Essays on Heterogeneity in Labor Markets

Ilse Lindenlaub

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, June 2014

European University Institute
Department of Economics

Essays on Heterogeneity in Labor Markets

Ilse Lindenlaub

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Nicola Pavoni, Università Bocconi (Supervisor)
Prof. Jérôme Adda, European University Institute
Prof. Jan Eeckhout, University College London
Prof. Omiros Papaspiliopoulos, Universitat Pompeu Fabra

Abstract

In my thesis, I study the effects of agents' heterogeneity on labor market outcomes, with particular focus on sorting, performance, wages, and inequality.

Chapter one 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 in worker-job *complementarities* in cognitive relative to manual inputs leads to more pronounced sorting and wage inequality across cognitive relative to manual skills. This can trigger wage polarization and boost aggregate wage dispersion. I then estimate the model for the US during the 1990s and show that cognitive-biased technological change (i.e. increases in worker-job complementarities in cognitive inputs and in cognitive skill-bias) can account for observed changes in worker-job sorting, wage polarization and a significant part of the increase in US wage dispersion.

Chapter two develops a theory that links differences in men's and women's social networks to disparities in their labor market performance. We are motivated by our empirical finding that men's and women's networks differ. Men have a higher degree (more network links) than women, but women have a higher clustering coefficient (a woman's friends are also friends among each other). In our model, a worker with a higher degree has better access to information. In turn, a worker with higher clustering faces more peer pressure. Both peer pressure and access to information can attenuate a team moral hazard problem in the work place. But whether peer pressure or access to information is more important depends on the work environment. We find that, in environments where uncertainty is high, information is crucial and, therefore, men outperform women – in line with findings from sectors with high earnings' uncertainty like the financial or film industry.

Acknowledgements

I would like to express my special appreciation and thanks to my supervisors and outstanding economists, Jan Eeckhout and Nicola Pavoni. Thanks, Jan, for adopting me as your student at UPF, for your unconditional support and guidance in my research, during the job market and in everything else – you are a tremendous mentor. Thanks, Nicola, for your tireless efforts and support despite the geographical distance, and especially for your dedication – I have learned a lot from you. I am also extremely grateful to my committee member, Omiros Papaspiliopoulos from UPF, for countless hours of discussion (the main project of my thesis on multidimensional matching would have not been possible without). I thank my second supervisor at the EUI, Jérôme Adda, especially for his support during the first years of my PhD that I spent at the EUI. I am grateful to David Levine for his support during the job market, and to Vasco Carvalho and Hector Chade for very helpful comments and discussions.

I would like to thank my co-author of the second chapter, Anja Prummer, for her excellent work. I thank my parents for being there – always – and all the great people and close friends who were part of this episode.

Thanks also to the staff of the EUI Economics Department, especially to Jessica Spataro and Lucia Vigna, for great support even during my time at UPF in Barcelona.

Finally, I would like to acknowledge the financial support provided by the German Academic Exchange Service and the Italian Ministry for Education.

Introduction

In my thesis, I analyze the role of individual-level *heterogeneity* in labor market outcomes. Heterogeneity of workers and jobs is an important aspect of the labor market. Individuals are born with different abilities, they are raised in distinct environments, have diverse educational backgrounds and work experiences as well as different social network structures. The jobs that are available to these individuals also differ in a range of attributes, namely in skill requirements, productivity, riskiness and in the stream of income they generate. Everyone chooses the job that best suits his characteristics, giving rise to many heterogeneous worker-job matches with varying performance. In my thesis, I study the effects of agents' (multidimensional) heterogeneity on labor market outcomes, with particular focus on worker-job sorting, performance, wages, and inequality.

Chapter one 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 in worker-job *complementarities* in cognitive relative to manual inputs leads to more pronounced sorting and wage inequality across cognitive relative to manual skills. This can trigger wage polarization and boost aggregate wage dispersion. I then estimate the model for the US and identify sizeable technology shifts: during the 1990s, worker-job complementarities in cognitive inputs increased by 15% whereas complementarities in manual inputs decreased by 41%. In addition to 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.

Chapter two develops a theory that links differences in men's and women's social networks to disparities in their labor market performance. We are motivated by our empirical finding that men's and women's networks differ. Men have a higher degree (more network links) than women, but women have a higher clustering coefficient (a woman's friends are also friends among each other). In our model, a worker with a higher degree has better access to information. In turn, a worker with a higher clustering coefficient faces more peer pressure. We show that both features have a significant impact in labor settings where information and incentives are important and highlight the trade-off between them in a model of teams: Someone with more information can better adjust his effort to the expected project value. In turn, more peer pressure always induces higher effort, thereby mitigating the team moral hazard problem independent of the expected reward. One of our main results is that information is more beneficial for performance when the uncertainty about the project value is large while peer pressure is more valuable in the opposite case. We therefore expect men to outperform women especially in jobs that are characterized by high earnings uncertainty, for instance in the financial sector or film industry – in line with the evidence.

Contents

1 Sorting Multidimensional Types: Theory and Application

1.1 Introduction

Technological progress has drastically changed the task composition of work and hence the structure of labor demand. Across the board, workers spend less time performing *manual* tasks such as assembling cars and more time performing *cognitive* tasks such as computer programming or selling products and services.[1] During the 1980s, a blue-collar worker in the car industry might have spent some time on cognitive tasks such as reporting to his supervisor, but he mainly engaged in manual labor on the assembly line. Ten years later, a newly-developed machine carries out his manual task. Programming the machine requires more cognitive than manual skills, and thus a different skill mix than 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 that extends the unidimensional notion of positive assortative matching. I derive the equilibrium allocation as well as equilibrium wages in closed form. I use this explicit solution to analyze the impact on equilibrium outcomes as cognitive (as opposed to manual) inputs become more important in production, capturing one of the main recent technological shifts. I then take this model to the data to study technological change in the US during the 1990s. Using this theoretical framework of multidimensional sorting, I can infer from data on observed equilibrium outcomes the degree to which underlying technological determinants have changed over time, and I can study their effects 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 cognitive versus manual tasks.[2] *Task-biased technological change*, which increases the *level* of complementarities between cognitive skills and skill demands (relative to those in the manual dimension), puts this trade-off to work. Sorting improves along the cognitive

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

[2] The meaning of complementarities in this context is that workers with high cognitive skills are particularly productive in jobs that put significant weight on cognitive ability, and similar in the manual dimension.

dimension but the opposite is true in the manual dimension, where matches are characterized by a poorer fit between workers' skills and job demands. In light of the previous example, the blue-collar worker who was replaced by a machine may now be employed as a car salesman. This new job is tailored to his cognitive skills but a poor fit with his manual abilities. The new allocation benefits workers with high cognitive abilities but harms those with manual know-how. This makes wages more convex in cognitive but less convex in manual skills, thereby fueling wage inequality along the cognitive dimension but compressing inequality in the manual dimension.

I estimate this model for the US and identify sizeable technological shifts: I find that during the 1990s, 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 affected the productivity of skills *independent* of the task, leaving worker-job complementarities unchanged.

The key findings are that these technological shifts may account for both observed wage polarization (i.e. stagnant lower tail but expanding upper tail wage inequality) and much of the increase in wage dispersion. More precisely, counterfactual exercises show that task-biased technological change can account for wage polarization. The reason this technology shift affects upper and lower tail wage inequality differently is that winners (i.e. workers with high cognitive skills) are clustered in the upper part of the wage distribution while those adversely affected (workers with mainly manual skills) are concentrated in the lower part. In turn, cognitive skill-biased technological change, which does not affect the curvature of the wage schedule, fuels inequality across *the whole* distribution. It can account for a significant part of the increase in US wage dispersion over the 1990s.

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

One advantage of these one-dimensional models is their tractability. However, it is important to note that collapsing agents' multiple characteristics into a single index is not innocuous. A notable study that rejects the single index model is by Willis and Rosen (1979). They show that worker performance depends on a *bundle* of different skills including

[3]See, for instance, Costinot and Vogel (2010) and Acemoglu and Autor (2011). In the literature, task-biased technological change is often referred to as *routinization*, meaning that new machines replace those workers performing routine tasks (e.g. Autor et al. (2003), Autor et al. (2006), Autor and Dorn (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.

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 agents are ranked and matched.[4]

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, existing studies provide little insights into the characteristics of the equilibrium and comparative statics.[5] This paper makes a first attempt at developing a tractable framework that allows for both.

Section 1.2 introduces the general theoretical framework. I develop an assignment model where workers and jobs match in pairs. Workers possess manual and cognitive skills. Each worker performs two tasks, a manual and a cognitive one. Jobs, in turn, differ in productivities or skill demands for each task. Within this task-based framework, I propose a generalization of positive assortative matching (PAM) and negative assortative matching (NAM) to the 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 similarly 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 forms of the production technology.

To study biased technological change, one ideally has a closed form solution that is amenable to comparative statics and estimation. Toward this goal, Section 1.3 specifies the environment to Gaussian distributions and linear-quadratic technology. Using this notion of assortative matching, I develop a technique to solve for equilibrium assignment and wage function in closed form.

It is important to note that notwithstanding 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 technology is not sufficient to pin down the output-maximizing PAM allocation. Instead, the

[4]More recently, results by Papageorgiou (2013) also favor the specialization hypothesis over a single index model.

[5]In non-technical terms, the optimal transport problem involves finding 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).

parametric specification of the production function (i.e. the *relative level* of complementarities across tasks and not only their *signs*) 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 focuses on *complementarities*. Worker-job complementarities determine the optimal PAM allocation from many existing ones. They range from strong assortativeness to significant mismatch between worker and job traits in one or both task(s), capturing a much richer set of assignments than one-dimensional PAM.

Section 1.4 uses the closed form to analyze task-biased technological change, which demonstrates how these matching patterns (and ulitmately wages) are shaped by technology. I also contrast these results with those for more standard skill-biased technological change. The latter only increases the relative productivity of workers' cognitive *skills* without affecting worker-job complementarities.

Section 1.5 brings this model to the data. I focus on the US economy during the 1990s. I first construct 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 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 adverse shocks to manual skills.

Section 1.6 places the main contribution of the paper into the literature. Section 1.7 concludes. The Appendix contains all proofs, data details and estimation results.

1.2 Theoretical Framework for Multidimensional Sorting

Toward the goal of developing a theoretical framework for multidimensional sorting, this section outlines the general model absent specific assumptions about 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.[6]

1.2.1 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

[6] All proofs for this section are given for N-dimensional heterogeneity.

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 of cognitive task C (respectively manual task M). Points in Y represent firm types. Denote the joint c.d.f. of (y_C, y_M) by $G(y_C, y_M)$, which is also assumed to be absolutely continuous. Assume that overall masses of firms and workers coincide.[7]

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

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

1.2.2 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 $\nu : \mathbb{R}^2_+ \to \mathbb{R}^2_+$, which can be uniquely characterized by its inverse $\mu \equiv \nu^{-1}$. I call μ 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 better firms. This concept captures two aspects: (a) purity of matching (i.e. $\mu(x)$ is one-to-one), and (b) direction of sorting. Here, I aim to define a multidimensional version of assortative matching that also incorporates these two features. As in the one-dimensional setting, here assortativeness involves properties of the first derivative of the matching function (i.e. of its Jacobian), given by:

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

I define multidimensional positive and negative assortative matching as follows:

Definition 1.1 (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*

$$[i] \quad \frac{\partial y_C^*}{\partial x_C} > (<) \quad 0 \qquad [ii] \quad \frac{\partial y_M^*}{\partial x_M} > (<) \quad 0 \qquad [iii] \quad Det(J_\mu) = \frac{\partial y_C^*}{\partial x_C}\frac{\partial y_M^*}{\partial x_M} - \frac{\partial y_C^*}{\partial x_M}\frac{\partial y_M^*}{\partial x_C} > 0 \tag{1.1}$$

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.[8] In economic terms, PAM means that intellectual types work in firms where workers need to perform complex intellectual tasks

[7] Otherwise, there is equilibrium unemployment or idle firms, which unnecessarily complicates the model.

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

(part [i]). Similarly, workers with strong manual skills work in firms that attach considerable weight to the manual task (part [ii]). Moreover, inequality [iii] dictates that these *within*-task matching forces dominate *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 1.1 also captures the fact that the assignment is *pure*, defined as follows:

Definition 1.2 (Pure Matching)**.** *Matching is pure if μ 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.[9] Gale and Nikaido (1965) link the P-matrix property of the Jacobian of a function to the function's injectivity, giving a sufficient condition for purity in the current setting: if $D_x\mathbf{y}^*$ is a P-matrix (or P^--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 ν.[10]

Definition 1.1 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 (1.1)) and *purity* of the assignment (guaranteed by the determinant condition [iii]). In both the one-dimensional and multidimensional settings, PAM implies purity.

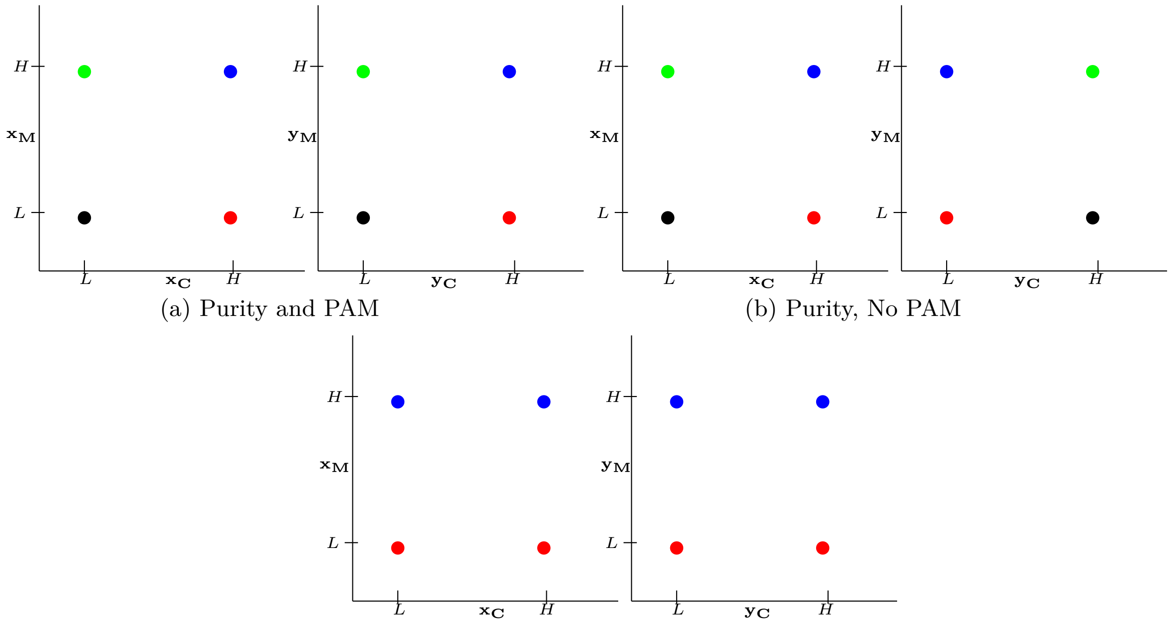
The figure below provides a graphical illustration of multidimensional PAM, using a discrete 2x2 example: Each side of the market has two attributes that can be high (H) or low (L). Hence, there are four worker and four firm types. In each subfigure, the left panel represents worker types and the right panel 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.

1.2.3 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

[9] 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).

[10] 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 $\mathbb{R}^n$.



(a) Purity and PAM

(b) Purity, No PAM

(c) No Purity, No PAM

this property. The firm's problem is given by:

$$\max_{(x_C, x_M) \in X} F(x_C, x_M, y_C, y_M) - w(x_C, x_M) \tag{1.2}$$

The FOCs of this maximization problem read

$$F_{x_C}(x_C, x_M, y_C, y_M) - w_{x_C}(x_C, x_M) = 0 \tag{1.3}$$

$$F_{x_M}(x_C, x_M, y_C, y_M) - w_{x_M}(x_C, x_M) = 0 \tag{1.4}$$

where subscripts denote derivatives. Equations (1.3) and (1.4) hold only at the equilibrium assignment.

1.2.4 The Equilibrium

I focus on a competitive equilibrium, which is defined as follows.

Definition 1.3 (Equilibrium)**.** *An equilibrium is characterized by a matching function* $\mu : X \to Y$*, and a wage function* $w : X \to \mathbb{R}_+$*, satisfying:*
(i) Optimality: Price-taking firms maximize profits (1.2) by choosing (x_C, x_M) *for a given* $w(x_C, x_M)$*. (ii) Market Clearing: Feasibility of* μ *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 a Walrasian equilibrium in the continuous assignment problem was proven in Gretsky et al. (1992) (Theorem 4) and, for slightly modified environments, in Chiappori et al.

(2010) and Ekeland (2010).[11] Notice that these papers do not guarantee that the equilibrium will be *differentiable*. Nevertheless, in what follows, I focus on a differentiable equilibrium where μ is C^1 and w is C^2. I show below that the differentiable equilibrium exists in various multivariate environments with absolutely continuous distributions. I will discuss some of them, with particular emphasis on the case with Gaussian distributions.

1.2.5 The Equilibrium Assignment

This section relates properties of the production technology to 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 technology features the following complementarities

$$D^2_{xy}F = \begin{bmatrix} \overset{+}{F_{x_C y_C}} & 0 \\ 0 & \overset{+}{F_{x_M y_M}} \end{bmatrix} \tag{1.5}$$

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.[12] Moreover, under the same condition, the assignment is a global maximum:

Proposition 1.1 (Assortativeness and Global Maximum)**.** *If $D^2_{xy}F$ is a diagonal P-matrix (P^--matrix), then the equilibrium assignment satisfies PAM (NAM), and is globally unique.*

The proof is in Appendix 1.8.1. 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 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 in a positive assortative way: Agents with strong intellectual skills work in firms that value these skills (and similarly for the manual dimension).[13]

This sorting result, which ensures a positive relation between skills and productivities along *natural* dimensions (i.e. within cognitive and manual tasks), 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 manual productivity and cognitive skill, $\frac{\partial y_M^*}{\partial x_C} \neq 0$, and also between cognitive skill demands and 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

[11] 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.

[12] The presented condition is related to the *twist condition* from optimal transport but is not equivalent. See Section 1.6.

[13] Similarly, in the case of NAM, assortative matching within tasks dominates assortativeness across tasks, only in this case high productivity workers are matched with low productive firms.

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 substitutatabilities ($F_{x_C y_M}, F_{x_M y_C} \neq 0$). In Section 1.8.5 of the Appendix, I show that a weaker version of (1.5) 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 strict monotonicity of matching function μ and the sorting direction by its positive slope. Similarly in this model, I impose conditions on the matrix of cross-partials $D^2_{xy}F$ to obtain PAM. The difference is that with multiple dimensions not only the signs but also the relative magnitudes of different complementarities need to be restricted in order to ensure assortative matching.

1.2.6 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, (1.3) and (1.4), 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 partial derivatives and obtained by cross-differentiating (1.3) and (1.4), when evaluated at (y^*_C, y^*_M):

$$\begin{aligned} & w_{x_C x_M} = w_{x_M x_C} \\ \Leftrightarrow \quad & F_{x_C y_C}\frac{\partial y^*_C}{\partial x_M} + F_{x_C y_M}\frac{\partial y^*_M}{\partial x_M} = F_{x_M y_C}\frac{\partial y^*_C}{\partial x_C} + F_{x_M y_M}\frac{\partial y^*_M}{\partial x_C}. \end{aligned} \tag{1.6}$$

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 1.2 (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 (1.6).*

The proof relies on *Frobenius' Theorem*. Both theorem and proof are stated in the Appendix 1.8.1. Integrability condition (1.6) has technical and economic implications. Technically, given

(1.6), there exists a C^2 wage function w, justifying the differentiation-based approach above.

Condition (1.6) also carries an important economic message. It highlights a crucial difference between multidimensional and one-dimensional settings. With multiple dimensions, there is a stronger link between technology and assignment. The equilibrium assignment (i.e. the Jacobian of the matching function) does not only depend on the signs of the cross partial derivatives, $F_{x_i y_j}, i, j \in \{C, M\}$, 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.

In the one-dimensional setting, there is no integrability condition because the wage is the solution to a *single* ordinary differential equation. In such a setting, the assignment depends only on the sign of $F_{x_i y_i}$, not on its level: supermodularity (submodularity) of the technology implies PAM (NAM). Given PAM (NAM), there exists a *unique* measure-preserving increasing (decreasing) map of skills to productivities, which can be pinned down by labor market clearing alone. Under PAM, this map is given by $y = G^{-1}(H(x))$. 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.

1.3 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 takes an important step toward achieving this objective. It specifies the environment to Gaussian distributions and quadratic technology and develops a technique to compute the multidimensional assignment and corresponding wage explicitly. The closed form solution then allows me to focus on the economics of multidimensional sorting, characterizing equilibrium properties and analyzing 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.

1.3.1 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

$$F(x_C, x_M, y_C, y_M) = \alpha x_C y_C + \beta x_M y_M = \alpha(x_C y_C + \delta x_M y_M) \tag{1.7}$$

where α and β 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.[14]

Technology (1.7) captures that there is *within*-task complementarity but *between*-task complementarity is shut down. Based on the results in Section 1.2, certain properties of the equilibrium assignment are already known at this point without having to check second-order conditions of the firm's problem.[15] Under (1.7), $D^2_{xy}F$ 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 for

$$F(x_C, x_M, y_C, y_M) = \alpha x_C y_M + \beta x_M y_C + \gamma x_C y_C + \delta x_M y_M \tag{1.8}$$

allowing for non-zero between-task complementarity (Appendix 1.8.5 with the assignment given by (1.87)-(1.90)). Moreover, I can allow for non-standard normally distributed variables or even arbitrary marginal distributions that are linked via Gaussian copulas (see Appendix 1.8.7). 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. Appendix 1.8.2 provides the details.

1.3.2 The Equilibrium Assignment Functions

The objective is to compute 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 δ. 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 as follows:

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 μ and nonsingular covariance ma-

[14] Nothing hinges on this restriction but it simplifies interpretation, and moreover, is in line with the data.

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

trix Σ. Then,

$$\mathbf{z} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu) \tag{1.9}$$

has mean $\mathbf{0}$ and covariance matrix $\mathbf{I_p}$. 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 Σ_x (respectively Σ_y) the covariance matrix of skills (respectively productivities). Apply (1.9) to the standard bivariate normal skills and productivities

$$\mathbf{z_x} = \begin{bmatrix} z_{x_C} \\ z_{x_M} \end{bmatrix} = \Sigma_x^{-\frac{1}{2}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \quad \text{and} \quad \mathbf{z_y} = \begin{bmatrix} z_{y_C} \\ z_{y_M} \end{bmatrix} = \Sigma_y^{-\frac{1}{2}} \begin{bmatrix} y_C \\ y_M \end{bmatrix} \tag{1.10}$$

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 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*

$$(1 - \Phi(z_{y_C}))(1 - \Phi(z_{y_M})) = (1 - \Phi(z_{x_C}))(1 - \Phi(z_{x_M})) \tag{1.11}$$

where Φ again denotes the standard normal c.d.f. The interpretation of (1.11) is that if 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).[16]

The market clearing condition (1.11) implicitly defines the vector-valued matching function of transformed variables, denoted by $\mu^z : \mathbb{R}^2 \to \mathbb{R}^2$. 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

$$\Phi(z_{y_i}) = \Phi(z_{x_i}) \qquad \forall \quad i \in \{C, M\} \tag{1.12}$$

which gives a system of two equations. In principle, there are many possible ways to match up the marginals in (1.11) but due to PAM (i.e. positive diagonal elements of $D_x\mathbf{y}^*$), this is the only sensible way. System (1.12) can be 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

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \tag{1.13}$$

where $D_x\mathbf{y}^* = \Sigma_y^{\frac{1}{2}} \Sigma_x^{-\frac{1}{2}}$ is the Jacobian of the matching function. System (1.13) is the candidate

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

equilibrium assignment, mapping bivariate skills into bivariate productivities. By (1.11), 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 a firm's optimal choice, which depends on the relative level of skill-productivity complementarities, captured by δ.[17] The Appendix shows how $\Sigma_y^{\frac{1}{2}}\Sigma_x^{-\frac{1}{2}}$ can be parameterized by δ, such that the resulting assignment is consistent with the firm's optimality for any level of complementarities across tasks.

Proposition 1.3 (Equilibrium Assignment)**.** *The equilibrium assignment μ is given by*

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \Sigma_y^{\frac{1}{2}}\Sigma_x^{-\frac{1}{2}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} = \underbrace{\begin{bmatrix} J_{11}(\rho_x,\rho_y,\delta) & J_{12}(\rho_x,\rho_y,\delta) \\ J_{21}(\rho_x,\rho_y,\delta) & J_{22}(\rho_x,\rho_y,\delta) \end{bmatrix}}_{D_x\mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \tag{1.14}$$

where $J_{11}, J_{12}, J_{21}, J_{22}$ are given in (1.56) in the Appendix.

Remark 1.1. *For $\delta = 1$: $J_{11} = J_{22}$ and $J_{12} = J_{21}$. For $\delta = 0$: $J_{22} \neq J_{11} = 1$ and $J_{21} \neq J_{12} = 0$. For $\delta \in (0,1)$: the assignment lies in between these 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 1.8.2 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 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 matching function only depends on underlying distributions (through skill correlation ρ_x and productivity correlation ρ_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 δ.

[17]This is done by taking into account the integrability condition (1.6), under which a wage schedule exists that induces firms to choose this assignment. With this bilinear technology, (1.6) collapses to $\frac{\partial y_C^*}{\partial x_M} = \delta \frac{\partial y_M^*}{\partial x_C}$.

1.3.3 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 1.4 (Equilibrium Wage Schedule). *The equilibrium wage function is given by*

$$w(\mathbf{x}) = \frac{1}{2}\alpha \mathbf{x}^T \tilde{J}\mathbf{x} + w_0 \tag{1.15}$$

where w_0 is the constant of integration.

See Appendix 1.8.3 for the proof and explicit expression. $\tilde{J}$ is a matrix of parameters closely related to the equilibrium assignment. It contains the assignment coefficients from the Jacobian of the matching function, J_μ. 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 me to compute the moments of the wage distribution in closed form. The next sections extensively discuss the properties of the wage function and how they depend on distributions and technology through the assignment.

As a final note on the equilibrium, notice that, not surprisingly, the one-dimensional equilibrium is subsumed by the two-dimensional model as a special case: For perfect correlations, $\rho_x, \rho_y \to 1$, the Jacobian of the matching function J_μ in (1.56) becomes the identity matrix (i.e. the assignment is independent of the complementarity weights in the production function) and the wage collapses to a simple quadratic function in skill.

1.3.4 Properties of the Equilibrium

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

Definition 1.4 (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 a firm's skill requirements for 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 1.5 (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 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 skills needed and skills supplied ($|\rho_x - \rho_y| \to 2$), then the labor market can only clear under considerable mismatch, with every worker being in a job for which he is either under or overqualified.

These results are illustrated in Figure 1.1, which displays contour plots of two standard normal distributions for various skill and productivity correlations. For the sake of illustration, assume that workers are represented by blue contour lines and firms by 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.

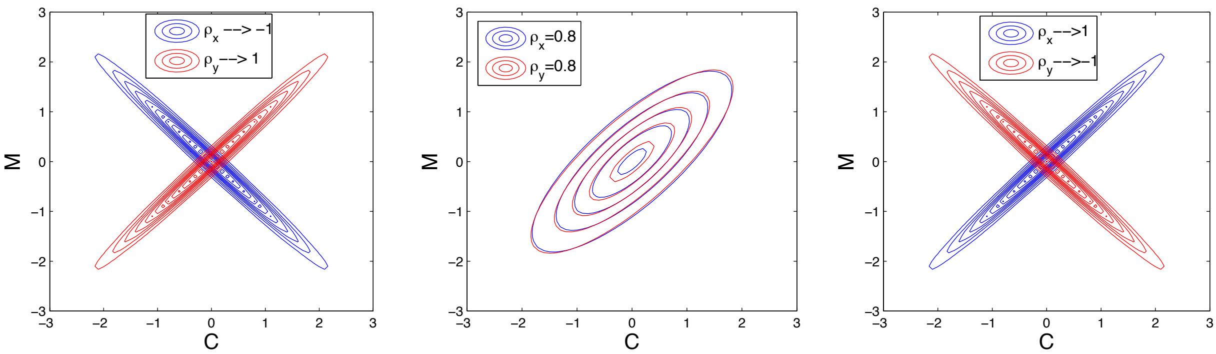


Figure 1.1: Contour Plots of Skill and Productivity Distributions

The Jacobian of the matching function offers an alternative graphical way of looking at equilibrium sorting properties. In the graph below, I plot 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). Assortativeness forces are given by the diagonal elements of the Jacobian J_μ (J_{11} and J_{22}), and 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 underlying distributions are identical ($\rho_x = \rho_y$). The lower panel displays the other extreme. Here 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 panels 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 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.

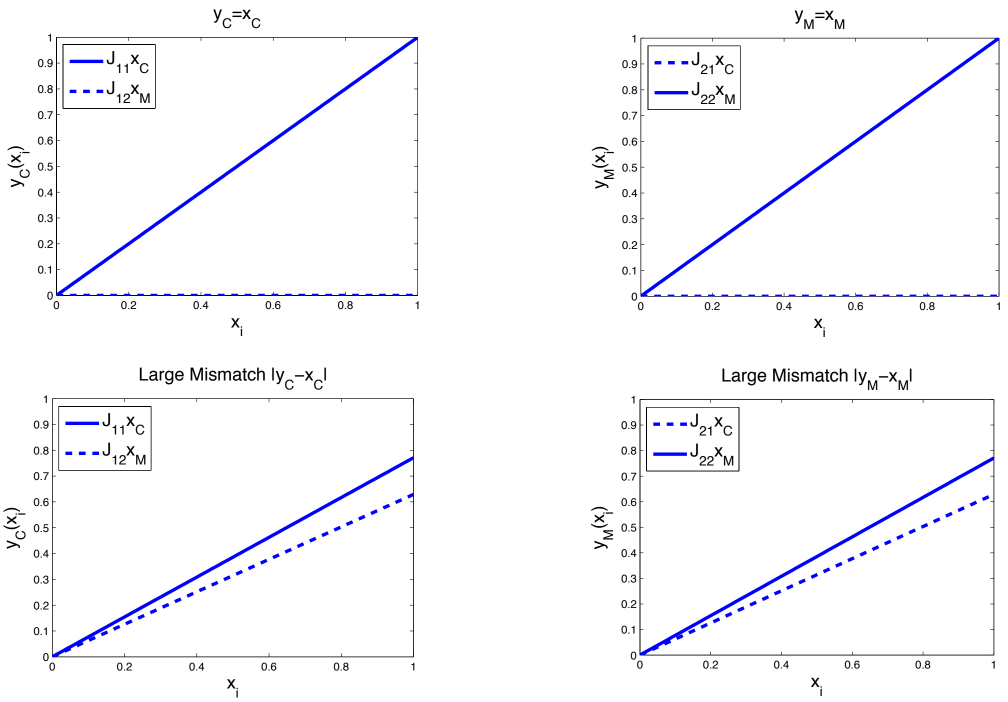


Figure 1.2: Perfect Assortativeness (upper panel) and Mismatch (lower panel) within PAM

The next result summarizes a selected set of properties of the wage function.

Proposition 1.6 (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 (1.15) is a symmetric P - matrix or positive definite) wages are convex. Convex wages mean that workers with large (absolute) quantities of skills earn disproportionally more than workers with small (absolute) quantities 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 were negative assortative, the wage function would be concave.[18]

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 disproportionally much. The force behind positive skewness is again PAM, which is the driving factor of wage inequality in an economy.

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

This section illustrated how sorting depends on an 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 1.3 that the assignment not only depends on distributions but also on *technology* (through the relative level of firm-worker complementarities across tasks). I will use the closed form and the developed sorting framework to examine 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?

1.4 Biased Technological Change

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

1.4.1 Task-Biased Technological Change

Task-biased TC is viewed as 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 performance of manual tasks but created stronger complementarities between skills and job attributes in cognitive tasks. The literature also refers to this technological change as *routinization*, where workers performing routine tasks are increasingly substituted by computers and machines.[19] Notice that task-biased TC does not imply that the prevalence of routine tasks in the production process has diminished over time – quite the opposite (Acemoglu and Autor (2011)). What has changed is the technology to perform them.

[18] 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)).

[19] 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.

Even though two intrinsically different skills are involved (manual and cognitive), task-biased TC is analyzed in the literature 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.[20]

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 relative complementarities in the manual task. Consider a change from δ to δ' such that $\delta' < \delta = 1$. Then, δ' is called *task-biased* relative to δ, with the bias favoring the cognitive task. Moreover, to obtain clean analytical results, I will focus on cases where $\rho_x, \rho_y \leq 0$ or $\rho_x, \rho_y \geq 0$.[21] The next result summarizes the effect of task-biased TC on the equilibrium assignment.

Proposition 1.7 (Task-Biased TC and Sorting)**.** *Suppose there is cognitive task-biased TC ($0 < \delta' < \delta = 1$): (i) Sorting becomes more (less) pronounced in the cognitive (manual) task (i.e. $|y_C - x_C|$ decreases, $|y_M - x_M|$ increases). (ii) As $\delta \to 0$, perfect assortativeness is achieved in the cognitive task ($y_C = x_C$) but manual mismatch becomes maximal. (iii) 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 $|\rho_x - \rho_y|$), 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.

For most underlying distributions, the equilibrium assignment will be such that workers do not obtain their perfect job matches. This is because such a situation 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 technology and in particular on relative levels 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, computers perform several manual tasks, replacing 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 (i)). The amount of worker reallocation depends on both, the size of the shock (part(ii)) and underlying distributions (part(iii)). How the size of the shock matters is illustrated in the figure below, which has a similar structure to that in Figure 1.2. The upper panels plot cognitive sorting, i.e. y_C as a function of x_C and x_M before (left panel)

[20] This is similar to the *skill weights approach* by Lazear (2009).

[21] This can be relaxed but I would have to rely more on simulations. Moreover, the restriction captures the empirically relevant case for the US. See below.

and after task-biased TC (as δ goes to zero). The lower panels plot manual sorting, i.e. y_M as a function of x_M and x_C and have the same structure.

