Supply and Demand in a Two-Sector Matching Model*

Paweł Gola†

This version: November 18, 2018

Abstract

This paper merges workers' self-selection (across sectors) with an assignment model (within sectors). First, I show that as two sectors (manufacturing and services) start using skill sets that are more similar, the overall supply of skill falls. If said fall is not biased toward either sector, the distribution of wages becomes more polarized. Second, I show that if a manufacturing-specific technological improvement favors both high-skilled workers and high-productivity firms, it increases the number of productive workers (and output produced) in manufacturing and might rise wage inequality in both sectors. If it favors low-skilled workers or low-productivity firms, manufacturing might contract.

Keywords: two-sector matching, self-selection, imperfect substitution, demand for skill, supply of skill, wage inequality, wage polarization

JEL codes: C78, D31, J24, J31

*I am very grateful to my doctoral advisor, Ian Jewitt, for his support and guidance. I would like to thank Peter Eso for feedback and encouragement, as well as Matej Bajgar, Leonie Baumann, Juan Block, River Chen, Paweł Dziewulski, Aytek Erdil, Edoardo Gallo, Axel Gottfries, Claudia Herresthal, Ed Hopkins, Joshua Lanier, Karan Nagpal, Bartosz Redlicki, Mikhail Safronov, Federico Torracchi, and Łukasz Woźny for detailed comments on this and earlier drafts. I would also like to thank Michał Burzyński, Vasco Carvalho, Meg Meyer, Climent Quintana-Domeque, Larry Samuelson, Coen Teulings, and seminar participants at Oxford, Cambridge, the University of Luxembourg, the University of Nottingham, the BI Norwegian Business School, the 2014 Ce^2 Workshop, the 2014 WIEM Conference, the 2015 RES Conference, the 2016 SSES Congress, and the 2016 North American Summer Meeting of the Econometric Society for helpful discussions and suggestions. The paper benefited greatly from insightful comments by James J. Heckman and four anonymous referees. All errors are mine.

†BI Norwegian Business School. E-mail: pawel.gola@bi.no

1 Introduction

Sectors differ in the technologies they use and, consequently, in the skill sets they demand. Within sectors, workers differ in the ability to perform their jobs, whereas firms differ in their organizational structures and production processes. In this paper, I show how this inherent heterogeneity in technology and skill affects wages, profits, and output when workers are free to choose the sector they work in.

To address this, I propose a new model that introduces assignment of workers to firms (within sectors) in the vein of Becker (1973) and Sattinger (1979) into a model of self-selection (across sectors) in the vein of Roy (1951). Workers' self-selection implies that within-sector distributions of skill are determined endogenously and depend, among other factors, on the technology used in each sector. The assignment of workers to firms introduces imperfect substitution and (in an extension in which firms' entry is endogenous) complementarity between workers with different skill levels.[1] This is in contrast to existing models of self-selection which, starting with Heckman and Sedlacek (1985), assume that workers are perfect substitutes. However, perfect substitution is inconsistent with empirical evidence (Katz and Murphy, 1992).[2]

I introduce a model with two sectors, manufacturing and services, and derive sharp monotone comparative statics results. Firstly, I show that if the two sectors start demanding skill sets that are more similar than before (e.g., if each sector starts using a dimension of skill that was previously specific to the other sector), the overall supply of skill in the economy *de facto* declines. As a result, the overall distribution of wages becomes more polarized as long as the decline affects wage functions in both sectors symmetrically, that is, if it is *unbiased*.[3] This stands in stark contrast to the literature on task-biased technological change (TBTC) (Costinot and Vogel, 2010; Acemoglu and Autor, 2011; Lindenlaub, 2017), in which wage polarization increases only if the change in demand or technology is biased toward the task/sector that employs predominantly medium-earning workers.

Secondly, I show that a technological improvement in one of the sectors (say manufac-

[1] To see this, note that in Sattinger's (1979) model, firms and workers match in a fixed proportion, that is, every firm hires the same, exogenously determined number of workers. Hence a firm cannot costlessly replace a skilled worker with two or more less skilled ones. If firms' entry is exogenous, neither can sectors, because highly productive firms are scarce. For example, suppose that two high-skilled workers leave manufacturing and that one of them is replaced by $x > 1$ workers of low skill. To replace the second high-skilled worker, more than x low-skilled workers are needed, as the second worker's replacements match with less productive firms than the workers replacing the first worker. The reasons why some workers will be complements if firms' entry is endogenous are explained later in this introduction and in Section 5.1.2.

[2] Perfect substitution of workers implies that if some high-skilled workers leave manufacturing, wages increase by the same proportion for low- and high-skilled workers (the "proportionality hypothesis"). Katz and Murphy (1992) provide evidence that changes in the relative supply of high- and low-skilled workers significantly affect their relative wages.

[3] The effect of the decline could be symmetric even if just one sector started requiring different skill sets than before.

turing) does not necessarily result in a larger number of high-skilled workers joining that sector. This is guaranteed only if the increase in output is greater for high-skilled workers and (with endogenous entry) high-productivity firms. If this is the case, it creates a force for an increase in wage inequality, not just in manufacturing but also in services.[4] If, however, the increase in output favors low-skilled workers or low-productivity firms, then some high-skilled workers might leave manufacturing for services. As a consequence of that, output in manufacturing might even contract.

These results differ qualitatively from comparative statics that arise in existing sorting models, which further underscores the importance of properly accounting for within-sector substitution. It is worth noting that I am able to derive the results in the absence of functional form assumptions and despite the fact that wages in my model depend on the entire distribution of skills in a sector (which is due to workers' imperfect substitution/complementarity). I accomplish this by leveraging the well-known relationship between the distribution of skill and wages that arises under positive and assortative matching.

Overview. The paper is organized in four main sections.

Section 2 sets up the baseline model and characterizes the unique equilibrium. In the model, there is a continuum of heterogeneous workers and two sectors, manufacturing (M) and services (S), each populated by a continuum of heterogeneous firms.[5] Each worker is endowed with a vector of basic skills $\mathbf{x}$, and each firm in sector $i \in \{M, S\}$ is endowed with a scalar productivity h_i. A match between a firm and a worker produces some surplus, which—in the absence of other inputs—can be interpreted as its output expressed in monetary terms.[6] The surplus produced by a match is determined by a surplus function that depends on the sector, the firm's productivity, and the worker's skill. In particular, the two sectors use workers' skills in different proportions. In the competitive equilibrium, both workers and firms take wages as given, each worker sorts into the sector that pays a higher wage for her skill endowment, and each firm hires at most one worker to maximize profits. Wages are set to clear the market.

To make the model tractable, I assume that in each sector the vector of basic skills $\mathbf{x}$ can be aggregated into a univariate index v_i in such a way that the vector (v_M, v_S) contains all the relevant information about $\mathbf{x}$.[7] The indices v_M, v_S are normalized to

[4]Wage inequality is certain to increase in both sectors if the number of workers and firms in each sector is unchanged; this will occur, for example, if entry is exogenous and there are fewer firms than people looking for work. An expansion or contraction in the number of workers has an ambiguous effect on wage inequality and can counteract the increase in some cases.

[5]This means that the distribution of firms' productivity in each sector is exogenously determined, although the productivity distribution among firms that end up matched is not. In the baseline model the measure of firms in each sector is also exogenous but this is relaxed in Section 5.

[6]Alternatively, this model can be seen as a reduced form of a model in which there are further inputs and there exists an explicit output function. In such a case, the surplus is the revenue less the cost of non-labor inputs for an optimal choice of non-labor inputs.

[7]This is a two-sector version of the separability assumption from Chiappori, Oreffice, and Quintana-

have standard uniform marginal distributions and hence are referred to as *relative skills*.[8] I assume, as is standard in the matching literature, that the surplus is increasing in productivity, strictly increasing in relative skill, and supermodular in productivity and relative skill.

Section 3 investigates whether an increase in the interdependence of relative skills can make the overall wage distribution more *polarized*, that is, decrease wage inequality in the left-tail while increasing the difference between the highest and lowest wages.[9] I start by observing that higher interdependence reduces the gains from self-selection, as more workers become similarly skilled in both dimensions. This leads to a downward shift in the production-possibility frontier of the economy and acts as if the overall supply of skill decreased in a sense equivalent to first-order stochastic dominance (FOSD). Any change in the wage distribution can be decomposed into the *wage effect*, in which wages change but the skill distribution remains constant, and the *composition effect*, in which the skill distribution changes but wages remain constant. The wage effect of increased interdependence resembles task-biased technological change, in that it can increase wage polarization only if it is *biased*, that is, if wages increase in one sector more than in the other. However, the composition effect always results in a distribution of wages that is worse in the FOSD sense and thus increases wage polarization unambiguously: The highest and lowest wages remain unchanged but the wages corresponding to all other quantiles decrease.

If the increase in interdependence is *unbiased*, that is, if the change in wages is identical across sectors, then under mild conditions wage polarization increases unambiguously.[10] The fall in left-tail inequality, which is necessary for an increase in polarization, happens purely due to the composition effect: In the case of an unbiased increase in interdependence, the wage effect increases wage inequality in both tails. In particular, the increase in interdependence is unbiased in the symmetric case (i.e., when the two sectors differ only in the dimension of relative skill they use), which is formally equivalent to the single-sector assignment model of Sattinger (1979).

Finally, I argue that it is plausible that an increase in relative skill interdependence contributed to the increase in wage polarization that was recorded in the 1990s in the

Domeque (2011), and it makes the relative skills an analogue of the tasks from Heckman and Sedlacek (1985).

[8]For example, consider a worker with relative skills vector $(0.25, 0.5)$. This means that 25% of the population is more skilled than this worker in manufacturing and that half of the population is more skilled than she is in services.

[9]I use the concordance ordering (Scarsini, 1984) as the notion of interdependence. Keeping marginal distributions unchanged, two random variables become more concordant if their joint distribution shifts upward. For jointly normally distributed variables, concordance increases if and only if correlation increases.

[10]The critical (in the sense used by Rodrik (2015)) condition needed for an increase in polarization is that the change in skill requirements does not change the support of skills among employed workers. This must be the case if, for example, the number of firms is greater than the number of workers.

US.[11] This is because, as first noted by Gould (2002), the interdependence of relative skills depends not only on the distribution of basic skills but also on the similarity of *skill content* of the sectors, that is, on the degree to which manufacturing and services use the same basic skills in the production process. Therefore, the interdependence of relative skills can change as a result of technological advances.[12] There exists strong evidence that the skill content of certain occupations changed in the decades preceding the 1990s (Autor, Levy, and Murnane, 2003; Spitz-Oener, 2006) and, in fact, became more similar across sectors (Gould, 2002). A likely reason for these changes is the process of computerization, which resulted in very different sectors coveting the same skills. From that point of view, an increase in relative skill interdependence constitutes an alternative mechanism through which routinization, that is, the change in task and skill content of occupations caused by computerization (Autor et al., 2003), could have caused an increase in wage polarization.

