \frac{\partial J_{12}}{\partial \delta} &= \frac{\left( \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \right) (1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2)^{\frac{3}{2}}} \geq 0 \quad \text{for } \rho_x \leq \rho_y \\ \frac{\partial J_{21}}{\partial \delta} &= -\frac{\left( \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \right) (\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2} + \delta)}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2)^{\frac{3}{2}}} \geq 0 \quad \text{for } \rho_x \geq \rho_y \\ \frac{\partial J_{22}}{\partial \delta} &= \frac{1 - \delta \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} + \left( \delta - \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1 \right) (\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})}{(1 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2)^{\frac{3}{2}}} \\ \frac{\partial J_{22}}{\partial \delta} \Big|_{\delta=1} &= \frac{\left( 1 - \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \right) \left( 1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2} \right)}{(2 + 2(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}))^{\frac{3}{2}}} \geq 0 \quad \text{for } |\rho_x| \leq |\rho_y| \end{aligned}$$

which together with Step 1 establishes the result.

(iii) Sorting when  $\delta = 0$ . In the cognitive task,  $y_C = x_C$ , follows from (53) in proof of Proposition 3. In turn, in the manual task, there is maximal mismatch  $|y_M - x_M|$ . To see this, notice that

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

since

$$\begin{aligned} \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} &\geq \frac{\delta + \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2}} \\ (1-\rho_y^2)(1+2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2) &\geq (1-\rho_x^2)\delta^2 + 2\delta\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2} + (1-\rho_y^2) \\ (1-\rho_y^2)(2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2) &\geq (1-\rho_x^2)\delta^2 + 2\delta\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2} \\ 2(\rho_x\rho_y(1-\rho_y^2) - \rho_y^2\sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2} + \delta^2(\rho_x^2 - \rho_y^2)) &\geq 0 \quad \text{for } |\rho_x| \geq |\rho_y|. \end{aligned}$$

Since  $J_{22}|_{\delta \neq 0} \geq 1$  if  $|\rho_x| \geq |\rho_y|$ , the result follows. Also,

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

since for  $\rho_x < \rho_y$ ,

$$\begin{aligned} \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} &> \frac{\delta \left( \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \right)}{\sqrt{1+2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2}} \\ &\Leftrightarrow \sqrt{1+2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) + \delta^2} > \delta \\ &\Leftrightarrow 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}) > 0. \end{aligned}$$

Clearly, for  $\rho_x = \rho_y$ ,  $J_{21}|_{\delta \neq 0} = J_{21}|_{\delta=0}$ . Since  $J_{21}|_{\delta \neq 0} \geq 0$  if  $\rho_x \leq \rho_y$ , the result follows.

(iv) Task-biased TC has no effect on sorting for  $\rho_x = \rho_y$ . Follows from (ii) in proof of Proposition 5. Furthermore, task-biased TC has maximal effects on reallocation (i.e.,  $\frac{\partial J_{11}}{\partial \delta}|_{\delta=1}$ ,  $\frac{\partial J_{12}}{\partial \delta}|_{\delta=1}$ ,  $\frac{\partial J_{21}}{\partial \delta}|_{\delta=1}$  and  $\frac{\partial J_{22}}{\partial \delta}|_{\delta=1}$  are largest) when  $|\rho_x - \rho_y|$  is maximal. To see this, I first show that these expressions do not have an interior maximum. Notice that  $\frac{\partial J_{11}}{\partial \delta}|_{\delta=1} = \frac{\partial J_{22}}{\partial \delta}|_{\delta=1}$  and  $\frac{\partial J_{12}}{\partial \delta}|_{\delta=1} = \frac{\partial J_{21}}{\partial \delta}|_{\delta=1}$ . So, it

suffices to show this for two of the four expressions. Recall that

$$r \equiv \frac{\partial J_{11}}{\partial \delta} \Big|_{\delta=1} = \frac{\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1}{2^{\frac{3}{2}}(2 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}))^{\frac{1}{2}}} \quad (75)$$

$$s \equiv \frac{\partial J_{12}}{\partial \delta} \Big|_{\delta=1} = \frac{\rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{2^{\frac{3}{2}}(2 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}))^{\frac{1}{2}}} \quad (76)$$

Then,

$$\frac{\partial r}{\partial \rho_x} = \frac{\rho_x \frac{\sqrt{1-\rho_y^2}}{(1-\rho_x^2)^{\frac{3}{2}}} (1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) - \frac{1}{2} \left( \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1 \right) \left( \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \right)}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \quad (77)$$

$$\frac{\partial r}{\partial \rho_y} = \frac{-\frac{\rho_y}{\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}} (1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) - \frac{1}{2} \left( \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1 \right) \left( \rho_x - \rho_y \frac{\sqrt{1-\rho_x^2}}{\sqrt{1-\rho_y^2}} \right)}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \quad (78)$$

$$\frac{\partial s}{\partial \rho_x} = \frac{-\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \left( 1 + \frac{\rho_x^2}{1-\rho_x^2} \right) (1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) + \frac{1}{2} \left( \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \right)^2}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \quad (79)$$

$$\frac{\partial s}{\partial \rho_y} = \frac{\left( 1 + \frac{\rho_y\rho_x}{\sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}} \right) (1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) + \frac{1}{2} \frac{(\rho_x\sqrt{1-\rho_y^2} - \rho_y\sqrt{1-\rho_x^2})^2}{\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}}}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \quad (80)$$