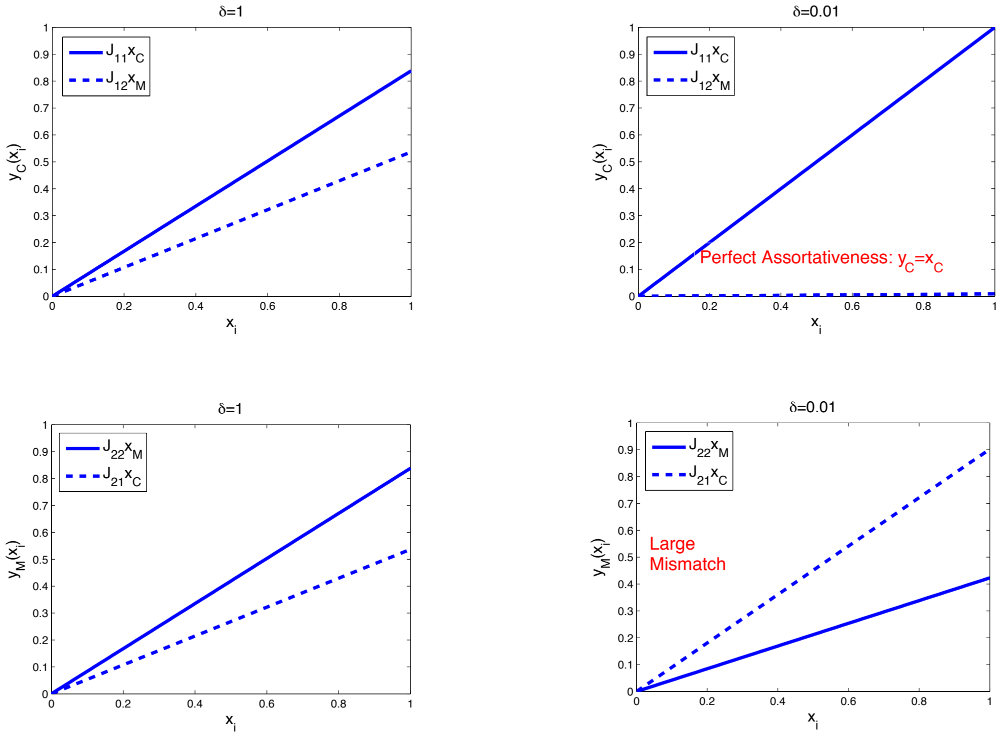


Figure 1.3: Effects of TBTC on Sorting in Cognitive (upper panel) and Manual Dimension (lower panel)

The slope of the straight lines indicates how strong the sorting forces are *within* tasks. 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 left to right, 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 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 (iii)). 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 δ 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 much to gain from improving cognitive sorting in response to task-biased TC.

It follows from this discussion that there are *two* sources of structural mismatch in the economy, technology and distributions. The first source stems from asymmetries in production technology (Proposition 1.7). The second is due to discrepancy between skill and productivity distributions or, in other words, between supply and demand ($\rho_x \neq \rho_y$, Proposition 1.5). 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 1.8 (TBTC and Wages)**.** *Suppose there is task-biased TC (*$0 < \delta' < \delta = 1$*):*
(i) Aggregate Wage Inequality: The effect on the wage variance is ambiguous.
(ii) Wage Curvature: If $|\rho_x| < |\rho_y|$*, wages become more convex in cognitive but less convex in manual skills. For a perfect fit of supply and demand (*$\rho_x = \rho_y$*), changes in the curvature are smallest. In turn, for the poorest fit (i.e. maximal* $|\rho_x - \rho_y|$*), 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 α and β. Since cognitive task-biased TC can either be driven by an increase in cognitive task-weight α or by a decrease in manual task-weight β, 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 (for the empirically relevant case $|\rho_y| \geq |\rho_x|$, see below). 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 many manual skills are hit most severely, compressing 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 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.[22] 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 from the one-dimensional setting is that in my model, there exist *generalists*. Generalists have a second (i.e. cognitive) skill, which offer them a shield against shocks to manual skills. They gain over manual specialists, who additionally lose relative to low-skilled workers and cognitive specialists in the economy (see Appendix 1.8.6 for formal statements).

[22] I think of manual (cognitive) specialists as workers who have manual (cognitive) but low cognitive (manual) skills.

1.4.2 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.[23] To make the model more suitable for empirical analysis, I augment the production technology by non-interaction skill terms and a constant, given by

$$F(x_C, x_M, y_C, y_M) = x_C(\alpha y_C + \lambda) + x_M(\beta y_M + \eta) + f_0 = \alpha(x_C y_C + \delta x_M y_M) + \lambda(x_C + \kappa x_M) + f_0 \tag{1.16}$$

where $\delta = \frac{\alpha}{\beta}$ is the relative manual task weight, λ, η are skill weights, $\kappa = \frac{\eta}{\lambda}$ is the relative manual skill weight and f_0 is a constant.[24] The assignment is unaffected by this new technology but the wage becomes a *non-homogenous* quadratic form in standard normal variables,

$$w(x_C, x_M) = \frac{1}{2}\alpha(\mathbf{x}-h)'\tilde{J}(\mathbf{x}-h) + C = \alpha\left(\frac{1}{2}J_{11}x_C^2 + J_{12}x_C x_M + \frac{1}{2}\delta J_{22}x_B^2\right) + \lambda(x_C + \kappa x_M) + w_0. \tag{1.17}$$

See Appendix 1.8.4 for the derivation and expressions $h, \tilde{J}, C$. Non-interaction skill terms can shift the location of the minimum wage to the left, allowing for a wage schedule that is increasing $\forall x_C, x_M \geq \underline{x}_C, \underline{x}_M$, where $\underline{x}_C, \underline{x}_M$ are, for instance, the lowest observed skills in the data. 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.[25]

Technology (1.16) gives rise not only to a more realistic wage schedule, it also allows for a sensible definition of *skill-biased technological change* (skill-biased TC hereinafter), independently of task-biased TC that works through complementarities in production. Consider a change in relative manual skill weight from κ to κ' with $\kappa' < \kappa$. Then, κ' is called *skill-biased* relative to κ, with the bias favoring cognitive skills. This shift increases the productivity of cognitive skills independent of a job's cognitive skill demands. For instance, advancements in communication technology (e.g. google) benefit both the secretary and the CEO even though their tasks require different levels of cognitive skill.

Proposition 1.9 (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 (1.17) it is clear that it also has no impact on the curvature of the wage function, which solely depends on task-bias parameters. Finally, similar to task-biased TC, the effect of skill-biased TC on the wage variance is ambiguous. However, for the empirically relevant case below with negative skill correlation ($\rho_x < 0$) and relatively large cognitive skill weight ($-\rho_x\lambda > \eta$), wage dispersion unambiguously increases with skill-biased TC. In this case, the decline in κ affects the variance

[23] Under the previous technology, the wage is folded around $(0, 0)$, e.g. workers (-1,-1) and (1,1) earn the same.

[24] Notice that including additional non-interaction productivity terms in the technology would not affect wages.

[25] With the previous technology, wages were always positive (see Appendix 1.8.6). However, when including non-interaction terms, wages can become negative. Hence, the inclusion of the constant.

mainly through the increase in cognitive skill weight λ, shifting the wage schedule without affecting its curvature. Compared to task-biased TC, skill-biased TC does not compress lower tail wage inequality but fuels inequality across the whole distribution. 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.

1.5 Quantitative Analysis

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

1.5.1 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, interviewed every year until 1994 and since then biennially. The reason for using the NLSY is that it contains detailed information on respondents' occupations, training and degrees, which I will use to construct a skill supply distribution. I supplement the NLSY by O*NET data to learn about occupational skill requirements. This data will be crucial for constructing a skill demand distribution, where I interpret *occupations* as the empirical counterpart of my model's firms.[26] The analysis in this paper covers the period 1992-2000.[27] 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 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 occupations' cognitive and manual skill requirements (y_C, y_M).

To construct these bivariate distributions, I rely heavily on the O*NET data, which provide 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 manual and cognitive tasks, which I interpret as the (y_C, y_M)-bundle from my model (see Table 1.4 in Appendix 1.8.8 for examples).[28] I then merge these scores into occupations of employed workers in the NLSY, which yields the bivariate skill demand distribution. Constructing

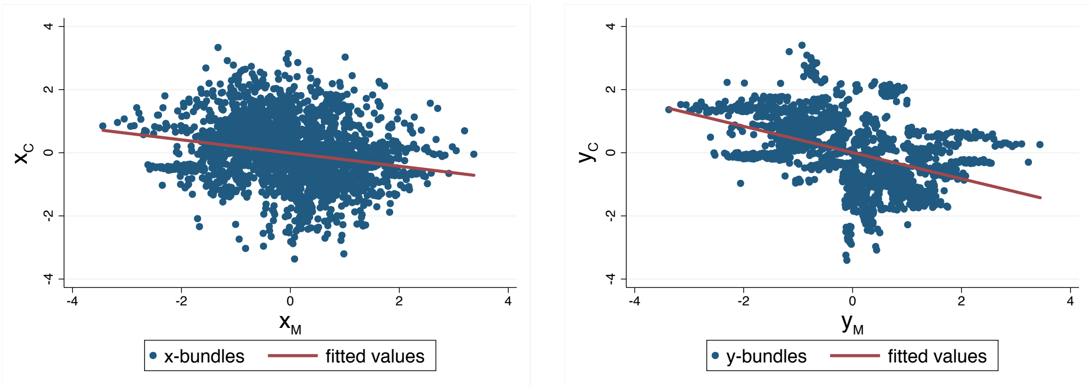
[26] The O*NET is the U.S. Department of Labor Occupational Characteristics Database.

[27] 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 manual and towards cognitive task inputs in production. Second, years beyond 2000 are excluded because NLSY occupations are recoded in 2002, which complicates the measurement of sorting.

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

the bivariate skill distribution is involved. Data on manual skills are not readily available. Moreover, the literature provides little guidance on this issue.[29] To impute agents' manual and cognitive skills, I use information on their college degrees, apprenticeships and vocational degrees, degrees of government programs and training on-the-job paid for by firms, provided by the NLSY.[30] From this information, I can proxy a manual and cognitive skill for each agent (see Appendix 1.8.8 for details). After data cleaning and sample restrictions, I am left with around 2700 yearly observations.

The Appendix provides summary statistics of bivariate skill and productivity distributions in 1992. In order to align the data with the model, I transform empirical skill and productivity distributions into *Gaussian copulas*, which takes out marginal characteristics (means and variances) and leaves the correlation as the only distributional parameter (see Appendix 1.8.7). The correlations between the transformed variables are plotted below (standard errors in parentheses). 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 higher degrees of specialization than 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.



$\underset{(0.0184)}{\rho_x = -0.2079}$ $\underset{(0.017)}{\rho_y = -0.415}$

Figure 1.4: US Skill and Productivity Correlations in 1992

1.5.2 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 $\theta = ((J_{11}, J_{12}, J_{21}, J_{22}), (\alpha, \beta, \lambda, \eta, w_0), (s, t, u))$,

[29] 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.

[30] I only consider training paid by a firm because it is presumably related to the occupation performed by the worker.

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 $z = (z_1, ..., z_n)$ where $\forall \quad i = 1, ...n, z_i = (w_i, y_{Ci}, y_{Mi}, x_{Ci}, x_{Mi})$. The log-likelihood function for this model is given by:

$$\ln L(\theta|z) = -\sum_{i=1}^{n} \frac{(w_i - (\frac{1}{2}\alpha J_{11}x_{Ci}^2 + \alpha J_{12}x_{Ci}x_{Mi} + \frac{1}{2}\beta J_{22}x_{Mi}^2 + \lambda x_{Ci} + \eta x_{Mi} + w_0))^2}{2s^2}$$

$$-\sum_{i=1}^{n} \frac{(y_{Ci} - (J_{11}x_{Ci} + J_{12}x_{Mi}))^2}{2t^2} - \sum_{i=1}^{n} \frac{(y_{Mi} - (J_{21}x_{Ci} + J_{22}x_{Mi}))^2}{2u^2} - n\ln(stu) - \frac{3n}{2}\ln 2\pi \tag{1.18}$$

See Appendix 1.8.8 for details. Notice that another advantage of this model is that all parameters are identified.[31] I estimate the model year by year. Appendix 1.8.8 reports detailed estimation results.

1.5.3 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.[32] 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 α, β are complementarity weights, λ, η are skill weights, and f_0 is a constant. Table 1.1 contains the ML-estimates of these technology parameters for the years 1992 and 2000. The estimation results suggest that production technology features complementarities between worker and job attributes in both tasks (α and β are positive; see Appendix 1.8.8 for the results of the remaining years). Moreover, the 1990s were characterized by task-biased TC in favor of cognitive tasks: Complementarities between cognitive worker and job attributes have gone up by 15% whereas complementarities in manual inputs have decreased by 41%. Relative manual complementarities, $\delta = \frac{\beta}{\alpha}$, dropped from 0.55 to 0.29 – a decline of 47%.

Besides these shifts in relative task complementarities, there was also a change in the skill-bias of technology, indicated by the skill weights η and λ. Over the 1990s, the US economy was characterized by a strong cognitive skill-biased TC. The cognitive skill weight λ

[31]My model circumvents 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 additional quadratic term in 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 technology $F(x, y)$ in x.

[32]See Abowd et al. (1999) and also Eeckhout and Kircher (2011).

	α	β	λ	η	w_0
1992	2.7291	1.5009	2.7962	0.2079	15.1680
	(0.6090)	(0.7244)	(0.1668)	(0.1677)	(0.2380)
2000	3.1358	0.8954	4.7352	−0.1773	18.4752
	(1.0472)	(0.8028)	(0.2641)	(0.2632)	(0.3792)

Standard errors in parenthesis.

Table 1.1: Maximum Likelihood Estimates of Technology Parameters

increased sharply (+68%) whereas the manual weight η decreased (however, η is statistically insignificant). In sum, these estimates suggest that during the 1990s, 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 distributions: workers were less specialized in 2000 compared to 1992 ($\rho_x = -0.2079$ in 1992 and $\rho_x = -0.05$ in 2000, see Appendix 1.8.8). On the other hand, the change in skill demand was negligible.

1.5.4 The Role of Technological Change in US Wage Inequality Shifts

Observed wage inequality shifts in the data can occur for many reasons. The advantage of estimating a structural model is that the effects of various sources can be disentangled. This section conducts counterfactual exercises to decompose the impact of task-biased TC, skill-biased TC and changes in underlying distributions on wage inequality. For instance, to study how much of the change in wage inequality is due to task-biased TC alone, I keep both skill-bias parameters λ and η as well as distributional parameters ρ_x and ρ_y at their 1992-levels and only feed the estimated changes in the task-bias into the model (given by α, β); similarly, for skill-biased TC and the change in distributions.

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 1.5 plots hourly wages by wage percentile for 1992 and 2000 (solid and dashed line, respectively), illustrating that inequality disproportionally increased in the upper part of the distribution with little action in the lower part. What might have caused this specific change in the wage distribution? Panels (a), (b) and (c) in Figure 1.6 analyze whether wage polarization can possibly be triggered by estimated technology shifts in task and skill-bias or distributional changes when fed into the model (blue curves).[33] 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 in panel (b) can only match the expanding upper tail inequality. It fails to account for stagnating inequality in the lower part of the distribution, overpredicting the increase in lower tail wage inequality. Finally, had only distributional

[33] Notice that I the model wage distribution exactly matches the data wage distribution in 1992, which is why I do not display the blue curves for 1992. They lie on top of the red solid lines.

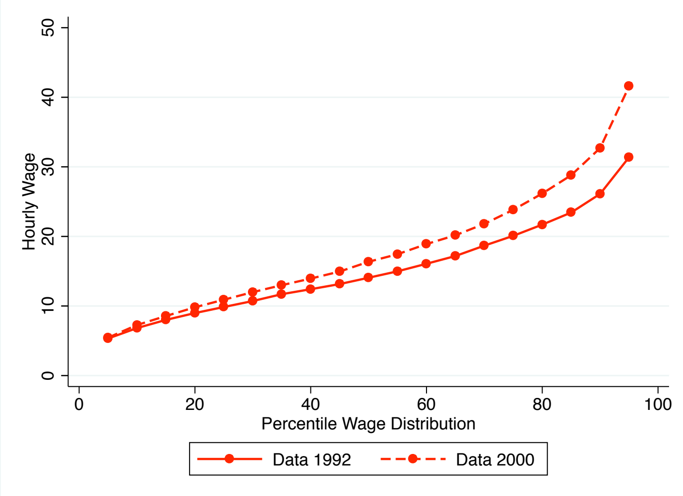


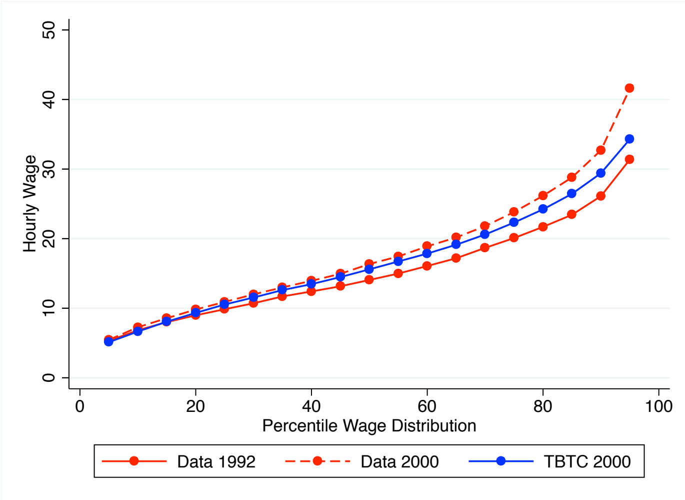
Figure 1.5: Wage Polarization in the US (Data)

changes taken place over the period of the 1990s, then the wage distribution would have simply shifted but not polarized (panel (c)).

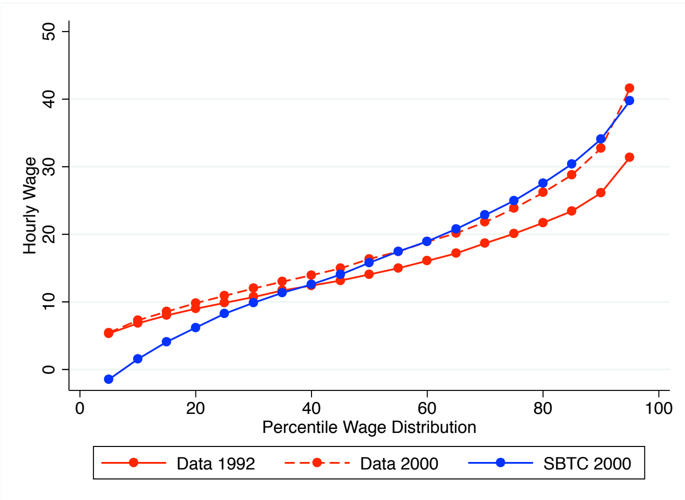
The model offers an explanation for why only task-biased TC can account for wage polarization. Through an increase in cognitive input complementarities (α goes up) and a decrease in manual input complementarities (β goes down), task-biased TC affects the *curvature* of the wage schedule. Wages become 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 a disproportionate increase in upper tail inequality.

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 1.7).[34] 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 1990s, 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 wage distributions of cognitive specialists and generalists over 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.[35]

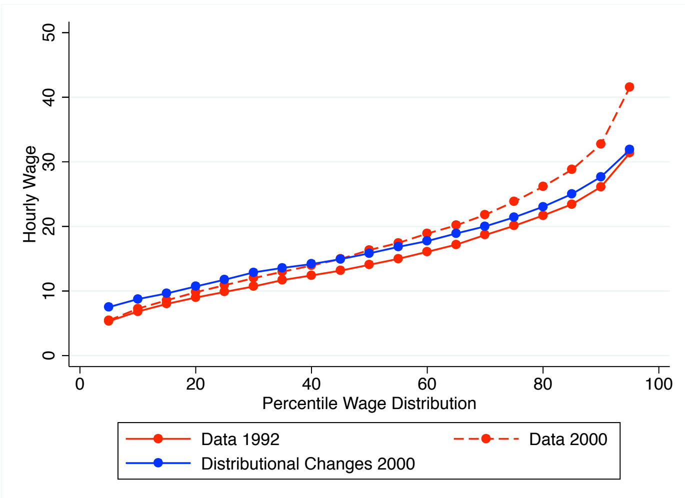
[34] Low-skilled are defined as $x_C < E(x_C), x_M < E(x_M)$, manual specialists as $x_C < E(x_C), x_M > E(x_M)$ etc.
[35] Figure 1.12 in Appendix 1.8.8 makes the same point with wage densities.



(a) Data and Model (Task-Biased TC)

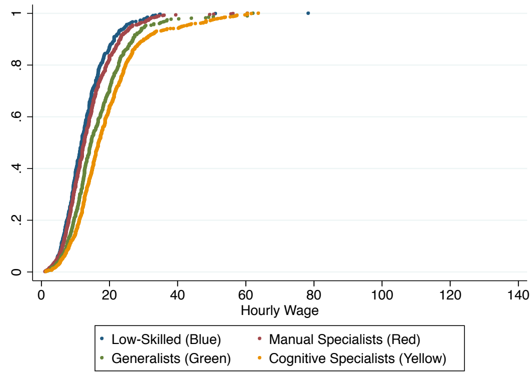


(b) Data and Model (Skill-Biased TC)

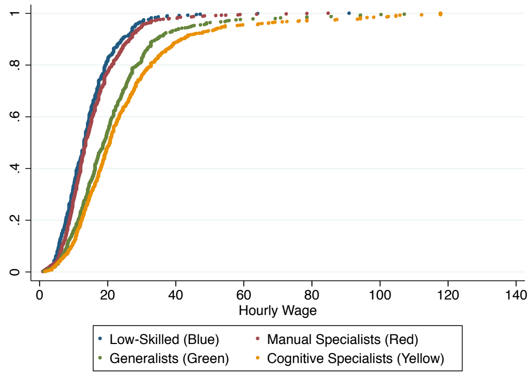


(c) Data and Model (Distributional Changes)

Figure 1.6: Wage Polarization 1992-2000: Data, Task-Biased TC and Skill-Biased TC, Distributional Shifts



(a) 1992



(b) 2000

Figure 1.7: Wage Distributions by Worker Group (Data)

Recall from Proposition 1.8 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_μ 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 1.8, 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 1.8.8 for the estimates. The structure is as in Figure 1.3: the slope of the solid lines indicates assortativeness (diagonal elements of J_μ, J_{11} and J_{22}) whereas the steepness of the dashed lines indicates mismatch (off-diagonal elements of J_μ, J_{12} and J_{21}).

Both in 1992 and 2000, sorting satisfies PAM with a positive relationship between skills and skill requirements in both tasks (given by positively sloped solid lines). This is in line with estimated worker-job *complementarities* (i.e. $\alpha, \beta > 0$).[a] 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. ρ_x and ρ_y) are fairly well aligned in 1992, my model predicts minor effects of task-biased TC on sorting.

[a]Moreover, the solid lines are steeper than the dashed lines, fulfilling the condition on the determinant of J_μ.

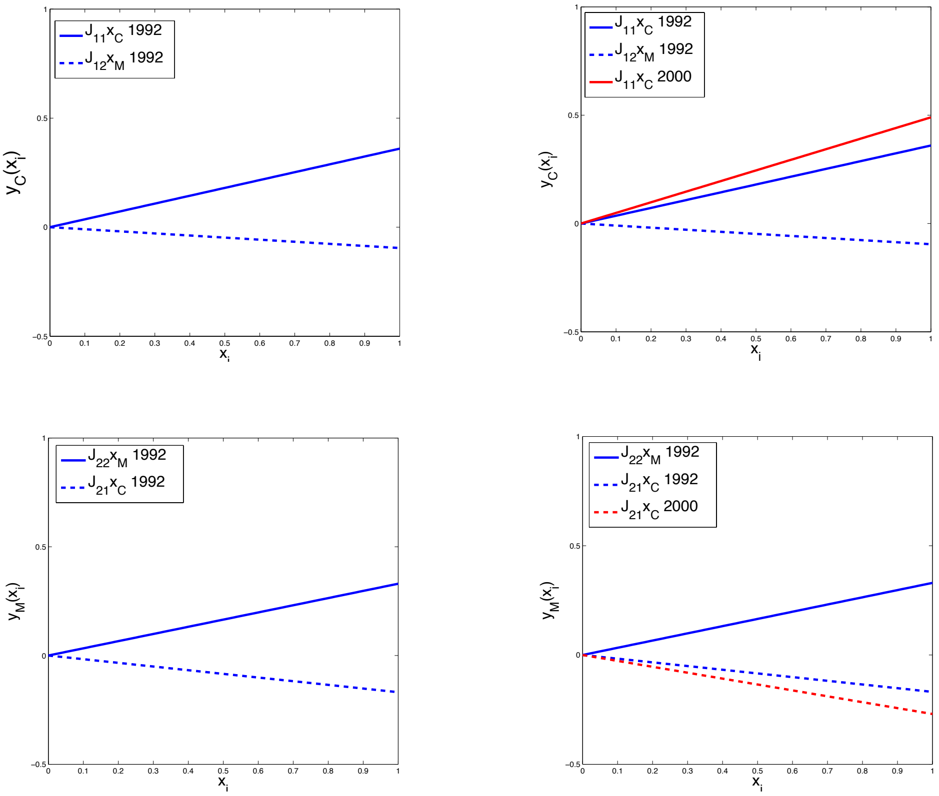


Figure 1.8: Maximum Likelihood Assignment Estimates 1992 and 2000: Cognitive Dimension (upper panel), Manual Dimension (lower panel)

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 1.13 in Appendix 1.8.8.) Finally, the distributional changes over the 1990s are small and were no driving force behind polarization.

Wage Dispersion

There has been a substantial increase in US *wage dispersion* during the 1990s (see Figure 1.9).The proposed model – despite being frictionless – does a decent job in matching this increase. It generates an 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 variance is the model off (the data variance in 1992 is 6.7 times higher than the model's 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 make a better comparison between the variance change in data and model, I normalize the model variance so that it has the same level as the data variance in 1992.

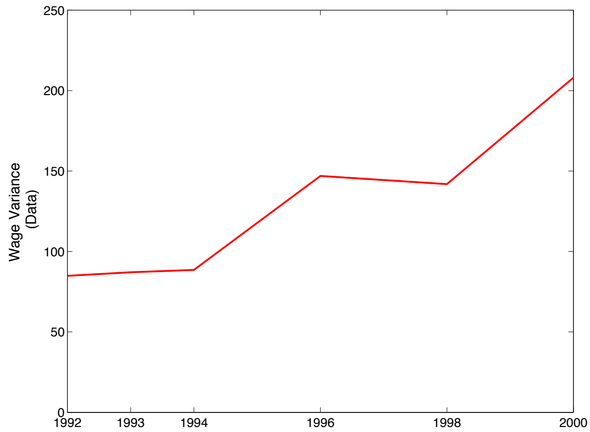
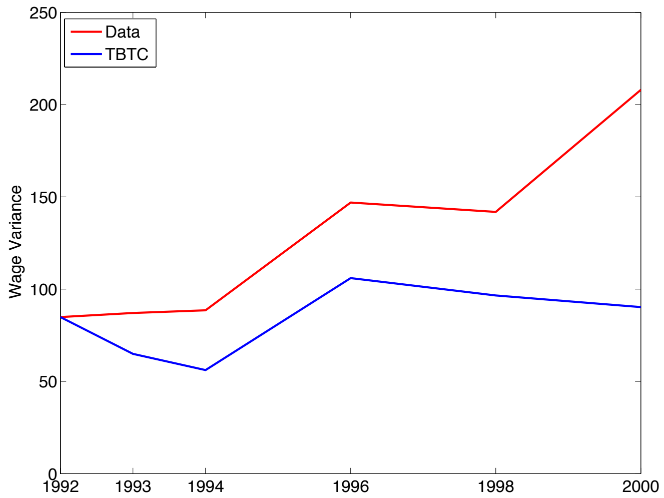


Figure 1.9: Evolution of Wage Dispersion in the US (Data)

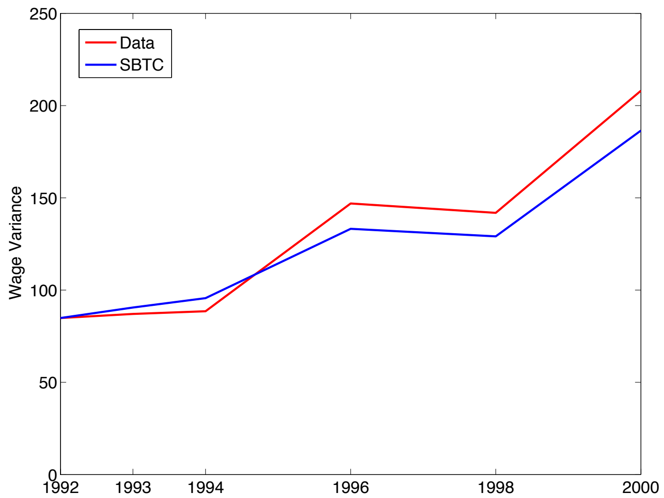
It can be shown that the driving force in wage inequality increase is the *within*-variance of manual workers, $E[Var(w|x_B)]$, as opposed to their *between*-variance, $Var(E[w|x_B])$. This multidimensional model offers a natural way to think about these concepts since, for instance, every group of workers with similar manual skills has a whole distribution of wages due to differences in their cognitive skills. Technological change in favor of cognitive inputs exacerbates this within-group wage dispersion. In contrast, between-wage dispersion of manual types contributes little to overall variance or its shift (indicating that the driving force of wage inequality is the cognitive and not the manual skill).[36]

Figure 1.10 gives clues into what drives the increase in wage dispersion: task-biased technological change, skill-biased technological or distributional changes. Table 1.2 reports the corresponding numerical decomposition for the shift in wage variance.

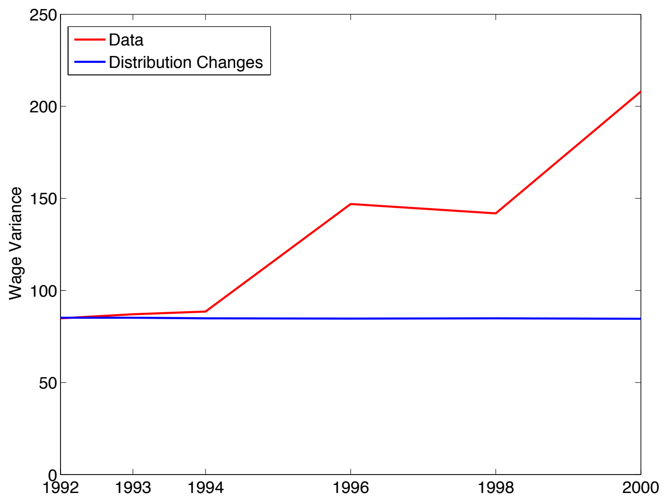
[36] The within and between variance predictions of the model can be computed since closed forms are available. Regarding the data, I first categorize workers into thirty bins depending on their manual skills, and then compute the wage variance within and across bins. The results are robust using more or fewer than thirty bins.



(a) Data and Model (Task-Biased TC)



(b) Data and Model (Skill-Biased TC)



(c) Data and Model (Distributional Changes)

Figure 1.10: Wage Dispersion 1992-2000: Data, Task-Biased TC, Skill-Biased TC, Distributional Shifts

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 of 7% over this period. The model offers two explanations to why task-biased TC played a minor role for wage dispersion. First, over the 1990s, the increase in cognitive task-weight α is accompanied but a strong decrease in manual task weight β. The first force fuels wage inequality while the second one compresses it, which is why the net effect of task-biased TC on wage dispersion is small. To the contrary, the changes in the skill bias are mainly due to the increase in cognitive skill weight λ, which shifts the wage schedule (instead of impacting its curvature) and fuels inequality across the whole distribution. 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 triggered more inequality.

	Data	Model	Task-Bias (α, β)	Skill-Bias (η, λ)	Distributions (ρ_x, ρ_y)	Trend (w_0)
$\Delta Var(w)$	+ 145%	+126%	+7%	+119%	-4%	–

Table 1.2: Change in Wage Variance over 1992-2000 (Data versus Model)

Besides technological progress, distributions also changed during the 1990s. There was a shift in skill supply, with workers becoming less specialized, but skill demand remained constant (see Appendix 1.8.8). At odds with the observed increase in inequality, the change in skill distribution had a negative effect on 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.[37] Skill-biased TC accounts for a significant portion 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.

1.5.5 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, commonly the single worker characteristic in one-dimensional settings.[38]

To ensure comparability of the two models, I assume standard normal distributions $x_C, y_C \sim N(0,1)$ and 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

$$w(x_C) = \alpha \frac{x_C^2}{2} + \lambda x_C + w_0 \tag{1.19}$$

[37]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.

[38]This is justified since I construct skills from educational attainment (i.e. degrees) and training data.

where w_0 is the constant of integration. I estimate parameters α, λ, w_0 by OLS, using (1.19). During the 1990s, α increased by 88% and λ by 74% (see the Appendix 1.8.8 for details), suggesting that technological change favored workers with high cognitive skills.

The one-dimensional model captures well the *convexification* in cognitive returns, indicated by an increase in α. 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 the fact that not all workers with manual skills suffer from cognitive-biased technological change. Looking at the data through a two-dimensional lens suggests that *generalists* (who hold above average skills in both dimensions) experienced a substantial real wage increase of 27% over the 1990s. Their second skill offers a buffer against shocks to manual skills. Notice that generalists form a sizeable group, almost one fifth of the US workforce in 1992 (see Appendix 1.8.8). 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 technology shifts so drastically that negative instead of positive assortative matching becomes optimal.[39] For changing (but still positive) α, 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 1990s.

1.6 Literature Review

This work contributes to literature of two types: that concerning multidimensional matching under transferable utility (including hedonic models and optimal transport); and that concerning task-biased technological change. I will now discuss those papers that are most relevant to my research.[40]

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 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. 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.

[39] This would be true if α switches from positive to negative, which is according to the estimates not the case.

[40] I do not discuss papers with non-transferable utility because there is little relation.

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 1.8.5). 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 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).[41] 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 complementarities in the surplus function from 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 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 characteristics that depends on 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.[42] 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 1.11 (b) in Appendix 1.8.5. Similarly, this model would capture sorting results from the environment specified in Eeckhout and

[41] 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.

[42] 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.

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 1.11 (a), Appendix 1.8.5). This suggests that the developed sorting framework is useful for deriving closed forms beyond the quadratic-Gaussian case.