Section 4 focuses on changes in technology and skill distribution that increase (decrease) the *difference* in surplus produced by manufacturing workers of any two relative skill levels. I call this an increase (decrease) in the *vertical differentiation* of manufacturing workers. Let me first use an example to illustrate the importance of vertical differentiation for sorting and its implications for output. Subsequently, I will discuss the implications for wage inequality.

Suppose there exist fewer firms than workers (jobs are *scarce*) and the vector of basic skills has three components: x_1 is manufacturing specific, x_2 is services specific and x_3 is a general-purpose skill. The surplus in manufacturing is strictly increasing in the manufacturing-specific and general-purpose skills but does not depend on the services-specific skill; and analogously for services. Suppose that the government wants to boost the total output produced in manufacturing and, to this end, decides to invest in population-wide training in x_1 in a way that shifts its distribution upward. This increases the supply of a basic skill that is used *only* in manufacturing while leaving the supply of basic skills used in services unchanged. As a result, the output produced by a worker of any relative skill (v_M, v_S) increases for matches with manufacturing firms but remains the same for matches with firms in services. In a model with perfect substitutes, this would necessarily lead to an expansion in manufacturing output. With imperfect substitution, however, it can easily backfire.

[11] See, for example, Figure 7 in Acemoglu and Autor (2011).

[12] Of course, a technological change would likely affect not only the relative skill interdependence but also the surplus produced by a worker of any given relative skill level. In fact, for a technological change to constitute an advance, it must also increase the surplus produced by workers of given skill, as otherwise an increase in interdependence would, on its own, decrease total output. This is not a problem, however, as the link between increased interdependence and polarization implies that technological changes that would have no effect on polarization in a one-dimensional world will increase polarization in my model. In particular, a technological advance that affects both sectors symmetrically, raises surplus multiplicatively, and increases interdependence would result in both higher output and more polarized wages.

Such a fall in manufacturing output can happen if the government's investment makes manufacturing workers less vertically differentiated.[13] How does this change firms' hiring choices in the short term, before wages have time to adjust? As the difference in *surplus* produced by different workers decreases but the difference in *wages* does not, high-skilled workers become relatively overpaid. As a result, all firms want to hire a worker of lower relative skill than in the old equilibrium. Further, because I assume that the surplus produced in any match always covers the reservation payoffs of both parties, scarcity of jobs implies that all firms want to hire some worker. As a result, more high-skilled workers and fewer low-skilled workers want to work in manufacturing than are demanded by manufacturing firms. In equilibrium, therefore, wages rise for workers of low manufacturing skill and fall for workers with high manufacturing skill, forcing some marginal high-skilled workers out of manufacturing but drawing in additional low-skilled workers from services. Thus in the sense of first-order stochastic dominance the distribution of relative skill improves in manufacturing and worsens in services. Overall, services expand, whereas the impact on the total output produced in manufacturing is ambiguous.

Further, I show that a shock that directly affects only manufacturing, might well increase both absolute and relative wage inequality in services.[14] Several attempts have been made in recent years to infer the causes of increases in wage inequality by using cross-sector comparisons (e.g., Bakija, Cole, and Heim, 2010; Kaplan and Rauh, 2010, 2013). Ideally, such analyses should take transmission of wage inequality into account. For example, Kaplan and Rauh (2013) concluded that the fact that "the increase in pay at the highest income levels is broad based" is more consistent with the superstar and scale-effects (Rosen, 1981) explanations of rising wage inequality than with an increase in managerial power or a weakening of social norms. If an increase in wage inequality that originates in a narrow subset of sectors could spread across the economy, however, then such conclusions would seem premature. This is less of a concern in models in which workers are perfect substitutes, where any change in the sectoral supply of skills will have the same relative impact on the wages of all workers. With imperfect substitution of skills, however, this is not the case. If jobs are scarce, then an increase in vertical differentiation in manufacturing draws in high-skilled workers from services. The latter increases the wages of high-skilled services workers compared to those of low-skilled workers. As a result, wage inequality increases in both sectors.[15]

Although there are no strategic interactions in my model, my results do suggest that

[13]This would be the case if, for example, basic surplus is multiplicatively separable in the basic skills and concave in x_1, and the distribution of x_1 is shifted upward by a constant. Intuitively, think of a situation in which the government's policy provides more resources for teaching of all students but focuses on the less able ones.

[14]It should be noted that both types of increases would be stronger in the directly affected sector.

[15]In particular, the range of wages increases in both sectors, and the top wages in manufacturing rise proportionately more than the lowest wages. In services, wages increase proportionately more for high-skilled workers than for low-skilled workers as long as reservation wages are not too small.

there might be strategic reasons for increases in wage inequality. I demonstrate this in Section 4.1.4, where I provide an example in which jobs are scarce and the two sectors (or regions in which they are concentrated) can make an investment in infrastructure which increases the vertical differentiation of their workers. Rising vertical differentiation increases wage inequality and creates a negative sorting externality for the other sector. Therefore, both sectors are willing to make the investment even in cases where this is not socially optimal.[16] In equilibrium, both sectors over-invest, which cancels out the sorting effect. Thus the sectors end up with lower output (net of the cost of the investment) and higher wage inequality.

Section 5 endogenizes firms' entry decisions. I follow Hopenhayn (1992) and Melitz (2003) in assuming that there exists an unlimited number of *ex ante* homogeneous firms. To enter a sector, a firm pays a sector-specific cost of entry and draws productivity from an exogenous distribution. Section 5.1 shows that the equilibrium of the extended model exists and that it is unique and efficient. Because the expected profits are equal to the cost of entry in equilibrium, workers of very high and very low skill are Hicks complements (rather than imperfect substitutes), that is, the arrival of additional low-skilled workers increases the wages of high-skilled workers.[17] This is also a feature of the Costrell and Loury (2004) assignment model with hierarchical firms. In fact, the comparative statics for the extended model are derived only for specifications that are equivalent to a two-sector extension of the Costrell and Loury (2004) model.[18]

Section 5.2 revisits the link between the interdependence of relative skills and wage polarization. The results are qualitatively very similar to the baseline model: The composition effect increases wage polarization unambiguously, whereas the wage effect has an ambiguous impact on polarization. In the symmetric case, which is formally equivalent to the single-sector Costrell and Loury (2004) model, the composition effect dominates and wage polarization increases.[19] This demonstrates that the insights from Section 3 hold even if workers of different skill levels are complements.

Section 5.3 generalizes the results of Section 4. In the extended model, sorting depends also on *firms'* vertical differentiation, which is defined analogously to workers' vertical differentiation.[20] In particular, if firms become less vertically differentiated, then man-

[16] That is, if it decreases the sum of the surpluses produced in the economy, net of the cost of the investment.

[17] If additional low-skilled workers join manufacturing, the wages of incumbent low-skilled workers must fall. However, wages of some workers must increase, as otherwise all firms would pay lower wages and expected profits would increase. As workers of very high skill do not compete with workers of very low skill, additional firm entry improves their matches and they are the ones whose wages increase.

[18] Proposition 13 (in Section 5.3) is an exception, as it holds in full generality.

[19] To be precise, polarization increases in absolute terms. That is, the increase in the lowest wages is greater than in wages in the interior of the distribution in absolute terms. An increase in relative terms is ensured only if wages were high enough originally.

[20] That is, firms become more vertically differentiated if the difference in surplus produced by firms of any two productivity levels increases.

ufacturing's output can contract even if all matches produce more output and workers' differentiation increases. The intuition behind this perverse output effect is different than the intuition behind the perverse output effects of the baseline model and is provided in Section 5.3.1, together with a simple example.

In the extended model, the impact of an increase in the vertical differentiation of manufacturing workers on wage inequality in services can be decomposed into two effects. If we hold the numbers of firms in both sectors constant, the impact on inequality is the same as in the baseline model (Section 4); I call this the *baseline effect*. However, if—in addition to the increase in workers' differentiation—both the vertical differentiation of firms and the surplus levels increase, then the equilibrium numbers of firms and workers will increase in manufacturing and fall in services. This gives rise to a new *entry effect*, which affects wage inequality in services ambiguously. In particular, if the contraction in services is sufficiently large, then the entry effect decreases wage inequality in services, thus counteracting the increase caused by the baseline effect.

Section 6 reviews further the related literature and places my main contributions into it. Section 7 concludes. Proofs can be found in the Appendix. The Online Appendix provides additional results. In particular, Online Appendix OA.6 develops a dynamic, overlapping-generations version of the extended model and shows that the steady-state equilibrium of the dynamic model corresponds one-to-one to the equilibrium of the static model.

2 The Model

In this section I set up the model, characterize the equilibrium, and prove its uniqueness.

2.1 The Setup

There are two sectors—manufacturing and services—and two populations: workers and firms.

Workers There is a unit measure of workers, each endowed with a vector of basic skills $\mathbf{x} = (x_1, x_2 \dots x_N) \in I_{\mathbf{x}} \subset \mathbf{R}^N$. Denote the distribution of $\mathbf{x}$ by F. Workers can either work for a firm and receive a market wage or remain unemployed and receive a reservation wage (normalized to 0).

Firms There is a measure R of firms, each endowed with vector $(z, i) \in I_z \times \{M, S\}$, where z denotes the firm's productivity, $i \in \{M, S\}$ denotes the sector in which the firm operates (manufacturing or services) and $I_z \subset \mathbf{R}$. The (exogenous) distribution of (z, i) is denoted by H_Z, whereas the measure of firms in sector i is denoted by $R_i > 0$, with

$R_M + R_S = R$.[21] Each firm hires at most one worker. A worker-firm pair produces surplus according to the basic surplus function $\Pi : I_{\mathbf{x}} \times I_z \times \{M, S\} \to \mathbf{R}_{\geq \mathbf{0}}$. For example, if a manufacturing firm with productivity h hires a worker with skill $\mathbf{x}$, they produce surplus $\Pi(\mathbf{x}, z, M)$. The fact that the surplus function depends on the sector means that workers' skill and firms' productivity might be used differently in each sector. If a firm does not hire a worker, it receives a reservation profit normalized to 0.