Expression (77) is negative for all  $\rho_x, \rho_y < 0$  and positive for all  $\rho_x, \rho_y > 0$ . At  $\rho_x = \rho_y = 0$ , the expression is zero, indicating a minimum (i.e. no reallocation when  $\rho_x = \rho_y$ ). Expression (78) is positive for all  $\rho_x, \rho_y < 0$  and negative for all  $\rho_x, \rho_y > 0$ . At  $\rho_x = \rho_y = 0$ , the expression is zero, indicating a minimum (i.e. no reallocation when  $\rho_x = \rho_y$ ). Hence, (75) has no interior maximum. Also, (80) is positive  $\forall \rho_x, \rho_y$ . Evaluating (79) at the corners  $\rho_x \rightarrow 1$  and  $\rho_x \rightarrow -1$  yields a strictly positive and a strictly negative expression. Hence, (76) has no interior maximum. The supremum of (75) and (76) must be in the corner. In a second step, it can be shown that (75) and (76) are most positive or negative for  $\rho_x = \pm 1$  and  $\rho_y = 0$  as well as  $\rho_x = 0$  and  $\rho_y = \pm 1$ . ■

### Proof of Proposition 8. .

- (i) The results follow from simulations of the closed forms (71) and (73). Available upon request.
- (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  $\beta$  decreases). For  $|\rho_x| < |\rho_y|$ ,  $J_{11} < 1$  and  $\frac{\partial J_{11}}{\partial \delta} < 0$  as well as  $J_{22} < 1$  and  $\frac{\partial J_{22}}{\partial \delta} > 0$ , where  $J_{11}, J_{22}$  are defined as in (74). It follows that for  $|\rho_x| < |\rho_y|$ ,

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

Hence, due to task-biased TC, wages become more convex in  $x_C$  but less convex in  $x_M$ . Notice that additionally decreasing  $\beta$  reinforces the effects. Finally, the result that the curvature changes are largest follows from Proposition 7 part (iv). ■

## C.2 Skill-Biased Technological Change (Section 4.2)

The wage function under the augmented technology

$$F(x_C, x_M, y_C, y_M) = \alpha(x_C y_C + \delta x_M y_M + \lambda x_C + \delta \eta x_M) + f_0 \quad (81)$$

is given by

$$w(x_C, x_M) = \alpha \left( \frac{1}{2} \mathbf{x}' \tilde{J} \mathbf{x} + \theta' \mathbf{x} \right) + w_0 = \frac{1}{2} \alpha (\mathbf{x} - h)' \tilde{J} (\mathbf{x} - h) + C \quad (82)$$

where

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

**Proof of Proposition 9.** (i) It is immediate that assignment (57) satisfies the first-order conditions of the firm under (81). (ii) (82) satisfied integrability condition (6), i.e. is the unique wage schedule supporting the assignment. From (82), skill-biased TC parameters  $\lambda$  and  $\eta$  do not affect the curvature of the wage function. (iii) Under (82), the variance of the wage distribution is given by

$$Var(w) = \alpha^2 (1 - \rho_x^2) \left( (J_{21} + J_{11} \rho_x)^2 + \frac{J_{11}^2}{2} (1 - \rho_x^2) \right) + (1 - \rho_x^2) \lambda^2 + \alpha^2 2 \left( \frac{J_{11}}{2} \rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2} \right)^2 + (\rho_x \lambda + \eta)^2 \quad (83)$$

where  $\lambda$  and  $\eta$  enter positively, which proves the result. To derive (83), notice that

$$\begin{aligned}E(w|x_M) &= \alpha \left( \frac{J_{11}}{2} E(x_C^2|x_M) + J_{12} x_M E(x_C|x_M) + \frac{\delta J_{22}}{2} x_M^2 \right) + \eta x_M + \lambda E(x_C|x_M) \\ &= \alpha \left( \frac{J_{11}}{2} (\rho_x^2 x_M^2 + (1 - \rho_x^2)) + \frac{\delta J_{22}}{2} x_M^2 \right) + \rho_x x_M (\alpha J_{12} x_M + \lambda) + \eta x_M\end{aligned}$$

$$\begin{aligned}
\text{Var}(E(w|x_M)) &= \alpha^2 \text{Var}(x_M^2) \left( \frac{J_{11}}{2} \rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2} \right)^2 + \text{Var}(x_M) (\rho_x \lambda + \eta)^2 \\
&= \alpha^2 2 \left( \frac{J_{11}}{2} \rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2} \right)^2 + (\rho_x \lambda + \eta)^2
\end{aligned} \tag{84}$$

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

$$\text{Var}(w|x_M) = \text{Var}(x_C|x_M)(\alpha^2 J_{12} x_M^2 + \lambda^2) + \text{Var}(x_C^2|x_M) \frac{\alpha^2 J_{11}^2}{4} + \text{cov}(x_C^2, x_C|x_M)(\alpha^2 J_{11} J_{12} x_M + \lambda)$$

where

$$\begin{aligned}
\text{Var}(x_C|x_M) &= E(x_C^4|x_M) - (E(x_C^2|x_M))^2 = 4\rho_x^2 x_M^2 (1 - \rho_x^2) + 2(1 - \rho_x^2)^2 \\
\text{cov}(x_C^2, x_C|x_M) &= E(x_C^3|x_M) - E(x_C^2|x_M)E(x_C|x_M) = 2\rho_x x_M (1 - \rho_x^2).
\end{aligned}$$

Hence,

$$\begin{aligned}
\text{Var}(w|x_M) &= \alpha^2 (1 - \rho_x^2) (x_M^2 (J_{12} + \rho_x J_{11})^2 + \frac{J_{11}^2}{2} (1 - \rho_x^2)) + (1 - \rho_x^2) (\rho_x J_{11} \lambda x_M + \lambda^2) \\
E(\text{Var}(w|x_M)) &= \alpha^2 (1 - \rho_x^2) ((J_{12} + \rho_x J_{11})^2 + \frac{J_{11}^2}{2} (1 - \rho_x^2)) + (1 - \rho_x^2) \lambda^2.
\end{aligned} \tag{85}$$

(83) follows from adding (84) and (85), i.e.  $\text{Var}(w) = E(\text{Var}(w|x_M)) + \text{Var}(E(w|x_M))$ . ■

## D 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 turn, 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 with independent uniform distributions or arbitrary but overlapping skill and productivity distributions.