Finally, this paper relates to the literature on optimal transport. In non-technical terms, the optimal transport problem involves finding a measure-preserving map that carries one distribution into another at minimal cost, using linear programming.[43] 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.[44]

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.[45]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 $D^2_{xy} F(\mathbf{x}, \mathbf{y})$ from my paper is sufficient for the twist condition to hold. Since $D^2_{xy} F(\mathbf{x}, \mathbf{y})$ 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*.[46]

Instead of implicitly assuming that manual skills are only used by medium-skilled workers, I make the assumption that every worker has both skills, yet in different proportions. This makes it possible to distinguish between generalists and different types of specialists, thereby capturing that generalists can shield against adverse shocks to 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 dimension. It is noteworthy that this *assortativeness-mismatch*

[43]Optimal transport has a long tradition in mathematical theory. See Villani (2009) for a recent reference book.

[44]Additionally, both show the equivalence to a third formulation, namely the market game. Notice that the examples provided in Gretsky et al. (1992) are restricted to one-dimensional types. See also Dizdar and Moldovanu (2012) for recent work on the intersection of multidimensional matching and mechanism design that makes use of the twist condition.

[45]In that setting, Chiappori et al. (2010) establish a similar equivalence, namely between hedonic pricing, stable matching and the optimal transport problem.

[46]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.

trade-off across tasks takes place despite pairwise matching and without violating positive assortative matching.[47] Finally, my paper adds a *unified* framework of task-biased TC and the more standard skill-biased TC to the literature, allowing me to distinguish between their effects both theoretically and quantitatively.

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 undergone a relative decrease in the medium-skill price and wage polarization.

1.7 Conclusion

Technological change has drastically changed the structure of production in favor of cognitive relative to manual inputs in the developed world. How does this shift affect worker-job assignments, wages and inequality? This is a *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 this issue 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 equilibrium wages in closed form. I then analyze the impact on these equilibrium outcomes as cognitive (as opposed to manual) inputs become more prevalent in production, capturing one of the main recent technological shifts. Finally, I take this model to the data to study technological change in the US during the 1990s. The empirical analysis reveals that technological change was strongly biased toward cognitive inputs. Counterfactual exercises suggest that this technology shift (as opposed to changes in skill and skill demand distributions) can account for observed changes in worker-job sorting, wage polarization and wage dispersion.

It is worth pointing out that the theoretical framework developed here is of independent interest and can be used beyond this paper's application to technological change. It could be applied to a variety of matching problems that involve multidimensional heterogeneity, not only in the labor but also in the marriage or education markets. To broaden the applicability of this theory even further, it would be important to extend this framework to settings with search frictions (see the Appendix 1.8.7 for first insights on sorting conditions in such an environment) and to settings where the two sides of the market have different numbers of characteristics, preventing pure matching. These are challenging problems that form part of my ongoing research agenda.

[47] Contrary to Costinot and Vogel (2010) and Acemoglu and Autor (2011), there is no intensive margin here.

1.8 Appendix

1.8.1 Proofs General Model (Section 1.2)

The Equilibrium Assignment

In order to prove Proposition 1.1, the following Lemma and Corollary are useful. Notice that the proofs will be given for N-dimensional heterogeneity where firm are characterized by $\mathbf{y} = (y_1, ..., y_N) \in Y \subset \mathbb{R}^N_+$ (with abs. continuous c.d.f. $G(\mathbf{y})$) and workers are characterized by $\mathbf{x} = (x_1..., x_N) \in X \subset \mathbb{R}^N_+$ (with abs. continuous c.d.f. $H(\mathbf{x})$).

Lemma 1.1 (P-Matrix Property)**.** *If $D^2_{xy}F(\mathbf{x}, \mathbf{y})$ 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 1.1. *$D^2_{xy}F$ is a diagonal P-matrix.*

Proof. It will be shown that under Assumption 1.1, 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 $D^2_{xy}F$ is a P^--matrix is analogous and therefore omitted. I proceed in several steps.

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

2. $Det(-\mathcal{H}^*) \neq 0$, i.e. $-\mathcal{H}^*$ is positive definite. To show this, differentiate the first order conditions (1.3) and (1.4), evaluated at the optimal assignment $\mathbf{y}^* = \mu(\mathbf{x})$, with respect to skill vector $\mathbf{x}$, which gives

$$\mathcal{H}^* = D^2_{xx}F(\mathbf{x}, \mathbf{y}^*) - D^2_{xx}w(\mathbf{x}) = -(D^2_{xy}F(\mathbf{x}, \mathbf{y}^*))(D_x\mathbf{y}^*) \tag{1.20}$$

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

$$D^2_{xx}F(\mathbf{x}, \mathbf{y}^*) = \begin{bmatrix} F_{x_C x_C}(\mathbf{x}, \mathbf{y}^*) & F_{x_C x_M}(\mathbf{x}, \mathbf{y}^*) \\ F_{x_M x_C}(\mathbf{x}, \mathbf{y}^*) & F_{x_M x_M}(\mathbf{x}, \mathbf{y}^*) \end{bmatrix}$$

Since $D^2_{xy}F$ is a P-matrix everywhere (and, hence, also along the equilibrium allocation $\mathbf{y}^*$), it is non-singular and hence the inverse $(D^2_{xy}F(\mathbf{x}, \mathbf{y}^*))^{-1}$ exists. From (1.20), it is given by

$$(D^2_{xy}F(\mathbf{x}, \mathbf{y}^*))^{-1} = -(D_x\mathbf{y}^*)(D^2_{xx}F(\mathbf{x}, \mathbf{y}^*) - D^2_{xx}w(x))^{-1}. \tag{1.21}$$

It follows that $(D^2_{xx}F(\mathbf{x}, \mathbf{y}^*) - D^2_{xx}w(x))^{-1}$ exists, and thus $Det(\mathcal{H}^*) \neq 0$ and also $Det(-\mathcal{H}^*) \neq 0$. Then, by Step 1, it must be $Det(-\mathcal{H}^*) > 0$. Hence, $-\mathcal{H}^*$ is a positive definite.

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. $\frac{\partial y_i^*}{\partial x_j}\frac{\partial y_j^*}{\partial x_i} > 0, \forall i,j \in \{1,2,...,N\}, i \neq j$. 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 (1.20) $-\mathcal{H}^* = (D^2_{xy}F(\mathbf{x},\mathbf{y}^*))(D_x\mathbf{y}^*)$, where $-\mathcal{H}^*$ has all positive eigenvalues (Step 2). Denote $M = D^2_{xy}F(\mathbf{x},\mathbf{y}^*)$, $J = D_x\mathbf{y}^*$. Denote the eigenvalues of $-\mathcal{H}^*$ by $\lambda^{\mathcal{H}}$. They must obey the characteristic equation $det(MJ - \lambda^{\mathcal{H}}I) = 0$. Since M is a P-matrix (Assumption 1.1), it is invertible and the characteristic equation can be reformulated as $det(R - \lambda^{\mathcal{H}}M^{-1}) = 0$, where λ^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.[48] The (real) generalized eigenvalues can be computed as $\lambda_i^{\mathcal{H}} = \frac{S_{ii}}{T_{ii}}$. 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 $\lambda_i^{\mathcal{H}} > 0$, it must be that $S_{ii} > 0$, i.e. $J = D_x\mathbf{y}^*$ has positive eigenvalues, i.e. is stable.

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

$$\begin{aligned} & \mathcal{H}^*_{ij} = \mathcal{H}^*_{ji} \\ \Leftrightarrow \quad & F_{x_iy_i}\frac{\partial y_i^*}{\partial x_j} = F_{x_jy_j}\frac{\partial y_j^*}{\partial x_i} \quad \forall \quad i,j \in \{1,2,...,N\}, i \neq j, \end{aligned} \tag{1.22}$$

and hence $D_x\mathbf{y}^*$ is sign-symmetric, i.e. $\frac{\partial y_i^*}{\partial x_j}\frac{\partial y_j^*}{\partial x_i} > 0 \quad \forall i,j \in \{1,2,...,N\}, i \neq j$. 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 1.1 (Assortativeness and Local Maximum)**.** .
If $D^2_{xy}F(\mathbf{x},\mathbf{y})$ is a diagonal P-matrix ($P^{(-)}$-matrix), then the assignment μ (i) satisfies PAM (NAM) and (ii) is a strict local maximum.

Proof. .
(i) Assortativeness: Follows from the definition of assortativeness (Definition 1.1) and Lemma 1.1.
(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 1.1 and the Gale-Nikaido theorem that the solution to the firm's problem a strict *local* maximum.

Proof of Proposition 1. .

(i) Assortativeness: Follows directly from Corollary 1.1.
(ii) Global Maximum: It will be shown that the solution to the firm's problem is a *global*

[48]If J has complex eigenvalue, S is quasi-upper triangular.

maximum. I proceed by contradiction. Consider a firm $\mathbf{y}$ which optimally chooses worker $\mathbf{x}$, i.e. $\mathbf{y} = \mu(\mathbf{x})$.[49] 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 1.1. 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 of both firms:

$$F_x(\mathbf{x}', \mathbf{y}) = w_x(\mathbf{x}') \tag{1.23}$$

$$F_x(\mathbf{x}', \mathbf{y}') = w_x(\mathbf{x}'). \tag{1.24}$$

I will show that, under Assumption 1.1, (1.23) and (1.24) 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 1.1, $D^2_{xy}F(\mathbf{x}, \mathbf{y}^*)$ is a P-matrix. Moreover, $F_{\mathbf{x}}$ is defined over a rectangular region on $\mathbb{R}^4$. It follows from the Gale-Nikaido Theroem (Gale and Nikaido (1965)) that F_x is injective with respect to $\mathbf{y}$. Thus, (1.23) and (1.24) cannot hold simultaneously because

$$F_x(\mathbf{x}', \mathbf{y}) = F_x(\mathbf{x}', \mathbf{y}') \tag{1.25}$$

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 1.1 is not only a local but also a global maximum.

The Wage Function

In technical terms, Proposition 1.2 states: Given a continuously differentiable assignment $\mathbf{y}^* = \mu(\mathbf{x})$, condition (1.6) is necessary and sufficient for the existence of a unique solution to the system (1.3) and (1.4), given by $w(\mathbf{x})$, such that $w(\underline{\mathbf{x}}) = w_0$.[50]

Proof of Proposition 1.2. The proof is based on *Frobenius Theorem.* Consider a system of linear first-order partial differential differential equations

$$\frac{\partial u^\rho}{\partial x^i} = \psi_i^\rho(x, u) \quad i = 1, ..., N; \rho = 1, ..., n \tag{1.26}$$

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

Theorem 1.1 (Frobenius Theorem)**.** *The necessary and sufficient conditions for the unique solution $u^\alpha = u^\alpha(x)$ to the system (1.26) such that $u(x_0) = u_0$ to exist for any initial data*

[49] 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 μ.

[50] 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.

$(x_0, u_0) \in \mathbb{R}^{N+n}$ *is that the relations*

$$\frac{\partial \psi_i^\alpha}{\partial x^j} - \frac{\partial \psi_j^\alpha}{\partial x^i} + \sum_\beta \left(\frac{\partial \psi_i^\alpha}{\partial u^\beta} \psi_j^\beta - \frac{\partial \psi_j^\alpha}{\partial u^\beta} \psi_i^\beta \right) = 0 \quad \forall i, j = 1, ..., N, \quad \alpha, \beta = 1, ..., n. \tag{1.27}$$

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: $u = w$, $x = (x_1, x_2, ...x_N)$ and $\psi_i(x, u) = F_{x_i}(\mathbf{x}, \mathbf{y}(\mathbf{x}))$. Notice that $n = 1$ because w is a real-valued function. Then, (1.27) 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

$$F_{x_C x_M} + F_{x_C y_C} \frac{\partial y_C^*}{\partial x_M} + F_{x_C y_M} \frac{\partial y_M^*}{\partial x_M} - F_{x_M x_C} - F_{x_M y_C} \frac{\partial y_C^*}{\partial x_C} + F_{x_M y_M} \frac{\partial y_M^*}{\partial x_C} = 0. \tag{1.28}$$

(1.28) coincides with condition (1.6) from the main text since $F_{x_C x_M} = F_{x_M x_C}$. Hence, given (1.6), the involutivity condition from Frobenius theorom is satisfied. A unique (local) solution to the system of linear partial differential equations (1.3) and (1.4) exists.

1.8.2 Proofs of Quadratic-Gaussian Model (Section 1.3)

Labor Market Clearing under PAM (or NAM)

Having applied the measure-preserving transformation (1.9) to skills and productivities, the labor market clearing of transformed variables under PAM reads

$$\int_{z_{y_C}}^{\infty} \int_{z_{y_M}}^{\infty} g(\hat{z}_{y_C}, \hat{z}_{y_M}) d\hat{z}_{y_M} d\hat{z}_{y_C} = \int_{z_{x_C}}^{\infty} \int_{z_{x_M}}^{\infty} h(\hat{z}_{x_C}, \hat{z}_{x_M}) d\hat{z}_{x_M} d\hat{z}_{x_C} \tag{1.29}$$

where h and g denote the standard normal p.d.f.'s of the uncorrelated skills and productivities, respectively. Equation (1.11) follows immediately, taking into account that the $z's$ are independent and standard normally distributed. Similarly, under NAM, the market clearing would read

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

The Equilibrium Assignment

The following two lemmas are building blocks for the proof of Proposition 1.3.

Lemma 1.2 (Continuum of Square Roots)**.** *(i) There exists a continuum of square roots of the covariance matrix* Σ*, 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, $\Sigma^{\frac{1}{2}}R(\Sigma^{\frac{1}{2}}R)^T = \Sigma^{\frac{1}{2}}RR^T(\Sigma^{\frac{1}{2}})^T = \Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T = \Sigma$.*

Proof. .

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

$$\Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \Sigma \tag{1.30}$$

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 Σ is positive-definite, one square root can be computed using the spectral square root decomposition

$$\begin{aligned} & \Sigma = CDC' \\ \Leftrightarrow \quad & \Sigma^{\frac{1}{2}} = CD^{\frac{1}{2}}C' \end{aligned} \tag{1.31}$$

where D is a diagonal matrix with the eigenvalues of Σ as diagonal entries and C is a matrix of orthonormal eigenvectors of Σ. Since the *spectral square root* is one solution to (1.30), 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 δ.

Lemma 1.3 (Orthogonal Transformation Matrices). *The system of equations to be solved is given by:*

$$\alpha_x^2 + \beta_x^2 = 1 \tag{1.32}$$
$$\alpha_y^2 + \beta_y^2 = 1 \tag{1.33}$$
$$\frac{\partial y_C^*}{\partial x_M} = \delta \frac{\partial y_M^*}{\partial x_C} \tag{1.34}$$

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 (1.32)-(1.34) is given by $\alpha_x = \pm 1, \beta_x = 0$ and

$$\alpha_y = \pm \frac{(1+\delta)\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}}\right)}{\sqrt{(1-\delta)^2\left(\sqrt{\frac{1+\rho_y}{1+\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_x}}\right)^2 + (1+\delta)^2\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}}\right)^2}} \tag{1.35}$$

$$\beta_y = \sqrt{1-\alpha_y^2}. \tag{1.36}$$

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

Proof. To solve (1.32)-(1.34), 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 (1.13) where I use *rotations* of the spectral square root (given by (1.31)) to uncorrelate skills and productivities. They are given by:

$$\Sigma_i^{\frac{1}{2}} = \begin{bmatrix} \frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) \\ \frac{1}{2}(\sqrt{1+\rho_i} - \sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i} + \sqrt{1-\rho_i}) \end{bmatrix} \begin{bmatrix} \alpha_i & -\beta_i \\ \beta_i & \alpha_i \end{bmatrix}, \forall \quad i \in \{x, y\}. \tag{1.37}$$

Using (1.37), the candidate equilibrium assignment can be computed from (1.13) as:

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{2}\left((\alpha_y\alpha_x + \beta_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) + (\beta_y\alpha_x - \alpha_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}}\right)\right) & \frac{1}{2}\left((\alpha_y\alpha_x + \beta_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) - (\beta_y\alpha_x - \alpha_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}}\right)\right) \\ \frac{1}{2}\left((\alpha_y\alpha_x + \beta_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) + (\beta_y\alpha_x - \alpha_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}}\right)\right) & \frac{1}{2}\left((\alpha_y\alpha_x + \beta_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) - (\beta_y\alpha_x - \alpha_y\beta_x)\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1-\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1+\rho_x}}\right)\right) \end{bmatrix}}_{D_x\mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \tag{1.38}$$

(i) The underdetermined system (1.32)-(1.34) has one degree of freedom. I exploit it by setting $\beta_x = 0$, which immediately gives $\alpha_x = \pm 1$ from equation (1.32). It remains to determine two unknowns, α_y, β_y, from two equations (1.33) and (1.34). From (1.33), $\beta_y = \pm\sqrt{1-\alpha_y^2}$. Using this relation along with $\alpha_x = \pm 1, \beta_x = 0$ and candidate assignment (1.38), integrability condition (1.34) reads:

$$\begin{aligned} &\left(\alpha_y\left(\sqrt{\frac{1+\rho_y}{1+\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_x}}\right) - \sqrt{1-\alpha_y^2}\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}}\right)\right) = \\ &\delta\left(\alpha_y\left(\sqrt{\frac{1+\rho_y}{1+\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_x}}\right) + \sqrt{1-\alpha_y^2}\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}}\right)\right) \end{aligned} \tag{1.39}$$

Reorganizing terms and solving for α_y yields:

$$\alpha_y = \pm \frac{(1+\delta)\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}}\right)}{\sqrt{(1-\delta)^2\left(\sqrt{\frac{1+\rho_y}{1+\rho_x}} - \sqrt{\frac{1-\rho_y}{1-\rho_x}}\right)^2 + (1+\delta)^2\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}} + \sqrt{\frac{1-\rho_y}{1+\rho_x}}\right)^2}} \tag{1.40}$$

Using (1.40), β_y can be backed out from (1.33)[51]

$$\beta_y = \sqrt{1-\alpha_y^2}. \tag{1.41}$$

[51]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.

(ii) Rearranging (1.39) yields:

$$\alpha_y(1-\delta)\left(\sqrt{\frac{1+\rho_y}{1+\rho_x}}-\sqrt{\frac{1-\rho_y}{1-\rho_x}}\right)=\sqrt{1-\alpha_y^2}(\delta+1)\left(\sqrt{\frac{1+\rho_y}{1-\rho_x}}+\sqrt{\frac{1-\rho_y}{1+\rho_x}}\right) \tag{1.42}$$

While $RHS \geq 0, \forall \rho_x, \rho_y$, $LHS \lesseqqgtr 0$ for $\rho_y \lesseqqgtr \rho_x$. 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 1.3. .
Computing the Assignment:

(i) For $\delta = 1$, from (1.35)

$$\alpha_y = \pm 1. \tag{1.43}$$

The orthogonal transformation (1.37) delivers:

$$\begin{bmatrix} \frac{1}{2}(\sqrt{1+\rho_i}+\sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i}-\sqrt{1-\rho_i}) \\ \frac{1}{2}(\sqrt{1+\rho_i}-\sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i}+\sqrt{1-\rho_i}) \end{bmatrix}\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}=\pm\begin{bmatrix} \frac{1}{2}(\sqrt{1+\rho_i}+\sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i}-\sqrt{1-\rho_i}) \\ \frac{1}{2}(\sqrt{1+\rho_i}-\sqrt{1-\rho_i}) & \frac{1}{2}(\sqrt{1+\rho_i}+\sqrt{1-\rho_i}) \end{bmatrix} \quad \text{for} \quad i \in \{x, y\}. \tag{1.44}$$

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

$$\begin{aligned} & \Sigma = CDC' \\ \Leftrightarrow \quad & \Sigma^{\frac{1}{2}} = CD^{\frac{1}{2}}C' \end{aligned} \tag{1.45}$$

where D is a diagonal matrix with the eigenvalues of Σ as diagonal entries and C is a matrix of orthonormal eigenvectors of Σ. The matrix $\Sigma^{\frac{1}{2}}$ in (1.45) is called the *spectral square root* of Σ. Notice that for $\Sigma^{\frac{1}{2}}$ to be positive-definite, the positive square roots of the diagonal entries of D are used. From (1.45) it follows that $\Sigma_y^{\frac{1}{2}}$ is given by:

$$\Sigma_y^{\frac{1}{2}}=\begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \sqrt{1+\rho_y} & 0 \\ 0 & \sqrt{1-\rho_y} \end{bmatrix}\begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix}=\begin{bmatrix} \frac{1}{2}\left(\sqrt{1+\rho_y}+\sqrt{1-\rho_y}\right) & \frac{1}{2}\left(\sqrt{1+\rho_y}-\sqrt{1-\rho_y}\right) \\ \frac{1}{2}\left(\sqrt{1+\rho_y}-\sqrt{1-\rho_y}\right) & \frac{1}{2}\left(\sqrt{1+\rho_y}+\sqrt{1-\rho_y}\right) \end{bmatrix} \tag{1.46}$$

Moreover, since

$$\Sigma^{-\frac{1}{2}} = CD^{-\frac{1}{2}}C' = C\begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{bmatrix}C'. \tag{1.47}$$

where λ_1, λ_2 are the eigenvalues of Σ and C is a matrix of the corresponding orthonormal eigenvectors, the matrix $\Sigma_x^{-\frac{1}{2}}$ is given by

$$\Sigma_x^{-\frac{1}{2}}=\begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{1+\rho_x}} & 0 \\ 0 & \frac{1}{\sqrt{1-\rho_x}} \end{bmatrix}\begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix}=\begin{bmatrix} \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}}+\frac{1}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}}-\frac{1}{\sqrt{1-\rho_x}}\right) \\ \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}}-\frac{1}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}}+\frac{1}{\sqrt{1-\rho_x}}\right) \end{bmatrix} \tag{1.48}$$

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

$$D_x\mathbf{y}^* = (\Sigma_y^{\frac{1}{2}}R_y)(\Sigma_x^{\frac{1}{2}}R_x)^{-1} = \Sigma_y^{\frac{1}{2}}\Sigma_x^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{2}\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) \\ \frac{1}{2}\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} - \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}\left(\frac{\sqrt{1+\rho_y}}{\sqrt{1+\rho_x}} + \frac{\sqrt{1-\rho_y}}{\sqrt{1-\rho_x}}\right) \end{bmatrix} \tag{1.49}$$

The assignment is then computed using (1.13).

(ii) For $\delta = 0$, it follows from Lemma 1.3 that R_y and R_x are respectively given by

$$R_y = \begin{bmatrix} \alpha_y & -\sqrt{1-\alpha_y^2} \\ \sqrt{1-\alpha_y^2} & \alpha_y \end{bmatrix} = \begin{bmatrix} \pm\frac{1}{2}\left(\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}\right) & -\frac{1}{2}\left(\sqrt{(1+\rho_y)(1-\rho_x)} - \sqrt{(1-\rho_y)(1+\rho_x)}\right) \\ \frac{1}{2}\left(\sqrt{(1+\rho_y)(1-\rho_x)} - \sqrt{(1-\rho_y)(1+\rho_x)}\right) & \pm\frac{1}{2}\left(\sqrt{(1+\rho_y)(1+\rho_x)} + \sqrt{(1-\rho_y)(1-\rho_x)}\right) \end{bmatrix} \tag{1.50}$$

$$R_x = \begin{bmatrix} \alpha_x & -\sqrt{1-\alpha_x^2} \\ \sqrt{1-\alpha_x^2} & \alpha_x \end{bmatrix} = \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}. \tag{1.51}$$

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 (1.46) and by the inverse of (1.48), respectively. Then,

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

$$(\Sigma_x^{\frac{1}{2}}R_x)^{-1} = \begin{bmatrix} \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}}\right) \\ \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_x}}\right) & \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_x}}\right) \end{bmatrix}. \tag{1.52}$$

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

$$D_x\mathbf{y}^* = \Sigma_y^{\frac{1}{2}}R_y(\Sigma_x^{\frac{1}{2}}R_x)^{-1} = \begin{bmatrix} 1 & 0 \\ \rho_y - \rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} & \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \end{bmatrix} \tag{1.53}$$

In the following, it is shown that (1.53) 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 Σ_i. Under the assumption of standard normality, L_i is given by:

$$L_i = \begin{bmatrix} 1 & 0 \\ \rho_i & \sqrt{1-\rho_i^2} \end{bmatrix} \quad \forall \quad i \in \{x, y\} \tag{1.54}$$

Hence,

$$L_y(L_x)^{-1} = \begin{bmatrix} 1 & 0 \\ \rho_y & \sqrt{1-\rho_y^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho_x & \sqrt{1-\rho_x^2} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ \rho_y - \rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} & \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} \end{bmatrix} \tag{1.55}$$

which coincides with (1.53). The equilibrium assignment is then given by (1.13)

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

(iii) The equilibrium assignment is obtained by using the suitable rotation matrices from Lemma 1.3 for each $\delta \in (0,1)$ together with the candidate equilibrium assignment (1.13).

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 (1.11) and because the transformation (1.9) is measure-preserving. (b) Verifying the PAM-property amounts to checking that $D_x\mathbf{y}^*$ is a P-matrix. Using Lemmas 1.2 and 1.3 equilibrium assignment (1.38) can be simplified by substituting in the expressions for α's and β's:

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \underbrace{\begin{bmatrix} J_{11}(\rho_x,\rho_y,\delta) & J_{12}(\rho_x,\rho_y,\delta) \\ J_{21}(\rho_x,\rho_y,\delta) & J_{22}(\rho_x,\rho_y,\delta) \end{bmatrix}}_{D_x\mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1+\delta\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \\ \frac{\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta+\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \end{bmatrix}}_{D_x\mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \tag{1.56}$$

Taking derivatives yields:

$$\frac{\partial y_C^*}{\partial x_C} > 0 \tag{1.57}$$

$$\frac{\partial y_M^*}{\partial x_M} > 0 \tag{1.58}$$

$$\frac{\partial y_C^*}{\partial x_C}\frac{\partial y_M^*}{\partial x_M} - \frac{\partial y_C^*}{\partial x_M}\frac{\partial y_M^*}{\partial x_C} = \sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}} > 0 \quad \text{for} \quad \rho_y << 1 \tag{1.59}$$

where (1.57) and (1.58) follow immediately from (1.56). Hence, $D_x\mathbf{y}^*$ is a P-matrix. (c) The assignment was derived under the integrability condition (1.34), i.e. it is satisfied.

1.8.3 The Equilibrium Wage Function

Proof of Proposition 1.4. .

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

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

where w_0 is the constant of integration, which I set to zero.[52] $J_{11}, J_{12}, J_{21}, J_{22}$ are the elements of the matching function's Jacobian, which are given by (1.56). Using the assignment (1.56) explicitly, this is

$$
w(x_C, x_M) = \tfrac{1}{2}\alpha [x_C \quad x_M] \underbrace{\begin{bmatrix} \frac{1+\delta \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\rho_y-\rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \\ \frac{\delta\left(\rho_y-\rho_x \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\delta+\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \end{bmatrix}}_{\equiv \tilde{J}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} = \tfrac{1}{2}\alpha [x_C \quad x_M] \underbrace{\begin{bmatrix} J_{11} & J_{12} \\ \delta J_{21} & \delta J_{22} \end{bmatrix}}_{\equiv \tilde{J}} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \tag{1.60}
$$

which is equivalent to expression (1.15) in Proposition 1.4, taking into account $w_0 = 0$. The guess needs to be verified. Given (1.60), the partial derivatives of the wage with respect to skills x_C, x_M are given by

$$
\frac{\partial w(x_C, x_M)}{\partial x_C} = \alpha (J_{11} x_C + J_{12} x_M) \tag{1.61}
$$

$$
\frac{\partial w(x_C, x_M)}{\partial x_M} = \alpha \delta \left(J_{22} x_M + J_{21} x_C \right) \tag{1.62}
$$

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

$$
\frac{\partial w(x_C, x_M)}{\partial x_C} = \alpha y_C^* \tag{1.63}
$$

$$
\frac{\partial w(x_C, x_M)}{\partial x_C} = \alpha \delta y_M^* \tag{1.64}
$$

evaluated at the equilibrium assignment (1.56). Moreover, the integrability condition

$$
\begin{aligned}
\frac{\partial^2 w(x_C, x_M)}{\partial x_C \partial x_M} &= \frac{\partial^2 w(x_C, x_M)}{\partial x_M \partial x_C} \\
\Leftrightarrow \quad J_{12} &= \delta J_{21}
\end{aligned} \tag{1.65}
$$

is satisfied by construction of the equilibrium assignment, which gives uniqueness of (1.60) by Proposition 1.2.

Proof of Proposition 1.5. .
(i) See Proof of Proposition 1.3 (last part, under *Consistency of the Assignment Functions from (i)-(iii) with the Equilibrium.*).

[52]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.

(ii) For $\rho_x = \rho_y$, (1.35) 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 (1.13) 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 1.4.

The second result is that the maximum mismatch is achieved at $|\rho_x - \rho_y| \to 2$, 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:

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

$$\rho_y : \left(\frac{1}{\sqrt{1+\rho_y}\sqrt{1+\rho_x}} - \frac{1}{\sqrt{1-\rho_y}\sqrt{1-\rho_x}}\right) = 0 \tag{1.66}$$

$$\rho_x : \frac{1}{2}\left(-\frac{\sqrt{1+\rho_y}}{(1+\rho_x)^{\frac{3}{2}}} + \frac{\sqrt{1-\rho_y}}{(1-\rho_x)^{\frac{3}{2}}}\right) = 0 \tag{1.67}$$

Solving the system of first order conditions (1.66) and (1.67) (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 ρ_x and ρ_y:

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

$$\rho_y : \left(\frac{1}{\sqrt{1+\rho_y}\sqrt{1+\rho_x}} + \frac{1}{\sqrt{1-\rho_y}\sqrt{1-\rho_x}}\right) = 0 \tag{1.68}$$

$$\rho_x : \frac{1}{2}\left(-\frac{\sqrt{1+\rho_y}}{(1+\rho_x)^{\frac{3}{2}}} - \frac{\sqrt{1-\rho_y}}{(1-\rho_x)^{\frac{3}{2}}}\right) = 0 \tag{1.69}$$

where it is immediate that (1.68) 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 ρ_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 $\rho_x = -1+\epsilon, \epsilon > 0$.

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

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

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

Morever, 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 ρ_x to (i) $\rho_x = 1-\epsilon$ and (ii) $\rho_x = -1+\epsilon$, and maximizing with respect to ρ_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 $|\rho_x - \rho_y| \to 2$.

Proof of Proposition 1.6. .
(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 (1.56). 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

$$E(w(x_C, x_M)) = tr(\tilde{J}\Sigma_x) = \frac{1}{2}J_{11} + J_{12}\rho_x + \frac{1}{2}\delta J_{22} \tag{1.70}$$

$$Var(w(x_C, x_M)) = 2tr(\tilde{J}\Sigma_x\tilde{J}\Sigma_x) = \frac{1}{2}((J_{11} + J_{12}\rho_x)^2 + 2(J_{11}\rho_x + J_{12})(J_{12} + \delta J_{22}\rho_x) + (J_{12}\rho_x + \delta J_{22})^2) \tag{1.71}$$

$$E\left[\left(\frac{w - E(w)}{\sqrt{Var(w)}}\right)^3\right] = \frac{E(w^3) - 3E(w)Var(w) - E(w)^3}{Var(w)^{\frac{3}{2}}} = \frac{8tr(\tilde{J}\Sigma_x\tilde{J}\Sigma_x\tilde{J}\Sigma_x)}{(2tr(\tilde{J}\Sigma_x\tilde{J}\Sigma_x))^{\frac{3}{2}}} \tag{1.72}$$

where $J_{11}, J_{12}, J_{21}, J_{22}$ are defined in (1.56), $\tilde{J}$ denotes the Hessian of the wage function and where

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

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

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

where, as before, J_μ denotes the Jacobian of the matching function.

Skewness: $J_\mu\Sigma_xJ_\mu$ and $\Sigma_xJ_\mu\Sigma_x$ 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 ρ_y and ρ_x yields:

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

$$\begin{aligned} \rho_x:&\quad \rho_x=\rho_y \\ \rho_y:&\quad \rho_x=\rho_y \end{aligned}$$

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 $H_1 1(E(w)) < 0$ and $Det(H(E(w)))$ simplifies to

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

Hence, the Hessian is negative semi-definite $\forall \rho_x, \rho_y$. Hence, $E(w(x_C, x_M))$ is concave. At $\rho_x = \rho_y$, $Det(H(E(w))) = 0$, 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 $\rho_y = -1, \rho_x = 1$.

1.8.4 Proofs on Technological Change

Task-Biased Technological Change (Section 1.4.1)

Proof of Proposition 1.7. .
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) 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 (1.56):

$$\underbrace{\begin{bmatrix} J_{11}(\rho_x,\rho_y,\delta) & J_{12}(\rho_x,\rho_y,\delta) \\ J_{21}(\rho_x,\rho_y,\delta) & J_{22}(\rho_x,\rho_y,\delta) \end{bmatrix}}_{D_x\mathbf{y}^*} = \underbrace{\begin{bmatrix} \frac{1+\delta\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \\ \frac{\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta+\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \end{bmatrix}}_{D_x\mathbf{y}^*} \tag{1.73}$$

I will show this claim in two steps:

1. Step:

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

This follows from (1.111), since

$$
\begin{aligned}
J_{11} =& \frac{1+\delta\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta'^2}} \geq 1 \\
\Leftrightarrow \quad & \left(1+\delta\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)^2 \geq 1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^{'2} \\
\Leftrightarrow \quad & 2\delta\frac{\sqrt{1-\rho_y^2}\rho_x^2-\rho_x\rho_y\sqrt{1-\rho_x^2}}{\sqrt{1-\rho_x^2}}+\delta^{'2}\left(\frac{1-\rho_y^2}{1-\rho_x^2}-1\right) \geq 0 \quad \text{for} \quad |\rho_x| \geq |\rho_y| \\
J_{12} =& \frac{\delta\left(\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \geq 0 \\
\Leftrightarrow \quad & \rho_y\sqrt{1-\rho_x^2}-\rho_x\sqrt{1-\rho_y^2} \geq 0 \quad \text{for} \quad \rho_x \leq \rho_y .
\end{aligned}
$$

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

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

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

where

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

$$\frac{\partial J_{22}}{\partial \delta} = \frac{1-\delta\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} + \left(\delta - \frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1\right)(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})}{(1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2)^{\frac{3}{2}}}$$

$$\frac{\partial J_{22}}{\partial \delta}|_{\delta=1} = \frac{\left(1-\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)\left(1+\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}\right)}{(2+2(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}))^{\frac{3}{2}}} \geq 0 \quad \text{for} \quad |\rho_x| \leq |\rho_y|$$

which together with Step 1 establishes the result.