2.1.1 Assumptions

Following Heckman and Sedlacek (1985), I assume that the basic surplus functions in each sector are *separable* in basic skills and productivity.[22]

Assumption 1 (Properties of the Surplus)**.** In both sectors, basic surplus Π is separable in skills and productivity, that is, for each sector there exist mappings $v_i : I_{\mathbf{x}} \to [0, 1]$ (*relative skill*), $h_i : I_z \to [0, 1]$ (*relative productivity*) and $\pi_i : [0, 1]^2 \to \mathbf{R}_{\geq \mathbf{0}}$ (*the reduced surplus*) such that the following hold:

A1.1 *Separability*: $\pi_i(v_i(\mathbf{x}), h_i(z)) = \Pi(\mathbf{x}, z, i)$

A1.2 *Differentiability*: π_i is twice continuously differentiable

A1.3 *Increasing surplus*: $\frac{\partial}{\partial v_i}\pi_i > 0$, $\frac{\partial}{\partial h_i}\pi_i \geq 0$

A1.4 *Supermodular surplus*: $\frac{\partial^2}{\partial v_i \partial h_i}\pi_i \geq 0$

Separability means that the impact of a basic skill $\mathbf{x}$ on the surplus in sector i is fully captured by the one-dimensional index $v_i = v_i(\mathbf{x})$. Together with separability, the increasingness assumption (A1.3) implies that workers and firms can be totally ordered within each sector with respect to the surplus they produce.[23] Supermodularity (A1.4) implies that highly productive firms benefit more from hiring high-skilled workers. This ensures that within-sector matching is positive and assortative and makes the model tractable. The comparative statics results would be unchanged if surplus functions were submodular.[24]

Assumption 2 (Properties of the Copula)**.** The distributions F and H_Z and the surplus Π are such that the joint distribution C of relative skill $(v_M, v_S) \in [0, 1]^2$ and the distributions H_i of relative productivity $h_i \in [0, 1]$ have the following properties:

[21] The measure of firms in each sector is endogenized in Section 5.

[22] In the context of matching, Chiappori, Oreffice, and Quintana-Domeque (2012) assume that there exists a single one-dimensional index that summarizes agents' preferences. My assumption is weaker, in that I only require firms' "preferences" over workers to be the same within a sector but allow them to differ across sectors.

[23] That is, if one manufacturing firm produces more surplus by hiring worker $\mathbf{x}'$ than worker $\mathbf{x}$, then all manufacturing firms do; and analogously for workers.

[24] Whether my results would hold even for surpluses that are neither super- nor submodular is an open question but it seems likely, given that the two extreme cases yield the same results.

A2.1 *Differentiability*: They are twice continuously differentiable.

A2.2 *Full support*: They have strictly positive, finite density on their respective supports.[25]

The full support assumption allows me to normalize the indices v_M, v_S and h_M, h_S in such a way that their marginal distributions are standard uniform, using the fact that Assumption 1 defines them only up to a monotone transformation.[26] This is why I refer to them as relative skills and productivities, respectively. Note that because the marginal distributions of v_M, v_S are standard uniform, C is a copula (Sklar, 1959). Further, the full support assumption precludes perfect (positive or negative) correlation but otherwise allows for very general dependence structures. For example, for C belonging to the family of Gaussian copulas, Assumption 2 allows for any correlation parameter $\rho \in (-1, 1)$.

The formulation of the model in terms of uniformly distributed relative skills will be referred to as *the canonical formulation.*[27] Apart from examples and applications, I will be working with the canonical formulation exclusively. For that reason, most of the time I will refer to v_i, h_i, and π_i as skill, productivity and surplus, respectively, dropping the adjectives relative and reduced.

Assumption 3 (Non-Degenerate Solutions)**.** For any $i, j \in \{M, S\}$ with $j \neq i$, either $R_i < 1$ or $\pi_i(0, 1 - \frac{1}{R^i}) < \pi_j(1, 1)$.

This assumption is necessary and sufficient for all equilibria of this model to be non-degenerate, so that a positive measure of workers is employed in each sector.

2.1.2 Supply, Demand, and Equilibrium

Supply of Relative Skills A worker with skill (v_M, v_S) who joins sector i receives wage $w_i(v_i)$, where $w_i : [0, 1] \to \mathbf{R}$. Workers sort into the sector that maximizes their wages. A worker with skill (v_M, v_S) joins manufacturing if and only if

$$w_M(v_M) \geq \max\{w_S(v_S), 0\}, \tag{1}$$

joins services if and only if

$$w_S(v_S) > \max\{w_M(v_M), 0\}, \tag{2}$$

[25] For the joint distribution C, it suffices for both conditions to hold just on $(0, 1)^2$. In particular, all results hold for the Gaussian copula.

[26] To see that this is a normalization, consider any v'_i and π'_i that meet Assumptions 1 and 2. Denote the marginal distribution of V'_i by F_i. Then let $V_i = F_i(V'_i)$ which ensures that the marginal distribution is standard uniform; this gives $\pi_i(v_i, h_i) = \pi'_i(F_i^{-1}(v_i), h_i)$.

[27] The canonical formulation defines equivalence classes: Any two models with the same canonical formulation will give rise to the same outcomes (i.e., wage and output distributions).

and remains unemployed otherwise.[28]

The sectoral *supply of relative skill* of level t, $S_i(t)$, is defined cumulatively, as the measure of workers with sector-specific skill of at least t who join sector i, for given wage functions w_M, w_S:

$$S_M(t) = \Pr\Big(V_M \geq t,\ w_M(V_M) \geq w_S(V_S),\ w_M(V_M) \geq 0\Big), \tag{3}$$

$$S_S(t) = \Pr\Big(V_S \geq t,\ w_M(V_M) < w_S(V_S),\ w_S(V_S) > 0\Big). \tag{4}$$

Note that $S_i(0)$ gives us the total measure of workers who joined sector i. Further, together with the joint distribution C, either of S_M, S_S determines the other.[29]

Demand for Relative Skills The demand for skills in each sector is determined by the firms' hiring decisions, which in turn are driven by profit maximization, with firms taking the wage function as given. Firm h_i earns profit $r_i(h_i)$ and hires worker $v_i^*(h_i)$, where $r_i : [0,1] \to \mathrm{R}$ and $v_i^* : [0,1] \to [0,1]$, with

$$r_i(h_i) = \max_{v \in [0,1]} \ \pi_i(v, h_i) - w_i(v), \tag{5}$$

$$v_i^*(h_i) \in \operatorname*{arg\,max}_{v \in [0,1]} \ \pi_i(v, h_i) - w_i(v). \tag{6}$$

Demand for skills is defined analogously to skill supply. The sectoral *demand for relative skill* of level t, $D_i(t)$, is equal to the measure of sector i firms that hire workers with sector-specific skill of at least t, for a given wage function w_i:

$$D_i(t) = R_i \Pr\Big(v_i^*(H_i) \geq t,\ r_i(H_i) \geq 0\Big). \tag{7}$$

This definition assumes that profits are strictly increasing in productivity, which is the case as long as $\frac{\partial}{\partial h_i}\pi_i > 0$. A more general definition, which holds even when surplus does not depend on productivity, is provided in Appendix A.[30]

The Competitive Equilibrium I focus on the competitive equilibrium, which is defined as follows.

Definition 1 (Equilibrium)**.** An equilibrium is characterized by:
(i) two *sectoral relative skill supply functions* $S_i : [0,1] \to [0,1]$, consistent with workers' sorting decisions and given by Equations (3) and (4);

[28] Of course, a worker for whom $w_S(v_S) = w_M(v_M)$ is indifferent and could join either sector; however, the set of all such workers will be of zero measure in equilibrium, hence we can assign all of them to manufacturing without loss of generality.

[29] The exact relation between them will become clear later (see Figure 1 and Equation (15)).

[30] Additionally, my definition of v_i^* implies that it is a function, which excludes the possibility of impure matchings. This greatly simplifies notation and is without loss of generality, because all matchings—even impure ones—will result in the same wage functions.

(ii) two *sectoral relative skill demand functions* $D_i : [0,1] \to [0,1]$, consistent with firms' profit maximization and given by Equation (7);
(iii) two *sectoral wage functions* $w_i : [0,1] \to \mathbf{R}$, which clear the markets: $S_i(t) = D_i(t)$ for $i \in \{M, S\}$ and all $t \in [0,1]$.

It is worth noting that because this model is an assignment game, the competitive equilibrium coincides with the core (Gretsky, Ostroy, and Zame, 1992).

2.2 Characterization Strategy

To characterize the competitive equilibrium, I employ a two-step strategy. In the first step, I treat sectoral supply functions S_i as given and derive the wage functions that equate supply with demand in each sector. This is very similar to the problem first solved by Sattinger (1979). In the second step, I use those wages to find the sectoral supply functions in a manner somewhat similar to the way they are found in Roy's model.

2.2.1 First Step

In this part, I treat the sectoral supply functions as given and find the wage functions for which demand will equal supply. Let the *critical skill* v_i^c be the relative skill of the least skilled worker who joins sector i:

$$v_i^c = \sup\{v \in [0,1] : S_i(v) = S_i(0)\}. \tag{8}$$

In equilibrium, $S_i(0)$ cannot be greater than $R_i = D_i(0)$, as otherwise the market would never clear; I will restrict attention to supply functions that meet this condition.

Proposition 1. In equilibrium, the wage functions are such that the following hold:

$$w_i(v_i) = \int_{v_i^c}^{v_i} \frac{\partial}{\partial v_i} \pi_i\Big(v, 1 - \frac{S_i(v)}{R_i}\Big) \mathrm{d}v + w_i(v_i^c) \qquad \text{for } v_i \geq v_i^c \tag{9}$$

$$w_i(v_i) \geq w_i(v_i^c) + \pi_i\Big(v_i, 1 - \frac{S_i(v_i^c)}{R_i}\Big) - \pi_i\Big(v_i^c, 1 - \frac{S_i(v_i^c)}{R_i}\Big) \qquad \text{for } v_i < v_i^c, \tag{10}$$

where $w_i(v_i^c) \in [0, \pi_i(v_i^c, 1 - \frac{S_i(0)}{R_i})]$. If, however, $S_i(0) < R_i$, then $w_i(v_i^c) = \pi_i(v_i^c, 1 - \frac{S_i(0)}{R_i})$.

It is well-known that if the surplus function is supermodular, workers and firms match positively and assortatively; that is, the most productive firm matches with the worker of highest relative skill, the second most productive firm matches with the second most skilled worker, and so on. Accordingly, my Proposition 1 is essentially a restatement of Sattinger's (1979) famous result on wage functions under positive and assortative matching.[31] The reason why the relationship between wages and the supply of skills needs to be of the form specified in Proposition 1 can be easily understood from the first-order condition of the firm's hiring decision:

$$\frac{\partial}{\partial v_i} w_i(v_i^*(h_i)) = \frac{\partial}{\partial v_i}\pi_i\Big(v_i^*(h_i), 1 - \frac{S_i(v_i^*(h_i))}{R_i}\Big).$$

Thus the difference in wages paid to workers of marginally different skill is equal to the difference in the surplus they produce. The value of this marginal surplus, however, depends on the firm the worker is matched with. This depends in turn on the supply of relative skills in that sector: The fewer high-skilled workers available, the better the match that can be secured by any worker. The wage paid to the worker with critical skill v_i^c depends on whether workers are in short supply in that sector. If this is the case, then competition drives the profits of the least productive matched firm to 0.