### D.0.1 Gaussian Distributions

Suppose the skill and productivity distribution are bivariate standard Gaussian and the technology is given by:

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

In this setting, with non-zero *between*-task complementarities, the sufficient condition for PAM/NAM is stated in the following proposition.

**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*

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

*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 diagonally dominant if  $|m_{ii}| > \sum_{j \neq i} |m_{ij}|, i = 1, 2, \dots, n$  and row diagonally dominant if  $|m_{ii}| > \sum_{j \neq i} |m_{ji}|, i = 1, 2, \dots, n$ . In this setting,  $D_{xy}^2$  is strictly diagonally dominant if  $1 \geq \delta > |\alpha|$  and  $1 \geq \delta > |\beta|$ , 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 cdf's in a decreasing instead of increasing way). The extension to non-standard Gaussian variables is given in the Online Appendix.

Under (86), integrability condition (6), which needs to be satisfied in order for a unique wage schedule to exist, is given by:

$$\begin{aligned} \frac{\partial^2 w(x_C, x_M)}{\partial x_C \partial x_M} &= \frac{\partial^2 w(x_C, x_M)}{\partial x_M \partial x_C} \\ \Leftrightarrow J_{12} + \alpha J_{22} &= \beta J_{11} + \delta J_{21} \end{aligned} \quad (87)$$

where  $J_{11} \equiv \frac{\partial y_C^*}{\partial x_C}, J_{12} \equiv \frac{\partial y_C^*}{\partial x_M}, J_{21} \equiv \frac{\partial y_M^*}{\partial x_C}, J_{22} \equiv \frac{\partial y_M^*}{\partial x_M}$  denote the elements of the matching function's Jacobian.

Using (87), the equilibrium assignment follows Proposition 3. It is given by

$$\begin{aligned} J_{11} &= \frac{4}{2Z\sqrt{1-\rho_x^2}} \left( (1 + \alpha\rho_y)\sqrt{1-\rho_x^2} + (\delta + \alpha\rho_x)\sqrt{1-\rho_y^2} \right) \\ J_{12} &= \frac{4}{2Z\sqrt{1-\rho_x^2}} \left( (\beta + \delta\rho_y)\sqrt{1-\rho_x^2} - (\delta\rho_x + \alpha)\sqrt{1-\rho_y^2} \right) \\ J_{21} &= \frac{4}{2Z\sqrt{1-\rho_x^2}} \left( (\alpha + \rho_y)\sqrt{1-\rho_x^2} - (\rho_x + \beta)\sqrt{1-\rho_y^2} \right) \\ J_{22} &= \frac{4}{2Z\sqrt{1-\rho_x^2}} \left( (\delta + \beta\rho_y)\sqrt{1-\rho_x^2} + (1 + \beta\rho_x)\sqrt{1-\rho_y^2} \right) \end{aligned}$$

where

$$z \equiv \sqrt{\left( (1 + \delta) \left( \frac{\sqrt{1 + \rho_y}}{1 - \rho_x} + \frac{\sqrt{1 - \rho_y}}{1 + \rho_x} \right) + (\alpha + \beta) \left( \frac{\sqrt{1 + \rho_y}}{1 - \rho_x} - \frac{\sqrt{1 - \rho_y}}{1 + \rho_x} \right) \right)^2 + \left( (1 - \delta) \left( \frac{\sqrt{1 + \rho_y}}{1 + \rho_x} - \frac{\sqrt{1 - \rho_y}}{1 - \rho_x} \right) + (\alpha - \beta) \left( \frac{\sqrt{1 + \rho_y}}{1 + \rho_x} + \frac{\sqrt{1 - \rho_y}}{1 - \rho_x} \right) \right)^2}.$$

PAM holds since  $\forall \rho_x, \rho_y$ ,

$$\begin{aligned} J_{11} &> 0 \quad \text{if } 1 \geq \delta > |\alpha| \\ J_{22} &> 0 \quad \text{if } 1 \geq \delta > |\beta| \\ \text{Det}(J_\mu) &= J_{11}J_{22} - J_{12}J_{21} > 0 \quad \text{if } 1 \geq \delta > |\beta|, |\alpha|, \end{aligned}$$

where determinant of the matching's function Jacobian reads

$$\begin{aligned} J_{11}J_{22} - J_{12}J_{21} &= \frac{16}{2Z\sqrt{1-\rho_x^2}} \underbrace{[(1-\rho_x^2)((1+\alpha\rho_y)(\delta+\rho_y\beta) - (\rho_y+\alpha)(\rho_y\delta+\beta))]}_{\equiv X} \\ &\quad + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2} \underbrace{[(1+\alpha\rho_y)(1+\rho_x\beta) + (\delta+\alpha\rho_x)(\delta+\rho_y\beta) + (\rho_y+\alpha)(\delta\rho_x+\alpha) + (\rho_x+\beta)(\rho_y\delta+\beta)]}_{\equiv Y} \\ &\quad + (1-\rho_y^2) \underbrace{[(\delta+\alpha\rho_x)(1+\rho_x\beta) - (\rho_x+\beta)(\delta\rho_x+\alpha)]}_{\equiv Z} \end{aligned}$$

where  $X$  and  $Z$  are positive under diagonal dominance and where  $Y$  can be expressed as:

$$Y = 1 + \alpha^2 + \beta^2 + \delta^2 + 2\beta\rho_x + 2\delta(\beta + \rho_x)\rho_y + 2\alpha(\delta\rho_x + \rho_y + \beta\rho_x\rho_y)$$