(ii) Sorting when $\delta = 0$. In the cognitive task, $y_C = x_C$, follows from (1.53) in proof of Proposition 1.3. In turn, in the manual task, there is maximal mismatch $|y_M - x_M|$. To see this, notice that

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

since

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

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

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

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

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

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

since for $\rho_x < \rho_y$,

$$\rho_y - \rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} > \frac{\delta\left(\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}}$$

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

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

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.

(iii) Task-biased TC has no effect on sorting for $\rho_x = \rho_y$. Follows from (ii) in proof of Proposition 1.5. 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

$$r \equiv \frac{\partial J_{11}}{\partial \delta}|_{\delta=1} = \frac{\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1}{2^{\frac{3}{2}}(2 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}))^{\frac{1}{2}}} \tag{1.74}$$

$$s \equiv \frac{\partial J_{12}}{\partial \delta}|_{\delta=1} = \frac{\rho_y - \rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{2^{\frac{3}{2}}(2 + 2\delta(\rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}))^{\frac{1}{2}}} \tag{1.75}$$

Then,

$$\frac{\partial r}{\partial \rho_x} = \frac{\rho_x\frac{\sqrt{1-\rho_y^2}}{(1-\rho_x^2)^{\frac{3}{2}}}(1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) - \frac{1}{2}\left(\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1\right)\left(\rho_y - \rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \tag{1.76}$$

$$\frac{\partial r}{\partial \rho_y} = \frac{-\frac{\rho_y}{\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) - \frac{1}{2}\left(\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}} - 1\right)\left(\rho_x - \rho_y\frac{\sqrt{1-\rho_x^2}}{\sqrt{1-\rho_y^2}}\right)}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \tag{1.77}$$

$$\frac{\partial s}{\partial \rho_x} = \frac{-\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\left(1 + \frac{\rho_x^2}{1-\rho_x^2}\right)(1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) + \frac{1}{2}\left(\rho_y - \rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)^2}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \tag{1.78}$$

$$\frac{\partial s}{\partial \rho_y} = \frac{\left(1 + \frac{\rho_y\rho_x}{\sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}}\right)(1 + \rho_x\rho_y + \sqrt{1-\rho_x^2}\sqrt{1-\rho_y^2}) + \frac{1}{2}\frac{(\rho_x\sqrt{1-\rho_y^2} - \rho_y\sqrt{1-\rho_x^2})^2}{\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2}}}{2^{\frac{3}{2}}(1 + \rho_x\rho_y + \sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})^{\frac{3}{2}}} \tag{1.79}$$

Expression (2.21) 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 (2.22) 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, (1.74) has no interior maximum. Also, (1.79) is positive $\forall \rho_x, \rho_y$. Evaluating (2.23) at the corners $\rho_x \to 1$ and $\rho_x \to -1$ yields a strictly positive and a strictly negative expression. Hence, (1.75) has no interior maximum. The supremum of (1.74) and (1.75) must be in the corner. In a second step, it can be shown that (1.74) and (1.75) 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 1.8. .
(i) The results follow from simulations of the closed forms (1.70) and (1.72). 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 (δ decreases), triggered by an increase in α (one could additionally assume β 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 (1.111). It follows that for $|\rho_x| < |\rho_y|$,

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

Hence, due to task-biased TC, wages become more convex in x_C but less convex in x_M. Notice that additionally decreasing β reinforces the effects. Finally, the result that that the curvature changes are largest follows from Proposition 1.7 part (iv).

Skill-Biased Technological Change (Section 1.4.2)

The wage function under the augmented technology

$$F(x_C, x_M, y_C, y_M) = \alpha x_C y_C + \beta x_M y_M + \lambda x_C + \eta x_M + f_0 \tag{1.80}$$

is given by

$$w(x_C, x_M) = \alpha \left(\frac{1}{2} \mathbf{x}' \tilde{J} \mathbf{x} + \theta' \mathbf{x} \right) + w_0 = \frac{1}{2} \alpha (\mathbf{x} - h)' \tilde{J} (\mathbf{x} - h) + C \tag{1.81}$$

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 1.9. (i) It is immediate that assignment (1.56) satisfies the first-order conditions of the firm under (1.80). (ii) (1.81) satisfied integrability condition (1.6), i.e. is the unique wage schedule supporting the assignment. From (1.81), skill-biased TC parameters λ and η do not affect the curvature of the wage function. (iii) Under (1.81), the variance of the wage distribution is given by:

$$Var(w) = \alpha^2 (1-\rho_x^2) \left((J_{21} + J_{11}\rho_x)^2 + \frac{J_{11}^2}{2}(1-\rho_x^2) \right) + (1-\rho_x^2)\lambda^2 + \alpha^2 2 \left(\frac{J_{11}}{2}\rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2} \right)^2 + (\rho_x \lambda + \eta)^2 \tag{1.82}$$

For $\rho_x > 0$, $Var(w)$ positively depends on λ and η (i.e. the effect of skill-biased TC on $Var(w)$ is ambiguous). For $\rho_x < 0$ and $-\rho_x \lambda > \eta$, $Var(w)$ positively depends on λ and negatively on η (i.e. in this case, the effect of skill-biased TC on $Var(w)$ is unambiguously positive). To derive (1.82), notice

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

$$\begin{aligned}
Var(E(w|x_M)) &= \alpha^2 Var(x_M^2)\left(\frac{J_{11}}{2}\rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2}\right)^2 + Var(x_M)(\rho_x\lambda + \eta)^2 \\
&= \alpha^2 2\left(\frac{J_{11}}{2}\rho_x^2 + \rho_x J_{12} + \frac{\delta J_{22}}{2}\right)^2 + (\rho_x\lambda + \eta)^2 \qquad (1.83)
\end{aligned}$$

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

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

where

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

Hence,

$$\begin{aligned}
Var(w|x_M) &= \alpha^2(1-\rho_x^2)(x_M^2(J_{12}+\rho_x J_{11})^2 + \frac{J_{11}^2}{2})(1-\rho_x^2)) + (1-\rho_x^2)(\rho_x J_{11}\lambda x_M + \lambda^2) \\
E(Var(w|x_M)) &= \alpha^2(1-\rho_x^2)((J_{12}+\rho_x J_{11})^2 + \frac{J_{11}^2}{2}(1-\rho_x^2)) + (1-\rho_x^2)\lambda^2. \qquad (1.84)
\end{aligned}$$

(1.82) follows from adding (1.83) and (1.84), i.e. $Var(w) = E(Var(w|x_M)) + Var(E(w|x_M))$.

1.8.5 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.

Gaussian Distributions

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

$$F(x_C, x_M, y_C, y_M) = \gamma(x_C y_C + \alpha x_C y_M + \beta x_M y_C + \delta x_M y_M) \qquad (1.85)$$

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

Proposition 1.10 (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 (1.85). If*

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

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 $|m_{ii}| > \sum_{j\neq i}|m_{ij}|, i = 1, 2, ..., n$ and row diagonally dominant if $|m_{ii}| > \sum_{j\neq i}|m_{ji}|, i = 1, 2, ..., n$. In this setting, D^2_{xy} 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 Appendix 1.8.7.

Under (1.85), integrability condition (1.6), which needs to be satisfied in order for a unique wage schedule to exist, is given by:

$$\begin{aligned} \frac{\partial^2 w(x_C, x_M)}{\partial x_C \partial x_M} &= \frac{\partial^2 w(x_C, x_M)}{\partial x_M \partial x_C} \\ \Leftrightarrow \quad J_{12} + \alpha J_{22} &= \beta J_{11} + \delta J_{21} \end{aligned} \tag{1.86}$$

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 (1.86), the equilibrium assignment follows Proposition 1.3. It is given by

$$J_{11} = \frac{4}{2Z\sqrt{1-\rho_x^2}}((1+\alpha\rho_y)\sqrt{1-\rho_x^2} + (\delta+\alpha\rho_x)\sqrt{1-\rho_y^2}) \tag{1.87}$$

$$J_{12} = \frac{4}{2Z\sqrt{1-\rho_x^2}}((\beta+\delta\rho_y)\sqrt{1-\rho_x^2} - (\delta\rho_x+\alpha)\sqrt{1-\rho_y^2}) \tag{1.88}$$

$$J_{21} = \frac{4}{2Z\sqrt{1-\rho_x^2}}((\alpha+\rho_y)\sqrt{1-\rho_x^2} - (\rho_x+\beta)\sqrt{1-\rho_y^2}) \tag{1.89}$$

$$J_{22} = \frac{4}{2Z\sqrt{1-\rho_x^2}}((\delta+\beta\rho_y)\sqrt{1-\rho_x^2} + (1+\beta\rho_x)\sqrt{1-\rho_y^2}) \tag{1.90}$$

where

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

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

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

where determinant of the matching's function Jacobian reads

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

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, ρ_x and ρ_y. Hence, the infimum of Y must be in a corner. If Y is positive in all corners, then $Det(J_\mu) > 0$. 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.

Non-Gaussian Distributions

This section states the sufficient condition for PAM

Proposition 1.11. *For (a) independent uniform skills* $\mathbf{x} \sim U([\underline{x}, \overline{x}]^N)$ *and productivities* $\mathbf{y} \sim U([\underline{y}, \overline{y}]^N)$ *or (b) whenever G=H, if* $D^2_{xy}F(\mathbf{x}, \mathbf{y})$ *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

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

which gives the assignment functions:

$$y_C = \frac{\overline{y} - \underline{y}}{\overline{x} - \underline{x}} x_C - \underline{x} \frac{\overline{y} - \underline{y}}{\overline{x} - \underline{x}} + \underline{y} \tag{1.91}$$
$$y_M = \frac{\overline{y} - \underline{y}}{\overline{x} - \underline{x}} x_M - \underline{x} \frac{\overline{y} - \underline{y}}{\overline{x} - \underline{x}} + \underline{y} \tag{1.92}$$

Both (1.91) and (1.92) 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 $D^2_{xy}F$. Recall from the proof of Lemma 1.1 that the Hessian of the firm's problem is given by:

$$H^* = D^2_{xx}F(\mathbf{x}, \mathbf{y}^*) - D^2_{xx}w(\mathbf{x}) = -(D^2_{xy}F(\mathbf{x}, \mathbf{y}^*))(D_x\mathbf{y}^*) \tag{1.93}$$

In the PAM allocations from Step 1, $D_x\mathbf{y}^*$ is a diagonal matrix. Since $D^2_{xy}F$ is a P-matrix, the matrix product $(D^2_{xy}F(\mathbf{x}, \mathbf{y}^*))(D_x\mathbf{y}^*)$ is positive-definite and, hence, the Hessian (1.93) is negative-definite.

3. Step: The PAM allocation from Step 1 satisfies the integrability condition (1.6). Hence, there exists a unique wage schedule supporting this allocation. To see this, first focus on (a). Since $D_x\mathbf{y}^*$ is diagonal, (1.6) collapses to

$$F_{x_C y_M} \frac{\partial y_M^*}{\partial x_M} = F_{x_M y_C} \frac{\partial y_C^*}{\partial x_C} \tag{1.94}$$

which must hold at the equilibrium path. Using (1.91) and (1.92), this simplifies condition (1.94) to

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

which holds under the assumption of positive-definiteness of $D^2_{xy}F$ ($F_{x_C y_M} = F_{x_M y_C}$) along the equilibrium path.
Consider (b). Under the assignment $y_C = x_C$ and $y_M = x_M$, (1.6) collapses to

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

which again holds under the assumption of positive-definiteness of $D^2_{xy}F$ ($F_{x_C y_M} = F_{x_M y_C}$)

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

1.8.6 Additional Results

Lemma 1.4 (Non-Negative Output and Wages)**.** *Both equilibrium output $F(x_C, x_M, y_C^*, y_M^*)$ and equilibrium wages $w(x_C, x_M)$ are non-negative for all $x_C, x_M \in X$.*

Proof. Both equilibrium output $F(x_C, x_M, y_C^*, y_M^*) = x_C y_C^* + x_M y_M^* = J_{11}x_C^2 + 2J_{12}x_M x_C + \delta J_{22}x_M^2 = \mathbf{x}^T J_\mu \mathbf{x} > 0$ and equilibrium wages $w(x_C, x_M) = \frac{1}{2}J_{11}x_C^2 + J_{12}x_M x_C + \frac{1}{2}\delta J_{22}x_M^2 = \frac{1}{2}\mathbf{x}^T J_\mu \mathbf{x} > 0$ are positive-definite quadratic forms in standard normal variables and hence positive for all non-zero column vectors $\mathbf{x} = (x_C, x_M)$. (Notice that I used $J_{12} = \delta J_{21}$ by (1.6)).[53]

Lemma 1.5 (Task-Biased TC and Relative Wages)**.** *The return to manual specialists decreases relative to cognitive specialists, generalists and low-skilled workers.*

Proof of Result 1.5. Let a worker with $(x_C, x_M) = (|x|, 0), |x| < \infty$ be a specialist in task C and the worker $(x_C, x_M) = (0, |x|), |x| < \infty$ be a specialist in task C. Notice that their relative wage is given by:

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

Differentiating the relative wage with respect to δ gives:

$$\frac{\partial\frac{w(|x|,0)}{w(0,|x|)}}{\partial\delta} = \frac{-\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}(2\delta^2+1)-2\delta}{\left(\delta\left(\delta+\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)\right)^2} < 0 \tag{1.95}$$

Hence, the relative wage increases as δ drops.
Let a worker with $(x_C, x_M) = (|x|, |x|), |x| < \infty$ be a generalist. Then, using (1.15):

$$\frac{w(|x|,|x|)}{w(0,|x|)} = 1 + \underbrace{\frac{\frac{1}{2}\left(1+\delta\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)+\delta\left(\rho_y-\rho_x\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)}{\frac{1}{2}\delta\left(\delta+\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)}}_{\equiv S}. \tag{1.96}$$

[53] For the copula model below it holds that, for all (x_C, x_M) such that $\tilde{x}_C = \Phi^{-1}(H_C(x_C)) \neq 0$ or $\tilde{x}_M = \Phi^{-1}(H_M(x_M)) \neq 0$, $F(x_C, x_M, y_C^*, y_M^*) > 0$ and $w(x_C, x_M) > 0$, which can be verified by simply plugging $\tilde{x}_C, \tilde{x}_M$ into wage and output.

Since,

$$\frac{\partial S}{\partial \delta}=\frac{\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\overbrace{(-\frac{1}{4}\delta^2-\frac{1}{4}+\frac{1}{2}\delta^2\rho_x)}^{<0 \quad \text{for} \quad \rho_x<<1}-\frac{1}{2}\delta-\frac{1}{2}\delta^2\rho_y}{\left(\frac{1}{2}\delta\left(\delta+\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)\right)^2}<0 \tag{1.97}$$

it follows that (1.96) increases as δ drops.

Finally, let a worker with $(x_C, x_M) = (|\epsilon|, |\epsilon|)$ be a low-skilled worker. Then, using (1.15): Then,

$$\frac{w(|\epsilon|,|\epsilon|)}{w(0,|x|)}=\frac{\epsilon^2}{x^2}\left(1+\underbrace{\frac{\frac{1}{2}\left(1+\delta\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)+\delta\left(\rho_y-\rho_x\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)}{\frac{1}{2}\delta\left(\delta+\sqrt{\frac{1-\rho_y^2}{1-\rho_x^2}}\right)}}_{=S}\right) \tag{1.98}$$

It follows from (1.97) that this relative wage is increasing as δ drops.

1.8.7 Generalizations and Extensions

Equilibrium for Non-Standard Normal Distributions

The continuum of square roots used to un-correlate skills and productivities cannot only deal with asymmetries in the technology (i.e. $\delta < 1$) but also with asymmetries in the distributions (i.e. different means and variances of skill and productivity distributions). This section solves the model in closed form for non-standard normally distributed skills and productivities.

Proposition 1.12 (Equilibrium Assignment and Wages under Normality)**.** *Denote the non-standard normally distributed skills by $(\hat{x}_C, \hat{x}_M)$ and the productivities by $(\hat{y}_C, \hat{y}_M)$, where:*

$$\begin{bmatrix}\hat{x}_C\\ \hat{x}_M\end{bmatrix}\sim N\left(\begin{bmatrix}\mu_{\hat{x}_C}\\ \mu_{\hat{x}_M}\end{bmatrix},\begin{bmatrix}\sigma^2_{\hat{x}_C} & \sigma_{\hat{x}_C}\sigma_{\hat{x}_M}\rho_{\hat{x}}\\ \sigma_{\hat{x}_C}\sigma_{\hat{x}_M}\rho_{\hat{x}} & \sigma^2_{\hat{x}_M}\end{bmatrix}\right),\quad \begin{bmatrix}\hat{y}_C\\ \hat{y}_M\end{bmatrix}\sim N\left(\begin{bmatrix}\mu_{\hat{y}_C}\\ \mu_{\hat{y}_M}\end{bmatrix},\begin{bmatrix}\sigma^2_{\hat{y}_C} & \sigma_{\hat{y}_C}\sigma_{\hat{y}_M}\rho_{\hat{y}}\\ \sigma_{\hat{y}_C}\sigma_{\hat{y}_M}\rho_{\hat{y}} & \sigma^2_{\hat{y}_M}\end{bmatrix}\right)$$

(i) Assignment: (i)-(iii) of Proposition 3 apply with minor modifications (see Proof,main paper).

(ii) Wage Schedule: If skills and productivities are normally distributed, the wage is given by

$$\begin{aligned}w(\hat{x}_C,\hat{x}_M)=&\frac{1}{2}J_{11}\frac{\sigma_{\hat{y}_C}}{\sigma_{\hat{x}_C}}\hat{x}_C^2+J_{12}\frac{\sigma_{\hat{y}_C}}{\sigma_{\hat{x}_M}}\hat{x}_C\hat{x}_M+\frac{1}{2}\delta J_{22}\frac{\sigma_{\hat{y}_M}}{\sigma_{\hat{x}_M}}\hat{x}_M^2 \\ &+\left(\mu_{\hat{y}_C}-\frac{\sigma_{\hat{y}_C}}{\sigma_{\hat{x}_C}}\mu_{\hat{x}_C}J_{11}-\frac{\sigma_{\hat{y}_C}}{\sigma_{\hat{x}_M}}\mu_{\hat{x}_M}J_{12}\right)\hat{x}_C+\delta\left(\mu_{\hat{y}_M}-\frac{\sigma_{\hat{y}_M}}{\sigma_{\hat{x}_C}}\mu_{\hat{x}_C}J_{21}-\frac{\sigma_{\hat{y}_M}}{\sigma_{\hat{x}_M}}\mu_{\hat{x}_M}J_{22}\right)\hat{x}_M+w_0\end{aligned} \tag{1.99}$$

where $J_{11}, J_{12}, J_{21}, J_{22}$ are the elements of the Jacobian of the matching function under standard normality (see Proposition 3, main paper).

Proof. .
(i) Recall the equilibrium assignment under *standard normality*:

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1+\delta\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \\ \frac{\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta+\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \end{bmatrix}}_{D_x\mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \tag{1.100}$$

Denote $J_\mu(x_C, x_M) = D_x\mathbf{y}^*$ the Jacobian of the matching function with *standardized* variables, where:

$$J_\mu(x_C, x_M) = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \tag{1.101}$$

Non-standardized skills, denoted by $\hat{x}_i$, and productivities $\hat{y}_i$ need to be standardized in order for (1.100) to hold: $x_i = \frac{\hat{x}_i-\mu_{\hat{x}_i}}{\sigma_{\hat{x}_i}}$ and $y_i = \frac{\hat{y}_i-\mu_{\hat{y}_i}}{\sigma_{\hat{y}_i}}$. This transformation is reversed below to express the assignment in terms of the non-standardized variables $\hat{x}_i, \hat{y}_i$. The equilibrium assignment with normally distributed skills and productivities is given by

$$\begin{aligned} \begin{bmatrix} y_C \\ y_M \end{bmatrix} &= J_\mu \begin{bmatrix} x_C \\ x_M \end{bmatrix} \\ \Leftrightarrow \quad \begin{bmatrix} \frac{\hat{y}_C^*-\mu_{\hat{y}_C}}{\sigma_{\hat{y}_C}} \\ \frac{\hat{y}_M^*-\mu_{\hat{y}_M}}{\sigma_{\hat{y}_M}} \end{bmatrix} &= J_\mu \begin{bmatrix} \frac{\hat{x}_C-\mu_{\hat{x}_C}}{\sigma_{\hat{x}_C}} \\ \frac{\hat{x}_M-\mu_{\hat{x}_M}}{\sigma_{\hat{x}_M}} \end{bmatrix} \end{aligned} \tag{1.102}$$

where J_μ is given in (1.101). (1.102) can be solved explicitly for $(\hat{y}_C^*, \hat{y}_M^*)$ as a function of $(\hat{x}_C, \hat{x}_M)$. Parts (i)-(iii) of Proposition 3 readily apply.

(ii) The wage guess needs to be verified. Denote the parameter in (1.100) by $\hat{\delta}$ instead of δ and set

$$\hat{\delta} = \delta \frac{\sigma_{\hat{y}_M}}{\sigma_{\hat{x}_C}} \frac{\sigma_{\hat{x}_M}}{\sigma_{\hat{y}_C}} \tag{1.103}$$

where $\delta \in [0, 1]$ is again the relative task weight in the production function. Given (1.99), one obtains:

$$\frac{\partial w(\hat{x}_C, \hat{x}_M)}{\partial \hat{x}_C} = \overbrace{J_{11}\frac{\sigma_{\hat{y}_C}}{\sigma_{\hat{x}_C}}(\hat{x}_C - \mu_{\hat{x}_C}) + J_{12}\frac{\sigma_{\hat{y}_C}}{\sigma_{\hat{x}_M}}(\hat{x}_M - \mu_{\hat{x}_M}) + \mu_{\hat{y}_C}}^{\hat{y}_C^*} \tag{1.104}$$

$$\frac{\partial w(\hat{x}_C, \hat{x}_M)}{\partial \hat{x}_M} = \delta \underbrace{\left(J_{21}\frac{\sigma_{\hat{y}_M}}{\sigma_{\hat{x}_C}}(\hat{x}_C - \mu_{\hat{x}_C}) + J_{22}\frac{\sigma_{\hat{y}_M}}{\sigma_{\hat{x}_M}}(\hat{x}_M - \mu_{\hat{x}_M}) + \mu_{\hat{y}_M} \right)}_{\hat{y}_M^*}. \tag{1.105}$$

which coincide with the first-order conditions of the firm, (1.3) and (1.4), evaluated at the equilibrium assignment (1.102). To derive (1.104) and (1.105), I made use of the integrability condition (1.6).

$$\begin{aligned} \frac{\partial^2 w(\hat{x}_C, \hat{x}_M)}{\partial \hat{x}_C \partial \hat{x}_M} &= \frac{\partial^2 w(\hat{x}_C, \hat{x}_M)}{\partial \hat{x}_M \partial \hat{x}_C} \\ \Leftrightarrow \quad J_{12} \frac{\sigma_{\hat{y}_C}}{\sigma_{\hat{x}_M}} &= \delta J_{21} \frac{\sigma_{\hat{y}_M}}{\sigma_{\hat{x}_C}} \\ \Leftrightarrow \quad J_{12} &= \hat{\delta} J_{21} \end{aligned} \tag{1.106}$$

which is satisfied by construction of the equilibrium assignment. Moreover, (1.106) implies uniqueness of $w(\hat{x}_C, \hat{x}_M)$ by Proposition 2 (main paper).

The Gaussian Copula Model

Deriving the assignment under standard normality can be generalized to *any* continuous marginal distributions via *Gaussian copulas*. The idea is to transform skills and productivities from arbitrary marginal distributions into variables that are marginally Gaussian and then use a copula to bind them together. This approach enables me to apply a second transformation that un-correlates *Gaussian* random variables and transforms them into independent standard normal ones. This transformation is crucial for the tractability of the model. The occurrence when the original variables are marginally Gaussian is captured as a special case.

Besides tractability, the advantage of copulas is their flexibility. Copulas are multivariate distributions where the margins cannot only come from different families of distributions but can also include positive and negative variables at the same time. The construction of copulas allows for separating considerations about the marginal distributions on the one hand and dependence of the data on the other. The *Gaussian* copula is particularly tractable[54] and has been widely used for applications when the data exhibit little or no tail dependence. For a discussion on copulas in general and on the Gaussian copula in particular, see, for instance, Cherubini et al. (2004).

In what follows, I formally define copulas with focus on the Gaussian copula and show why they prove useful in this model. I focus on the case of $N = 2$ although this section can be generalized to arbitrary N. A two-dimensional copula is a c.d.f. whose support is contained in $[0, 1]^2$ and whose one-dimensional margins are uniform distributions $U(0, 1)$.[55] Consider a bivariate distribution function $Q(x_1, x_2)$ with univariate marginal distributions $Q_1(x_1), Q_2(x_2)$ and quantile functions Q_1^{-1}, Q_2^{-1}. The copula associated with $Q(x_1, x_2)$ is a distribution function $C : [0, 1]^2 \to [0, 1]$ such that

$$Q(x_1, x_2) = C(Q_1(x_1), Q_2(x_2); \theta) \tag{1.107}$$

where θ is the dependence parameter of the copula, which measures the dependence between

[54](a) The only constraint on the correlation matrix is that it has to be positive definite. (b) One can specify different levels of correlation between the margins.

[55]Copulas were initially introduced by Sklar (1959).

the marginals Q_1 and Q_2. The copula is unique if the marginal c.d.f.'s Q_1 and Q_2 are continuous (Sklar's Theorem). Equation (1.107) captures the relation between distribution functions and copulas. Denote $u_1 = Q_1(x_1), u_2 = Q_2(x_2)$, where u_1, u_2 are uniformly distributed on $[0, 1]$. Then, $x_1 = Q_1^{-1}(u_1), x_2 = Q_2^{-1}(u_2)$ and (1.107) can be expressed as

$$C(u_1, u_2; \theta) = Q(Q_1^{-1}(u_1), Q_2^{-1}(u_2)). \tag{1.108}$$

To construct the *Gaussian copula*, skills and productivities from arbitrary marginal distributions are converted into Gaussian variables using the inverse transform method and then bound together via (1.108). To illustrate this, consider the case of bivariate skills. Assume that the two skills x_C and x_M have marginal distributions $H_C(x_C)$ and $H_M(x_M)$, respectively. Then, $\tilde{x}_C = \Phi^{-1}(H_C(x_C))$ and $\tilde{x}_M = \Phi^{-1}(H_M(x_M))$ are standard normally distributed whose dependence is modeled using the Gaussian copula, given by

$$C(H_C(x_C), H_M(x_M); \rho_{\tilde{x}}) = \Phi_2(\Phi^{-1}(H_C(x_C)), \Phi^{-1}(H_M(x_M))).$$

Then, $(\tilde{x}_C, \tilde{x}_M)$ are standard bivariate normal with correlation $\rho_{\tilde{x}}$.

The equilibrium of the Gaussian copula model can be derived in the exact same way as in Section 3 (main paper): Compute the closed form assignment and wage function, using the transformed variables $(\tilde{x}_C, \tilde{x}_M)$ and $(\tilde{y}_C, \tilde{y}_M)$ (which are bivariate standard Gaussian) as well as the technology (1.7) defined in terms of the transformed variables. Then, the equilibrium assignment is linear and the wage function convex in *transformed* variables.

Notice that important properties of equilibrium assignment and wages in transformed variables $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ can be shown to also hold for the original variables $(\mathbf{x}, \mathbf{y})$. The assignment in terms of *original* variables is given by

$$\begin{aligned}
\tilde{y}_C^* = \tilde{J}_{11}\tilde{x}_C + \tilde{J}_{12}\tilde{x}_M \quad &\Leftrightarrow \quad \Phi^{-1}(G_C(y_C^*)) = \tilde{J}_{11}\Phi^{-1}(H_C(x_C)) + \tilde{J}_{12}\Phi^{-1}(H_M(x_M)) \\
\tilde{y}_M^* = \tilde{J}_{21}\tilde{x}_C + \tilde{J}_{22}\tilde{x}_M \quad &\Leftrightarrow \quad \Phi^{-1}(G_M(y_M^*)) = \tilde{J}_{21}\Phi^{-1}(H_C(x_C)) + \tilde{J}_{22}\Phi^{-1}(H_M(x_M))
\end{aligned}$$

where $\tilde{J}_{11}, \tilde{J}_{12}, \tilde{J}_{21}, \tilde{J}_{22}$ are the elements of the matching function's Jacobian (in transformed variables). The variables $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ were obtained by monotone transformations of $(\mathbf{x}, \mathbf{y})$. Since the sign of the first partial derivative is invariant under monotone transformations of the variables, it holds that $J_{11} \equiv \frac{\partial y_C^*}{\partial x_C} > 0$ and $J_{22} \equiv \frac{\partial y_M^*}{\partial x_M} > 0$ (in original variables). Moreover, the determinant of the matching function's Jacobian in *original* variables is positive if and only if the determinant of the Jacobian in *transformed* variables is positive since

$$Det(D_x\mathbf{y}^*) = R(\Phi, H_C, H_M, G_C, G_M)Det(D_{\tilde{x}}\tilde{\mathbf{y}}^*)$$

where $R(.)$ is a function that takes positive values. Hence, matching in original variables satisfies PAM. The next Lemma summarizes this result.

Lemma 1.6 (PAM in Original Variables). $(y_C^*, y_M^*) = \mu(x_C, x_M)$ *satisfies PAM.*

Proof. $(y_C^*, y_M^*) = \mu(x_C, x_M)$ is implicitly given by

$$\Phi^{-1}(G_C(y_C^*)) = \tilde{J}_{11}\Phi^{-1}(H_C(x_C)) + \tilde{J}_{12}\Phi^{-1}(H_M(x_M)) \tag{1.109}$$

$$\Phi^{-1}(G_M(y_M^*)) = \tilde{J}_{21}\Phi^{-1}(H_C(x_C)) + \tilde{J}_{22}\Phi^{-1}(H_M(x_M)) \tag{1.110}$$

where $\tilde{J}_{11}, \tilde{J}_{12}, \tilde{J}_{21}, \tilde{J}_{22}$ are the elements of the Jacobian of the matching function in transformed variables given by:

$$\begin{bmatrix} y_C^* \\ y_M^* \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{J}_{11} & \tilde{J}_{12} \\ \tilde{J}_{21} & \tilde{J}_{22} \end{bmatrix}}_{D_x\mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1+\delta\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta\left(\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}\right)}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \\ \frac{\rho_y-\rho_x\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} & \frac{\delta+\frac{\sqrt{1-\rho_y^2}}{\sqrt{1-\rho_x^2}}}{\sqrt{1+2\delta(\rho_x\rho_y+\sqrt{1-\rho_y^2}\sqrt{1-\rho_x^2})+\delta^2}} \end{bmatrix}}_{D_x\mathbf{y}^*} \begin{bmatrix} x_C \\ x_M \end{bmatrix} \tag{1.111}$$

It has to be shown that $J_\mu \equiv D_x\mathbf{y}^*$ is a P-matrix. $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ are obtained by monotone transformations of $(\mathbf{x}, \mathbf{y})$. Since the sign of the first partial derivative is invariant under monotone transformations of the variables and $\tilde{J}_{11} = \frac{\partial \tilde{y}_C^*}{\partial \tilde{x}_C} > 0, \tilde{J}_{22} = \frac{\partial \tilde{y}_M^*}{\partial \tilde{x}_M} > 0$ (by Proposition 3, main paper), it holds that $J_{11} = \frac{\partial y_C^*}{\partial x_C} > 0, J_{22} = \frac{\partial y_M^*}{\partial x_M} > 0$. It remains to show that $Det(D_x\mathbf{y}^*) > 0$. Implicit differentiate (1.109) and (1.110) with respect to x_C and x_M, which gives four derivatives. For instance:

$$J_{11} = \frac{\frac{\partial \Phi^{-1}}{\partial H_C}\frac{\partial H_C}{\partial x_C}}{\frac{\partial \Phi^{-1}}{\partial G_C}\frac{\partial G_C}{\partial y_C}}\tilde{J}_{11}$$

Then compute $Det(D_x\mathbf{y}^*)$ as:

$$Det(D_x\mathbf{y}^*) = J_{11}J_{22} - J_{12}J_{21} = \underbrace{\frac{\frac{\partial \Phi^{-1}}{\partial H_C}\frac{\partial H_C}{\partial x_C}\frac{\partial \Phi^{-1}}{\partial H_M}\frac{\partial H_M}{\partial x_M}}{\frac{\partial \Phi^{-1}}{\partial G_C}\frac{\partial G_C}{\partial y_C}\frac{\partial \Phi^{-1}}{\partial G_M}\frac{\partial G_M}{\partial y_M}}}_{\equiv R(\Phi, H_C, H_M, G_C, G_M)} \underbrace{(\tilde{J}_{11}\tilde{J}_{22} - \tilde{J}_{12}\tilde{J}_{21})}_{Det(D_{\tilde{x}}\tilde{\mathbf{y}}^*)}$$

where $R(\Phi, H_C, H_M, G_C, G_M)$ takes only positive values because it involves derivatives of strictly increasing c.d.f.'s and where $Det(D_{\tilde{x}}\tilde{\mathbf{y}}^*)$ is positive from the proof of Proposition 3. Hence, $Det(D_x\mathbf{y}^*) > 0$ if and only if $Det(D_{\tilde{x}}\tilde{\mathbf{y}}^*) > 0$.

To find the wage as a function of the original variables, $w(x_C, x_M)$, substitute $\tilde{x}_i = \Phi^{-1}(H_i(x_i)), i \in \{C, M\}$ into (1.15). Notice that (1.15) is a positive-definite quadratic form in standard normal variables *for all* original skills (x_C, x_M). Hence, even though original skills and productivities are allowed to be negative, wages and output are non-negative for all matches that form in equilibrium (see Section 1.4 in this Appendix). The intuition is that supermodularity of skills and productivities within tasks induces PAM, which is a force towards matching firms and workers of similar types.

This section closes with an illustrative example of the copula approach. Consider standard normal marginal distributions of skills and productivities, $H_i = \Phi, G_i = \Phi$, so that $\tilde{x}_i = x_i$ and $\tilde{y}_i = y_i, \forall i \in \{C, M\}$. Then, the assignment is linear in the original skills (x_C, x_M)

and productivities (y_C, y_M)

$$\begin{aligned} \tilde{y}_C^* &= \tilde{J}_{11}\tilde{x}_C + \tilde{J}_{12}\tilde{x}_M \quad &\Leftrightarrow \quad y_C^* &= J_{11}x_C + J_{12}x_M \\ \tilde{y}_M^* &= \tilde{J}_{21}\tilde{x}_C + \tilde{J}_{22}\tilde{x}_M \quad &\Leftrightarrow \quad y_M^* &= J_{21}x_C + J_{22}x_M \end{aligned}$$

where $J_{11}, J_{12}, J_{21}, J_{22}$ are the elements of the matching function's Jacobian, given by (1.111). Hence, the example with (standard) normally distributed skills and productivities is captured as a special case of the Gaussian copula model.