2.2.2 Second Step

In the second step, I treat the sectoral wage functions w_M, w_S as given and derive the corresponding sectoral supply functions.[32] Note that by Proposition 1, wages are strictly increasing in each sector for $v_i \geq v_i^c$. This has two important implications for sorting. Firstly, any worker with relative services skill $v_S > v_S^c$ can earn a strictly positive wage and therefore will never choose to remain unemployed. Secondly, for any such worker there will exist a cut-off value $\psi(v_S)$ of the relative manufacturing skill such that she will strictly prefer to join services if $v_M > \psi(v_S)$ and strictly prefer to join manufacturing if $v_M < \psi(v_S)$. Therefore, the sorting of workers to sectors can easily be expressed by the means of the critical skills v_M^c, v_S^c and the *separation function* $\psi : [v_S^c, 1] \to [v_M^c, 1]$, which takes the services skill as an argument and returns the corresponding cut-off value of the manufacturing skill.

Lemma 1. In equilibrium, the critical skills in manufacturing and services are respectively

$$v_M^c = \sup\{v_M \in [0,1] : w_M(v_M) \leq \max\{w_S(0), 0\} \text{ or } v_M = 0\}, \tag{11}$$
$$v_S^c = \sup\{v_S \in [0,1] : w_S(v_S) \leq \max\{w_M(0), 0\} \text{ or } v_S = 0\}. \tag{12}$$

[31]The only difference is that my result explicitly allows for surplus to be weakly supermodular, that is, it allows for $\frac{\partial^2}{\partial v_i \partial h_i}\pi_i \geq 0$. Legros and Newman (2002) call "famous" the result that under weakly supermodular surpluses any stable matching can be supported only by payoff schemes that support PAM. However, they provide no references and their Proposition 3 holds only for one-sided matching markets. The argument in Sattinger (1979) implies only that if the cross-derivative of the surplus function is strictly positive than matching is positive and assortative but is silent on what happens if the cross-derivative is weakly positive.

[32]Technically, Equations 3 and 4 already do that. The challenge, however, is to express S_M, S_S as functions of wages in a way that will allow me to characterize the equilibrium.

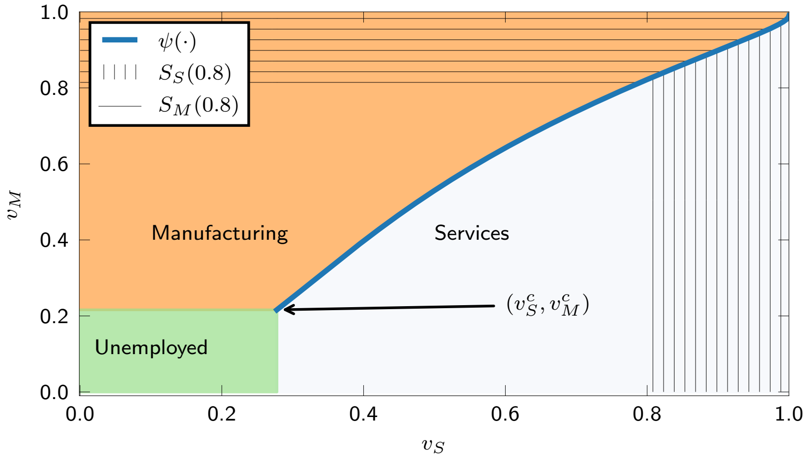

Figure 1: Graphical representation of the relation between the separation function and sorting. The hatched areas represent the space of workers with skill $V_i \geq 0.8$ who join occupation i; their supply $S_i(0.8)$ depends on how many workers reside in this space (which depends on the copula).

Provided that $v_M^c, v_S^c < 1$, it is the case that $w_M(v_M^c) = w_S(v_S^c)$.

For $v_S \geq v_S^c$ the separation function depends on the sectoral wage functions as follows:

$$\psi(v_S) = \max\{v_M \in [v_M^c, 1] : w_M(v_M) \leq w_S(v_S)\}. \tag{13}$$

Note that for v_S's such that $w_S(v_S) \leq w_M(1)$ this implies that

$$w_S(v_S) = w_M(\psi(v_S)). \tag{14}$$

The critical skills and the separation function are sufficient to characterize the sorting of workers to sectors. This is depicted in Figure 1. By the definitions of v_M^c and v_S^c, all but a zero measure of workers with $(v_M, v_S) < (v_M^c, v_S^c)$ remain unmatched. Manufacturing is populated by workers with $v_M \geq v_M^c$ and $v_M \geq \psi(v_S)$. Services are populated by workers with $v_S > v_S^c$ and $v_M < \psi(v_S)$.[33] Thus, v_M^c, v_S^c and ψ fully determine the sectoral supply functions.

Lemma 2. Given the critical skills v_M, v_S and the separation function ψ, the supply of relative skill in manufacturing and services is respectively

$$S_M(v) = \begin{cases} \int_v^1 \frac{\partial}{\partial v_M} C(r, \phi(r)) \mathrm{d}r, & v \geq v_M^c \\ S_M(v_M^c) & v < v_M^c, \end{cases} \quad S_S(v) = \begin{cases} \int_v^1 \frac{\partial}{\partial v_S} C(\psi(r), r) \mathrm{d}r, & v \geq v_S^c \\ S_S(v_S^c) & v < v_S^c, \end{cases} \tag{15}$$

[33] As in Section 3.1, it does not matter where the indifferent workers are assigned, as they are of measure zero.

where $\phi : [v_M^c, 1] \to [v_S^c, 1]$ depends on ψ as follows:

$$\phi(v_M) = \sup\{v_S \in [v_S^c, 1] : \psi(v_S) < v_M\}.$$

In equilibrium, the separation function determines the sectoral supply of skill, the sectoral supply of skill determines wages, and wages determine the separation function. Any separation function that corresponds to supply and wage functions that hold in some equilibrium will be called an *equilibrium separation function*. Equilibrium separation functions can be found by substituting the sectoral supply functions from Equation (15) into the results in Proposition 1 and then substituting the resulting wage functions into Equations (11)–(13).

Theorem 1. An equilibrium exists. The equilibrium separation function, the equilibrium supply functions, and the equilibrium demand functions are unique.

The proof entails constructing a map the fixed point of which is equivalent to the solution of (13) and finding a norm for which this map is a *contraction mapping*.[34] This proves that $\psi(\cdot)$ is unique *given* (v_M^c, v_S^c)—and that it is continuous in both v_M^c and v_S^c. Then showing existence and uniqueness is merely a matter of proving that Equations (11) and (12) have a unique solution given the function $\psi(\cdot, v_M^c, v_S^c)$. Of course, the existence of a unique equilibrium separation function implies trivially that there exists an equilibrium. It means further that the equilibrium is unique for most practical purposes, in that the equilibrium supply and demand functions are unique. By Proposition 1, however, the equilibrium wage functions are uniquely determined only for $v_i \geq v_i^c$, and even then possibly only up to the lowest wage $w_i(v_i^c)$.[35]

2.2.3 Sattinger and Roy

The first step in my characterization strategy is very similar to Sattinger (1979), the second to Roy (1951). This is not a coincidence: The model nests both one-sector assignment models and Roy-like models of self-selection.[36]

Sattinger (1979) Firstly, the model reduces to the one-sector assignment model if one of the two sectors does not employ any workers. This can happen either if Assumption 3

[34] The norm I use is Bielecki's norm for a high-enough parameter λ.

[35] This is the case if $R_M + R_S = 1$; otherwise $w_i(v_i^c)$ is uniquely determined.

[36] Formally, it does not fully nest the actual models by Sattinger and Roy. In the former case, the reason is that Sattinger allows for cases in which both firms and workers are unemployed, which is ruled out here by the assumption of positive surpluses, so only certain special cases of his model are nested. In the latter, the reason is that Roy uses bivariate log-normal distribution of skills, which is not defined over a rectangle—however, we can get an arbitrarily good approximation of Roy's model by using bivariate log-normal distribution, truncated arbitrarily high and arbitrarily close to zero. This is done in Section 4.1.1 and Online Appendix OA.5.

does not hold or, alternatively, if there exist no firms in services ($R_S = 0$). In either case, wages in manufacturing are simply

$$w_M^{\mathrm{SAT}}(v_M) = \int_{v_M^c}^{v_M} \frac{\partial}{\partial v_i} \pi_M\Big(v, 1 - \frac{1-v}{R_M}\Big) \mathrm{d}v + w_M(v_M^c), \tag{16}$$

where $v_i^c = \max\{0, 1 - R_i\}$, with $w_M(v_M^c) = 0$ if $R_M < 1$ and $w_M(v_M^c) = \pi_M(0, 1 - \frac{1}{R_M})$ if $R_M > 1$.

More interestingly, the two sector-model is also equivalent to the one-sector model when the two sectors are symmetric.

Definition 2. The model is *symmetric* iff (i) $C(v_M, v_S) = C(v_S, v_M)$ for all $(v_M, v_S) \in [0,1]^2$, (ii) $\pi_M(v,h) = \pi_S(v,h)$ for all $(v,h) \in [0,1]^2$, and (iii) $R_M = R_S$.

If the sectors are symmetric, workers choose the sector in which their relative skill is higher. In the symmetric case, wages are then

$$w_i^{\mathrm{SYM}}(v_i) = \int_{v_i^c}^{v_i} \frac{\partial}{\partial v_i} \pi_i\Big(v, 1 - \frac{C(v,v)}{R_i}\Big) \mathrm{d}v + w_i(v_i^c),$$

where $v_i^c = \max\{0, 1 - 2R_i\}$, with $w_i(v_i^c) = 0$ if $R_i < \frac{1}{2}$ and $w_i(v_i^c) = \pi_i(0, 1 - \frac{1}{R_i})$ if $R_i > \frac{1}{2}$. This is exactly equivalent to an assignment model in which there is just one sector with $R = 2R_i$ firms but the surplus produced in each match depends on the maximum of the workers' services and manufacturing relative skills, so that $v = \max\{v_M, v_S\}$.

Roy (1951) The model reduces to Roy's model in two related ways. First, suppose that firms in both sectors are identical, so that the surplus produced by any match depends on the worker's skill only.[37] If, in addition, there is an abundance of firms in each sector ($R_i > 1$), then firms have no market power. Hence workers receive the entire surplus, and their wage does not depend on the sectoral supply of skill: $w_i(v_i) = \pi_i(v_i)$. This is exactly as in Roy's model. In other words, Roy-like models can be seen as two-sector matching models in which all firms from the same sector are homogeneous.[38]

Alternatively but similarly, the model reduces to Roy's model if the number of firms in each sector is unlimited, that is, if $R_i \to \infty$. In this case all workers match with a firm of highest productivity, as $1 - \frac{S_i(v)}{R_i} \to 1$, and so workers again receive the entire surplus.[39]

[37] This implies that $\Pi_z(\bullet) = 0$, which is allowed by my assumptions.