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

$$\begin{aligned} Y|_{\rho_x=\rho_y=1} &= ((1+\delta) + (\alpha+\beta))^2 > 0 \\ Y|_{\rho_x=\rho_y=-1} &= ((1+\delta) - (\alpha+\beta))^2 > 0 \\ Y|_{\rho_x=1, \rho_y=-1} &= ((1-\delta) - (\alpha-\beta))^2 > 0 \\ Y|_{\rho_x=-1, \rho_y=1} &= ((1-\delta) - (\alpha-\beta))^2 > 0 \end{aligned}$$

which proves the result. ■

## D.0.2 Non-Gaussian Distributions

This section states the sufficient condition for PAM

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

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

1. Step: 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 in line

with PAM can be specified as

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

Because of PAM, match up the marginals *within* each dimension

$$\begin{aligned} H_{x_C}(x_C) &= G_{y_C}(y_C) \\ H_{x_M}(x_M) &= G_{y_M}(y_M) \end{aligned}$$

which gives the assignment functions:

$$y_C = \frac{\bar{y} - \underline{y}}{\bar{x} - \underline{x}} x_C - x \frac{\bar{y} - \underline{y}}{\bar{x} - \underline{x}} + \underline{y} \quad (88)$$

$$y_M = \frac{\bar{y} - \underline{y}}{\bar{x} - \underline{x}} x_M - x \frac{\bar{y} - \underline{y}}{\bar{x} - \underline{x}} + \underline{y} \quad (89)$$

Both (88) and (89) 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  $\frac{\partial y_C}{\partial x_C} \frac{\partial y_M}{\partial x_M} - \frac{\partial y_C}{\partial x_M} \frac{\partial y_M}{\partial x_C} > 0$ . Consider (b). A PAM allocation is given by  $y_C = x_C$  and  $y_M = x_M$ , which is clearly feasible.

2. Step: The PAM allocation from Step 1 satisfies the firms' necessary second-order conditions for optimality under the P-matrix property of  $D_{xy}^2 y$ . Recall from the proof of Lemma 1 that the Hessian of the firm's problem is given by:

$$H^* = D_{xx}^2 F(\mathbf{x}, \mathbf{y}^*) - D_{xx}^2 w(\mathbf{x}) = -(D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*))(D_x \mathbf{y}^*) \quad (90)$$

In the PAM allocations from Step 1,  $D_x \mathbf{y}^*$  is a diagonal matrix. Since  $D_{xy}^2 F$  is a P-matrix, the matrix product  $(D_{xy}^2 F(\mathbf{x}, \mathbf{y}^*))(D_x \mathbf{y}^*)$  is positive-definite and, hence, the Hessian (90) is negative-definite.

3. Step: The PAM allocation from Step 1 satisfies the integrability condition (6). Hence, there exists a unique wage schedule supporting this allocation. To see this, first focus on (a). Since  $D_x \mathbf{y}^*$  is diagonal, (6) collapses to

$$F_{x_C y_M} \frac{\partial y_M^*}{\partial x_M} = F_{x_M y_C} \frac{\partial y_C^*}{\partial x_C} \quad (91)$$

which must hold at the equilibrium path. Using (88) and (89), this simplifies condition (91) to

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

which holds under the assumption of positive-definiteness of  $D_{xy}^2 F$  ( $F_{x_C y_M} = F_{x_M y_C}$ ) along the equilibrium path.



Consider (b). Under the assignment  $y_C = x_C$  and  $y_M = x_M$ , (6) collapses to

$$F_{x_C y_M} = F_{x_M y_C}$$

which again holds under the assumption of positive-definiteness of  $D_{xy}^2 F$  ( $F_{x_C y_M} = F_{x_M y_C}$ ) along the equilibrium path. Hence, for both (a) and (b), there exists a unique wage schedule that support the PAM allocation from Step 1.

4. Step. Since  $D_{xy}^2 F$  is a P-matrix everywhere, the equilibrium is globally unique (Proposition 1). Hence, the PAM allocation from Step 1 is the unique equilibrium. ■

## E Quantitative Analysis

### E.1 Construction of Skill and Productivity Distributions

As far as the productivity distribution is concerned (i.e. the distribution of  $y$ 's), I use the data by Sanders (2012) who classifies occupational skill requirement into two categories, manual and cognitive. He then aggregates this large amount of information, using Principal Component Analysis, to get two task scores for each occupation (i.e.  $y_C$  and  $y_M$ ). Using the described procedure, 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  $(y_C, y_M)$ -bundle from my model. I drop the observations whose  $(y_C, y_M)$ -bundles are missing. Table 3 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. Summary statistics of the  $(y_C, y_M)$ -bundles are given in the Online Appendix.