Labor Market with Search Frictions

In this section, the model of multidimensional heterogeneity is embedded into a model with search frictions on the labor market and directed search. The search frictions stem from the lack of coordination of a large number of agents when applying for jobs. As in the baseline model with competitive labor market, I derive a sufficient condition on the production technology for purity of the equilibrium and assortative matching. This condition is a straightforward generalization of *root-supermodularity* from the one-dimensional setting (Eeckhout and Kircher (2010)) to the setting with multidimensional heterogeneity. I briefly outline the model, which is identical to theirs except that it allows for two-dimensional skills and productivities.

The frictional hiring process of firms can be described by a static game with three stages: In the first stage, every firm, characterized by some productivity bundle $\mathbf{y} = (y_C, y_M)$, posts a wage $w(\mathbf{y})$. In a second stage, unemployed workers observe these wages. They anticipate that different wages $w(\mathbf{y})$ are associated with different applicant-vacancy ratios $q(\mathbf{y}) \in [0, \infty]$, which I will refer to as the *queue length*. They choose to apply at firm $\mathbf{y}$, characterized by a pair $(w(\mathbf{y}), q(\mathbf{y}))$, to maximize their expected income. In the last stage, firms that receive at least one application hire one worker and production takes place. If a firm receives more than one applicant, then it hires one of them at random. Unmatched workers remain unemployed and unmatched firms will end up with a vacant job. Unmatched agents produce nothing and have zero payoff.

Denote the probability that firm $\mathbf{y}$ fills a vacancy by $m(q(\mathbf{y}))$ and the probability that a worker is hired by that firm by $\frac{m(q(\mathbf{y}))}{q(\mathbf{y})}$ where the matching technology has the following properties: $m(.)$ is twice differentiable and satisfies $m_q(.) > 0$, $m_{qq}(.) < 0$, $\frac{\partial \frac{m(q(\mathbf{y}))}{q(\mathbf{y})}}{\partial q(\mathbf{y})} < 0$, $m(0) = 0$ and $m(\infty) = 1$. In words, the vacancy-filling probability is strictly increasing and strictly concave in the queue length whereas the worker's hiring probability is strictly decreasing in the number of other workers queueing up for the same job.

In order for a firm to attract a certain worker type $\mathbf{x} = (x_C, x_M)$ it needs to offer him an expected payoff that is at least as high as his expected equilibrium market utility $U(\mathbf{x})$. $U(\mathbf{x}) = \frac{m(q(\mathbf{y}))}{q(\mathbf{y})} w(\mathbf{y})$ is what he would get at his best alternative job, where $w(\mathbf{y})$ is the worker's actual wage when hired. The expected market wage $U(x)$ implicitly defines the queue length $q(\mathbf{y})$ as a function of the actual wage $w(\mathbf{y})$, which can be shown to be an increasing function. Notice that the firms (and workers) take $U(\mathbf{x})$ and hence the relationship between $q(\mathbf{y})$ and $w(\mathbf{y})$ as given, which is justified by the large number of both workers and firms.

Contrary to the competitive labor market, however, here the firm chooses the wage. The firm's problem is:

$$\max_{w,q} \quad m(q)(F(\mathbf{x},\mathbf{y}) - w) \tag{1.112}$$

$$s.t. \quad \frac{m(q)}{q} w \geq U(\mathbf{x}) \tag{1.113}$$

Before analyzing the equilibrium properties, it is useful to define a pure equilibrium in this setting.

Definition 1.5 (Pure Equilibrium with Search)**.** *A pure symmetric competitive search equilibrium consists of wages, queue length, and market utility* $(w(\mathbf{y}), q(\mathbf{y}), U(\mathbf{x}))$ *as well as a mapping* $\mathbf{y}^* = \mu(\mathbf{x})$ *s.t.:*
(i) Firm optimality: Given $U(\mathbf{x})$ *and other firms' strategies, each firm* $\mathbf{y}$ *solves (1.112) s.t. (1.113);*
(ii) Worker optimality: A worker $\mathbf{x}$ *applies to a job at* $\mathbf{y}$ *only if* $\frac{m(q(\mathbf{y}))}{q(\mathbf{y})} w(\mathbf{y}) \geq U(\mathbf{x})$*;*
(iii) Purity: $\mathbf{y}^* = \mu(\mathbf{x})$ *is a one-to-one function;*
(iv) Market clearing: The queue length q *satisfies the market clearing for applications.*

In equilibrium, the firm will not offer more utility to the worker than necessary to attract him. Hence, the constraint (1.113) holds with equality. The maximization problem can be reformulated as:

$$\max_{\mathbf{x},q,w} \quad m(q)(F(\mathbf{x},\mathbf{y}) - w)$$

$$s.t. \quad \frac{m(q)}{q} w = U(\mathbf{x}) \tag{1.114}$$

In words, the firm can choose the wage and the trading probabilities in order to attract a certain worker type (to whom it needs to offer his market utility). The constraint is similar to a participation constraint. When substituting (1.114) into the firm's objective function, the maximization problem reads:

$$\max_{\mathbf{x},q} \quad m(q)F(\mathbf{x},\mathbf{y}) - qU(\mathbf{x})$$

The first oder conditions with respect to $q, \mathbf{x}$ are respectively given by

$$m_q(q)F(\mathbf{x},\mathbf{y}) - U(\mathbf{x}) = 0$$
$$m(q)F_{x_C}(\mathbf{x},\mathbf{y}) - qU_{x_C}(\mathbf{x}) = 0$$
$$m(q)F_{x_M}(\mathbf{x},\mathbf{y}) - qU_{x_M}(\mathbf{x}) = 0$$

where subscripts denote derivatives. An assignment is consistent with the equilibrium only if the second-order conditions are satisfied, i.e. if the Hessian is negative semi-definite. Denote the decision variables of the firm by the vector $\mathbf{z} = (\mathbf{x}, q)$, where $\mathbf{x} = (x_C, x_M)$. Hence, the Hessian is a 3x3 matrix. Differentiating the FOCs (evaluated at $\mathbf{y}^*$) with respect to $\mathbf{z} = (\mathbf{x}, q)$

yields the Hessian evaluated at the equilibrium assignment $\mathbf{y}^*$

$$D^2_{zz}(m(q(\mathbf{y}^*))F(\mathbf{x},\mathbf{y}^*) - q(\mathbf{y}^*)U(\mathbf{x})) = D^2_{qq}\left(m(q(\mathbf{y}^*))F(\mathbf{x},\mathbf{y}^*)\right)\left(\frac{F(\mathbf{x},\mathbf{y}^*)F_{xy}(\mathbf{x},\mathbf{y}^*)}{F_x(\mathbf{x},\mathbf{y}^*)F_y(\mathbf{x},\mathbf{y}^*)} - \epsilon_M(q(\mathbf{y}^*))\right) D_x\mathbf{y}^* \tag{1.115}$$

where $\epsilon_M(q) \equiv \frac{m_q(q)(m_q(q)q-m(q))}{m(q)m_{qq}(q)q} = \frac{M_u(q)M_v(q)}{M_{uv}(q)M(q)}$ is the elasticity of the aggregate matching function, as in Eeckhout and Kircher (2010).[56] A necessary condition for optimality is that (1.115) is negative semi-definite. Notice that by assumption $m_{qq}(q) < 0$ and hence $D^2_{qq}\left(m(q)F(\mathbf{x},\mathbf{y}^*)\right)$ is a negative scalar. (1.115) is the multidimensional extension of expression (12) in Eeckhout and Kircher (2010). Optimality requires that

$$\left(\frac{F(\mathbf{x},\mathbf{y}^*)F_{xy}(\mathbf{x},\mathbf{y}^*)}{F_x(\mathbf{x},\mathbf{y}^*)F_y(\mathbf{x},\mathbf{y}^*)} - \epsilon_M(q(\mathbf{y}^*))\right)(D_x\mathbf{y}^*) \tag{1.116}$$

is positive semi-definite. The following proposition states a sufficient condition for purity and assortativeness of the equilibrium assignment in the multidimensional setting.

Proposition 1.13 (Pure and Assortative Equilibrium under Search)**.** *If*

$$\left[\frac{FF_{xy}}{F_yF_x} - \epsilon_M(q)\right] = \begin{bmatrix} \frac{FF_{x_Cy_C}}{F_{y_C}F_{x_C}} - \epsilon_M(q) & \frac{FF_{x_Cy_M}}{F_{y_M}F_{x_C}} - \epsilon_M(q) \\ \frac{FF_{x_My_C}}{F_{y_C}F_{x_M}} - \epsilon_M(q) & \frac{FF_{x_My_M}}{F_{y_M}F_{x_M}} - \epsilon_M(q) \end{bmatrix} \tag{1.117}$$

is a diagonal P-matrix (P^--matrix), then $D_x\mathbf{y}^*$ is a P-matrix (P^--matrix). The assignment $\mu(\mathbf{x}) = \mathbf{y}^*$ is positive (negative) assortative. The equilibrium (if it exists) is globally unique.[57]

Proof of Proposition 1.13. When $\left[\frac{FF_{xy}}{F_yF_x} - \epsilon_M(q)\right]$ is diagonal, then the symmetry of the Hessian, given by (1.116), requires that $D_x\mathbf{y}^*$ is sign-symmetric. The proof of Lemma [1] applies, which proves that $D_x\mathbf{y}^*$ is a P-matrix. The result on assortativeness follows. The proof for global uniqueness is stated by Eeckhout and Kircher (2010) (p. 569) and applies here as well.

In the setting with one-dimensional heterogeneity, the matrix condition of Proposition 1.13 reduces to the technological condition of root-supermodularity, $\frac{FF_{xy}}{F_yF_x} - \epsilon_M(q) \geq 0$, which ensures both assortativeness and uniqueness. Root-supermodularity is a stronger notion of complementarity than supermodularity. This concept is extensively discussed in Eeckhout and Kircher. When there are search frictions, high skilled workers and high productivity firms have strong incentives to secure a match during search. The types who provide this trading insurance are the low types because matching is less important to them. Hence, search frictions are a force towards negative assortative matching. In order for PAM to obtain, the complementarities between skills and productivities must be stronger than supermodularity (which is required under the competitive labor market). If the production function is root-submodular, NAM obtains in equilibrium.

[56] The aggregate matching function is defined as the total number of matches that form when u workers and v vacancies are in the market $M(u,v) = vm(u/v) = vm(q)$.

[57] The focus here is on the characterization of the equilibrium. Existence is dealt with in Eeckhout and Kircher (2010).

1.8.8 Quantitative Analysis

The Data

I drop observations with missing wage data or those with hourly wages smaller than one euro or one dollar. Notice that labor income in the NLSY is truncated. The NLSY truncation algorithm takes the top two percent of respondents with valid values and averages them. The averaged value replaces the value for all cases in the top range.

Also notice that due to the panel structure of the NLSY that starts surveying and following young people over time starting in 1979, only a certain age group is present in the US sample in each year. As a result of the data limitation of the NLSY, I am following one (young) cohort over the 1990s.

Table 1.3: Age of the Observations by Year

year	**min**	**max**
1992	27	35
1993	28	36
1994	29	37
1996	31	39
1998	33	41
2000	35	43
Total	27	43

To increase the number of observations, I do *not* only consider observations from the balanced panel but treat the data as subsequent cross sections. The results are qualitatively robust to considering the panel.

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 this 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 1.4 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.

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

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 1.4: Examples of Occupations' Manual and Cognitive Skill Requirements

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.[58] To obtain a single cognitive and a single manual skill from the skill vectors, I take the *maximum* skills from the vectors.[59] 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.[60] On the other hand, if the worker with

[58] 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.

[59] Taking averages leads to similar results in the analysis.

[60] I might over-estimate the skills of those workers with missing data if their skills are missing because they

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.[61] 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.

Tables 1.5 and 1.6 provide summary statistics of skill and productivity distributions in 1992 and 2000. Table 1.5 shows that jobs in the US require on average a higher level of cognitive than manual skills in both years. In line with this demand, workers hold more cognitive than manual skills. Over time, both skill and skill demand distributions are relatively stable in terms of means and variances of the marginal distributions.

	μ	σ	min	max
x_C	.41	.88	-2.05	2.67
y_C	.08	.86	-2.18	3.55
x_M	-.14	.95	-1.79	2.20
y_M	-.11	1.04	-1.69	2.19

	μ	σ	min	max
x_C	.49	.89	-1.97	3.55
y_C	.16	.90	-2.06	2.26
x_M	-.07	1.01	-1.74	2.16
y_M	-.16	1.06	-1.69	2.10

Table 1.5: Summary Statistics of Skills and Skill Demand Distributions in 1992 (left) and 2000 (right)

The skill correlation indicates how specialized the workforce is, with a more negative correlation pointing to more specialized workers who are either good in the cognitive or in the manual task. Similarly, a strongly negative productivity correlation indicates that most jobs require either in manual dexterity or cognitive ability and few jobs require a balanced skill set. Table 1.6 shows that in the US in 1992 most jobs are specialized, indicated by a negative productivity correlation of around -.5. Notice that the demand for skills is considerably more specialized than skill supply.

	1992	2000
ρ_y	-0.4741	-0.5081
ρ_x	-0.2428	-0.1283

Table 1.6: Estimated Skill and Productivity Correlations in 1992 (untransformed)

The distributional parameters are estimated from the empirical distributions. The correlations of the transformed data (i.e. after transforming the skill and productivity distributions to Gaussian copulas; see Section 1.8.7), are given by Table 1.7:

To obtain the proportions of generalists, manual specialists, cognitive specialists and low-skilled workers in the US, I define cognitive (manual) specialists as worker with cognitive (manual) skills above the mean and manual (cognitive) skills below the mean. Similarly, generalists are workers who have above-average skills on both accounts whereas low-skilled

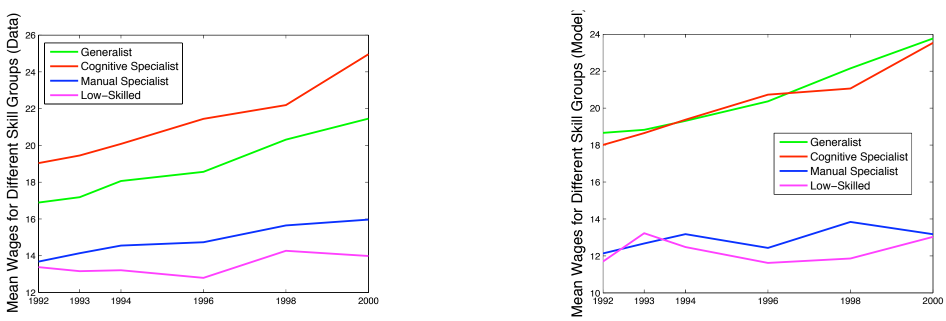
have not acquired any skills after high school. However, drawing the missing skills from the *lower* part of the distribution of other high school graduates leads to similar results in the following analysis.

[61] 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.

year	ρ_x	ρ_y
1992	-0.2079	-0.4156
1993	-0.1893	-0.4263
1994	-0.1028	-0.4391
1996	-0.0721	-0.4459
1998	-0.0992	-0.4457
2000	-.0535	-0.4337

Table 1.7: Estimated Skill and Productivity Correlations over Time (transformed)

workers hold below-average skills. In 1992, the proportion of generalists, cognitive specialists, manual specialists and low-skilled workers is respectively, 18.69%, 34,51%, 30,92%, 15,88%. The evolution of their hourly wages (CPI adjusted) during the 90s is plotted in the Figure below (left panel: data, right panel: model).



Data

Model

Figure 1.11: US Wage Polarization over 1992-2000 (Data and Model)

Maximum Likelihood Estimation

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

$$\begin{aligned} w &= \frac{1}{2}\alpha J_{11}x_C^2 + \alpha J_{12}x_C x_M + \frac{1}{2}\beta J_{22}x_M^2 + \lambda x_C + \eta x_M + w_0 + \epsilon_w \\ y_C &= J_{11}x_C + J_{12}x_M + \epsilon_C \\ y_M &= J_{21}x_C + J_{22}x_M + \epsilon_M \end{aligned}$$

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,

$$w|x_C, x_M \sim N(\frac{1}{2}\alpha J_{11}x_C^2 + \alpha J_{12}x_C x_M + \frac{1}{2}\beta J_{22}x_B^2 + \lambda x_C + \eta x_M + w_0 + \epsilon_w, s^2) \tag{1.118}$$

$$y_C|x_C, x_M \sim N(J_{11}x_C + J_{12}x_M, t^2) \tag{1.119}$$

$$y_M|x_C, x_M \sim N(J_{21}x_C + J_{22}x_M, u^2). \tag{1.120}$$

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 $z = (z_1, ...z_n)$ where $\forall i = 1, ...n, z_i = (w_i, y_{Ci}, y_{Mi}, x_{Ci}, x_{Mi})$; 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

$$\begin{aligned} L(\theta|z) &= \Pi_{i=1}^{n}[w_i, y_{Ci}, y_{Mi}|x_{Ci}, x_{Mi}, \alpha, \beta, J_{11}, J_{12}, J_{21}, J_{22}, \lambda, \eta, w_0] \\ &= \Pi_{i=1}^{n}[w_i|x_{Ci}, x_{Mi}, \alpha, \beta, J_{11}, J_{12}, J_{21}, J_{22}, \lambda, \eta, w_0] \\ &\times \Pi_{i=1}^{n}[y_{Ci}|x_{Ci}, x_{Mi}, J_{11}, J_{12}, J_{21}, J_{22}] \times \Pi_{i=1}^{n}[y_{Mi}|x_{Ci}, x_{Mi}, J_{11}, J_{12}, J_{21}, J_{22}] \end{aligned} \tag{1.121}$$

From (1.121), one obtains (1.18) when using (1.118)-(1.120) and taking logs. The parameter estimates are obtained by maximizing (1.18) with respect to θ.

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

Standard errors in parentheses

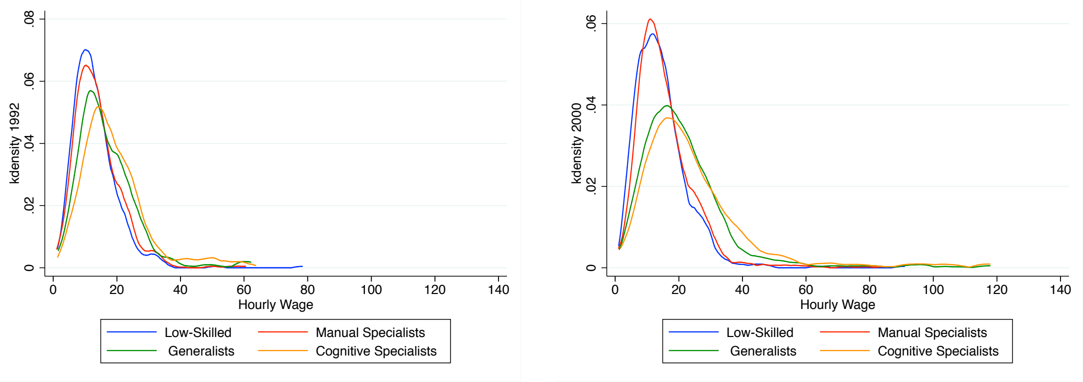
Table 1.8: Maximum Likelihood Estimates for Years 1992-2000

Estimation Results One-Dimensional Model

The following results are obtained from an OLS regressions of [19].

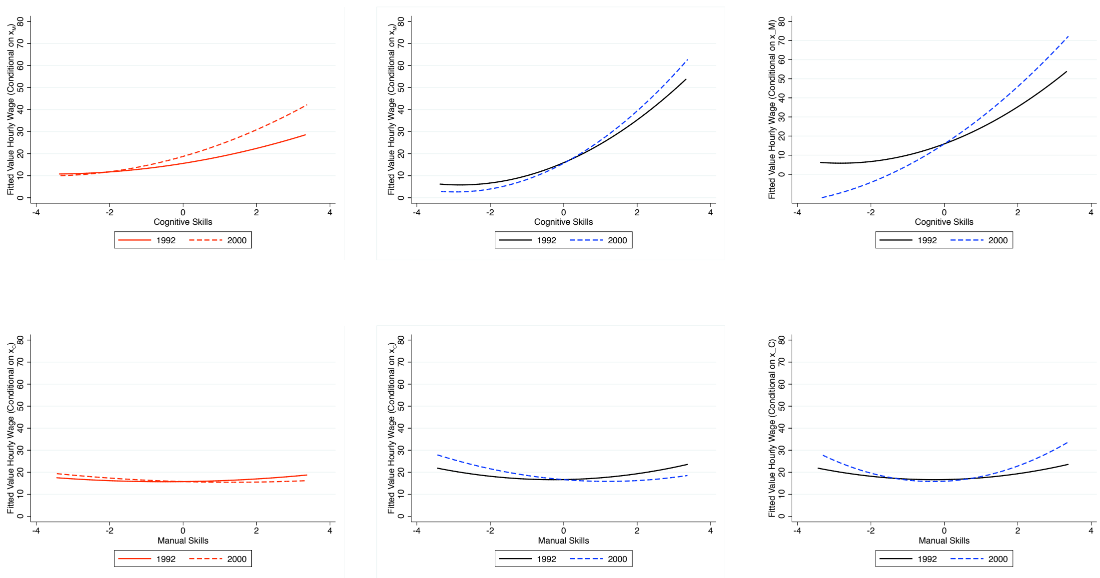
	α	λ	w_0
1992	0.7289	2.7454	15.5910
	(0.2357)	(0.1633)	(0.2006)
1993	0.4309	2.9150	16.0110
	(0.2382)	(0.1654)	(0.2021)
1994	0.2704	3.0447	16.5016
	(0.2385)	(0.1666)	(0.2040)
1996	0.9897	3.8227	16.6513
	(0.3461)	(0.2384)	(0.2912)
1998	1.3519	3.7958	17.6281
	(0.3115)	(0.2156)	(0.2636)
2000	1.3724	4.7944	18.7158
	(0.3777)	(0.2612)	(0.3203)

Standard errors in parentheses



(a) 1992 (b) 2000

Figure 1.12: Wage Densities by Worker Group (Data)



(a) Data (b) Model (TBTC) (c) Model (SBTC)

Figure 1.13: (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.

Bibliography

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999, March). High wage workers and high wage firms. *Econometrica 67*(2), 251–334.

Acemoglu, D. and D. Autor (2011). *Skills, Tasks and Technologies: Implications for Employment and Earnings*, Volume 4 of *Handbook of Labor Economics*, Chapter 12, pp. 1043–1171. Elsevier.

Autor, D. H. and D. Dorn (2012). The growth of low skill service jobs and the polarization of the u.s. labor market. *American Economic Review (forthcoming)*.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006, May). The polarization of the u.s. labor market. *American Economic Review 96*(2), 189–194.

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

Becker, G. S. (1973, July-Aug.). A theory of marriage: Part i. *Journal of Political Economy 81*(4), 813–46.

Boehm, M. J. (2013). The wage effects of job polarization: Evidence from the the wage effects of job polarization: Evidence from the allocation of talents. *Working Paper*.

Bojilov, R. and A. Galichon (2013). Closed-form solutions for multivariate matching.

Brown, J. N. and H. S. Rosen (1982, May). On the estimation of structural hedonic price models. NBER Technical Working Papers 0018, National Bureau of Economic Research, Inc.

Chiappori, P.-A., R. McCann, and L. Nesheim (2010, February). Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory 42*(2), 317–354.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012). Matching with a handicap: The case of smoking in the marriage market. Working Paper.

Choo, E. and A. Siow (2006, February). Who marries whom and why. *Journal of Political Economy 114*(1), 175–201.

Chua, L. and Y. Lam (1972). Global homeomorphism of vector-valued functions. *Journal of Mathematical Analysis and Applications 39*(3), 600–624.

Costinot, A. and J. Vogel (2010, 08). Matching and inequality in the world economy. *Journal of Political Economy 118*(4), 747–786.

Decker, C., E. H. Lieb, R. J. McCann, and B. K. Stephens (2013). Unique equilibria and substitution effects in a stochastic model of the marriage market. *Journal of Economic Theory 148*(2), 778–792.

Dizdar, D. and B. Moldovanu (2012). Surplus division and efficient matching. Working Paper, University of Bonn.

Dupuy, A. and A. Galichon (2012, October). Personality traits and the marriage market. IZA Discussion Papers 6943, Institute for the Study of Labor (IZA).

Eeckhout, J. and P. Kircher (2011). Identifying sorting–in theory. *Review of Economic Studies 78*(3), 872–906.

Eeckhout, J. and P. Kircher (2012). Assortative matching in large firms: Span of control over more versus better workers. *Working Paper*.

Ekeland, I. (2010). Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types. *Economic Theory 42*, 275–315.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004, February). Identification and estimation of hedonic models. *Journal of Political Economy 112*(S1), S60–S109.

Gale, D. and H. Nikaido (1965). The jacobian matrix and global univalence of mappings. *Mathematische Annalen 159*(2), 81–93.

Galichon, A. and B. Salanié (2010, June). Matching with trade-offs: Revealed preferences over competing characteristics. CEPR Discussion Papers 7858, C.E.P.R. Discussion Papers.

Gretsky, N. E., J. M. Ostroy, and W. R. Zame (1992, January). The nonatomic assignment model. *Economic Theory 2*(1), 103–27.

Hershkowitz, D. and N. Keller (2005). Spectral properties of sign symmetric matrices. *Electronic Journal of Linear Algebra 13*, 90–110.

Lazear, E. P. (2009, October). Firm-specific human capital: A skill-weights approach. *Journal of Political Economy 117*(5), 914–940.

Magnus, J. R. (1978). The moments of products of quadratic forms in normal variables. *Statistica Neerlandica 32*(4), 201–210.

McCann, R. J., X. Shi, A. Siow, and R. P. Wolthoff (2012, April). Becker meets ricardo: Multisector matching with social and cognitive skills. IZA Discussion Papers 6533, Institute for the Study of Labor (IZA).

Olkin, I. and F. Pukelsheim (1982). The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications 48*(0), 257 – 263.

Papageorgiou, T. (2013). Learning your comparative advantage. *Review of Economic Studies forthcoming.*

Sanders, C. (2012). Skill uncertainty, skill accumulation, and occupational choice. *Working Paper*.

Shapley, L. and M. Shubik (1971). The assignment game i: The core. *International Journal of Game Theory 1*, 111–130.

Simsek, A., A. Ozdaglar, and D. Acemoglu (2005). On the uniqueness of solutions for nonlinear and mixed complementarity problems. Working Paper.

Tinbergen, J. (1956). On the theory of income distribution. *Weltwirtschaftliches Weltwirtschaftliches Archiv 77*(2), 155–73.

Uren, L. and G. Virag (2011, 05). Skill requirements, search frictions, and wage inequality. *International Economic Review 52*(2), 379–406.

Villani, C. (2009). *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften. Springer London, Limited.

Willis, R. J. and S. Rosen (1979, October). Education and self-selection. *Journal of Political Economy 87*(5), S7–36.

Yamaguchi, S. (2012). Tasks and heterogeneous human capital. *Journal of Labor Economics 30*(1), 1 – 53.

2 Gender, Social Networks and Performance

"Loose connections are the connections you need. It's the No. 1 rule of business."
Sallie Krawcheck, owner of the global women's network *85 Broads*[1]

2.1 Introduction

Gender differences in labor market outcomes remain striking. In the US, women's earnings in 2012 were 80.9% of men's earnings.[2] Even though part of it can be explained by occupational sorting, *within* occupations wage gaps are considerable. Management occupations, such as financial manager and chief executive, are particularly affected, whereas healthcare support and administrative occupations show much smaller gaps.[3] Similar patterns were found for the UK, where full-time working women in the financial sector earn 55% less than full-time male workers – a gap twice as large as the gap in the economy as a whole.[4] What these high-wage-gap occupations have in common is that they are characterized by a large amount of uncertainty, commonly measured by earnings variability. Earnings of both executives and financial managers are largely based on performance pay and thus not constant. Women's lower earnings in these occupations are mainly due to large differences in performance pay and bonuses, suggesting that men perform better. At the same time, and possibly as a logical consequence, more men than women sort into occupations with high earnings volatility.[5] But why do women perform relatively poorly in "high-risk" occupations and avoid them?

In this paper, we offer a novel answer to this question, which is based on social network heterogeneity between men and women.[6] We argue that men's network structures allow them to better perform in uncertain environments compared to women and our model clarifies how this works. This approach is motivated by our empirical finding that men's and women's social networks differ. We show in the AddHealth Data Set that women have less friends than

[1]Krawcheck at Marie Claire's luncheon for the *New Guard*, November 2013.

[2]See http://www.bls.gov/cps/cpswom2012.pdf

[3]See Reports (2013), http://www.bls.gov/cps/cpswom2012.pdf

[4]Wage differences are considerable even when controlling for hours of work (full time) and type of job. See the report by the Equality and H.R.Commission (2009).

[5]See Dohmen and Falk (2011).

[6]Common explanations for these patterns are discrimination against women in male-dominated environments, or differences in preferences and risk aversion. See Eckel and Grossman (2008) for an overview of the literature that finds women to be more risk averse than men. Other explanations involve differences in bargaining strength, which can account for part of the gender wage gap (Card et al. (2013)) as well as future fertility concerns which leads women to self-select in different occupations (Adda et al. (2011)).

men, but their friends are more likely to be friends among each other.[7] Thus, women have smaller but tighter networks, whereas men have larger but looser ego networks. To the best of our knowledge, we are the first to document this empirical fact.

We argue that tight and loose networks provide different types of social capital: A tight network fosters trust or peer pressure among agents, as it prevents them from shirking because they fear repercussions not only from the individual they affect directly with their behavior but also from other members of their network. As a result, closed networks help overcome free-riding problems (Coleman (1988a)).[8] But network closure comes at a cost. Networks with high closure do not allow individuals to access as much information and other low-value resources as networks with lower closure. Being in a loose network with links to individuals that are not connected themselves is particularly valuable for information acquisition. This is what the literature has referred to the "strength of weak ties" (Granovetter (1973)).[9] We are interested under what circumstances tightly connected female networks and thus high peer pressure are more important for *performance on the job* and in what environments the opposite is the case.[10]

We focus on differences in the network *structure* of workers and develop a theory that connects workers' networks to their outcomes at work. In our model, workers repeatedly form partnerships to complete projects. Project success positively depends on the partners' efforts. Effort is unobservable and only the project outcome is public information. If the project is completed successfully, the project payoff is shared between the team members. Because output is split but costs are not, there is a team moral hazard problem at work. As a result, the project partners exert inefficiently low effort.[11] We will show how networks can help attenuate this moral hazard problem by increasing effort.

We are interested in the effort levels of the project partners as a proxy for their performance and specifically in the factors influencing this choice. First, the choice of effort depends on *information* about the value of the project, which can be high or low, depending on the state of the world. Workers receive signals about the state and form expectations about the project value. These expectations influence effort. The more information a worker has, the more precise is his belief about the state of the world. This allows for a better judgment whether high effort (in case the signals point towards the good state) or low effort (if the signals make the low state more probable) should be exerted. Second, effort positively depends on the amount of *peer pressure* individuals face.

How much information a worker has and how much peer pressure he faces depends on

[7]We are using data of teenagers, not of men and women who are already employed. We do this as we are interested in the informal networks, not the formal ones. To our knowledge there does not exist a data set that contains information on *informal* networks at the workplace.

[8]Specifically, closed networks mitigate free-riding problems through the creation of norms and punishment systems. Coleman (1988b) emphasizes the importance of this mechanism for diamond traders in New York.

[9]These two types of social capital can also be related to the concepts of bonding versus bridging social capital defined in Putnam (2000).

[10]We do not aim to address the question of job *search*, as has been done, for example, in Arrow and Borzekowski (2004), Calvo-Armengol and Jackson (2004), Calvó-Armengol and Jackson (2007), but we are interested in how the network structure matters once the job has been found.

[11]See Holmstrom (1982) for moral hazard problems in teams.

his network structure. Workers with a higher degree (i.e. with more friends) hold more information, as they receive a higher number of signals about the state of the world. In turn, workers with higher clustering (i.e. agents whose friends are also friends among each other) face more peer pressure through the following mechanism: A failed project leads to discord between the project partners. But this discord also affects their common friends, that is their disagreement spreads through the entire group – an idea that is based on the *structural balance theory*.[12] Since an intact friendship is necessary for a successful project, repercussions of a failure are worse for a worker with high clustering compared to someone with a looser network. Therefore, higher clustering leads to higher effort in order to be on good terms with future potential project partners.

Our model then allows us to rank the networks of workers regarding their benefit for job performance. We show under which circumstances a network with higher clustering is more beneficial for performance and ultimately wages and when a network with a higher degree is more advantageous. Our main findings are as follows: A higher degree is more beneficial for performance in environments where the uncertainty about the project value is considerable, which is particularly true when (i) overall information (that is information coming from sources unrelated to the network) is scarce, (ii) when signals are noisy and (iii) when project rewards differ significantly across the two states. In these cases, uncertainty about the state of the world is large and the benefits of a purely information-based, loose network outweigh the benefits of a closed network that leads to more peer pressure. In turn, peer pressure leads to higher effort and thus project completion in environments characterized by certainty where additional information has no value. In general, someone with more information can better fine-tune his effort to the expected project reward, exerting high effort only when there is something at stake. In turn, a worker facing high peer pressure exerts extra effort even if the project reward is expected to be low.

Effort choices directly translate into wages. Someone with higher clustering earns more than someone with higher degree when uncertainty about the state is negligible (in both states of the world). Such a worker also has a comparative advantage in jobs whose outcomes are more certain compared to jobs with less certain outcomes. Finally, we show that, due to the dynamic effect of clustering, there is a strong persistence of wage patterns across time, consolidating early career wage gaps.

We then model a man's network as one that is characterized by a relatively high degree and a woman's network as one that is characterized by relatively high clustering. We provide a mechanism of how this social network heterogeneity relates to differences in labor market outcomes of men and women and show that our theory is consistent with a variety of empirical facts: (1) Wage gaps within occupations are large and especially within those occupations that characterized by uncertainty.[13] (2) More men than women choose occupations with high earnings volatility (Dohmen and Falk (2011)). In our model, this would happen even

[12] This is a concept first proposed by Heider (1946) who has spawned a field of research that remains active even today. For an overview on the numerous works on structural balance theory, see Easley and Kleinberg (2010), chapters 3 and 5.