[38] An example would be the Gaussian-exponential specification, which will be introduced in Section 4.1.1, with $R_M, R_S > 1$ and $\gamma_M = \gamma_S = 0$.

[39] Formally, $w_i(v_i) \in (w_i^{\mathrm{SAT}}(v_i), \pi_i(v_i, 1))$, because the highest feasible supply of skill v_i in any sector is $1 - v_i$, and the wages in sector i are bounded from above by the wages holding in the case of an empty sector j. Inspection of Equation (16) shows that $\lim_{R_i \to \infty} w_i(v_i) = \pi(v_i, 1)$.

3 Skill Interdependence and Wage Polarization

In this section, I investigate the link between an increase in the interdependence of relative skills and changes in the polarization of the economy-wide wage distribution. Wage distribution becomes more polarized if overall inequality increases (highest wages increase more than lowest) but left-tail inequality falls. Empirically, an increase in wage polarization has been recorded in the US in the 1990s and 2000s (Acemoglu and Autor, 2011).

Firstly, I show that—keeping the reduced surplus functions unchanged—an increase in the interdependence of relative skills is *equivalent* to a fall in the supply of skill. Further, I demonstrate that the interdependence of relative skills depends not just on the distribution of basic skill but also on the skill content of the sectors, that is, on the proportion in which basic skills are used in each sector.

Secondly, I decompose the change in the distribution of wages into the *wage effect*, which captures the change in wages paid to workers of given relative skill, and the *composition effect*, which captures the change in the number of workers with a given level of skill. The composition effect results in a worse distribution of wages in the sense of first-order stochastic dominance, because of the fall in skill supply; this increases wage polarization unambiguously. The impact of the wage effect is ambiguous. In the symmetric case, the composition effect is certain to dominate under mild conditions and polarization increases. It follows that a technological change that is not biased toward any particular sector (skill, occupation) but instead homogenizes skill content across sectors can plausibly increase wage polarization.

Note that in all comparative statics exercises in this paper, I compare the equilibria of two specifications of the model: the *old* one and the *new* one.[40] The old specification is denoted by θ_1 and the new one by θ_2. For example, $C(\bullet; \theta_1)$ is the old copula of relative skills and $C(\bullet; \theta_2)$ is the new one.

3.1 Interdependence and the Overall Supply of Skill

I will use the *concordance ordering* as the notion of interdependence.

Definition 3 (Scarsini, 1984)**.** Copula $C(\bullet, \theta_2)$ is more concordant than copula $C(\bullet, \theta_1)$ if $C(v_M, v_S, \theta_2) \geq C(v_M, v_S, \theta_1)$ for all $(v_M, v_S) \in [0, 1]^2$.

The concordance ordering formalizes the idea of greater interdependence, as higher concordance implies that large values of V_M are more likely to go with large values of V_S. For example, for bivariate random variables with a Gaussian copula, an increase in concordance is equivalent to an increase in the correlation parameter ρ (Joe, 1997).[41]

[40] With both specifications meeting all conditions from Section 2, including Assumption 3.

[41] I call ρ the correlation parameter, because for variables with the Gaussian copula and normal marginal distributions it is equal to linear correlation.

To establish the link between the concordance ordering and changes in the overall supply of skill, we need to define the latter. I will assume a definition analogous to *first-order stochastic dominance* and say that the supply of skill increases whenever the change in copula is certain to increase the maximal total surplus produced in the economy. Formally, denote by $S = (S_M, S_S)$ a pair of sectoral supply functions; then $\mathbf{S}$ denotes the set of all S that are *feasible* under a given copula, that is, all S such that $S_i(0) \leq R_i$ and

$$S_S(v_S) + S_M(\psi(v_S)) \leq 1 - C(\psi(v_S), v_M).^{42} \tag{17}$$

The total surplus produced in the economy for a given S is denoted by $T(S)$:

$$T(S) = \int_1^0 \pi_M\Big(v_M, 1 - \frac{S_M(v_M)}{R_M}\Big)\mathrm{d}S_M(v_M) + \int_1^0 \pi_S\Big(v_S, 1 - \frac{S_S(v_S)}{R_S}\Big)\mathrm{d}S_S(v_S).^{43} \tag{18}$$

Definition 4. The supply of skill under copula $C(\bullet, \theta_1)$ is *unambiguously higher* than under copula $C(\bullet, \theta_2)$ if

$$\max_{S \in \mathbf{S}(\theta_2)} T(S) \leq \max_{S \in \mathbf{S}(\theta_1)} T(S)$$

for all quadruples (π_M, π_S, R_M, R_S) that meet Assumptions 1 and 3.

It can be shown that if the reduced surplus functions are kept constant, a fall in concordance is *equivalent* to an upward shift in the overall supply of skill.

Proposition 2. Consider two copulas that meet Assumption 2. The supply of skill is unambiguously higher under copula $C(\bullet, \theta_1)$ than under copula $C(\bullet, \theta_2)$ if and only if $C(\bullet, \theta_2)$ is more concordant than $C(\bullet, \theta_1)$.

As concordance falls, the production possibility frontier shifts outward, in the sense that the set of feasible pairs of supply functions expands: $\mathbf{S}(\theta_2) \subset \mathbf{S}(\theta_1)$. Therefore, the maximal total surplus must increase. In other words, as more workers are highly skilled in at least one sector, it becomes possible to increase the supply of skill in one sector without decreasing it in the other sector. In the other direction, the implication holds for the usual reasons: If the two copulas are not ordered with respect to concordance, there always exist surplus functions that value particularly highly the skills that exist in greater supply under copula $C(\bullet, \theta_2)$.

Finally, let me stress that an increase in relative skill concordance is equivalent to a fall in the supply of skill only when the reduced surplus functions are held constant.[44] Note that any change in the marginal distributions of the components of $\mathbf{x}$ would affect the reduced surplus functions π_i. Therefore, assuming that the basic surplus function Π_i is increasing in all dimensions of $\mathbf{x}$, a change in the distribution of $\mathbf{x}$ constitutes a

[42] Note that $\psi(\cdot) \in [0, 1]$ and is implicitly defined through Equation (15).

[43] To see why, recall that S_i is decreasing for $v_i \geq v_i^c$ and constant otherwise.

[44] As is the case in Definition 4.

fall in the supply of skill if and only if relative skills become more concordant and the marginal distributions of all components of $\mathbf{x}$ improve in the sense of first-order stochastic dominance.[45]

3.2 Skill Content of Sectors

Recall that the relative skill $v_i(\mathbf{x})$ is equal to the measure of workers who would produce less surplus in sector i than a worker with basic skill vector $\mathbf{x}$ (if matched with the same firm). Therefore, the copula of relative skills depends not only on the distribution of basic skills but also on the skill content of the sectors.

This can be easily seen in an example with standard normally distributed basic skills, $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$.[46] Suppose that the surplus in each sector depends on a linear combination of cognitive (x_1) and non-cognitive (x_2) skills, so that $\Pi_i(\mathbf{x}, z, i) = \bar{\Pi}_i(\alpha_{i1}x_1 + \alpha_{i2}x_2, z)$, where $\alpha_{ji} \in [0, 1]$. To isolate the impact of changes in skill content on relative skill interdependence from their impact on the reduced surplus functions, let us assume that sector i uses the non-cognitive skill with weight $\sqrt{1 - \alpha_{i1}^2}$.[47] This results in relative skills of the form $v_i(\mathbf{x}) = \Phi^{-1}(\alpha_{i1}x_1 + \sqrt{1 - \alpha_{i1}^2}x_2)$ and distributed according to a Gaussian copula with correlation parameter $\rho = \alpha_{S1}\alpha_{M1} + \sqrt{(1 - \alpha_{M1}^2)(1 - \alpha_{S1}^2)}$.[48]

For Gaussian copulas, an increase in ρ is equivalent to an increase in concordance, which proves that the relative skill interdependence depends on the skill content of each sector. Differentiating ρ with respect to α_{M1} reveals that the relative skill interdependence will increase in response to a small increase in the weight of cognitive skills in manufacturing if and only if $\alpha_{M1} < \alpha_{S1}$. Thus if the importance of cognitive skills increases in a sector that is relatively non-cognitive skill intensive, the interdependence of relative skills increases. If, on the other hand, cognitive skills become more important in a relatively cognitive skill intensive sector, interdependence falls.

3.3 Wage Polarization

I will now address the impact of changes in copula on wage polarization. To the best of my knowledge, there does not exist a standard, formal definition of polarization. Therefore, I provide my own definition, which captures the most salient empirical facts about wage polarization: the decrease in wage inequality in the left tail of the distribution and the

[45] More generally, a change in the distribution of $\mathbf{x}$ constitutes an increase in skill supply if and only if it causes both a fall in the concordance of relative skills and a universal improvement in the surplus (Definition 7) in both sectors.

[46] $\mathbf{I}$ denotes the identity matrix.

[47] This ensures that a change in α_{i1} does not affect the reduced surplus functions, because the distribution of $\alpha_{i1}X_1 + \sqrt{1 - \alpha_{i1}^2}X_2$ does not depend on α_{i1}.

[48] To be precise, $C(v_M, v_S) = \mathbf{\Phi}_\rho\Big(\Phi^{-1}(v_M), \Phi^{-1}(v_S)\Big)$, where $\mathbf{\Phi}_\rho$ denotes the cdf of a standardized bivariate normal distribution with correlation and Φ denotes the cdf of the univariate standard normal distribution.

increase in the difference between the highest and lowest wages.[49] In doing so, it will be convenient to focus on the *inverse wage distribution* $W : [0,1] \to \mathbf{R}_{\geq 0}$, which maps the quantile of a wage into its actual level (e.g., $W(0.5)$ is the median wage). Any unmatched workers will be assigned the reservation wage, that is, $w_i(v_i) = 0$ for $v_i < v_i^c$.[50]

Definition 5 (Wage Polarization)**.** Wage polarization increases in absolute terms if wage inequality (i) falls in the left tail, that is, there exists $\bar{t} \in (0,1)$ such that $W(t) - W(0)$ falls for all $t \in (0, \bar{t})$, and (ii) increases overall, that is, the range of wages $(W(1) - W(0))$ increases. Wage polarization increases in relative terms if conditions (i) and (ii) hold for $\log W(t)$. If (i) holds strictly for some $t \in (0, \bar{t})$, then the increase is called strict.