Occupation	Cognitive score ( $y_C$ )	Manual score ( $y_M$ )
Ushers, Lobby Attendants, and Ticket Takers	.1846	.3149
Telephone Operators	.2994	.1383
File Clerks	.3190	.3099
Legal Secretaries	.3796	.0731
Brickmasons, Blockmasons, Stonemasons	.1705	.8360
Helpers–Pipelayers, Plumbers, Pipefitters, Steamfitters	.1759	.6792
Helpers–Carpenters	.1984	.7187
Dancers	.3374	1
Radiologic Technicians	.4280	.6470
Machinists	.4303	.7152
Physical Therapist Assistants	.4758	.5494
Electricians	.4879	.8146
Economists	.6149	.0334
Public Relations and Fundraising Managers	.6199	.0587
Judges, Magistrate Judges, and Magistrates	.6752	.0517
Physicists	1	.1113

Table 3: Examples of Occupations' Manual and Cognitive Skill Requirements

To construct the skill distribution, I proceed as follows. College, apprenticeships and training qualify workers for particular occupations. I match the data on college degrees and apprenticeships to occupations, using standard cross-walks. Then, the  $(y_C, y_M)$ -bundles from the O\*NET data can be used 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 on-the-job training), then he also holds the skills required for that occupation. For instance, if a worker holds a degree in economics and the occupation *economist* has skill requirements  $(y_C = 1.34, y_M = -1.58)$ , then he holds the skills  $(x_C = 1.34, x_M = -1.58)$ .

To construct the skill distribution in a given year, I use each worker's 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 the data.<sup>54</sup> To obtain a single cognitive and a single manual skill from the skill vectors, I take the *maximum* skills from the vectors.<sup>55</sup> For instance, if a worker who is economist by training (with skills  $(x_C = 1.34, x_M = -1.58)$ ) had a previous career as a dancer (with skills  $(x_C = -0.25, x_M = 2.2)$ ), then he holds the skill bundle  $(x_C = 1.34, x_M = 2.2)$ , implying that he is qualified for both jobs with high cognitive and jobs with high manual skill requirements.

In the NSLY, there are observations to whom I cannot assign any skills either because there is no information on training or degrees. These agents might be educated and the information is simply missing. Or they are low-skilled and do not have any degrees. In many cases, the latter is true: Skills are missing for low-skilled workers who have never received a degree or training. The number of observations that are not assigned any skills from education or training is non-negligible. During the period considered (1992-2000), this affects 20%-30% of the NLSY observations.

If the skill information is missing but if the workers have received *some* education, I assign them cognitive and manual skills through a random draw from the distribution of skills of similarly educated people. For instance, if the skill data of a worker with high school degree in the US is missing, he gets a random draw of cognitive and manual skills of the distribution of other high school graduates who do have skill data.<sup>56</sup> On the other hand, if the worker with missing skill information has no education at all (high school drop out), then I assign to him the lowest cognitive and the lowest manual skill from the data set.<sup>57</sup> Finally, to reduce the discreteness of the skill distributions and 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. Similarly, for the productivity distribution.

---

<sup>54</sup>I do not use the training received in the current year to avoid an extremely high correlation between skills and skill requirements, which might be mechanical if a worker receives some training in nearly every job.

<sup>55</sup>Taking averages leads to similar results in the analysis.

<sup>56</sup>I 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. However, drawing the missing skills from the *lower* part of the distribution of other high school graduates leads to similar results in the following analysis.

<sup>57</sup>To decrease the discreteness of the distribution, I let them randomly draw their manual and cognitive skills from below the tenth percentile of the economy's marginal skill distributions. Nothing in the analysis hinges on the *tenth* percentile. The reason why I chose it is to assign low skills to low-skilled workers.

## E.2 Maximum Likelihood Estimation

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

$$w = \frac{1}{2}\alpha J_{11}x_C^2 + \alpha J_{12}x_Cx_M + \frac{1}{2}\beta J_{22}x_M^2 + \lambda x_C + \eta x_M + w_0 + \epsilon_w$$

$$y_C = J_{11}x_C + J_{12}x_M + \epsilon_C$$

$$y_M = J_{21}x_C + J_{22}x_M + \epsilon_M$$

where I assume measurement errors  $\epsilon_w, \epsilon_C, \epsilon_M$  with  $\epsilon_w \sim N(0, s^2)$ ,  $\epsilon_C \sim N(0, t^2)$ ,  $\epsilon_M \sim N(0, u^2)$ . Then,

$$w|x_C, x_M \sim N\left(\frac{1}{2}\alpha J_{11}x_C^2 + \alpha J_{12}x_Cx_M + \frac{1}{2}\beta J_{22}x_M^2 + \lambda x_C + \eta x_M + w_0 + \epsilon_w, s^2\right) \quad (92)$$

$$y_C|x_C, x_M \sim N(J_{11}x_C + J_{12}x_M, t^2) \quad (93)$$

$$y_M|x_C, x_M \sim N(J_{21}x_C + J_{22}x_M, u^2). \quad (94)$$

Denote the parameter vector by  $\theta = (J_{11}, J_{12}, J_{21}, J_{22}, \alpha, \beta, \lambda, \eta, w_0, s, t, u)$  and the data vector  $z = (z_1, \dots, z_n)$  where  $\forall i = 1, \dots, n, z_i = (w_i, y_{Ci}, y_{Mi}, x_{Ci}, x_{Mi})$ ;  $n$  is the number of observations. Due to conditional independence of  $w, y_C, y_M$  given  $[x_C, x_M, \alpha, \beta, J_{11}, J_{12}, J_{21}, J_{22}, \lambda, \eta, w_0]$ , the likelihood function is given by