[13] See, for instance, http://www.bls.gov/cps/cpswom2012.pdf or Equality and H.R.Commission (2009).

though both men and women are risk-neutral and thus have the same attitude towards risk. The reason is that women have a comparative advantage in job environments characterized by little uncertainty. (3) Having women in the network is particularly beneficial high up in the organizational hierarchy (Lalanne and Seabright (2011)). In light of our model, we expect that having women in the network is particularly beneficial when information is abundant. We argue that this is the case at higher levels of the organizational hierarchy when networks have grown large rather than in low positions that are commonly held at the beginning of the career. (4) During recessions (i.e. when returns are low) men's unemployment exceeds women's unemployment (Albanesi and Sahin (2013)). Our model predicts that, incentivized by peer pressure, women put over-effort despite low expected rewards whereas men are more selective in their effort choice. (5) The beginning of the career is the crucial period for the wage gap (Babcock and Laschever (2003), Gerhart and Rynes (1991), Martell et al. (1996)). In our model, an initial wage gap is strongly persistent because women are deprived of more project opportunities over time due to their high clustering. This makes it difficult for them to catch up.

In sum, we expect that, based on their loose networks, men outperform women in work environments that are characterized by uncertainty but yield high expected returns – conditions that are typical for a large number of jobs in business and research. Our predictions are in line with the claim of various business leaders that loose and not deep connections are the key to success in business.

We add to the limited amount of network literature that evaluates the trade-off between network density and the network span. We explicitly model the impact of these network features on peer pressure and information acquisition at the work place. This allows us to gain novel insights on the impact of these network characteristics on labor market outcomes. The trade-off between network density and span has also been analysed in Karlan et al. (2009) where individuals use their network to borrow goods. They focus on the trust generated in networks and find that higher closure increases trust but reduces access. Although the impact of network closure on economic outcomes is analysed in both Karlan et al. (2009) and our work, the theoretical frameworks and applications are entirely different. Dixit (2003) also discusses the trade-off between sparse and closed networks in a trade setting. He focuses on the role of self governance, as an alternative to official institutions, in trading relationships. Trading with more distant individuals offers higher gains, but information flows about cheating are decreasing in this distance. There is a clear trade-off between networks that have a high closure, that is a local bias in trade, and networks that span a larger distance but this trade-off differs from ours that focusses on information and peer pressure.

Our work also contributes to a growing literature on the origins and effects of peer pressure. Kandel and Lazear (1992) incorporate peer pressure in their model through a simple function, where peer pressure depends on own effort, the effort of peers as well as other actions of the agents that do not affect firm output directly. Their finding is that peer pressure induces individuals to exert higher effort, which leads to a higher profit for the firm. They argue that firms can create peer pressure by establishing norms and mutual monitoring. In the case of mutual monitoring, the crucial issue is to define the relevant group, that is the team, the department or the entire firm. We put forward an alternative source of peer pressure (i.e. the social network), define the relevant group (i.e. friends and common friends) and provide a novel mechanism of how peer pressure operates.

The paper proceeds as follows: In section 2, we document empirically that men's and women's networks differ. In Section 3, we develop the general model, which we then solve for the static case in Section 4 and for the dynamic case in Section 5. Section 6 uses the model to make predictions on gender differences in labor market outcomes and connects these results to a variety of empirical facts. Section 7 discusses our equilibrium selection. Section 8 concludes.

2.2 Gender Differences in Networks

A main assumption of our model is that women have a higher clustering coefficient than men, but that men have a higher degree than women. This is based on our findings of different network properties of males and females in the AddHealth data set.

The AddHealth data set contains data on students in grades 7-12 from a nationally representative sample of roughly 140 U.S. schools in 1994-95. Every student attending the sampled schools on the interview day is asked to compile a questionnaire (in-school data) on respondents' demographic and behavioral characteristics, education, family background and friendships. The AddHealth website describes surveys and data in detail. This sample contains information on 90,118 students.[14]

Why AddHealth? Our main reason for using a dataset of *students* instead of employees in a firm is to circumvent the problem that networks can be shaped by the work environment. For instance, if men and women prefer to have friends of the same gender, some male-dominated work environments would cause women to have smaller networks. Moreover, if we were using data of firms or occupations we would be concerned that individuals with certain network types sort into those occupations and firms for which their network type is most beneficial. However, at the school level there is no such selection bias or constrained availability of same-sex individuals. We can therefore estimate male and female network structures more accurately with this data set.

Further, it is well documented that individuals are more likely to name others as their friends if these have a higher social status, see Marsden (2005). At the workplace social status is connected to a higher position in the hierarchy and therefore to formal power. However, here we think of a link between individuals as a friendship instead of trying to have a connection with someone superior. Additionally, as higher status is connected to formal power, it is difficult to distinguish between formal and informal networks. We believe that this is less of a problem at school as by definition the networks formed there are informal. There might be some misreporting in the sense that popular children will also be named as friends by individuals who would just like to be associated with them. But we believe that there is less of

[14] For more information on the AddHealth data set, see http://www.cpc.unc.edu/projects/addhealth.

an incentive for high school students to be strategic about their friendship nominations than for employees. A possible reason is that superiors might be able to access this nomination data and therefore employees have an incentive to name them. In contrast, from the design of AddHealth it is clear that students will not have access to the nomination data.

We are interested in these network characteristics of men and women as *exogenous types*, comparable to different ability or skill types commonly used in the literature, where this network type is stable over time. Burt (2011) provides evidence for the existence of different network types from a multi-role game in a virtual world.[15] He finds that people build a similar type of network, e.g. either a closed or a sparse network independently of what is required for the role.[16] Based on this, we argue that boys' and girls' networks at school, closely resemble the ones they will form as adults both in their private and work life.[17]

Friendship Network The friendship network constructed from the AddHealth dataset is a directed network, based on friendship nominations.[18] For this network, we compute both the directed and undirected clustering coefficients as well as the in-, out- and overall degree.[19]

The clustering coefficient is computed as the ratio of the number of links between a node's neighbors to the total possible number of links between the node's neighbors, both for the directed and undirected network. The in-degree denotes how often an individual was named, the out-degree gives how many friends this individuals named and the degree is then the sum of in- and out-degree.

We consider two subsamples, namely students that are at least 17 and students that are older than 17.[20] We do a t-test of the standardized variables and consider the differences between boys and girls.[21] The results are given in Table 2.1. We find that boys always have a lower clustering coefficient, independently of whether we consider the directed or undirected one. If we consider the sample that contains 17 year olds, we find that boys have a lower in-degree, but a higher out and overall degree. This changes when we consider the sample of students above the age of 17. Then for any measure of the degree, boys have a higher one than girls.

[15] This is a video game where players can play different roles and the different roles require different network structures. For some roles it is better to have sparse networks, for others dense networks.

[16] About a third of network variance is consistent with individuals across roles, but the correlation coefficient between the network formed and the network type is above 0.5.

[17] Unfortunately, there does not exist much further evidence of how persistent network types are or in general of how persistent differences between girls and boys are, i.e. whether this improves over time or not. A notable exception is Sutter and Rützler (2010) who shows that gender differences in competitive behavior emerge as early as age three and are quite persistent over time. The girls who exhibited a more competitive behavior earlier on, were more likely to be less competitive later on, those who were less competitive remained so. Therefore, the gender differences became more pronounced later in life.

[18] For more details on the friendship networks, see the Appendix.

[19] For the undirected clustering coefficient we assume that a link exists if at least one of the individuals named the other one as a friend.

[20] Our results for the entire sample are given in the Appendix.

[21] As we have standardized the clustering coefficients as well as degrees, the coefficients can be interpreted in terms of standard deviations.

Table 2.1: Difference in Network Characteristics Men-Women

	Age > 16	Age > 17
Cl. Coeff. (dir.)	-0.0799***	-0.0677***
	(0.0112)	(0.0144)
Cl. Coeff.	-0.0721***	-0.0618***
	(0.0112)	(0.0145)
In Degree	0.0261**	0.0222*
	(0.00859)	(0.00965)
Out Degree	-0.0635***	0.0208**
	(0.00847)	(0.00669)
Degree	-0.0236**	0.0259***
	(0.00827)	(0.00749)
Observations	42072	28259

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Standard errors in parentheses.

Male and Female Networks Beyond AddHealth To the best of our knowledge, differences in the clustering coefficient between men and women have not been documented in the literature. However, Fischer and Oliker (1983) look at the number of friends individuals have. They show that women have a lower number of friends than men, in particular at the workplace. We use part of the table from Fischer and Oliker (1983), p. 127, to document this. Their sample consists of employed men and women. They find that the number of friendships

	Under 36, unmarried, no children	Under 36, married, no children	Under 36, married, children	36-64, married, no children	36-64, married, children
Men	2.8	3.4	3.1	2.3	2.0
Women	2.5	2.1	0.9	1.4	1.4
Men (N)	113	50	70	63	54
Women(N)	76	51	98	67	60

Table 2.2: Friendships with Coworkers, see Fischer and Oliker (1983), p. 127

with co-workers differs greatly between them. For individuals under 36, who are unmarried and do not have children, the gender difference in the number of friends at the work place is small: men have on average 2.8 friends, women 2.5. But this difference increases, when men and women under 36 and married, with or without children, are compared. Without children, men have on average one more friend than women, with children they even have two more friends at the workplace. Therefore, our finding that older girls have a lower degree than older boys does not suddenly reverse, but is also documented for man and women at the workplace across all age groups.

Other studies also find network differences between the sexes. That girls and boys have different types of networks has been shown by Eder and Hallinan (1978) and is also documented in a survey by Belle (1989). The emphasis in this literature is on dyadic and triadic relationships, whereas we focus on the entire network. Gender differences in networks for adults have been shown by Kürtösi (2008), Tattersall and Keogh (2006) and Marsden (1987). These studies emphasize the number of friends and the content of the relationship, but do not contain precise information on gender differences in the network structure. Nevertheless, this literature shows that women form closed groups and emotional ties, whereas men's networks are sparser and characterized by instrumental ties, which is in line with our empirical findings.

Taking our estimation results together with the evidence in the literature, we feel confident to assume that women have a higher clustering coefficient but lower degree than men. This points to a new dimension of heterogeneity between men and women, which might help explain the gender wage gap or differences in occupational sorting. We do not have a causal argument since there might be an underlying factor that causes these network differences but also impacts labor market outcomes directly. Identifying the source of network differences is beyond the scope of this paper. Nor do we want to argue that differences in social networks is the whole story behind wage and performance gaps as well as occupational sorting. However, we do believe that networks play an important role and our model clarifies how these network differences can matter for job performance and wages.

In our setting, a higher clustering coefficient leads to higher peer pressure and a higher degree leads to more information. Both of these features are valuable and we characterize environments under which peer pressure is more beneficial and contrast them to settings where access to information is more important. We then obtain theoretical predictions for when we would expect men to perform better than women, which we connect in Section 6 to observed disparities in labor market outcomes. Our first step is to develop a model that translates clustering into peer pressure and the degree into access to information, highlighting our main theoretical mechanism.

2.3 Model

We consider an undirected network g of N workers. Two of those workers, $i, j \in N$, are selected in each period t. We focus here on a two period model, $t \in \{1, 2\}$, to keep our set up as simple as possible but note that it is straightforward to extend our setting to more periods. Once two workers are selected they have to complete a project. Whether they are successful or not depends on their network structure and past project outcomes as well as exerted effort. In order to highlight how each of these factors matter we first consider the game that is played in each period t.

1. Worker Selection At the beginning of each period, two workers are drawn at random from the set of workers to complete a project. These workers can be linked directly, where a link between i and j, denoted by $g_{ij} = g_{ji} = 1$, implies a good relationship between coworkers.

We assume that two workers can only complete their project successfully if there exists a direct link between them. If there is no link between two selected workers, their project fails with certainty, leading to a payoff of zero.[22] The number of links of worker i, his degree, is denoted by d_i. Then, the probability of being selected for a project and being partnered with a directly connected worker is given by (see Appendix)

$$s_i = \frac{2d_i}{N(N-1)}.$$

This probability is proportional to the degree of an individual, that is workers with higher degrees will be selected more often into potentially profitable projects.[23]

2. Information Every period is marked by a state of the world, θ, which can be high or low

$$\theta = \begin{cases} \theta_h & \text{with probability} \quad q \\ \theta_l & \text{with probability} \quad 1-q. \end{cases}$$

It is drawn after project teams are formed and is not observable to the workers. In the high (low) state, the project value is $2v_h$ ($2v_l$), with $v_h > v_l$. We assume that the payoff of the project is split equally among the project partners.[24]

In the following, we show how a worker's network structure affects his information about the state of the world. Each worker obtains a signal $x_i \in \{0, 1\}$ about the state, where $x_i = 1$ indicates the high state. Signals are informative in the sense that $Pr(x_i = 1|\theta_h) = p > \frac{1}{2}$ and $Pr(x_i = 1|\theta_l) = 1 - p$.[25]

Each worker receives one signal directly, but can also observe the signals of all workers he is directly or indirectly connected to. Note that the entire network might or might not be connected (where *connected* means that there are no isolated nodes). We denote the overall number of signals a worker receives by n_i. We allow for $n_i > N$ and interpret the additional signals (i.e. the signals beyond the number of workers in the network N) as basic information everyone possesses, which enables us to vary the baseline amount of information below.

Based on the observed signals, a worker can compute a sufficient statistic y_i, which is the number of high signals out of all observed signals, that is $y_i \in \{0, 1, \ldots, n_i\}$. Note that for two (directly or indirectly) connected workers, i and j, $y_i = y_j$. Let $n_{max} = max_i n_i$. Then $y_i, i = 1, ..., N$ takes values on $Y = \{0, 1, \ldots, n_{max}\}$.

As our focus is on the effects of ego-networks, we distinguish between the number of signals a worker obtains from himself and his direct friends, $n_{int,i} = d_i + 1$ and the signals he obtains indirectly from external sources, $n_{ext,i}$, with $n_i = n_{int,i} + n_{ext,i}$.[26]

[22] A link or rather a good relationship between workers makes them better team partners. To simplify, we set the payoff of projects between unlinked workers to zero.

[23] This is in line with Aral et al. (2012), who study project performance in a recruiting firm. They find that peripheral nodes, i.e. nodes that are not well connected, do fewer projects per unit of time than central nodes.

[24] We impose the equal split assumption as we aim for a model in which agents are perfectly symmetric except for their network. This allows to show the effects of network structures in the cleanest way possible.

[25] Put differently, if $\theta = \theta_h$, each signal $x_i \sim$ Bernoulli(p) and if $\theta = \theta_l$, this signal is $x_i \sim$ Bernoulli($1 - p$).

[26] Ego networks consist of a focal node (*ego*) and the nodes to whom ego is directly connected to (*friends*)

Based on y_i, the posterior probability of being in the high state, $Pr(\theta_h|y_i)$, is computed via Bayesian updating and thus having a higher number of signals gives a more precise posterior. The project value, $\pi(y_i)$, is then given by

$$\pi(y_i) = Pr(\theta_h|y_i)v_h + (1 - Pr(\theta_h|y_i))v_l.$$

To summarize, the network structure matters as a higher degree gives a higher number of internal signals, which in turn affects the expectation about the project value.

3. Choice of Effort The paired workers simultaneously choose what effort, $e_i \geq 0$, $\forall i$ to exert on the project. This effort is costly with all workers facing the same cost function $c(e)$. We impose the following assumptions:

Assumption 2.1. *Cost of Effort:* $c(e) = ke^2$, *where* $k > 0$.

Given that the project fails certainly if the two project partners are not connected, we focus on the effort choice of two directly linked project partners. Effort makes project success more likely. The probability that the project is completed is given by the success probability function $f(e_i, e_j) \in [0, 1)$. In order to ensure that $f(e_i, e_j)$ is strictly smaller than one, we assume that effort is bounded, $e_i \in E = [0, e_{max}]$ where $f(e_{max}, e_{max}) < 1$.[27] This implies that success cannot be guaranteed. We impose additional assumptions on the success function, summarized in Assumption 2.2.

Assumption 2.2. *Success Probability Function* $f(e_i, e_j)$*:*

(a) *Symmetry:* e_i *and* e_j *enter* $f(e_i, e_j)$ *symmetrically.*
(b) $f_1(e_i, e_j) = f_2(e_j, e_i) > 0$
(c) $f_{11}(e_i, e_j) = f_{22}(e_j, e_i) < 0$.
(d) *Strict Supermodularity:* $f_{12}(e_i, e_j) = f_{21}(e_i, e_j) > 0$.
(e) $f(e_i, 0) = f(0, e_j) = 0$.
(f) $f(\lambda e_i, \lambda e_j) = \lambda f(e_i, e_j)$, $\lambda e_i, \lambda e_j \leq e_{max}$.

The effort levels of the workers are complements. We focus on complements as with substitutes a worker should complete the project by himself. There is no reason to form a team. Additionally, if one team member chooses zero effort, the project fails for sure. We assume further that the success probability function exhibits constant returns to scale. We know that $e_i \in [0, e_{max}]$. If $\lambda \in [0, 1]$, then $\lambda e_i \leq e_{max}$, and for $\lambda > 1$ we impose the additional restriction that $\lambda e_i \leq e_{max}$, $\forall i$. After effort has been chosen, the project outcome – success or failure – is realized.

These three stages occur in both periods. What differs across periods is information (i.e. the signals workers obtain) and the effect of peer pressure (which impacts effort only if today's project outcome matters for tomorrow's). Effort depends on information through the sufficient

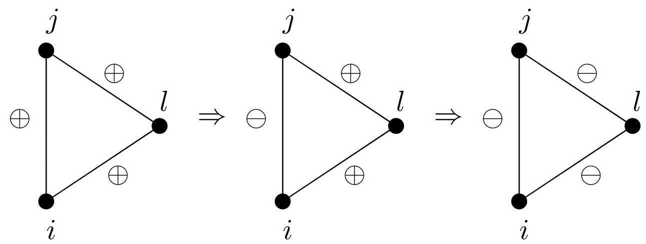
plus the ties, if any, among the friends.

[27]By choosing an appropriate bound on v_h, we can guarantee an interior solution $e \leq e_{max}$.

statistic y. It depends on peer pressure because publicly observable past project outcomes affect current relationships between coworkers, especially when the network is characterized by high clustering. We now outline the peer pressure channel and how past project outcomes matter rather informally, a formal description is in the Appendix.

We assume that a failure has an impact if the same team partners are chosen in two consecutive periods. We believe it is intuitively plausible that a project failure leads to discord among team partners and their relationship turns 'bad'. The failure has to be justified, which is disagreeable and affects their relationship. We further argue that this discord between team partners also spreads to common friends. This idea is based on the well-established *structural balance theory*. According to this theory, triads of friends are only stable as long as the relationships are balanced. Suppose that i, j and l are all directly connected. Initially, all their relationships are 'good'. Then, i and j work on a project together that fails, turning their relationship into a bad one. But a triad in which one relationship is bad and the other two are good is unstable. This instability has to be resolved, meaning the workers have to take sides. To simplify our analysis, we assume that *all* relationships in a triad will be bad after a project failure.[28] This is why project failures affect workers with high clustering more than those with low clustering. They are deprived of more future project opportunities. This sequence of events is depicted in Figure 2.1, where a plus (minus) signifies a good (bad) relationship.[29]

Figure 2.1: Structural Balance Theory



Each project failure induces some bad relationships, whereas a project success means that all directly connected workers have good relationships. We denote the quality of the relationship by $\gamma \in \{\gamma_b, \gamma_g\}$, that is the relationship can be good or bad. The relationship between i and j is bad after a project failure in the previous period if either (1) i and j were teamed in the previous period or (2) i or j were teamed with a common friend in the previous period. Otherwise, i and j have a good relationship. We assume that in period one the relationship between any two workers is good.

This relationship quality between two directly connected workers constitutes a state, $\gamma \in \Gamma$, and we can define a pure public strategy $\sigma(\gamma, y) : \Gamma \times Y \to E$, which maps from the relationship state and the signals into the action space.

[28]According to the (weak) structural balance theory a triad with three negative links is balanced (Davis (1967)). Our assumption is a simplification of the following idea: Given a project failed, a worker faces with a positive probability more than one negative connection if he and the project partner had common friends, but only has one negative connection if the project failed with someone he does not have a common friend with.

[29]Note that discord does not imply that links are cut. If there is no link between two workers, then they never get along. Once a link exists, we interpret this as two individuals getting along in principle. Thus, a bad link is transitory. Also, information is still transferred if the link is bad but is not if the link was cut.

Due to our restriction to public strategies, the equilibrium concept applied is that of a public perfect equilibrium. We index the variables in the second period by *prime*.

Definition 2.1. *A public perfect equilibrium (PPE) is a profile of public strategies σ that for any state $\gamma, \gamma' \in \Gamma$ and for any signal realization $y, y' \in Y$ specifies a Nash equilibrium for the repeated game, i.e. in the first period, $\sigma(\gamma_g, y)$ is a Nash equilibrium and in the second period $\sigma'(\gamma', y')$ is a Nash equilibrium.*

In our setting a higher degree leads to more signals, which allow for a more precise belief about the project value. Higher clustering, on the other hand, makes a bad relationship after a project failure more likely and therefore incentivizes effort through peer pressure. This is the basic trade-off we are focussing on. We will show in more detail how peer pressure influences effort choices in the dynamic setting but, before doing so, we want to discuss the static case, where only information matters. After presenting the model and our results, we will justify our equilibrium selection, comparing the workers' payoffs from choosing this strategy to the payoffs of other strategies.

2.4 Static Decision Problem

In the static setting, worker i chooses effort to maximize his expected payoff, given by

$$\max_{e_i \in E} f(e_i, e_j)\pi(y) - c(e_i). \tag{2.1}$$

Recall that $y_i = y_j = y$ since each worker observes not only his own signal but also the signals of all workers he is (in)directly connected to, so we write $\pi(y)$. Given our assumptions, the first order condition of (2.1) is both necessary and sufficient for a maximum. The same holds true for worker j. Based on the first order approach, we can determine the pure strategy public perfect equilibria of the game where, to simplify notation, we define $e(y)$ to denote the optimal strategy based on y.

Proposition 2.1 (Public Perfect Equilibria Static Game)**.** .
1. Every public perfect equilibrium is symmetric s.t. in equilibrium $e_i(y) = e_j(y) = e(y)$ $\forall y$.
2. For each y, there exist exactly two pure public perfect equilibria.

$$\begin{aligned} &\textit{(i) Zero effort:} && e(y) = 0 \\ &\textit{(ii) Strictly positive effort:} && e(y) = \frac{f_1(1,1)\pi(y)}{2k} \end{aligned} \tag{2.2}$$

All proofs are in the Appendix. Given the symmetry in our setting, in particular, the symmetry of $f(\cdot, \cdot)$, identical cost functions and equal split of the payoff, both workers will always exert the same effort in any equilibrium. We further show that the one-period problem has two pure strategy PPE. There always exists a PPE where both project partners exert zero effort independently of signal realizations. It is a best response to choose zero effort given the other worker has chosen zero effort as $f(e_i, 0) = f(0, e_j) = 0$. Each team member has to exert at least some effort for the project to be successful. But there also exists a PPE with strictly positive efforts. The uniqueness of the positive effort equilibrium follows from supermodularity and constant returns to scale property of $f(\cdot,\cdot)$, as well as the convexity of the cost function. In particular, we obtain a closed form expression for effort when taking into account symmetry across team members, where k is the multiplicative constant in the cost function and where $f_1(1,1)$ is a constant as well.

We are now interested in how network characteristics influence equilibrium effort through the information channel in the static model. All else equal, a worker with a higher degree receives more signals about the state of the world. We want to know how effort varies with the number of signals. It follows from (2.2) that effort positively depends on the project value $\pi(y)$.

We therefore focus on how the expected project value, $E(\pi(y))$, varies with the number of signals as this is the channel through which information affects effort. If additional signals increase the expected project value, then expected effort, $E(e(y))$, increases as well. A worker has an incentive to work harder if he believes the payoff for his work to be higher.

We first show that $\pi(y)$ has the martingale property, meaning that it is unaffected by the number of signals, which follows from Bayes' Rule. This is not true once we condition on the state. To emphasize that a worker receives n signals, we denote the project value by $\pi(y_n)$ instead of $\pi(y)$.

Lemma 2.1 (Information and Expected Project Value)**.** *$\pi(y_n)$ satisfies the martingale property: $\pi(y_n) = E(\pi(y_{n+1})|y_n)$. However, given that the state is realized, a worker with more signals holds a more accurate posterior belief about the state of the world and thus about the project value:*

$$v_h > E\left(\pi(y_{n+1})|\theta_h\right) > E\left(\pi(y_n)|\theta_h\right) \qquad v_l < E\left(\pi(y_{n+1})|\theta_l\right) < E\left(\pi(y_n)|\theta_l\right).$$

The impact of an additional signal vanishes, if uncertainty vanishes, i.e. $E\left(\pi(y_n)|\theta\right) = E\left(\pi(y_{n+1})|\theta\right)$, if either (i) $v_l \to v_h$ (ii) $p \to 1$, (iii) $q \to 1$ if $\theta = \theta_h$, $q \to 0$ if $\theta = \theta_l$, or (iv) $n_{ext} \to \infty$.

An additional signal does not contain further information about the state of the world, given the state of the world has *not* been realized, that is $E\left(\pi(y_n)\right) = E\left(\pi(y_{n+1})\right)$. But once the state of the world has been realized, this is no longer true. Since signals are informative, the more signals are available the more accurate is the posterior belief about the state of the world. The expected project value increases in the number of signals if the state of the world is high and decreases in the number of signals if the state of the world is low. This implies that, given the high state of the world, a worker with more information expects a higher project value compared to a worker with less information. If the low state has been realized, the reverse is true.

The expected project value becomes *independent* of the number of overall signals n when the uncertainty of the underlying environment vanishes. This can happen for four reasons: (i) There is no difference between high and low project values. (ii) The signals are completely informative.[30] (iii) A worker's prior reflects complete certainty about the state of the world.

[30] In fact, the expected project value also becomes independent of the number of overall signals n when

(iv) Moreover, if overall information becomes abundant, which happens when the number of external signals, n_{ext}, becomes large, then in the limit, all agents know the state of the world with certainty even if the number of signals obtained through their ego-networks, n_{int}, differs. The expected payoff converges to the high (low) value when the state is high (low). In sum, the effect of additional information on the expected project value is reinforced when the uncertainty of the underlying environment is considerable and dies out when uncertainty vanishes.

Taking Lemma 2.1 together with equation (2.2) we can shed light on the effect of information on expected effort, summarized by the following proposition.

Proposition 2.2 (Information and Expected Effort)**.** *A worker with more information, i.e. with a higher degree, exerts on average more (less) effort when the state of the world is high (low) compared to a worker with less information. The impact of additional signals on effort vanishes as the underlying uncertainty vanishes.*

A worker with a higher degree and thus more signals holds more accurate information about the state of the world. Therefore, if the state is high, more information leads to a higher expected project value (Lemma 2.1) which in turn leads to higher effort. The opposite is true for the low state. Intuitively, workers with more accurate information, i.e. more signals, can better fine-tune their effort to the expected project reward.

2.5 Dynamic Decision Problem

Having discussed the static game, we can now analyse the agents' effort choices and how they are impacted by their network characteristics in a dynamic setting. Here, not only agents' *degree* but also their *clustering* matters for their actions as they adjust their effort to their relationship quality, namely $\forall\ y'$

$$\sigma'(\gamma'_g, y') > 0 \qquad \text{and} \qquad \sigma'(\gamma'_b, y') = 0.$$

This implies that, when two workers have a bad relationship, they exert zero effort. We know from the static game that zero effort in every period constitutes a PPE, regardless of the signals. As $f(0,0) = 0$, such a project will fail for sure and yields zero payoffs. In turn, when two team partners have a good relationship they exert strictly positive effort. In what follows, we focus on the dynamic decision problem that pins down the high effort PPE in both periods. We are interested in what determines this choice.

The dynamic maximization problem of team partner i reads

$$\begin{aligned} \max_{e_i, e'_i} \quad & -c(e_i) + f(e_i, e_j)\left(\pi(y) + \beta s_i E\left(f(e'_i, e'_k)\pi(y') - c(e'_i)\right)\right) \\ & + (1 - f(e_i, e_j))\left(0 + \beta s_i (1 - r_{ij}) E\left(f(e'_i, e'_k)\pi(y') - c(e'_i)\right)\right) \end{aligned} \tag{2.3}$$

where the expectation is taken over all possible signal realizations in period two. Problem (2.3) is dynamic since workers choose today's effort not only based on the current project

signals are completely uninformative $p \to 0.5$.

payoff but also based on the second period expected payoff, taking into account that today's project outcome matters for tomorrow's through its impact on relationships. Therefore, this expected payoff of workers i and j, who are selected in period one to complete a project, depends not only on second period performance but also on
(i) the probability of being selected next period, as defined in the worker selection, s_i and s_j,
(ii) the probability of *first period* project success, $f(e_i, e_j)$, or failure, $1 - f(e_i, e_j)$, as well as
(iii) the probability that in the next period they are doing a project with someone who would be affected by the project failure, *given* that they are chosen for a project, r_{ij} and r_{ji}, where r_{ij} is given by

$$r_{ij} \equiv \frac{1 + \sum_{k,k\neq i,k\neq j} g_{ik}g_{jk}}{d_i} = \frac{C_{ij}}{d_i}. \tag{2.4}$$

The term $\sum_{k,k\neq i,k\neq j} g_{ik}g_{jk}$ denotes the number of common friends of i and j and therefore C_{ij} is a proxy for their common friends. Equation (2.4) gives the probability of workers having a bad relationship after a failure. They are only affected by their failure if they are chosen to do a project together or with a common friend in the second period.

We solve problem (2.3) by backward induction, starting in the second period. Clearly, the second period problem is identical to the static problem.[31] Recall that the high effort level is given by

$$\sigma_i'^*(\gamma_g', y') \equiv \arg\max_{e_i'} V_i(\gamma_g', y') = \arg\max_{e_i'} [f(e_i', e_j')\pi(y') - c(e_i')], \tag{2.5}$$

where $\sigma_i'^*(\gamma_g', y')$ is the optimal second period effort level if the project partners have a good history and observe signals y' (see equation (2.2) for the solution to this problem). We will denote this equilibrium effort by $e_i'(y') \equiv \sigma_i'^*(\gamma_g', y')$. Moreover, we denote the maximized second period payoff by $V_i^*(\gamma_g', y')$. Then, the maximization problem of agent i in the first period reads

$$\max_{e_i} \quad f(e_i, e_j)\pi(y) - c(e_i) + \beta s_i(f(e_i, e_j) + (1 - r_{ij})(1 - f(e_i, e_j)))EV_i^*(\gamma_g', y') \tag{2.6}$$

Similar to the static problem, we show that there exists a unique PPE in which both team partners exert positive effort. The solution to (2.6) is given by $e_i(y) \equiv \sigma_i^*(\gamma_g, y)$.

Proposition 2.3 (Public Perfect Equilibria Dynamic Game)**.** .

1. Both project partners always exert the same effort in any PPE, that is effort is symmetric.
2. In both periods, there exists a unique PPE in which both team partners exert strictly positive effort, $\forall y, y'$

$$\begin{aligned} e_i(y) = e_j(y) =& \frac{f_1(1,1)(\pi(y) + \beta s r EV^*(\gamma_g', y'))}{2k} \\ e_i'(y') = e_j'(y') =& \frac{f_1(1,1)\pi(y)}{2k} \end{aligned} \tag{2.7}$$

[31] As j and l belong to the same set, namely the friends of i, we can replace l in the second period by j.

We already know from Proposition 2.1 that in the second period there exists exactly one PPE with strictly positive effort, which is symmetric. But also in the first period, effort levels are symmetric. This is because two workers can only have the same number of common friends, implying that $s_i r_{ij} = \frac{C_{ij}}{\frac{1}{2}(N-1)N}$ is constant across project partners and, thus, $\beta s_i r_{ij} EV_i^*(\gamma_g', y') = \beta sr EV^*(\gamma_g', y')$.[32] Moreover, for $y = y'$ first period effort is higher than second period effort, stemming from the dynamic effort-enhancing effect of clustering: Having common friends creates particularly strong incentives for effort, reducing the team moral hazard problem that causes effort to be inefficiently low.

Again, we are interested in how the agents' network characteristics affect effort. We first discuss the effect of *degree*, which impacts effort through the information channel because someone with a higher degree receives more signals about the state of the world. We then turn to the effect of *clustering*, which influences effort through peer pressure.

Taking expectations over first period signals in (2.7), it follows that information impacts expected effort through the expected first period project value, $E(\pi(y))$, and second period value, $EV^*(\gamma_g', y')$. Both have a positive effect on effort. Recall from our discussion on the static game that $E(\pi(y))$ positively (negatively) depends on the number of signals if the state is high (low). To see how the expected value, $EV^*(\gamma_g', y')$, depends on the number of signals and thus information, we first establish that $V^*(\gamma_g', y')$ is a convex function of the second period project value, $\pi(y')$ (which, in turn, is a martingale). It immediately follows how the expected second period value depends on the number of signals and thus on information. To simplify notation, we write $EV^*(y')$ instead of $EV^*(\gamma_g', y')$.

Lemma 2.2 (Information and Second Period Expected Value). *$V^*(y_n')$ is a submartingale. And, thus, a worker with more signals has a higher second period expected value:*

$$E(V^*(y_n')) < E(V^*(y_{n+1}')),$$

The impact of an additional signal vanishes, if uncertainty vanishes, i.e. $E(V^*(y_n')) = E(V^*(y_{n+1}'))$, if either (i) $v_l \to v_h$ (ii) $p \to 1$ (iii) $q \to 1$ if $\theta = \theta_h$, $q \to 0$ if $\theta = \theta_l$, or (iv) $n_{ext} \to \infty$.

To gain some intuition for this result first suppose that the additional signal is high. This implies that effort increases, that the project value increases and that the overall payoff, $V^*(y)$, increases as well. If the additional signal is low, then effort decreases, the project value decreases and the overall payoff is lower. But due to the convexity of the payoff, an additional positive signal has a stronger effect than an additional negative signal. Therefore, having an additional signal increases the expected value in the second period unless uncertainty vanishes. As in the static game (see Lemma 2.1), here additional information has no impact if (i) there is no variance in the project value across states, (ii) if signals are completely informative, (iii)

[32] Note that the symmetry in efforts only holds in the two period model, as in the last period having common friends does not matter anymore and both project partners have the same level of information. Thus, if this game is played for T periods, then in T and $T-1$, effort levels are symmetric. But the network structure matters in period $T-1$ and is therefore taken into account in $T-2$ resulting in asymmetric effort levels.

if the prior is correct, or (iv) if overall information becomes abundant. Intuitively, information only matters under uncertainty.