The distribution of wages in the economy can easily be derived. Consider an arbitrary wage $w \in \mathbf{R}$. Self-selection implies that only workers with skill such that $w_M(v_M), w_S(v_S) \leq w$ will earn less than w. Therefore, the economy-wide distribution of wages is simply

$$F_W(w) = \Pr(W \leq w) = C(w_M^{-1}(w), w_S^{-1}(w)), \tag{19}$$

where $w_i^{-1}(w) = \max\{v_i \in [0,1] : w_i(v_i) \leq w\}$. The inverse wage distribution is thus

$$W(t) = \begin{cases} 0 & \text{for } t \in [0, C(v_M^c, v_S^c)], \\ F_W^{-1}(t) & \text{otherwise,} \end{cases} \tag{20}$$

where $F_W^{-1}(t)$ is the inverse of $F_W(w)$. Note that both the distribution and the inverse distribution of wages are fully determined by the wage functions and the copula of relative skills.[51] Of course, both w_M and w_S are endogenous and ultimately depend on C. The wage function in sector i that holds under copula $C(\bullet, \theta_j)$ will be denoted by $w_i(\cdot, \theta_j)$.

Consider any parametric family of copulas, and denote one of the parameters by θ. I will differentiate $W(t)$ with respect to θ, to find the effect of changes in the copula on the inverse wage distribution.[52] This is done for expositional ease; all of the reasoning given here holds unchanged for finite jumps in θ. Denote by $p_i(t)$ the probability that a worker who occupies rank t in the distribution of wages works in sector i.[53] Then for

[49] Note that my definition does not require wages to increase less for the median worker than for the least-earning workers, which did happen in the 1990s in the US. This will happen in my model for plausible parameterizations (see Figure 2) but this result is not universal.

[50] Alternatively, we might want to consider the distribution of wages among employed workers only. This does not matter for the overall effect on changes in concordance, as the number of employed workers is unchanged; it does change the decomposition, however.

[51] By Lemma 1, v_M^c and v_S^c are fully determined by the wage functions.

[52] Of course, this implicitly assumes that C is differentiable with respect to θ.

[53] The formula for $p_i(t)$ can easily be derived. First, note that as the density of all workers occupying rank t must be 1, $p_i(t)$ is simply equal to the density of workers with rank t who work in sector i.

$t \in (C(v_M^c, v_S^c), 1)$ we can write

$$\frac{\mathrm{d}}{\mathrm{d}\theta} W(t) = \underbrace{-\frac{\frac{\partial}{\partial\theta} C(v_M(t), v_S(t))}{\frac{\partial}{\partial w} F_W(W(t))}}_{\text{composition effect}} + \underbrace{p_M(t)\frac{\partial}{\partial\theta} w_M(v_M(t)) + p_S(t)\frac{\partial}{\partial\theta} w_S(v_S(t))}_{\text{wage effect}}, \tag{21}$$

where $v_i(t) = w_i^{-1}(W(t))$ denotes the relative skill in sector i of a worker occupying quantile t.

The composition effect holds constant the wage received by a worker with skill (v_M, v_S)—and thus also their sector—but it varies with the numbers of workers with each relative skill endowment that are available in the economy, whence the name.

As concordance increases, high skills are concentrated among fewer workers, hence the number of workers skilled enough to earn more than $w = W(t)$ falls. This means that workers who earn $w = W(t)$ occupy a higher rank than previously and, as a result, the wage corresponding to quantile t must fall. Thus the composition effect of increased concordance is unambiguously negative. Because wage functions are kept constant, the lowest and highest wages must remain unchanged, and wages increase by less (fall more) for workers in the middle of the wage distribution than at either extreme. This implies trivially that *the wage distribution becomes more polarized* in both absolute and relative terms. Note that this also implies trivially that the distribution of wages deteriorates in the FOSD sense.[54]

In the standard Roy's model, only the composition effect is present, because the wage paid to a worker with vector (v_M, v_S) is given exogenously. In consequence, any increase in skill concordance *unambiguously increases wage polarization in the model of* Roy (1951). To the best of my knowledge, this feature of Roy's model has not been pointed out before.[55] Finally, note that most models of self-selection differ only in how changes in supply affect the wage functions. This means that the composition effect should work similarly in all models of self-selection.

The wage effect captures how the inverse wage distribution is affected by changes in the wage functions which are induced by a change in the copula. It is equal to the sum of the changes in wages paid for skills $(v_M(t), v_S(t))$, weighted by the probability that a worker who occupies rank t works in sector i. Note that the number of workers with each

Therefore, it follows from Equation 15 that

$$p_S(t) = \lim_{h\to 0} \frac{\int_{w_S^{-1}(W(t))}^{w_S^{-1}(W(t+h))} \frac{\partial}{\partial v_S} C(\psi(r), r)\mathrm{d}r}{h} = \frac{\frac{\partial}{\partial v_S} C(w_M^{-1}(W(t)), w_S^{-1}(W(t)))}{\frac{\partial}{\partial v_S} w_S(w_S^{-1}(W(t)))\frac{\partial}{\partial w} F_W(W(t))},$$

and analogously for $p_M(t)$.

[54] This can be seen immediately by inspection of Equation (19).

[55] Gould (2002) pointed out the relationship between the correlation of skills and wage inequality in the model of Heckman and Sedlacek (1985), but not its effect on wage polarization.

skill endowment is held constant.

If the two sectors are symmetric (per Definition 2), then the wage effect increases the wages of all workers and, in the process, also the absolute wage inequality. This is the case, because under symmetry the fall in the overall supply of skill affects both sectors equally, which by Equation (9) must increases wages and wage inequality. Note that, by the definition of increased polarization, an increase in wage inequality counteracts the increase in wage polarization caused by the composition effect. This case is discussed in detail in Section 3.4.

In general, however, the fall in overall supply of skill can be *biased* toward one of the sectors, in which case the wage effect might increase wage polarization. Suppose, for example, that services employ the least and the most skilled workers ($w_S(0) > w_M(0)$, $w_S(1) > w_M(1)$) but outside of those two extremes they employ only workers who are highly skilled in services but have low manufacturing skill. This means that most of the workers who are equally skilled in both sectors ($v_M \approx v_S$) work in manufacturing. As concordance increases, the overall supply of skill falls but this is fueled mostly by a fall in supply in services, as this sector has been employing most of the workers that were highly skilled in just one dimension. In fact, the supply of skill in manufacturing might increase for at least some skill levels. In such a case, wages increase and become more unequal in services but fall and become less unequal in manufacturing. As highest and lowest paid workers work in services, wage polarization increases by Equation (21). A stylized example depicting such a scenario can be found in Online Appendix OA.1.

The overall effect, therefore, is ambiguous. There are circumstances in which both the composition and wage effects contribute to an increase in polarization; in many other cases, the wage effect will counteract the increase in wage polarization induced by the composition effect.

There exists a crucial difference in the way in which the wage and composition effects can cause wage polarization. The wage effect increases wage polarization only if the fall in skill supply is *biased* toward a sector that employs most of the highest and lowest earners. In that sense, the wage effect resembles the mechanism underpinning the most prominent explanation for wage polarization: *task-biased technological change* (TBTC). Very roughly, the task-biased technological change hypothesis posits that wage polarization is caused by a fall in demand for tasks that are performed predominantly by medium-earning workers, which decreases their relative wages. My wage effect differs from this explanation in that it shows that the fall in relative wages of medium-earning workers can also be caused by supply-side changes. Nevertheless, it still requires the change in supply to be biased toward one of the sectors.[56]

[56]The similarity between these mechanisms is underlined by the fact that my sectors bear a certain resemblance to tasks in the TBTC literature, in that they are both defined by the combination of skills required to work in them/perform them. In fact, Heckman and Sedlacek (1985) refers to what I call

The composition effect, however, is present even if the fall in the overall supply of skill is *unbiased*, that is, if the changes in $w_S(v_S)$ and $w_M(\psi(v_M))$ are identical. This suggests that it is fundamentally different from the existing explanations for wage polarization. In fact, as the symmetric case of the model is equivalent to a single-sector assignment model, an increase in skill interdependence can plausibly increase wage polarization even in a model with no sectors or tasks, contrary to the founding observation in the TBTC literature (Acemoglu and Autor, 2011; Lindenlaub, 2017). That this is possible follows immediately from the fact that the wage effect does not exist in Roy's model. In the next section I will demonstrate that under symmetry the composition effect must dominate the wage effect, even if the latter is present, as long as the support of employed skills does not change.

3.4 Polarization in the Symmetric Case

I defined the symmetric case of the model in Section 2.2.3. Recall that under symmetry, workers sort into the sector in which their relative skill is higher, implying that the distributions of skill and wages are identical in both sectors. As a consequence, the symmetric model is equivalent to a single-sector matching model. Thus, the results derived in this section hold for any decrease in skill supply in the FOSD sense in a single-sector assignment model, irrespective of what caused it.

In the symmetric case, an increase in skill concordance decreases the supply of skill in each sector (as $S_i(v_i) = \frac{1}{2}(1 - C(v_i, v_i))$), and so its wage effect increases wages and wage inequality. This counteracts the increase in wage polarization caused by the composition effect. Nevertheless, the overall effect of an increase in relative skill interdependence on wage polarization is positive under fairly general conditions. Importantly, the change in relative skill interdependence needs to meet a mild regularity condition. I will say that—for a given value of critical skill—the change in interdependence is regular if there exists some $v' > v_S^c$, such that $\frac{\partial}{\partial v_S}C(v', v'; \theta_2) - \frac{\partial}{\partial v_S}C(v', v'; \theta_1) \neq 0$ and

$$\operatorname{sgn}\Big(\frac{\partial}{\partial v_S}C(v, v; \theta_2) - \frac{\partial}{\partial v_S}C(v, v; \theta_1)\Big) = \operatorname{sgn}\Big(\frac{\partial}{\partial v_S}C(v', v'; \theta_2) - \frac{\partial}{\partial v_S}C(v', v'; \theta_1)\Big)$$

for all $v \in (v_S^c, v')$. This condition is satisfied by all commonly used families of copulas but excludes pathological cases, where the derivative oscillates between positive and negative values in any neighborhood of the critical skill.[57]

relative skills as tasks, and many TBTC models use Roy's model to capture workers' self-selection into tasks (Autor et al., 2003; Boehm, 2015).

[57] To be precise, this condition is satisfied for $v_S^c = 0$ by all of the one-parameter families listed in Chapter 4 of Joe (1997) that satisfy Assumption 2, that is, families B1–B8 and B10. Furthermore, I have verified that Gaussian, Plackett, Gumbel, Galambos, Hussler and Reiss, and Morgerstern copulas (families B1, B2, B6, B7, B8, and B10) satisfy the regularity condition for any v_S^c.