$$\begin{aligned} L(\theta|z) &= \prod_{i=1}^n [w_i, y_{Ci}, y_{Mi}|x_{Ci}, x_{Mi}, \alpha, \beta, J_{11}, J_{12}, J_{21}, J_{22}, \lambda, \eta, w_0] \\ &= \prod_{i=1}^n [w_i|x_{Ci}, x_{Mi}, \alpha, \beta, J_{11}, J_{12}, J_{21}, J_{22}, \lambda, \eta, w_0] \\ &\quad \times \prod_{i=1}^n [y_{Ci}|x_{Ci}, x_{Mi}, J_{11}, J_{12}, J_{21}, J_{22}] \times \prod_{i=1}^n [y_{Mi}|x_{Ci}, x_{Mi}, J_{11}, J_{12}, J_{21}, J_{22}] \end{aligned} \quad (95)$$

From (95), one obtains (18) when using (92)-(94) and taking logs. The parameter estimates are obtained by maximizing (18) with respect to  $\theta$ .

	$\alpha$	$\beta$	$\lambda$	$\eta$	$w_0$	$J_{11}$	$J_{12}$	$J_{21}$	$J_{22}$	s	t	u
1992	2.7291 (0.6090)	1.5009 (0.7244)	2.7962 (0.1668)	0.2079 (0.1677)	15.1680 (0.2380)	0.3627 (0.0178)	-0.0957 (0.0173)	-0.1687 (0.0175)	0.3309 (0.0176)	8.6916 (0.1151)	0.9150 (0.0121)	0.9137 (0.0121)
1993	1.9176 (0.5815)	0.6535 (0.7168)	2.9093 (0.1685)	-0.0327 (0.1677)	15.7141 (0.2405)	0.3824 (0.0175)	-0.1125 (0.0172)	-0.1954 (0.0173)	0.3311 (0.0172)	8.8081 (0.1163)	0.9051 (0.0120)	0.9025 (0.0119)
1994	1.1832 (0.5479)	0.3048 (0.8175)	3.0329 (0.1683)	-0.0196 (0.1682)	16.3309 (0.2386)	0.4202 (0.0170)	-0.1045 (0.0170)	-0.2409 (0.0172)	0.2957 (0.0173)	8.8550 (0.1176)	0.8929 (0.0119)	0.9096 (0.0121)
1996	3.4585 (0.7968)	2.6730 (1.1946)	3.8010 (0.2396)	-0.4512 (0.2379)	15.9802 (0.3398)	0.4238 (0.0188)	-0.1342 (0.0182)	-0.2573 (0.0190)	0.2935 (0.0192)	11.3905 (0.1670)	0.8890 (0.0130)	0.9088 (0.0133)
1998	3.3866 (0.6682)	1.2076 (0.9832)	3.8139 (0.2173)	0.2278 (0.2170)	17.2972 (0.3117)	0.4680 (0.0168)	-0.0773 (0.0164)	-0.2756 (0.0173)	0.3198 (0.0173)	11.1818 (0.1515)	0.8643 (0.0117)	0.8931 (0.0121)
2000	3.1358 (1.0472)	0.8954 (0.8028)	4.7352 (0.2641)	-0.1773 (0.2632)	18.4752 (0.3792)	0.4870 (0.0168)	-0.0986 (0.0167)	-0.2695 (0.0170)	0.3675 (0.0171)	13.4909 (0.1842)	0.8624 (0.0118)	0.8761 (0.0119)

Standard errors in parentheses

Table 4: Maximum Likelihood Estimates for Years 1992-2000

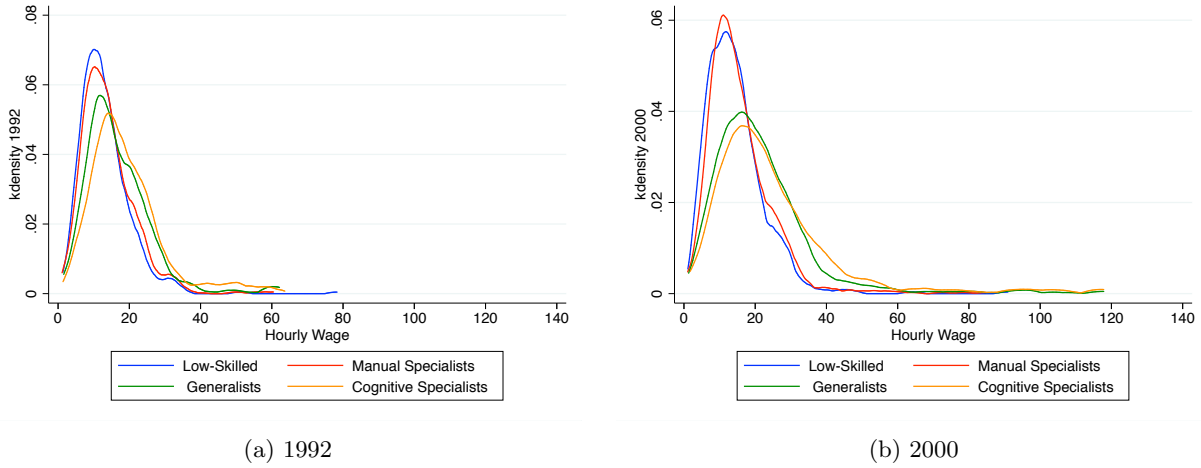


Figure 10: Wage Densities by Worker Group (Data)