In turn, the effect of peer pressure, $s_i r_i$, on first period effort (through clustering) is straightforward and unambiguously positive.

We summarize the effect of information and peer pressure (and thus of the agents' network characteristics) on first period effort in the next proposition.

Proposition 2.4 (Information, Peer Pressure and Expected First Period Effort)**.** *More information, i.e. a higher degree, unambiguously increases first period effort only if the state is high. Furthermore, higher peer pressure, i.e. higher clustering, increases first period effort independently of the state of the world. Finally, unless uncertainty vanishes, a worker with more information but less peer pressure better adjusts his effort to the expected project value compared to a worker with less information and more peer pressure.*

The proof follows immediately from Lemmas 2.1 and 2.2 and equation (2.7) and is therefore omitted. Both network characteristics, high degree and high clustering, affect first period effort and thus project completion. A higher *degree* improves information about the state of the world. This information is particularly beneficial when the true state is high and, at the same time, when the agents' uncertainty about the state is considerable. In this case, additional signals induce the agents to put significantly more weight on the high state, translating into higher effort. (The same logic also applies to second period effort, given by the static case in Proposition 2.2.)

In turn, *clustering* or common friends positively impact first period effort through a dynamic peer pressure effect. This channel is *independent* of the true state of the world and the underlying uncertainty. Peer pressure induces higher effort because a potential project failure today puts more friendships and thus future project opportunities in jeopardy. Since workers with more information are more selective with their effort choice (depending on the state) and workers facing peer pressure increase their effort no matter the expected payoff, it follows that workers with a higher degree are better able to fine-tune their effort to the project reward than workers with higher clustering. This means that their difference of efforts across states, $E(e(y)|\theta_h) - E(e(y)|\theta_l)$, is larger, which follows directly from (2.7).

We now turn to the agents' wages, which are tightly linked to their effort choices. Denote the probability of having a good relationship with the second period project partner given first period state by

$$Pr(\gamma_g'|\theta) \equiv E[f\left(e(y), e(y)\right) + (1-r_i)(1-f\left(e(y), e(y)\right))|\theta] = E(e(y)|\theta) r_i f(1,1) + 1 - r_i.$$

We define first and second period wages as follows:

Definition 2.2 (Equilibrium Wages)**.** *First and second period wages for a given state are respectively defined as*

$$w_i(\theta) \equiv E[f\,(e(y), e(y))\,v|\theta] = E(e(y)|\theta)f(1,1)v(\theta) \tag{2.8}$$

$$w_i'(\theta,\theta') \equiv s_i Pr(\gamma_g'|\theta)E[f\,(e'(y'), e'(y'))\,v'|\theta'] = s_i Pr(\gamma_g'|\theta)E(e'(y')|\theta)f(1,1)v(\theta') \tag{2.9}$$

where $\theta, \theta' \in \{\theta_l, \theta_h\}$ is the realized first (second) period state.

These are *expected* wages because even though positive effort is exerted there is no guarantee for project success. The wages reflect that the agents obtain their share of output in case the project is successful. We define these wages *given* that a certain state of the world has materialized. The expected wage across states can then be easily computed, e.g. $E(w_i) = qw_i(\theta_h) + (1-q)w_i(\theta_l)$.

Notice that the structure of both periods' wages is the same, only that in the second period, one also has to take into account the probability of being selected for a project with someone the agent is on good terms with (i.e. the probability of having a good friendship history with the project partner, given by $Pr(\gamma_g'|\theta)$). Since friendship histories matter, the second period expected payoff depends not only on contemporaneous but also on first period effort.

Both periods' wages are increasing in effort, highlighting the tight link between the agents' actions and their rewards. As a consequence, Propositions 2.2 and 2.4 on the effects of network characteristics on effort give insights into how degree and clustering affect the agents' wages. We summarize these results in the next proposition.

Proposition 2.5 (Information, Peer Pressure and Wages)**.** *More information, i.e. a higher degree, unambiguously increases first and second period wages only if the state is high in both periods. The effect vanishes as uncertainty vanishes. In turn, peer pressure, i.e. higher clustering, increases the first period wage independently of the state but has an ambiguous effect on second period wage.*

These results follow from Propositions 2.2 and 2.4 and wage Definition 2.2. Information (and thus a high degree) leads to a significant effort and wage boost if the underlying state of the world is high because agents want to reap the benefits of a high project value.[33] Through the effort channel, information only increases wages if there is uncertainty about the state of the world. In turn, when the agent faces a dynamic decision problem (i.e. in the first period), higher clustering unambiguously increases effort and wages through peer pressure, independent of the state and the underlying uncertainty. Only in the second period, the effect on wages is ambiguous: Peer pressure leads to higher first period effort (increasing $Pr(\gamma_g'|\theta)$), but many common friends also make a non-intact relationship with the second period team partner more likely (decreasing $Pr(\gamma_g'|\theta)$).

While this discussion has focussed on comparative statics effects of a single network characteristic holding other network characteristics fixed, we now turn to the more interesting but also more involved case of comparing two types of workers: one with higher degree

[33] If the state is low, it is ambiguous whether clustering or degree leads to a higher wage.

but lower clustering (denoted as D-worker) and one with lower degree but more clustering (denoted as C-worker).

Proposition 2.6 (Trade-Off Between Information and Peer Pressure)**.** *Suppose that $v_l = 0$. (i) Wage Dynamics: If a C-worker has a lower first period wage than a D-worker, then he also expects a lower wage in the second period, even if second period uncertainty vanishes. This wage gap arises even if both workers perform equally well in the first period. (ii) Comparative Advantage: If $E(\pi(y)|\theta_h)$ and $EV^*(y')$ are sufficiently concave in information n, signal precision p, and the prior belief q, C-Workers hold a comparative advantage in environments with less uncertainty, that is, $\frac{E(w^C)}{E(w^D)}$ increases as uncertainty becomes smaller.*

We make these statements precise in the Appendix where all the formal conditions are provided. Our model predicts a strong impact of early career wages on the future wage trajectory through peer pressure, which puts workers with high clustering but low information into disadvantage. Notice that wage gaps between C-workers and D-workers arise even if they exert the same effort in the first period. Moreover, if these wage gaps exist in the first period, they persist in the second period even if they perform equally well (i.e. even if uncertainty vanishes in the second period).

Our model also predicts that workers with higher clustering and less information have a *comparative advantage* in environments characterized by less uncertainty compared to workers with less clustering and more information. That is, the ratio of expected wages of C-workers to D-workers increases when uncertainty diminishes, which happens when the amount of overall information, n, increases, when signals become more informative (for p sufficiently large), and when the prior belief q becomes more correct.[34] This is consistent with our previous predictions that clustering gains importance as uncertainty vanishes.

Our framework allows us to rank networks according to effort choices and wages for different underlying environments. In the next section, we connect this theory with our empirical finding on gender networks that men have a higher degree and women more clustering.

2.6 Performance of Men versus Women

In this section, we use our model to analyse how peer pressure and information influence effort and wages of men versus women. We show that these results are consistent with various observed gender differences in labor market outcomes. Previously, we showed that women have a higher clustering coefficient and a lower degree than men, that is, they face more peer pressure but are less informed. We therefore want to compare agents with these features. To do so, we fix the number of links and nodes in the network, so additional clustering comes at the cost of a lower degree and vice versa.

Recall that in our model effort and wages are determined by first and second period expected payoffs as well as clustering. Women have a lower degree than men and therefore

[34] We also allow for $v_l > 0$ in simulations. When project values, v_l and v_h, become more similar across states, then the ratio of expected wages of C-workers to D-workers increases.

fewer signals. As a consequence, men have an informational advantage over women, implying that their expected first period payoff, $E(\pi(y)|\theta)$, more accurately reflects the state and their second period expected payoff, $EV^*(y')$, is always higher. In turn, women have more clustering, sr, which translates into higher peer pressure. Thus, the trade-off between degree and clustering translates into a trade-off between peer pressure and access to information and it is not a-priori clear which network characteristic is more conducive to project success and wages.

In what follows, we use our model to predict in which environments men outperform women and show that these predictions are in line with a variety of empirical facts.

1. Wage and performance gaps between men and women are especially large within occupations and tasks characterized by uncertainty like in the financial sector, film-industry and in basic research.

In most developed countries, the gender wage gap is still large. In the US in 2012, for instance, women's earnings were 80.9% of men's earnings.[35] Part of it can be explained by differences in occupational choices where women select into low-paying occupations while men go into high-paying jobs. However, even *within* occupations wage gaps are considerable. Notably, some occupations are more affected than others. In the US, the within-occupational wage gap is pronounced in management occupations, especially for financial managers and chief executives where female earnings are respectively 70.3% and 76% of male's, as well as in business and financial operations occupations where women earn 74% of men's earnings. In contrast, the wage gap is much smaller in healthcare support and administration where women's earnings are respectively 90.2% and 89.9% of men's.[36] A similar pattern was found in the UK, where full-time working women in the financial sector earn 55% less than full-time male workers – a gap twice as large as the gap in the economy as a whole.[37] The evidence suggests that women's lower earnings in financial management and executives occupations are especially due to large differences in performance pay and boni.[38]

Another well-studied sector where gender inequalities persist is the film industry (Lutter (2012) and Lutter (2013)). This industry is highly project-based where tasks involve little routine work and have uncertain outcomes. Ferriani et al. (2009) argue that the film market requires constant adjustment to new work environments since film ventures operate under constant uncertainty and have to foresee ex-ante whether the project opportunity is valuable. Women in this sector generate lower box revenues from movies, which is a direct measure of performance.

Last, an area well-known for gender disparities is the market for patents. Hunt et al. (2012) document that women in the US are much less likely to be granted a patent than men, with women holding only 5.5% of commercialized patents. This is not due to women's underrepresentation in science and engineering degrees but due to their underrepresentation in

[35] See http://www.bls.gov/cps/cpswom2012.pdf

[36] See Reports (2013), http://www.bls.gov/cps/cpswom2012.pdf

[37] Wage differences are considerable even when controlling for hours of work (full time) and type of job. See the report by the Equality and H.R.Commission (2009).

[38] See the report by the Equality and H.R.Commission (2009).

patent-intensive fields of study as well as patent-intensive job tasks like design and development. Again, patents can be seen as measures of performance.

This implies that the gender wage gap is particularly pronounced in occupations or tasks characterized by a large amount of uncertainty, commonly measured by earnings variability. In these occupations, earnings are uncertain. They are based on success which is difficult to foresee. Earnings of executives and financial managers are largely based on performance pay. Similarly, the success of research (and thus patents) as well as movies is difficult to foresee at the time of production.

Our model provides a new mechanism why men outperform women under uncertainty. The main prediction is that men's network structure is conducive to information acquisition which is more valuable in such environments than the undifferentiated effort-enhancing effect of women's peer pressure.

Our network-based view finds support in various empirical studies on financial and management occupations, the film industry and patenting. Forret and Dougherty (2004) analyse the impact of networking activities on career outcomes (promotions, total compensation and perceived career success) of male and female MBA graduates over 35 years in the U.S. Those graduates take on positions in management, finance, marketing and other professional jobs – occupations characterized by relatively large amounts of earnings variability. They find that only for men, network activities positively affect career outcomes. The authors speculate that the reason for this finding is that women network less effectively. We propose a theory *why* women's networks are less effective in these settings.

As far as the film industry is concerned, Ferriani et al. (2009) argue that information is crucial to identify potentially successful scripts and to assemble the right project team. Based on the finding that producers who are more central in their network (i.e. have more access to information) are more likely to increase the box revenue from a movie, the authors conclude that social networks provide crucial access to information. In a similar vein, Lutter (2013) documents that women with loose information-based networks perform better in the film-industry than women with dense networks, supporting our hypothesis that information is the key to success in uncertain environments.

With regards to research and development, Gabbay and Zuckerman (1998) document that in basic research, which is typically characterized by complex, uncertain tasks, scientists benefit from sparse networks with many holes, whereas in applied research, which is typically characterized by noncomplex, certain tasks, scientists benefit from dense networks. Supporting this view, Ding et al. (2006b) argue that an important reason for the gender wage gap in patenting is that women's networks are less effective: In relying more on close relationships, they lack reach to industry contacts.

Our theory offers a unified explanation for these findings. In uncertain environments information is crucial for success and men hold more of this type of social capital than women. We show next that this argument also provides a rationale for why occupational choices differ across gender.

2. More men than women choose occupations with high earnings volatility.

Dohmen and Falk (2011) show that men rather than women select into "risky" jobs that are characterized by performance pay and high earnings volatility. They explain this finding, arguing that men and women face *the same* mean-variance trade-off with regards to wages in all occupations but differ in their attitude towards risk (with men being less risk-averse). We offer an alternative explanation, which is based on comparative advantage. Our model predicts that women have a comparative advantage in environments characterized by lower uncertainty.

In such environments, women put relatively more effort than men, translating into relatively higher earnings, compared to more uncertain environments. In turn, in uncertain environments, men tend to face more income dispersion but are compensated for this risk by higher expected wages. We do not model occupational choice explicitly but this argument suggests that women would select into environments with low uncertainty whereas the opposite is true for men. Notably, this holds even though both men and women are risk-neutral and hence do not differ in their risk attitudes.

3. Having women in the network is particularly beneficial high up in the organizational hierarchy.

Lalanne and Seabright (2011) document empirically that having females in the network is beneficial to both male and female executives but not to agents at lower levels in the organizational hierarchy. We believe that networks at high levels of the hierarchy are considerably larger than at lower levels. Hence, information is particularly scarce at the beginning of the career but abundant at the executive level. The more information there is, the lower is the uncertainty about the true state, making men's additional information less valuable. To the contrary, women bring closure to the network, which is particularly beneficial once a sufficient amount of information is available. This is the case further up the hierarchy, but not initially. Therefore, in line with Lalanne and Seabright (2011), our model predicts that at management levels, it is especially profitable to have women in the team because, in environments saturated with information, women's peer pressure kicks in more strongly than men's additional information.

Walker et al. (1997) provides additional support for this view, arguing that sparse networks are most important at the beginning of the network formation process. They analyze the changing value of social capital over the life cycle of inter-firm networks and find that being at a position that bridges a structural hole is more valuable at early stages of network formation, since most tasks of the early networks are informational. However, as the network becomes established, densely connected network relationships and closure become more valuable than brokerage opportunities. In a similar vein, Ferriani et al. (2009) show that producers in the film industry who are more central in the network (i.e. have more access to information) are more likely to increase the box revenue from a movie but that returns to centrality become smaller the more central the producer is. Our theory sheds light on the diminishing returns to network reach (i.e. to degree) and provides a mechanism for why the different types of social capital that emerge from tight and lose networks are complementary.

4. During recessions men's unemployment exceeds women's unemployment.

Albanesi and Sahin (2013) find that this is true even when controlling for sectors. Employers seem to have a preference for keeping woman on their workforce during recessions, a pattern our model can help understand. Our model predicts that women do particularly well when rewards are low, which we believe is the case in economic downturns. Men's additional information leads to a particular advantage if the state of the world is high. In this case, they are more certain than woman that the true state is high, leading to extra effort. The opposite is true in the low state where men assign a higher probability to the low state than women, leading to lower effort. In turn, women take into account that project failures hit them particularly hard because, due to more common friends, failures destroy more second period project opportunities. This effect pushes up women's effort *independently* of the state of the world.

This is the mechanism why women are more likely to outperform men when the state of the world is low. We thus argue that women perform relatively better than men in recessions because they remain productive even if rewards are low. In contrast, men are more selective in their effort choice and better adjust their effort to the expected project value. They put low effort for low value.

5. The beginning of the career is the most decisive period for the gender wage gap formation.

Several studies point out the importance of the gender wage gap at early stages of the career for the future wage path (Babcock and Laschever (2003), Gerhart and Rynes (1991), Martell et al. (1996)). Bertrand et al. (2010a) document that, already 5 years into the career, the gender wage gap among MBAs in the US is substantial and keeps growing thereafter. Napari (2006) shows that in Finland, early years after labor market entry have the largest impact on gender wage differences. Thereafter, the wage gap simply persists. Similar findings are documented for Germany, where the entry wage gap is already 25% (Kunze (2003)).

Our model predicts a strong impact of the performance at the beginning of the career on the future income trajectory of men relative to women. This is because the second period wage does not only depend on the contemporaneous project outcome but also on first period performance. This effect is particularly important for women. Due to their higher clustering, they would lose more second period project opportunities in case of first period project failure even if first period performance is equal across gender. Because of this dynamic effect of peer pressure and since men outperform women in the high state, a first period wage gap in favor of men is persistent and can only be reverted if the second period state is low. Interestingly, a first period wage gap would persist even if there is no uncertainty in the second period. The reason is that women are more likely to be teamed up with someone who punishes them for a previous failure by exerting low effort.

2.7 Equilibrium Selection

In our analysis, we have selected the equilibrium that induces workers to play high effort if their relationship is good and zero effort if their relationship is bad. In an alternative equilibrium, agents could choose to play the static high effort PPE in each period, independently of their relationship.[39] Here we justify why we focus on the former equilibrium.

We first show that there are settings where payoffs from playing the static high effort PPE are lower than payoffs from our proposed strategy. We then argue that the theoretical predictions obtained under our strategy are similar to those we would obtain if the agents always chose the payoff-dominant strategy. To see this, we compare the overall welfare across time under these two strategies,

$$\begin{aligned} W_i^{stat} =& s_i(1+\beta)E\left[f(e'(y),e'(y))\pi(y)-c(e'(y))\right], && (2.10)\\ W_i^{dyn} =& s_iE\left[f(e(y),e(y))\pi(y)-c(e(y))\right] \\ &+ s_i\beta E\left[(1-r_i(1-f(e(y),e(y))))\right]E\left[f(e'(y'),e'(y'))\pi(y')-c(e'(y'))\right]. && (2.11) \end{aligned}$$

In equation (2.11), the expectation has to be taken not only with respect to y but also with respect to r_i. To avoid this, we assume that r_i is constant. The equilibrium we select yields a higher payoff than the static PPE whenever $W_i^{dyn} > W_i^{stat}$, which may or may not hold. To simplify notation, we let $EV_1 = E\left[f(e(y),e(y))\pi(y)-c(e(y))\right]$ and $EV_2 = E\left[f(e'(y'),e'(y'))\pi(y')-c(e'(y'))\right]$. Welfare under our strategy, W_i^{dyn}, is higher than welfare in the static high effort PPE, W_i^{stat}, whenever

$$EV_1 - EV_2 > \beta r_i(1-E[f(e(y),e(y))])EV_2 \qquad (2.12)$$

So, if $EV_1 - EV_2 > 0$ and $E[f(e(y),e(y))]$ is sufficiently large, then welfare is higher under our strategy.[40] An example for which equation (2.12) holds is given in the Appendix.

Notice that $E[f(e(y),e(y))]$ is large if effort is high under any signal realization. But effort is only stable if the project values across states are similar, implying little uncertainty in the environment. And we have shown that women exert higher effort than men in these environments.

Instead, suppose now that agents always play the payoff-dominating strategy. Then, in an environment with high uncertainty the static high effort PPE will be selected, whereas in an environment with low uncertainty and relatively high payoffs, our proposed strategy should be implemented. But this implies that the differences between men and women, which we discussed in Section 6, would be exacerbated if agents always choose the payoff dominant

[39] Obviously, there are other equilibria, such as whenever a project fails, all relationships in the network turn bad and then all players choose zero effort. Another possibility is that a good relationship leads to zero effort and a bad relationship to positive effort. We find these equilibria hard to justify and therefore use the static PPE as a benchmark.

[40] Note that $EV_1 - EV_2 > 0$ might not always be the case, although $e > e'$. To see this we consider the example given in Section 6, where $EV_1 < EV_2$. The reason is that workers choose very high effort in the first period even if the project does not yield a payoff in order to avoid having a bad relationship in the second period.

strategy.[41] Women would do even worse than men in uncertain environments than under our strategy and do even better in situations with low uncertainty and high payoffs. Essentially, our proposed strategy biases the results against us, which is why we feel comfortable about our equilibrium selection.

A different question is whether the payoff dominant strategy will always be played. There are two equilibria in the static game, where choosing zero effort is risk dominant. The evidence on whether the payoff dominant or risk dominant strategy is played is mixed at best (Van Huyck et al. (1990), Cooper et al. (1990), Cooper et al. (1992)). We give an intuitive explanation under which circumstances workers might risk to choose the high effort which can potentially result in a loss (namely when they trust their project partner after a good history) and when they go for the risk dominant strategy (which is after a loss and thus bad history).

2.8 Conclusion

We identify a new dimension of heterogeneity between men and women, namely differences in their networks structure, and connect these differences to discrepancies in their labor market outcomes. We first establish that men have a higher degree than women, whereas women have a higher clustering coefficient. Based on this, we build a model that sheds light on the relative advantages of having a male network (high degree, low clustering) versus a female network (low degree, high clustering). A higher clustering coefficient implies higher peer pressure, whereas a higher degree improves access to information. Both peer pressure and access to information can attenuate a team moral hazard problem in the work place. But whether peer pressure or access to information is more important depends on the work environment. We find that, in environments where uncertainty is high, information is crucial and, therefore, men outperform women. This uncertainty can either stem from large payoff variability, moderately informative signals, a small number of overall signals or little prior knowledge about the state of the world.

Our findings are in line with large gender wage gaps in occupations characterized by uncertainty and with the fact that more men than women choose occupations with high earnings volatility, where volatility can be interpreted as uncertainty. Additionally, it is documented that having women in the network is beneficial once there is an abundance of information. Our model suggests that this is due to women adding network closure which is more beneficial under these circumstances than additional information. Further, it is documented that women have a higher employment rate in recessions when rewards are low, which our model would also predict. Last, our model is consistent with empirical findings of how the gender wage gap changes over the career paths, with a strong impact of the early career wage gap on future wage trajectories of men and women.

We propose a novel, network-based explanation for gender differences in labor market outcomes. We see this approach complementary to other explanations, such as differences

[41] There is one exception concerning the benefit of having women in the network higher up in the organizational hierarchy. Here women would not outperform men anymore, but would do the the same unless the payoffs in both states are fairly close together.

in preferences, risk aversion, bargaining behavior and discrimination. Ideally, we would like to test our theory empirically in order to quantify the impact of network differences on wages is. However, data requirements are significant. We would need a dataset of informal networks *at* the workplace. We are aware of no such dataset at this moment and leave this question for future research.

It is beyond the scope of this paper to analyse the source of network differences between men and women. There could be an underlying trait that makes women choose more closed networks, such as risk aversion, which also leads them to choose different occupations. But the network structure could also emerge due to differences in games boys and girls play. Whereas boys tend to play in big groups, girls are encouraged to socialize in a different manner already from an early age onwards. So the question is whether friendship formation is guided by an innate trait or a trait that is learned.

Last, at its current stage, we do not use our model to study the optimal composition of a team. The optimal team composition should depend on the network structures of the team members. We believe that this is an interesting extension of our research, which we aim to address in future work.

2.9 Appendix

2.9.1 Data Appendix

Friendship Networks

The friendship information in the AddHealth data set is based upon actual friends nominations. Students were asked to name up to 5 male and female friends. Students named friends both from the school they attend as well as friends from outside the school. Some of the friends, who do not attend the same school attend a sister school[42] and can still be identified. The other friends cannot be identified and are dropped subsequently from the sample.[43]

Descriptive Statistics AddHealth

Table 2.3: Differences in Degree and Clustering for Men and Women

	Male Students				Female Students			
	Mean	Std Dev.	Min	Max	Mean	Std Dev.	Min	Max
Cl. Coeff.	0.117	0.195	0	1	0.130	0.195	0	1
Cl. Coeff. (dir.)	0.0876	0.151	0	1	0.0996	0.154	0	1
In Degree	3.597	3.554	0	37	3.917	3.482	0	34
Out Degree	3.417	3.660	0	10	4.100	3.651	0	10
Degree	7.014	5.921	0	44	8.017	5.970	0	43
Age	15.08	1.719	10	19	14.92	1.702	10	19
Size/1000	1.196	0.678	0.0290	2.982	1.200	0.686	0.0290	2.982
Observations	54881				53240			

Estimation Results

We estimate whether gender has a significant influence on degree as well as on the clustering coefficients. We standardize all of our measures in order to improve the interpretability of our results. Further, we normalize the age by subtracting 16. In all our regressions, we also control for school, which serves to capture location effects as well as time differences from when the data was collected. Note that we are not interested in determining which other factors influence these network characteristics, as is done e.g. in Conti et al. (2013).[44] The

[42] A sister school is a school in the same community. So, if in a community there is a high school and a middle school, then the high school is the sister school of the middle school and the middle school is the sister school of the high school.

[43] Overall, less than 10% of the observations are dropped. We believe this to not be a problem as we are interested in a proxy for the friendship network *at* the workplace, not for the entire friendship network of individuals.

[44] Conti et al. (2013) take the in-degree of high school students and find that wages 35 years are influenced by how often students were named as friends. They argue that a high in-degree is a measure of social skills, of how good someone is in building positive personal relationships and in adjusting to a certain environment and situation. They also provide evidence that the in-degree manages to capture something other than personality, by controlling for personality traits. Similarly, we use the network as a measure of social skills that can still impact outcomes later on.

Table 2.4: Differences in Degree and Clustering for Men and Women

	Cl. Coeff. (dir.)	Cl. Coeff. (dir.)	Cl. Coeff.	Cl. Coeff.	In Degree	In Degree	Out Degree	Out Degree	Degree	Degree
Female	0.110***	0.0729***	0.0922***	0.0612***	0.107***	0.146***	0.186***	0.193***	0.178***	0.205***
	(0.00565)	(0.0107)	(0.00563)	(0.0106)	(0.00683)	(0.0144)	(0.00609)	(0.0123)	(0.00634)	(0.0131)
Age-16	0.00440	-0.000773	0.00456	-0.000780	-0.0173***	0.00629	-0.0500***	-0.0375***	-0.0410***	-0.0193***
	(0.00249)	(0.00292)	(0.00252)	(0.00296)	(0.00271)	(0.00335)	(0.00247)	(0.00304)	(0.00255)	(0.00313)
Size/1000	0.0310	0.0235	-0.0170	-0.0224	0.0420	0.0329	0.0339	0.0240	0.0457	0.0342
	(0.0634)	(0.0631)	(0.0750)	(0.0748)	(0.0571)	(0.0578)	(0.0839)	(0.0831)	(0.0714)	(0.0707)
Female*Age 16-17		0.0668***		0.0594***		-0.102***		-0.0443***		-0.0878***
		(0.0111)		(0.0110)		(0.0121)		(0.0108)		(0.0113)
Female*Age 18-19		-0.0360		-0.0126		-0.356***		-0.214***		-0.342***
		(0.0253)		(0.0258)		(0.0221)		(0.0230)		(0.0225)
Female*Size/1000		0.0122		0.00858		0.0148		0.0169		0.0192*
		(0.00901)		(0.00896)		(0.00991)		(0.00911)		(0.00936)
Constant	0.0767	0.0842	0.186	0.190	-0.242*	-0.210	0.165	0.189	-0.0410	-0.00819
	(0.127)	(0.126)	(0.151)	(0.150)	(0.114)	(0.116)	(0.169)	(0.168)	(0.144)	(0.142)
Additional Controls: School Fixed Effects										
Observations	84792	84792	84792	84792	84792	84792	84792	84792	84792	84792
R^2	0.198	0.199	0.205	0.205	0.132	0.134	0.170	0.170	0.205	0.207

Standard errors in parentheses

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

purpose of this estimation is only to show that men's and women's networks differ. Our results are given in Table 2.4.

We find that girls have a significantly higher clustering coefficient, independently of how the clustering coefficient is measured. Both younger and older girls have a higher clustering coefficient, i.e. this characteristic does not change as students grow older. Girls also have a higher in and out degree as well as overall degree. But older girls have a lower absolute degree, out degree and in degree than younger girls, i.e. unlike with the clustering coefficients this property changes as girls mature. However, the degree does not change much for boys as they grow older. When just taking into consideration the oldest students, i.e. those aged 18 and 19, which are the students we are most interested in as we are interested in the network properties of men and women, girls have a lower in, out and overall degree.[45]

2.9.2 Technical Appendix A

Derivation of s_i

The probability that one agent is chosen is given by $Pr(K) = \frac{N-1}{\frac{1}{2}N(N-1)} = \frac{2}{N}$, and the probability that this agent i is linked to the suggested project partner j, given that he is selected by $Pr(g_{ij} = 1|K) = \frac{d_i}{N-1}$. Then, the probability of being chosen *and* being partnered with a friend is

$$s_i \equiv Pr(g_{ij} = 1 \wedge K) = Pr(g_{ij} = 1|K)Pr(K) = \frac{2d_i}{N(N-1)}.$$

Relationship Quality

We outline here formally how a project outcome affects the relationships of workers. As mentioned previously, whether the project of workers i and j was a success, S, or a failure, F is publicly observable and denoted by $\omega \in \Omega = \{S, F\} \times \{1, 2, \ldots, N\}^2$. As an example, if $\omega = S12$, this means that a project was successfully completed by workers 1 and 2. We condition also on the workers who carried out the project as we do not only care about whether the project was successful but also about the workers who were involved. Each project failure induces some bad relationships in the network g. The network that contains the links that signify a bad relationship is denoted by $g_b \subset g$. The specific network g_b that arises after Fij, that is a project failure between workers i and j, where $g_{ij} = 1$, is given by $g_b(Fij) = \{\{ij, il, jl\}|g_{il} = 1 \wedge g_{jl} = 1, \forall l\}$. Workers i and j have a bad relationship with each other if their joint project fails. But a worker l, who is connected to both i and j also has a bad relationship with both of them. Denote by $g_g(Fij) = g \backslash g_b(Fij)$ the good relationships in the network g. Let $\gamma_g \in g_g$ and $\gamma_b \in g_b$. Further, for any i, j $g_g(Sij) = g$.

[45] As we have standardized the clustering coefficients as well as degrees, the coefficients can be interpreted in terms of standard deviations.

Equilibrium Selection

An example for which equation (2.12) holds is given in Table 2.5. We assume $f(e_i, e_j) = \sqrt{e_i, e_j}$ and $c(e_i) = \frac{1}{2}e_i^2$. In this example, men exert on average lower effort than women, in both states of the world. This is not surprising given that the project value in both states of the world is fairly similar.

Table 2.5: Welfare Parameters

v_l	v_h	p	q	β	d^W	d^M	C^W	C^M	N
1.5	1.6	0.75	0.5	0.9	2	3	2	1	4

2.9.3 Technical Appendix B: Proofs

Proof of Proposition 2.1: Static Decision Problem

Given the assumptions on $f(.,.)$, there always exists an equilibrium where both project partners exert zero effort. It therefore remains to be shown that there exists exactly one equilibrium with $e_i = e_j > 0$.

We first show symmetry. From the first order conditions we obtain

$$\frac{f_1(e_i, e_j)}{f_2(e_i, e_j)} = \frac{c'(e_i)}{c'(e_j)} \tag{2.13}$$

Suppose, by contradiction, that effort levels are not symmetric and assume that $e_j > e_i$. Due to convexity of the cost functions, the RHS of (2.13) is smaller than one. Due to concavity and supermodularity of the effort function, we have $f_1(e_i, e_j) > f_2(e_i, e_j)$, which is why the LHS is larger than one, which gives the contradiction.

Further, there is exactly one equilibrium where both workers exert strictly positive effort. It suffices to show that the FOCs (which under symmetry become a function of one variable) have one zero under the condition that effort is strictly positive.

$$f_1(e, e)\pi(y) = c'(e) \tag{2.14}$$

Due to our assumption of constant returns to scale, $f_1(e, e)$ is constant in e. The first derivative of the cost function $c'(e)$ is linear in e and starts in the origin (due to Assumption 2.1). Hence, the two functions have a unique intersection, implying one symmetric equilibrium with strictly positive effort.

Proof of Lemma 2.1:

$\pi(y)$ has the martingale property:

$$\pi(y_n) = Pr(\theta_h|y_n)v_h + (1 - Pr(\theta_h|y_n))v_l$$

Define $\psi_n \equiv Pr(\theta_h|y_n)$. We know that the stochastic process $\{\psi_n\}$ is a martingale as

$$E(\psi_{n+1}|y_n) = E(E(\psi|y_{n+1})|y_n) = E(\psi|y_n) = \psi_n,$$

where the second equality follows from the *tower property* of conditional expectations. Then,

$$\begin{aligned} E(\pi(y_{n+1})|y_n) &= E(\psi_{n+1}v_h + (1-\psi_{n+1})v_l|y_n) = E(\psi_{n+1}v_h|y_n) + E((1-\psi_{n+1})v_l|y_n) \\ &= \psi_n v_h + (1-\psi_n)v_l = \pi(y_n) \end{aligned}$$

Properties of $E\left(\pi(y_n)\right)$ and $E\left(\pi(y_n)|\theta\right)$:

1. The number of signals do not matter for $E(\pi(y))$ due to the martingale property of $\pi(y)$,

$$E(\pi(y_{n+1})) = E(E(\pi(y_{n+1})|y_n)) = E(\pi(y_n)).$$

2. We note that the posterior is given by

$$\begin{aligned} Pr(\theta_h|y) &= \frac{Pr(y|\theta_h)Pr(\theta_h)}{Pr(\theta_h)Pr(y|\theta_h) + Pr(\theta_l)Pr(y|\theta_l)} = \frac{qp^y(1-p)^{n-y}}{qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y} \\ &= \frac{1}{1 + \frac{1-q}{q}\left(\frac{1-p}{p}\right)^{2y-n}} \end{aligned} \tag{2.15}$$

To simplify notation we define $\tilde{p} \equiv \frac{1-p}{p}$, $\tilde{q} \equiv \frac{1-q}{q}$ and $\hat{y} \equiv 2y-n$. Then, $\psi_n = Pr(\theta_h|y) = \frac{1}{1+\tilde{q}\tilde{p}^{\hat{y}}}$.