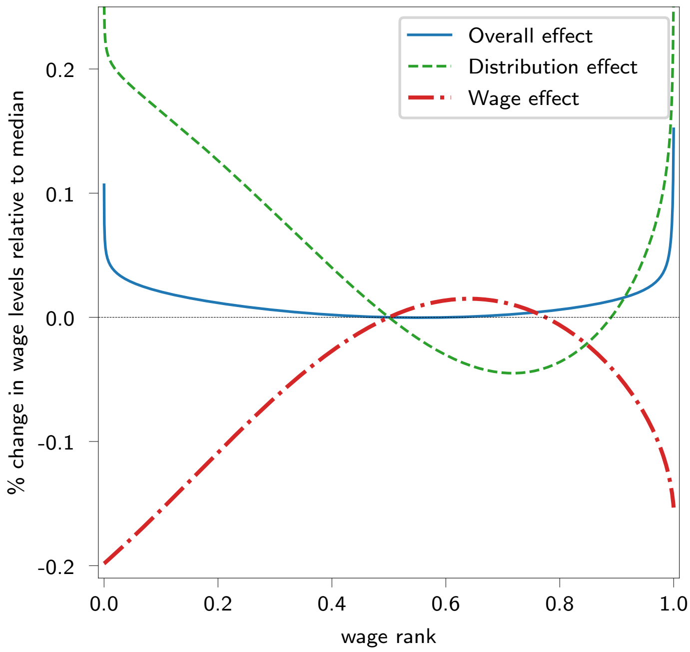

Figure 2: The effect of an increase in concordance on the distribution of log wages.
Computed for an increase in correlation from $\rho = 0.66$ to $\rho = 0.99$ in a Gaussian-exponential specification of the model (see Section 4.1.1) with $\sigma_i = 1$, $\mu_i = 0$, $\delta = -\frac{1}{3}$, $\beta_i = 4$, $A_i = 1.37$, and $\gamma = 1214$, $i \in \{1, 2\}$.

Proposition 3 (Wage Polarization)**.** Suppose the model is symmetric, the concordance of the relative skills distribution increases and this increase is regular. As long as the increase in concordance leaves the critical skill level (v_i^c) unchanged, wage polarization increases strictly in both absolute and relative terms.

Figure 2 depicts the impact of higher concordance on the inverse distribution of wages. The increase in wage range follows trivially from the fact that $W(0)$ is unchanged and $W(1)$ increases. However, if the critical skill levels remain unchanged, then the composition effect must dominate the wage effect in the left tail of the distribution, thus increasing wage polarization. I will explain this for the case in which the numbers of workers and firms are exactly equal and the value of the copula increases strictly for all $(v_M, v_S) \in (0, 1)^2$.[58] The former ensures that every worker is matched with a firm of productivity $h = G_i(v_i)$, where $G_i : [v_i^c, 1] \to [0, 1]$ denotes the distribution of relative skill in that sector:

$$G_i(v) = \Pr\Big(V_i \leq t | w_M(V_i) \geq w_S(V_j), w_M(V_i) \geq 0\Big),$$

with $j \neq i$. Note that $G_i(v_i) = C(v_i, v_i)$ in this case, and thus the critical skill is equal to 0, as required.

[58]The proof for the general case is based on analogous reasoning.

In the symmetric case the difference between the wage received by a worker at rank t and a worker at rank 0 is

$$W(t) - W(0) = \int_0^{G_i^{-1}(t)} \frac{\partial}{\partial v_i}\pi_i(s, G_i(s))\mathrm{d}s. \tag{22}$$

As the change in copula is regular, there exists some $\bar{t}_1$ such that $\frac{\partial}{\partial\theta}G_i(G_i^{-1}(s)) \leq \frac{\partial}{\partial\theta}G_i(G_i^{-1}(t))$ for all $s \leq t \leq \bar{t}$. Therefore, for $t < \bar{t}_1$ we can write

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\Big(W(t) - W(0)\Big) &= \frac{\partial}{\partial v_i}\pi_i(s, G_i(s))\frac{\mathrm{d}}{\mathrm{d}\theta}G_i^{-1}(t) + \int_0^{G_i^{-1}(t)} \frac{\mathrm{d}}{\mathrm{d}\theta}G_i(s)\frac{\partial^2}{\partial v_i \partial h_i}\pi_i(s, G_i(s))\,\mathrm{d}s \\ &\leq \frac{\partial}{\partial\theta}G_i(G_i^{-1}(t))\Big[\int_0^{G_i^{-1}(t)} \frac{\partial^2}{\partial v_i \partial h_i}\pi_i(s, G_i(s))\,\mathrm{d}s - \frac{\frac{\partial}{\partial v_i}\pi_i(s, G_i(s))}{g(G_i^{-1}(t))}\Big].\end{aligned}$$

Because a strict increase in concordance implies a strict fall in skill supply, $\frac{\partial}{\partial\theta}G_i(G_i^{-1}(t)) < 0$ for all $t \in (0,1)$ and, trivially, the RHS must be negative for small enough t, thus proving the fall in relative left-tail inequality. As the lowest wage is unchanged, left-tail inequality must also fall in relative terms.

The condition that the critical skill levels remain unchanged is actually quite mild. It boils down to the requirement that the fall in skill supply does not change the support of skills that are actually employed in the economy. In this model, this is ensured whenever jobs are weakly abundant ($R_M + R_S \geq 1$).[59] This might seem unappealing, as it implies full employment. It is important to stress, however, that full employment is just one (and a particularly stark) way of ensuring that skill support does not change, rather than a condition in its own right. The actual critical condition is that, prior to the change in concordance, workers of almost all skill levels have been employed with some positive probability.[60]

Finally, it is worth noting that Proposition 3 holds even when the model is not symmetric as long as the change in concordance is unbiased. I show this formally in Online Appendix OA.1.

3.5 Discussion

To the best of my knowledge, I am the first to point out that a fall in skill supply increases wage polarization in a single-sector assignment model.[61] There are good reasons why

[59] More generally, v_i^c does not change whenever $C(v_M^c(\theta_1), v_S^c(\theta_1))$ remains constant.

[60] For example, suppose that only some fraction of the entire population of workers is able to access the job market, perhaps due to some unspecified search process. In such an economy, involuntary unemployment exists even if jobs are abundant, in the sense that more firms than workers managed to enter the job market. Even this, however, is but a crude way of ensuring that the support of employed skills does not change.

[61] Costinot and Vogel (2010) consider the effect of a fall in skill supply on wages in an assignment model in the vein of Sattinger (1975) and Teulings (1995). However, they do not take the composition

this is the case. After all, in a one-dimensional world a fall in skill supply is a highly implausible cause for the increases in polarization that were recorded in the US, given the improvements in educational attainment in the preceding decades.[62] This is not, however, a major problem for the plausibility of the mechanism I propose. In my model, an improvement in educational attainment would be represented as an improvement in the marginal distribution of basic skills. This is where the point from Section 3.1 becomes crucial: An improvement in the marginals of the skill distribution does not necessarily increase the overall supply of skill in a multivariate world if it is accompanied by the skill content of different sectors becoming more similar. Note that there exists strong empirical evidence that the skill content of sectors, tasks, and occupations has been changing (Autor et al., 2003; Spitz-Oener, 2006) and, in fact becoming more similar across sectors (Gould, 2002).

Of course, it is nevertheless hard to imagine that the overall surplus produced in the economy has fallen, which would be implied by a technological (or other) change that results *only* in reduced supply of skill. Fortunately, this is not required for the mechanism proposed here to produce an increase in wage polarization. Consider any technological improvement that affects both sectors in the same way and, ignoring the effect it has on relative skill interdependence, either leaves the distribution of relative wages unchanged or even (slightly) decreases inequality. Under symmetry, such a technological advancement must increase wage polarization in relative terms as long as it increases relative skill interdependence.[63] Therefore, the link between skill interdependence, skill supply, and wage polarization expands significantly the set of changes in technology and/or marginal distributions of $\mathbf{x}$ that can *plausibly* explain wage polarization.[64]

An increase in relative skill interdependence is perfectly consistent with the "routinization" hypothesis by Autor et al. (2003), which posits that polarization has increased because routine jobs were replaced by machines. The advent of computers is a very plausible reason for why different sectors started coveting the same (cognitive) skills.[65] In

effect into account, and thus find only that wages become more unequal.

[62]For example, between 1950 and 1990, the proportion of US residents age 25 or over who completed high school rose from a little over 30% to over 70% (Figure 2 in Ryan and Bauman (2016)).

[63] To see this, suppose that the model is originally symmetric, the copula becomes more concordant, and surplus in both sectors increases multiplicatively: $\pi_M(\bullet, \theta_2) = \pi_S(\bullet; \theta_2) = A\pi_i(\bullet; \theta_1)$. Denote as $W(t; \pi_i)$ the inverse wage distribution holding under surplus functions π_i. Then $\log W(t; \pi_i(\theta_2)) = \log A + \log W(t; \pi_i(\theta_1))$, and thus $\log W(t; \pi_i(\theta_2)) - \log W(t; \pi_i(\theta_2)) = \log W(t; \pi_i(\theta_1)) - \log W(t; \pi_i(\theta_1))$. Thus by Proposition 3, the increase in concordance increases polarization as long as the support of employed skills is unchanged. Similarly, if $\pi_M(\bullet, \theta_2) = \pi_S(\bullet; \theta_2) = \pi_I(\bullet; \theta_1) + A$, then left-tail inequality will fall even more than if the reduced surplus function did not change, yet $\log W(1) - \log W(0)$ will increase as long as A is not too large.

[64]The word "plausibly" is key here, as we can always easily construct a surplus function that perfectly fits any distribution of wages. The problem is that the type of technological change implied by the new surplus function is often hard to reconcile with what we know about the nature of real-world technological change.

[65]In the example from Section 3.2, an increase in α_i can be thought of as a replacement of a worker who was performing a non-cognitive task with a machine that requires no non-cognitive skill to operate

fact, the impact this would have on relative skill interdependence provides an explanation for the lag between the advent of computerization and increases in wage polarization: Computerization started in the 1970s at the latest (Card and DiNardo, 2002), whereas increases in wage polarization have been first recorded in the 1990s (Acemoglu and Autor, 2011). However, if computers were first introduced in sectors that were already cognitive-skill intensive, then changes in interdependence would at first contribute to a fall in polarization—and not until later, when other sectors caught up, to an increase in polarization.