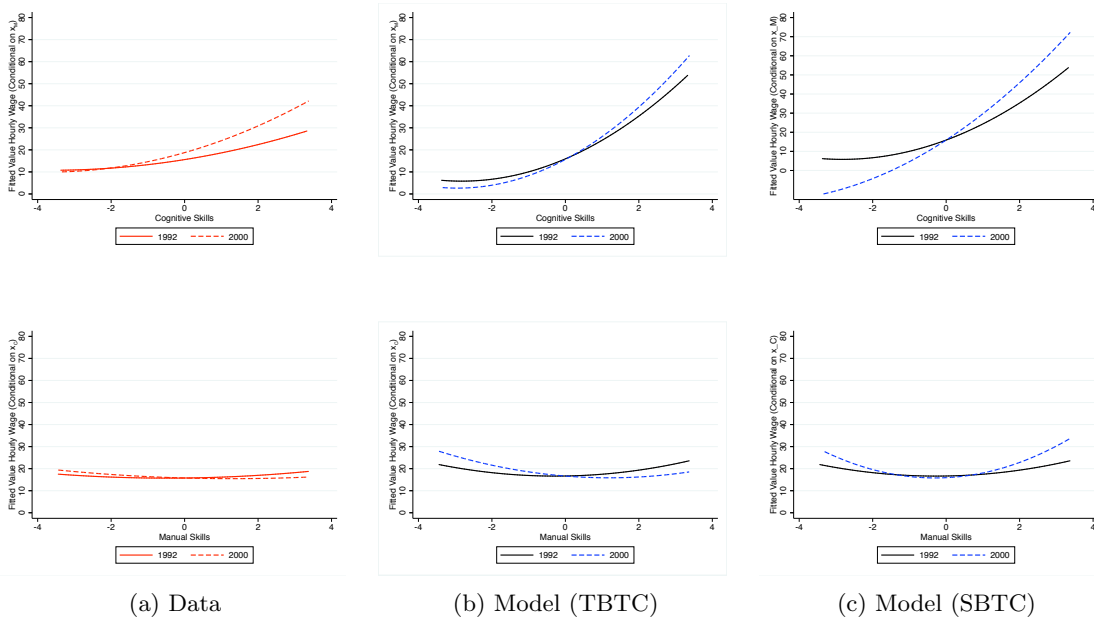


Figure 11: (De)Convexification of Wages 1992-2000: Data and Model (TBTC and SBTC). This figure displays the wage as a function of cognitive skills (upper panels) and manual skills (lower panels), comparing data and model, where the model is simulated separately for task-biased TC and skill-biased TC. To construct these figures, I use conditional wages. For instance, in the upper panels, the plotted wage is the residual wage from regression wages on manual skills. Task-biased TC gets the shifts in the curvature of the returns qualitatively right. In line with the data, task-biased TC makes cognitive (but not manual) returns more convex. To the contrary, skill-biased TC causes cognitive returns to become simply steeper (but not more convex) and manual returns to become more convex.

## References

- John M. Abowd, Francis Kramarz, and David N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–334, March 1999. URL <http://ideas.repec.org/a/ecm/emetrp/v67y1999i2p251-334.html>.
- Daron Acemoglu and David Autor. *Skills, Tasks and Technologies: Implications for Employment and Earnings*, volume 4 of *Handbook of Labor Economics*, chapter 12, pages 1043–1171. Elsevier, 2011. URL <http://ideas.repec.org/h/eee/labchp/5-12.html>.
- David H. Autor and David Dorn. The growth of low skill service jobs and the polarization of the u.s. labor market. *American Economic Review (forthcoming)*, 2012. URL <http://ideas.repec.org/p/nbr/nberwo/15150.html>.
- David H. Autor, Frank Levy, and Richard J. Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333, November 2003. URL <http://ideas.repec.org/a/tpr/qjecon/v118y2003i4p1279-1333.html>.
- David H. Autor, Lawrence F. Katz, and Melissa S. Kearney. The polarization of the u.s. labor market. *American Economic Review*, 96(2):189–194, May 2006. URL <http://ideas.repec.org/a/aea/aecrev/v96y2006i2p189-194.html>.
- Gary S Becker. A theory of marriage: Part i. *Journal of Political Economy*, 81(4):813–46, July-Aug. 1973. URL <http://ideas.repec.org/a/ucp/jpolec/v81y1973i4p813-46.html>.
- Michael J. Boehm. The wage effects of job polarization: Evidence from the the wage effects of job polarization: Evidence from the allocation of talents. *Working Paper*, 2013.
- Raicho Bojilov and Alfred Galichon. Closed-form solutions for multivariate matching. 2013.
- James N. Brown and Harvey S. Rosen. On the estimation of structural hedonic price models. NBER Technical Working Papers 0018, National Bureau of Economic Research, Inc, May 1982. URL <http://ideas.repec.org/p/nbr/nberte/0018.html>.
- Pierre-André Chiappori, Robert McCann, and Lars Nesheim. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, 42(2):317–354, February 2010. URL <http://ideas.repec.org/a/spr/joecth/v42y2010i2p317-354.html>.
- Pierre-André Chiappori, Sonia Orefice, and Climent Quintana-Domeque. Matching with a handicap: The case of smoking in the marriage market. Working Paper, 2012. URL <http://ideas.repec.org/p/hka/wpaper/2011-017.html>.
- Eugene Choo and Aloysius Siow. Who marries whom and why. *Journal of Political Economy*, 114(1): 175–201, February 2006. URL <http://ideas.repec.org/a/ucp/jpolec/v114y2006i1p175-201.html>.