We are interested in showing that

$$E\left(\pi(y_{n+1})|\theta_h\right) > E\left(\pi(y_n)|\theta_h\right) \tag{2.16}$$

$$E\left(\pi(y_{n+1})|\theta_l\right) < E\left(\pi(y_n)|\theta_l\right) \tag{2.17}$$

We will show that equation (2.16) holds and leave the proof of equation (2.17) to the reader. We can rewrite equation (2.16) and we obtain

$$(v_h - v_l)E\left((\psi_{n+1} - \psi_n)|\theta_h\right) > 0$$

As $(v_h - v_l) > 0$, by assumption, it remains to be shown that $E\left(\psi_{n+1} - \psi_n|\theta_h\right) > 0$. Given $\theta = \theta_h$, $\psi_{n+1} = \frac{1}{1+\tilde{q}\tilde{p}^{\hat{y}+1}}$ with probability p and $\psi_{n+1} = \frac{1}{1+\tilde{q}\tilde{p}^{\hat{y}-1}}$, with probability $(1-p)$. We can show that

$$\begin{aligned} & \frac{1}{1+\tilde{q}\tilde{p}^{\hat{y}}} < \frac{p}{1+\tilde{q}\tilde{p}^{\hat{y}+1}} + \frac{1-p}{1+\tilde{q}\tilde{p}^{\hat{y}-1}} \\ \Leftrightarrow \quad & p\tilde{p}^2 + (1-p) - \tilde{p} < \tilde{q}\tilde{p}^{\hat{y}}(p + (1-p)\tilde{p}^2 - \tilde{p}) \end{aligned}$$

Note that $p\tilde{p}^2 + (1-p) - \tilde{p} = 0$. Then, $0 < \tilde{q}\tilde{p}^{\hat{y}}(p + (1-p)\tilde{p}^2 - \tilde{p})$, which holds for $p > \frac{1}{2}$ and concludes the proof.

Additional signals do not matter in the following cases:

(i) For $v_l \to v_h$,

$$\lim_{v_l \to v_h} E\left(\pi(y_n)|\theta_h\right) = \sum_{y=0}^{n} \frac{(n)!}{y!(n-y)!} \left(p^y(1-p)^{n-y}\right) v_h = (p+1-p)^n v_h = v_h,$$

where the second step follows from the binomial formula. The expression is independent of n and therefore additional signals do not matter. Similarly, this also holds for $E\left(\pi(y)|\theta_l\right)$.

(ii) Assume $p \to 1$. Then,

$$\begin{aligned}
\lim_{p\to 1} E\left(\pi(y_n)|\theta_h\right) &= \lim_{p\to 1} \sum_{y=0}^{n} \frac{(n)!}{y!(n-y)!} \left(p^y(1-p)^{n-y}\right) \left(\frac{qp^y(1-p)^{n-y}v_h + (1-q)p^{n-y}(1-p)^y v_l}{qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y}\right) \\
&= \lim_{p\to 1} \frac{(n)!}{n!(n-n)!} \left(p^n(1-p)^{n-n}\right) \left(\frac{qp^n(1-p)^{n-n}v_h + (1-q)p^{n-n}(1-p)^n v_l}{qp^n(1-p)^{n-n} + (1-q)p^{n-n}(1-p)^n}\right) \\
&= \lim_{p\to 1} p^n \left(\frac{qp^n v_h + (1-q)(1-p)^n v_l}{qp^n + (1-q)(1-p)^n}\right) = v_h,
\end{aligned}$$

and analogue for $\theta = \theta_l$.

(iii) Assume $q \to 1$. Then,

$$\lim_{q\to 1} E\left(\pi(y_n)|\theta_h\right) = \sum_{y=0}^{n} \frac{(n)!}{y!(n-y)!} \left(p^y(1-p)^{n-y}\right) v_h = (p+1-p)^n v_h = v_h$$

which is independent of n. Similarly for $q \to 0$ and $E\left(\pi(y)|\theta_l\right)$.

(iv) Note that $y \sim \text{Binomial}(np, np(1-p))$ if $\theta = \theta_h$ and $y \sim \text{Binomial}(n(1-p), np(1-p))$ if $\theta = \theta_l$. Then, $\lim_{n\to\infty}(y-(n-y)) = \infty$ if $\theta = \theta_h$ and $\lim_{n\to\infty}(y-(n-y)) = -\infty$ if $\theta = \theta_l$. To see this note that $y-(n-y) = 2y-n$. By the central limit theorem, as $n \to \infty$,

$$\begin{aligned}
\text{if} \quad \theta = \theta_h \quad y \xrightarrow{p} np \quad &\Rightarrow \lim_{n\to\infty}(2np-n) = \infty \\
\text{if} \quad \theta = \theta_l \quad y \xrightarrow{p} n(1-p) \quad &\Rightarrow \lim_{n\to\infty}(2n(1-p)-n) = -\infty.
\end{aligned}$$

Then, $\lim_{n\to\infty} Pr(\theta_h|y) = 1$ if $\theta = \theta_h$ and $\lim_{n\to\infty} Pr(\theta_h|y) = 0$ if $\theta = \theta_l$ as

$$\lim_{n\to\infty} Pr(\theta_h|y) = \lim_{n\to\infty} \frac{1}{1 + \frac{1-q}{q}\left(\frac{1-p}{p}\right)^{2y-n}}$$

We have already shown that $Pr(\theta_h|y)$ is increasing in n if $\theta = \theta_h$ and decreasing in n if $\theta = \theta_l$. Thus we can apply the Monotone Convergence Theorem, which implies that $\lim_{n\to\infty} E(Pr(\theta_h|y)v_h) = E(\lim_{n\to\infty} Pr(\theta_h|y)v_h)$. From this it follows that $\lim_{n\to\infty} E\left(\pi(y)|\theta_h\right) = v_h$ and $\lim_{n\to\infty} E\left(\pi(y)|\theta_l\right) = v_l$.

Proof of Lemma 2.2:

$V^*(y')$ is a Submartingale: We can express $V^*(y')$ as a function of $\pi(y')$, and write

$$V^*(y') \equiv g(\pi(y')) \tag{2.18}$$

As $\pi(y')$ is a martingale, we know that when g is a convex function, then $g(\pi(y'))$ is a submartingale whenever $E(V^*(y'_n)) < \infty$, which is always fulfilled as $0 \leq E(V^*(y'_n)) < v_h \; \forall n$.

Note that the equilibrium effort depends the expected project payoff through the signals, or $e'('y)$. We mostly omit this dependence here in order to keep notation simple but write simply e'.

Applying the envelope theorem repeatedly, the first and second derivative of g are given by

$$\begin{aligned}
\frac{\partial g(\pi(y))}{\partial \pi(y)} &= f_2(e',e')\pi(y')\frac{\partial e'}{\partial \pi(y')} + f(e',e') \\
\frac{\partial^2 g(\pi(y'))}{\partial \pi(y')\partial \pi(y')} &= [f_{22}(e',e') + f_{12}(e',e')]\pi(y')\left(\frac{\partial e'}{\partial \pi(y')}\right)^2 + f_2(e',e')\pi(y')\frac{\partial^2 e'}{\partial \pi(y')\partial \pi(y')} + f_2(e',e')\frac{\partial e'}{\partial \pi(y')} \\
&+ (f_1(e',e') + f_2(e',e'))\frac{\partial e'}{\partial \pi(y')} \\
&= f_2(e',e')\pi(y')\frac{\partial^2 e'}{\partial \pi(y')\partial \pi(y')} + f_2(e',e')\frac{\partial e'}{\partial \pi(y')} + (f_1(e',e') + f_2(e',e'))\frac{\partial e'}{\partial \pi(y')}
\end{aligned}$$

From first order condition of the static problem, evaluated at the equilibrium effort, we can compute

$$\begin{aligned}
\frac{\partial e'}{\partial \pi(y')} &= \frac{f_1(e',e')}{c''(e')} > 0 \\
\frac{\partial^2 e'}{\partial \pi(y')\partial \pi(y')} &= \frac{(f_{11}(e',e') + f_{21}(e',e'))\frac{\partial e'}{\partial \pi(y)}}{c''(e')} = 0
\end{aligned}$$

It follows that

$$\frac{\partial^2 g(\pi(y'))}{\partial \pi(y')\partial \pi(y')} = f_2(e',e')\frac{\partial e'}{\partial \pi(y')} + (f_1(e',e') + f_2(e',e'))\frac{\partial e'}{\partial \pi(y')} > 0,$$

which implies that $V^*(y'_n)$ is a submartingale.

Properties of $E(V^*_n) = \sum_{y=0}^{n} \frac{n!}{(y)!(n-y)!}(qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y)\left(f(e',e')\pi(y) - c(e')\right)$

(i) $v_l \to v_h$.

We are interested in

$$\lim_{v_l \to v_h} E(V^*_n) = \lim_{v_l \to v_h} \sum_{y=0}^{n} \frac{n!}{(y)!(n-y)!}(qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y)\left(f(e'(y'),e'(y'))\pi(y') - c(e'(y'))\right),$$

where $e'(y')$ is the equilibrium effort for given y'. As the other terms are constant in

v_l, all that matters is

$$\begin{aligned}\lim_{v_l \to v_h} \left(f(e'(y'), e'(y'))\pi(y') - c(e'(y'))\right) &= \lim_{v_l \to v_h} f(e'(y'), e'(y')) \lim_{v_l \to v_h} \pi(y') - \lim_{v_l \to v_h} c(e'(y')) \\ &= \lim_{v_l \to v_h} f(e'(y'), e'(y'))v_h - \lim_{v_l \to v_h} c(e'(y'))\end{aligned}$$

Note that $\lim_{\pi(y') \to v_h} e'(y') = e'_{v_h}$, i.e. the effort converges to some constant as $\pi(y') \to v_h$ since $e'(y')$ is a linear function of $\pi(y')$ (see (2.2)). Also, due to constant returns to scale, $f(e'(y'), e'(y')) = e'(y')f(1,1)$ and thus $\lim_{e'(y') \to e'_{v_h}} f(e'(y'), e'(y')) = e'_{v_h} f(1,1)$, which again is constant in n. As $f(.,.)$ is continuous, i.e. $f(e'_{v_h}, e'_{v_h}) = e'_{v_h} f(1,1)$, we know that $\lim_{\pi(y') \to v_h} f(e'(y'), e'(y')) = e'_{v_h} f(1,1)$. The argument is similar for $c(.)$. Then, we can write

$$\lim_{v_l \to v_h} \left(f(e'(y'), e'(y'))\pi(y') - c(e'(y'))\right) = b_{v_l},$$

where b_{v_l} is constant and thus independent of n. Therefore, as v_l converges to v_h, the expected second period value converges to a constant and is independent of the number of signals,

$$\lim_{v_l \to v_h} E(V_n^*) = b_{v_l}.$$

(ii) $p \to 1$ for $\theta \in \{\theta_h, \theta_l\}$.

Note that

$$\begin{aligned}&\lim_{p \to 1} \pi(y) = v_h \quad \text{if} \quad n - 2y < 0 \\ &\lim_{p \to 1} \pi(y) = qv_h + (1-q)v_l \quad \text{if} \quad n - 2y = 0 \\ &\lim_{p \to 1} \pi(y) = v_l \quad \text{if} \quad n - 2y > 0\end{aligned}$$

As $\pi(y)$ converges to some constant (and, of course, the same holds for $\pi(y')$), so does $(f(e'(y'), e'(y'))\pi(y') - c(e'))$. We denote by $V^*(v_h)$ $(V^*(v_l))$ $[V^*(v)]$ the limit when $\pi(y)$ converges to v_h (v_l) $[qv_h + (1-q)v_l]$.

Note further that if $n - 2y < 0$, $\lim_{p \to 1}(qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y) = \lim_{p \to 1} qp^y(1-p)^{n-y}$. Then we know that

$$\lim_{p \to 1} = \begin{cases} q & \text{if} \quad y = n \\ 0 & \text{otherwise} \end{cases}$$

If $n - 2y > 0$, $\lim_{p \to 1}(qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y) = \lim_{p \to 1}(1-q)p^{n-y}(1-p)^y$. It follows that

$$\lim_{p \to 1} = \begin{cases} 1-q & \text{if} \quad y = 0 \\ 0 & \text{otherwise} \end{cases}$$

Last, if $n - 2y = 0$, $\lim_{p \to 1}(qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y) = \lim_{p \to 1} p^y(1-p)^{n-y} = 0$, as $y, n > 0$ From this it then follows that

$$\lim_{p \to 1} E(V_n^*) = qV^*(v_h) + (1-q)V^*(v_l),$$

which is independent of n.

(iii) $q \to 1$.

Notice that,

$$\lim_{q \to 1}(qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y) = p^{n-y}(1-p)^y,$$
$$\lim_{q \to 1} \pi(y) = v_h.$$

It follows that $\lim_{q \to 1} E(V_n^*)$ is a constant and independent of n.

Next, $q \to 0$.

$$\lim_{q \to 0}(qp^y(1-p)^{n-y} + (1-q)p^{n-y}(1-p)^y) = p^{n-y}(1-p)^y,$$
$$\lim_{q \to 0} \pi(y) = v_l,$$

and $\lim_{q \to 0} E(V_n^*)$ is constant.

(iv) Abundance of Information: $n_{ext} \to \infty$.

We want to show that

$$\lim_{n \to \infty} E(V_n^*) = E(V^*).$$

We know that for each n, $E(V_n^*) \leq E(V_{n+1}^*)$ as V_n^* is a submartingale and that $E(V_n^*) \leq v_h$ for all n. By the monotone convergence theorem, we know that a finite limit exists, which we denote by $E(V^*)$.

Proof of Proposition 2.6: Trade-Off Between Information and Peer Pressure

We assume that a D-worker has a higher degree and hence more signals n_{int} and has clustering $(sr)^D$. In turn, a C-worker has a lower degree and thus a lower number of signals (and therefore $s^D > s^C$) but higher clustering and therefore $(sr)^C > (sr)^D$. Further, assume $v_l = 0$.

(i) Wage Dynamics:

Claim 1: $\quad w^D(\theta) > w^C(\theta) \quad \Rightarrow \quad E(w'^D) > E(w'^C).$

From definition (2.2), it follows that the second period expected wage across states is defined as

$$E(w') = qw'(\theta, \theta_h') + (1-q)w'(\theta, \theta_l') = qw'(\theta, \theta_h')$$

where the second equality is due to $v_l = 0$ and where we dropped the subindex i for convenience. Also, recall

$$w'(\theta, \theta_h') \equiv sPr(\gamma_g'|\theta)E(e'(y')|\theta_h)f(1,1)v_h$$

where $Pr(\gamma_g'|\theta) \equiv E(e(y)|\theta)rf(1,1) + 1 - r$. Suppose that in the first period $w^D(\theta) > w^C(\theta)$, implying $E(e(y)^D|\theta) > E(e(y)^C|\theta)$. Moreover, by assumption, $s^C < s^D$ and $(sr)^C > (sr)^D$. Hence, $[sPr(\gamma_g|\theta)]^D > [sPr(\gamma_g|\theta)]^C$. Last, by Proposition 2, $E(e'(y')^D|\theta_h') > E(e'(y')^C|\theta_h')$ and therefore $w'^D(\theta, \theta_h') > w'^C(\theta, \theta_h')$. Thus, $w^D(\theta) > w^C(\theta)$ implies $E(w'^D) > E(w'^C)$, which proves the claim.

Claim 2: (a) $w^D(\theta) = w^C(\theta) \Rightarrow E(w'^D) > E(w'^C)$.
(b) $w^D(\theta) > w^C(\theta) \Rightarrow E(w'^D) > E(w'^C)$ even if $E(e'(y')^D|\theta_h') = E(e'(y')^C|\theta_h')$.

(a) Even if $w^D(\theta) = w^C(\theta)$ and thus $E(e(y)^D|\theta) = E(e(y)^C|\theta)$, we have $[sPr(\gamma_g|\theta)]^D > [sPr(\gamma_g|\theta)]^C$ due to $s^C < s^D$ and $(sr)^C > (sr)^D$. Also, by Proposition 2, $E(e'(y')^D|\theta_h') > E(e'(y')^C|\theta_h')$ and therefore $w'^D(\theta, \theta_h') > w'^C(\theta, \theta_h')$. It follows: $w^D(\theta) = w^C(\theta) \Rightarrow E(w'^D) > E(w'^C)$.

(b) We use a similar argument as in (a). Even if uncertainty vanishes in the second period, that is even if the D-worker loses his informational advantage, implying $E(e'(y')^D|\theta_h') = E(e'(y')^C|\theta_h')$, it holds that if $w^D(\theta) > w^C(\theta)$ then $E(w'^D) > E(w'^C)$, because $[sPr(\gamma_g|\theta)]^D > [sPr(\gamma_g|\theta)]^C$.

(ii) Comparative Advantage:

We want to show that $\frac{E(w^C)}{E(w^D)}$ increases in n, q and $p > p^*$ if $E(\pi(y)|\theta_h)$ and $EV^*(y')$ are sufficiently concave in n, q and p, respectively. First notice that, assuming $v_l = 0$,

$$\frac{E(w^C)}{E(w^D)} = \frac{qw^C(\theta_h)}{qw^D(\theta_h)} = \frac{[E(\pi(y)|\theta_h) + \beta srEV^*(\gamma_g', y')]^C}{[E(\pi(y)|\theta_h) + \beta srEV^*(\gamma_g', y')]^D} \tag{2.19}$$

where we used the definition of wages and the expression for equilibrium effort (2.7). We want to show that (2.19) is increasing as uncertainty vanishes. To illustrate the argument, we show this for the case of increasing n (strictly, speaking we let n_{ext} increase). We adopt the following notation

$$
\begin{aligned}
[E(\pi(y)|\theta_h)]^C &= E(\pi(y_n)|\theta_h) \\
[E(\pi(y)|\theta_h)]^D &= E(\pi(y_{n+1})|\theta_h) \\
[EV^*(\gamma_g', y')]^C &= EV^*(y_n') \\
[EV^*(\gamma_g', y')]^D &= EV^*(y_{n+1}') \\
(sr)^C &= \overline{sr} \\
(sr)^D &= \underline{sr}
\end{aligned}
$$

We will show that (2.19) is increasing in n, that is,

$$
\frac{E(\pi(y_n)|\theta_h) + \beta\overline{sr}EV^*(y_n')}{E(\pi(y_{n+1})|\theta_h) + \beta\underline{sr}EV^*(y_{n+1}')} > \frac{E(\pi(y_{n-1})|\theta_h) + \beta\overline{sr}EV^*(y_{n-1}')}{E(\pi(y_n)|\theta_h) + \beta\underline{sr}EV^*(y_n')} \tag{2.20}
$$

if $E(\pi(y)|\theta_h)$ and $EV^*(y')$ are sufficiently concave, i.e. if

$$
EV^*(y_n')^2 > EV^*(y_{n-1}')EV^*(y_{n+1}') \tag{2.21}
$$

$$
E(\pi(y_n)|\theta_h)^2 > E(\pi(y_{n+1}|\theta_h))E(\pi(y_{n-1})|\theta_h) \tag{2.22}
$$

$$
\frac{EV^*(y_n')}{EV^*(y_{n-1}')}\frac{E(\pi(y_n)|\theta_h)}{E(\pi(y_{n+1})|\theta_h)} > \frac{\overline{sr}}{\underline{sr}} > \frac{EV^*(y_{n+1}')}{EV^*(y_n')}\frac{E(\pi(y_{n-1})|\theta_h)}{E(\pi(y_n)|\theta_h)}. \tag{2.23}
$$

To show this, rearrange (2.20) to get:

$$
\begin{aligned}
&[E(\pi(y_n)|\theta_h)]^2 - E(\pi(y_{n+1})|\theta_h)E(\pi(y_{n-1})|\theta_h) + \\
&\underline{sr}\,\overline{sr}\beta^2([EV^*(y_n')]^2 - EV^*(y_{n+1}')EV^*(y_{n-1}')) + \\
&\beta\overline{sr}E(\pi(y_n)|\theta_h)EV^*(y_n') - \beta\underline{sr}E(\pi(y_{n-1})|\theta_h)EV^*(y_{n+1}') + \\
&\beta\underline{sr}E(\pi(y_n)|\theta_h)EV^*(y_n') - \beta\overline{sr}E(\pi(y_{n+1})|\theta_h)EV^*(y_{n-1}') > 0
\end{aligned}
$$

This expression is positive if (2.21)-(2.23) hold. To see that (2.21)-(2.23) are implied by sufficiently strong concavity note the following. A function $f(n)$ is *log-concave* if:

$$
f(n+1)f(n-1) < f(n)^2 \tag{2.24}
$$

Hence, for (2.21)-(2.23) to hold, $E(\pi(y)|\theta_h)$ and $EV^*(y')$ must be sufficiently log-concave. But concavity implies log-concavity: Concavity of an increasing discrete function means

$$
\frac{1}{2}(f(n+1) + f(n-1)) < f(n) \tag{2.25}
$$

Then (2.25) implies (2.24) since

$$
\frac{1}{2}(f(n+1) + f(n-1)) > (f(n+1)f(n-1))^{0.5}.
$$

Last, we established before that $E(\pi(y)|\theta_h)$ and $EV^*(y')$ are increasing in n and converge.

Consequently, for all $n > n^*$, $E(\pi(y)|\theta_h)$ and $EV^*(y')$ are concave as defined in (2.25). We focus on the part of the parameter space where $E(\pi(y)|\theta_h)$ and $EV^*(y')$ are sufficiently concave, i.e. where conditions (2.21)-(2.23) hold.

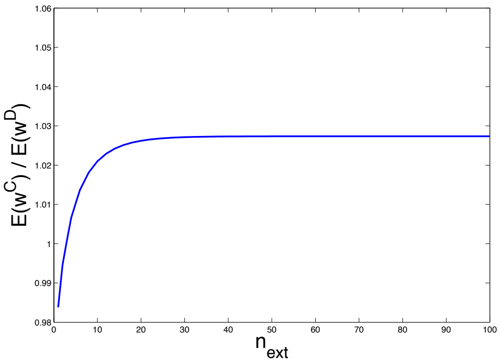
The arguments that (2.19) is increasing in p (for $p > p^*$) and q are analogous and slightly simpler because $E(\pi(y)|\theta_h)$ and $EV^*(y')$ are continuously differentiable in p and q. We omit them for brevity and instead highlight some of our simulation results.

To graphically illustrate the comparative advantage results, we compute a parametric example of this model and provide some simulations. Effort and cost functions are respectively given by $f(e_i, e_j) = \sqrt{e_i, e_j}$ and $c(e_i) = \frac{1}{2}e_i^2$. We set the parameters such that $e < e_{max}$ always holds (see Table 2.6).

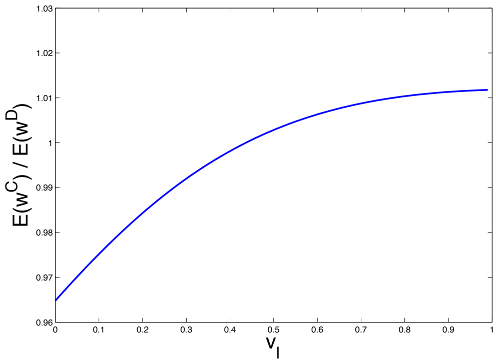
Table 2.6: Baseline Parameters

v_l	v_h	p	q	β	d^W	d^M	C^W	C^M	N
0	1	0.75	0.5	0.9	2	3	2	1	10

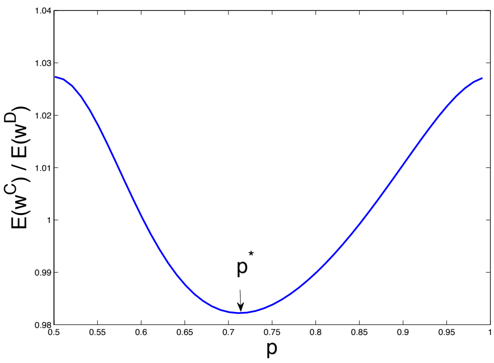
Figure 2.2: Expected Wage of Agent with Higher Clustering Relative to Agent with More Information



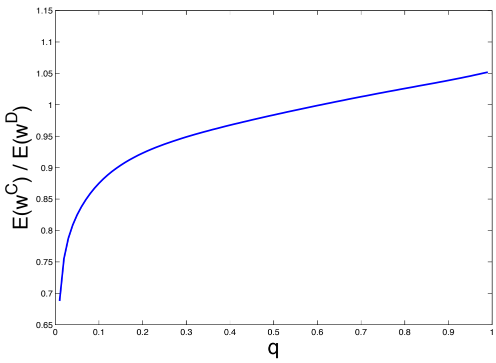
As a Function of External Info



As a Function of Low Value



As a Function of Signal Precision



As a Function of Prior Belief

Bibliography

Adda, J., C. Dustmann, and K. Stevens (2011). The Career Costs of Children.

Albanesi, S. and A. Sahin (2013). *The Gender Unemployment Gap.* Centre for Economic Policy Research.

Allcott, H., D. Karlan, M. Möbius, T. Rosenblat, and A. Szeidl (2007). Community Size and Network Closure. *The American Economic Review 97*(2), 80–85.

Aral, S., E. Brynjolfsson, and M. W. Van Alstyne (2012, September). Information, Technology and Information Worker Productivity. *Information Systems Research 23*(3), 849–867.

Arrow, K. and R. Borzekowski (2004). Limited Network Connections and the Distribution of Wages. *FEDS Working Paper No. 2004-41*.

Atkinson, S., S. Baird, and M. Frye (2003). Do female mutual fund managers manage differently? *Journal of Financial Research 26*(1), 1–18.

Babcock, L. and S. Laschever (2003). *Women Don't Ask: Negotiation and the Gender Divide.* Princeton University Press.

Beaman, L. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the US. *The Review of Economic Studies 79*(1), 128–161.

Beaman, L., N. Keleher, and J. Magruder (2013). Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi.

Belle, D. (1989). *Children's Social Networks and Social Supports*, Volume 136. John Wiley & Sons Inc.

Bertrand, M., C. Goldin, and L. F. Katz (2010a, July). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics 2*(3), 228–55.

Bertrand, M., C. Goldin, and L. F. Katz (2010b). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, 228–255.

Bierema, L. L. (2005). Women's Networks: A Career Development Intervention or Impediment? *Human Resource Development International 8*(2), 207–224.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources 8*(4), 436–455.

Bohnet, I. and F. Saidi (2011). Informational Differences and Performance: The Role of Organizational Demography.

Booth, A. L. (2009). Gender and competition. *Labour Economics 16*(6), 599–606.

Burt, R. (1992). *Structural Holes: The Social Structure of Competition.* Harvard University Press.

Burt, R. (2011). Structural holes in virtual worlds. unpublished.

Calvó-Armengol, A. and J. d. M. Beltran (2009). Information gathering in organizations: equilibrium, welfare, and optimal network structure. *Journal of the European Economic Association 7*(1), 116–161.

Calvo-Armengol, A. and M. Jackson (2004). Effects of Social Networks on Employment and Inequality. *The American Economic Review 94*(3), 426–454.

Calvó-Armengol, A. and M. Jackson (2007). Networks in Labor Markets: Wage and Employment Dynamics and Inequality. *Journal of Economic Theory 132*(1), 27–46.

Card, D., A. R. Cardoso, and P. Kline (2013). Bargaining and the Gender Wage Gap: A Direct Assessment. Technical report, Institute for the Study of Labor (IZA).

Coleman, J. (1988a). Free riders and Zealots: The Role of Social Networks. *Sociological Theory 6*(1), 52–57.

Coleman, J. (1988b). Social Capital in the Creation of Human Capital. *American journal of sociology*, 95–120.

Conti, G., A. Galeotti, G. Müller, and S. Pudney (2013). Popularity. *Journal of Human Resources 48*(4), 1072–1094.

Cooper, R., D. V. DeJong, R. Forsythe, and T. Ross (1990). Selection criteria in coordination games: Some experimental results. *American Economic Review 80*(1), 218–233.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross (1992). Communication in coordination games. *The Quarterly Journal of Economics 107*(2), 739–771.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricicans.* Oxford University Press, USA.

Davis, J. A. (1967). Clustering and structural balance in graphs. *Human relations.*

Ding, W., F. Murray, and T. Stuart (2006a). Gender Differences in Patenting in the Academic Life Sciences. *Science 313*(5787), 665–667.

Ding, W. W., F. Murray, and T. E. Stuart (2006b). Gender differences in patenting in the academic life sciences. *Science 313*(5787), 665–667.

Dixit, A. (2003). Trade expansion and Contract Enforcement. *Journal of Political Economy 111*(6), 1293–1317.

Dohmen, T. and A. Falk (2011, April). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review 101*(2), 556–90.

Easley, D. and J. Kleinberg (2010). *Networks, Crowds, and Markets.* Cambridge Univ Press.

Eckel, C. and P. Grossman (2008). Men, Women and Risk Aversion: Experimental Evidence. *Handbook of experimental economics results 1*, 1061–1073.

Eder, D. and M. Hallinan (1978). Sex Differences in Children's Friendships. *American Sociological Review*, 237–250.

Equality and H.R.Commission (2009). Financial services inquiry: Sex discrimination and gender pay gap report of the equality and human rights commission. Technical report.

Ferriani, S., G. Cattani, and C. Baden-Fuller (2009). The relational antecedents of project-entrepreneurship: Network centrality, team composition and project performance. *Research Policy 10*(38), 1545–1558.

Fischer, C. and S. Oliker (1983). A Research Note on Friendship, Gender, and the Life Cycle. *Social Forces 62*(1), 124–133.

Flabbi, L. (2012). Gender Differences in Education, Career Choices and Labor Market Outcomes on a Sample of OECD Countries. *World Development Report, Gender Equality and Development*.

Forret, M. L. and T. W. Dougherty (2004). Networking behaviors and career outcomes: differences for men and women? *Journal of Organizational Behavior 25*(3), 419–437.

Fortin, N. and M. Huberman (2002). Occupational Gender Segregation and Women's Wages in Canada: an Historical Perspective. *Canadian Public Policy/Analyse de Politiques*, 11–39.

Fortin, N. M. (2008). The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People. *Journal of Human Resources 43*(4).

Franzen, A. and D. Hangartner (2006). Social networks and Labour Market Outcomes: The Non-Monetary Benefits of Social Capital. *European Sociological Review 22*(4), 353–368.

Gabbay, S. M. and E. W. Zuckerman (1998). Social capital and opportunity in corporate r&d: The contingent effect of contact density on mobility expectations. *Social Science Research 27*(2), 189 – 217.

Gerhart, B. (1990). Gender Differences in Current and Starting Salaries: The Role of Performance, College Major and Job Title. *Industrial and Labor Relations Review 43*(4), 418–433.

Gerhart, B. and S. Rynes (1991). Determinants and Consequences of Salary Negotiations by Male and Female MBA Graduates. *Journal of Applied Psychology 76*, 256–262.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 1360–1380.

Heider, F. (1946). Attitudes and Cognitive Organization. *The Journal of Psychology 21*(1), 107–112.

Holmstrom, B. (1982). Moral Hazard in Teams. *The Bell Journal of Economics*, 324–340.

Huffman, M. and L. Torres (2002). It's Not Only "Who You Know" that Matters. *Gender & Society 16*(6), 793–813.

Hunt, J., J.-P. Garant, H. Herman, and D. J. Munroe (2012, March). Why Don't Women Patent? NBER Working Papers 17888, National Bureau of Economic Research, Inc.

Ibarra, H. (1997). Paving an Alternative Route: Gender Differences in Managerial Networks. *Social Psychology Quarterly*, 91–102.

Ioannides, Y. M. and L. D. Loury (2004). Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature 42*(4), 1056–1093.

Kandel, E. and E. Lazear (1992). Peer Pressure and Partnerships. *Journal of Political Economy*, 801–817.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009). Trust and Social Collateral. *The Quarterly Journal of Economics 124*(3), 1307–1361.

Kunze, A. (2003). Gender Differences in Entry Wages and Early Career Wages. *Annales d'Economie et de Statistique* (71-72), 245–265.

Kürtösi, Z. (2008). *Differences in Female and Male Social Networks in a Work Setting.* Ph. D. thesis, Sociology Department Budapest.

Lalanne, M. and P. Seabright (2011). *The Old Boy Network: Gender Differences in the Impact of Social Networks on Remuneration in Top Executive Jobs.* Centre for Economic Policy Research.

Lutter, M. (2012). Anstieg oder ausgleich? die multiplikative wirkung sozialer ungleichheiten auf dem arbeitsmarkt für filmschauspieler. *Zeitschrift für Soziologie* (41), 435–457.

Lutter, M. (2013). Is there a closure penalty? cohesive network structures, diversity, and gender inequalities in career advancement. *MPIfG Discussion Paper* (13/9).

Marmaros, D. and B. Sacerdote (2002). Peer and Social Networks in Job Search. *European Economic Review 46*(4), 870–879.

Marsden, P. (1987). Core Discussion Networks of Americans. *American Sociological Review*, 122–131.

Marsden, P. (2005). Recent Developments in Network Measurement. *Models and Methods in Social Network Analysis 8*, 30.

Martell, R., D. Lande, and C. Emrich (1996). Male-Female Differences: A Computer Simulation. *American Psychologist 51*, 157–158.

Napari, S. (2006, August). The Early Career Gender Wage Gap. CEP Discussion Papers dp0738, Centre for Economic Performance, LSE.

Ogden, S. M., D. McTavish, and L. McKean (2006). Clearing the way for gender balance in the management of the uk financial services industry: Enablers and barriers. *Women In Management Review 21*(1), 40–53.

Podolny, J. and J. Baron (1997). Resources and Relationships: Social Networks and Mobility in the Workplace. *American Sociological Review*, 673–693.

Polachek, S. W. (1981). Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure. *The Review of Economics and Statistics 63*(1), 60–69.

Ports, M. H. (1993). Trends in Job Search Methods, 1970-92. *Monthly Labor Review 116*(10), 63–67.

Putnam, R. (2000). *Bowling Alone: America's Declining Social Capital.* Simon and Schuster.

Reports, B. (2013). Highlights of women's earnings in 2012. Technical Report 1045, U.S. Bureau of Labor Statistics, http://www.bls.gov/cps/cpswom2012.pdf.

Rosenbaum, J. E., S. DeLuca, S. R. Miller, and K. Roy (1999). Pathways into Work: Short-and Long-Term Effects of Personal and Institutional Ties. *Sociology of Education*, 179–196.

Silva, C., N. M. Carter, and A. Beninger (2012). Good intentions, imperfect execution? women get fewer of the "hot jobs" needed to advance. *Catalyst*.

Smith, S. S. (2000). Mobilizing Social Resources: Race, Ethnic, and Gender Differences in Social Capital and Persisting Wage Inequalities. *The Sociological Quarterly 41*(4), 509–537.

Sutter, M. and D. Rützler (2010). Gender Differences in Competition Emerge Early in Life.

Tattersall, A. and C. Keogh (2006). Women, Networking and the Impact on Pay, Position and Status in the UK ICT Industry. unpublished.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review 80*(1), 234–248.

Walker, G., B. Kogut, and W. Shan (1997). Social capital, structural holes and the formation of an industry network. *Organization Science 8*(2), 109–125.