- L. Chua and Y. Lam. Global homeomorphism of vector-valued functions. *Journal of Mathematical Analysis and Applications*, 39(3):600–624, 1972.
- Arnaud Costinot and Jonathan Vogel. Matching and inequality in the world economy. *Journal of Political Economy*, 118(4):747–786, 08 2010. URL <http://ideas.repec.org/a/ucp/jpolec/v118y2010i4p747-786.html>.
- Colin Decker, Elliott H. Lieb, Robert J. McCann, and Benjamin K. Stephens. Unique equilibria and substitution effects in a stochastic model of the marriage market. *Journal of Economic Theory*, 148(2):778–792, 2013. URL <http://ideas.repec.org/a/eee/jetheo/v148y2013i2p778-792.html>.
- Deniz Dizdar and Benny Moldovanu. Surplus division and efficient matching. Working Paper, University of Bonn, 2012.
- Arnaud Dupuy and Alfred Galichon. Personality traits and the marriage market. IZA Discussion Papers 6943, Institute for the Study of Labor (IZA), October 2012. URL <http://ideas.repec.org/p/iza/izadps/dp6943.html>.
- Jan Eeckhout and Philipp Kircher. Identifying sorting—in theory. *Review of Economic Studies*, 78(3):872–906, 2011. URL <http://ideas.repec.org/a/oup/restud/v78y2011i3p872-906.html>.
- Jan Eeckhout and Philipp Kircher. Assortative matching in large firms: Span of control over more versus better workers. *Working Paper*, 2012.
- Ivar Ekeland. Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types. *Economic Theory*, 42:275–315, 2010.
- Ivar Ekeland, James J. Heckman, and Lars Nesheim. Identification and estimation of hedonic models. *Journal of Political Economy*, 112(S1):S60–S109, February 2004. URL <http://ideas.repec.org/a/ucp/jpolec/v112y2004is1ps60-s109.html>.
- David Gale and Hukukane Nikaido. The jacobian matrix and global univalence of mappings. *Mathematische Annalen*, 159(2):81–93, 1965.
- Alfred Galichon and Bernard Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. CEPR Discussion Papers 7858, C.E.P.R. Discussion Papers, June 2010. URL <http://ideas.repec.org/p/cpr/ceprdp/7858.html>.
- Neil E Gretskey, Joseph M Ostroy, and William R Zame. The nonatomic assignment model. *Economic Theory*, 2(1):103–27, January 1992. URL <http://ideas.repec.org/a/spr/joecth/v2y1992i1p103-27.html>.
- Daniel Hershkowitz and Nathan Keller. Spectral properties of sign symmetric matrices. *Electronic Journal of Linear Algebra*, 13:90–110, 2005.

- Edward P. Lazear. Firm-specific human capital: A skill-weights approach. *Journal of Political Economy*, 117(5):914–940, October 2009. URL <http://ideas.repec.org/a/ucp/jpolec/v117y2009i5p914-940.html>.
- Jan R. Magnus. The moments of products of quadratic forms in normal variables. *Statistica Neerlandica*, 32(4):201–210, 1978.
- Robert J. McCann, Xianwen Shi, Aloysius Siow, and Ronald P. Wolthoff. Becker meets ricardo: Multisector matching with social and cognitive skills. IZA Discussion Papers 6533, Institute for the Study of Labor (IZA), April 2012. URL <http://ideas.repec.org/p/iza/izadps/dp6533.html>.
- I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48(0):257–263, 1982. ISSN 0024-3795. doi: [http://dx.doi.org/10.1016/0024-3795\(82\)90112-4](http://dx.doi.org/10.1016/0024-3795(82)90112-4). URL <http://www.sciencedirect.com/science/article/pii/0024379582901124>.
- Theodore Papageorgiou. Learning your comparative advantage. *Review of Economic Studies*, forthcoming, 2013.
- Carl Sanders. Skill uncertainty, skill accumulation, and occupational choice. *Working Paper*, 2012.
- L.S. Shapley and M. Shubik. The assignment game i: The core. *International Journal of Game Theory*, 1:111–130, 1971. ISSN 0020-7276. doi: [10.1007/BF01753437](http://dx.doi.org/10.1007/BF01753437). URL <http://dx.doi.org/10.1007/BF01753437>.
- Alp Simsek, Asuman Ozdaglar, and Daron Acemoglu. On the uniqueness of solutions for nonlinear and mixed complementarity problems. Working Paper, 2005.
- Jan Tinbergen. On the theory of income distribution. *Weltwirtschaftliches Weltwirtschaftliches Archiv*, 77(2):155–73, 1956.
- Lawrence Uren and Gabor Virag. Skill requirements, search frictions, and wage inequality. *International Economic Review*, 52(2):379–406, 05 2011. URL <http://ideas.repec.org/a/ier/iecrev/v52y2011i2p379-406.html>.
- C. Villani. *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften. Springer London, Limited, 2009. ISBN 9783540710509. URL [http://books.google.es/books?id=hV8o5R7\\_5tkC](http://books.google.es/books?id=hV8o5R7_5tkC).
- Robert J Willis and Sherwin Rosen. Education and self-selection. *Journal of Political Economy*, 87(5): S7–36, October 1979. URL <http://ideas.repec.org/a/ucp/jpolec/v87y1979i5ps7-36.html>.
- Shintaro Yamaguchi. Tasks and heterogeneous human capital. *Journal of Labor Economics*, 30(1):1–53, 2012. URL <http://ideas.repec.org/a/ucp/jlabec/doi10.1086-662066.html>.