Even so, the mechanism through which routinization operates here is distinct from the mechanisms suggested in the literature thus far. In the existing literature, the change in technology must be *biased* toward particular tasks (or occupations) (e.g Costinot and Vogel, 2010; Acemoglu and Autor, 2011; Boehm, 2015; Lindenlaub, 2017). Here, the increase in relative skill interdependence increases wage polarization even if the two sectors remain symmetric. The difference goes beyond semantics; it follows from the reasoning in Section 3.2 that a technological change need not be biased toward a particular dimension of basic skill in order to increase polarization.[66] Further, if the increase in interdependence is caused by changes in skill content, then it is going to produce a rich pattern of rank switching, even under symmetry. In particular, in the example from Section 3.2, some of the workers most hurt by the change in skill content would have been earning very high wages previously.[67] Finally, note that the existing empirical literature has routinely relied on the assumption that the skill content of occupations is unchanged. This assumption is made for the purpose of identifying the impact of changes in task prices (or the wage effect, using my terminology) on wage polarization (see Boehm, 2015) and it likely leads to biased estimates.

4 Changes to the Reduced Surplus Functions

In this section I investigate the effects of changes to the reduced surplus functions. I focus on changes that are sector specific, as I am interested in the way in which shocks and policy interventions spread through the economy. I start the analysis by defining two key concepts: vertical differentiation of workers and an increase in surplus levels.

Definition 6 (Vertical Differentiation)**.** Workers in manufacturing become (strictly)

but needs to be occasionally reprogrammed and thus requires some degree of cognitive skill.

[66] To see this, note that if in the example from Section 3.2 services were originally more cognitive-intensive and manufacturing more non-cognitive intensive, then an increase in non-cognitive skill content in services and in cognitive skill content in manufacturing would clearly lead to increased polarization.

[67] For example, suppose that $\alpha_{M1} = 0$ and $\alpha_{S1} = 1$ originally, and then α_{M1} increases to ≈ 1. This would mean that some of the workers who were originally top earners have been moved to the very bottom of the wage distribution.

more *vertically differentiated* if, for any $h_M \in [0,1]$ and any $0 \leq v'_M < v''_M \leq 1$

$$\pi_M(v''_M, h_M; \theta_2) - \pi_M(v'_M, h_M; \theta_2)(>) \geq \pi_M(v''_M, h_M; \theta_1) - \pi_M(v'_M, h_M; \theta_1).$$

Workers become more vertically differentiated in manufacturing if the difference in the surplus they produce increases for all levels of relative skill and all firms.[68] This is equivalent to an increase in the *spread* of the distribution of $\Pi(\mathbf{X}, h, M)$ in the sense of Bickel and Lehmann (1979) (for all $h \in [0,1]$).

Definition 7 (Increase in Levels)**.** The level of surplus produced in manufacturing *increases universally* if, for all $(v_M, h_M) \in [0,1]^2$, $\pi_M(v_M, h_M; \theta_2) \geq \pi_M(v_M, h_M; \theta_1)$.

The level of surplus produced in manufacturing increases universally if any match in manufacturing produces more surplus than before. This is a very strong condition, but I will demonstrate in this section that, despite its strength, a universal increase in the manufacturing surplus level is not sufficient to generate an increase in the equilibrium supply of skill in manufacturing. Further, I will provide an example showing that this condition does not guarantee—in and of itself—that the total surplus produced in manufacturing (that is, the sum of surpluses produced by manufacturing firms) increases.

In Section 4.1 I first show that if jobs are scarce ($R_M + R_S \leq 1$), an increase in vertical differentiation of manufacturing workers is sufficient for an increase in the equilibrium supply of relative skill in that sector, regardless of what happens to the levels of surplus. Then I consider a series of structured examples, which (a) illustrate that vertical differentiation of workers can increase because of changes in either the distribution of basic skills $\mathbf{x}$, the distribution of basic productivities $(z|i)$, or changes in the basic surplus function Π; (b) demonstrate that a universal increase in manufacturing surplus levels accompanied by a fall in vertical differentiation can result in lower total surplus in manufacturing; and (c) explore the implications of my results. In Section 4.2, I then consider the abundant-jobs case ($R_M + R_S > 1$) and show that in this case an increase in both differentiation and surplus levels is needed to ensure a higher equilibrium supply of relative skill in manufacturing.

4.1 Scarce Jobs

Scarcity of jobs implies that all firms are matched, the measure of firms in each sector is fixed, and the positive and assortative matching function in each sector is equal to the distribution of the relevant relative skill in that sector, G_i.

Lemma 3. If $R_M + R_S \leq 1$, then $S_i(0) = R_i$, which implies that $G_i(v_i) = 1 - \frac{S_i(v_i)}{R_i}$. If jobs are strictly scarce ($R_M + R_S < 1$), then additionally $w_M(v^c_M) = w_S(v^c_S) = 0$.

[68] This is equivalent to an increase in the marginal surplus of relative skill for all matches in manufacturing: $\frac{\partial}{\partial v_M}\pi_M(v_M, h_M; \theta_2) \geq (>) \frac{\partial}{\partial v_M}\pi_M(v_M, h_M; \theta_1)$ for all $(v_M, h_M) \in [0,1]^2$.

All firms must be matched if jobs are scarce, because otherwise the unmatched firms would hire the unmatched workers. With strictly scarce jobs, the competition from the unemployed workers drives the wages of the least skilled employed workers down to their reservation wage.

Proposition 4. If jobs are scarce and workers in manufacturing become more vertically differentiated, then (i) the distribution of relative skill in manufacturing improves in the FOSD sense, and (ii) the distribution of relative skill in services deteriorates in the FOSD sense. If in addition, the increase in differentiation is strict, then strictly more relative skill is supplied to manufacturing and strictly less to services.

This result can best be understood by focusing on the impact on the demand for relative skills. How do manufacturing firms' hiring decisions change after an increase in vertical differentiation but before the wage functions have time to adjust?[69] Because the difference in *surplus* produced by different workers has increased but the difference in *wages* has not, high-skilled workers become relatively underpaid. As a result, all firms want to hire a worker of higher relative skill than in the old equilibrium: The demand for relative skill shifts upward.[70] This is depicted in the left panel of Figure 3. Note that with scarce jobs, all firms want to hire some worker, and hence their hiring decisions depend on the differences in surplus only; the levels play no role at all. The upward shift in skill demand draws in additional marginal high-skilled workers into manufacturing and causes some marginal low-skilled workers to leave for services, which is depicted in the right panel of Figure 3.

In addition to its impact on sorting, the shift in demand affects wages as well. To bring the most interesting results into focus, in the discussion on wages I focus on strict increases in vertical differentiation; all of those results extend easily to the more general case. It is also worth noting that in the case of strict supermodularity, results (i) and (iii) in the following proposition hold strictly.

Proposition 5. If jobs are scarce and workers in manufacturing become strictly more vertically differentiated, then (i) in services, wages increase for all workers, and the higher the relative skill the greater the increase; (ii) in manufacturing, wages increase strictly for a positive mass of the workers of highest relative skill; and (iii) in both sectors the

[69] That is, if firms still have to pay wage $w_M(v_M;\theta_1)$ for relative skill v_M.

[70] The reasoning here is the same as in the monotone comparative statics results in Milgrom and Shannon (1994), with vertical differentiation being a condition analogous to increasing differences. Accordingly, an increase in vertical differentiation is a stronger condition than needed for Proposition 4 to hold. What is sufficient is that the marginal surplus of relative skill increases for all existing matches, rather than globally (see Theorem 3 and proof thereof in Appendix C). For small changes in the surplus function, this weaker condition is equivalent to an upward shift in demand. In general, however, these two are not equivalent. As an example, consider a change in surplus from $v_M h_M + 1$ to 1 for all $v_M < 1$ and to 2 for $v_M = 1$. Under the old wages, all firms would like to hire workers with $v_M = 1$. However, the measure of such workers is zero, so in equilibrium essentially all manufacturing firms would be hiring only those workers whose services skill is too low to enable them to find employment in services.

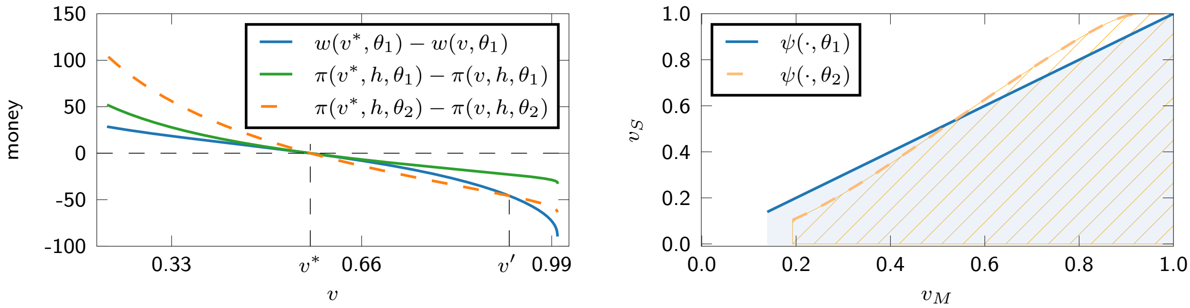

Figure 3: Left panel: v^* is the skill of the worker hired by firm h in the old equilibrium. All workers with skill $v \in (v^*, v')$ are strictly preferred by the firm after the increase in differentiation but before wage functions have time to adjust.
Right panel: change in the equilibrium separation function.

range of the wage distribution increases, strictly in manufacturing. If jobs are strictly scarce, the increases in wages and wage range in services are strict, and manufacturing wages fall strictly for a positive mass of workers with low relative skill.

Wages increase for manufacturing workers of high relative skill, as those workers are now in higher demand; similarly, wages fall for low-skilled workers in manufacturing, as those workers are now less in demand.[71] It is not true, in general, that the wage of a worker with higher skill increases by more than the wage of a worker with lower skill. This is because two effects are in play. On the one hand, the increase in vertical differentiation increases the wages of high-skilled workers by more. On the other hand, however, the increase in the supply of skill means that workers with higher skill face tougher competition than previously, which lowers their wages in comparison to manufacturing workers of lower skill. In the services sector, the fall in the supply of skill is the only force affecting wages and the difference in wages earned by workers of any two levels of relative skill increases.

In addition to the results in Proposition 5, in the case of services we can conclude something more about measures of inequality other than range. Keeping the relative skill distribution constant, the variance of wages increases.[72] This means that the variance increases for the *incumbents*, that is, the workers who work in services in both the old and new equilibria.[73] It does not, however, necessarily imply an overall increase in wage variance in services, as the distribution of relative skill changes.[74] Keeping the wage

[71] If $R_M + R_S = 1$, then it is possible that $v_M^c = 0$ in both the old and new equilibria, in which case $w_M(0) = w_S(0) = 0$ and there is no change in wage for the least skilled worker.

[72] The increase in the difference in wages earned by workers of different skill levels means that—keeping the distribution of relative skill constant—the wage distribution spreads out in the Bickel and Lehmann (1979) sense; hence variance increases by Equation (3.B.25) in Shaked and Shanthikumar (2007).

[73] Formally, the set of incumbents is defined as $I = \{(v_M, v_S) : w_S(v_S, \theta_1) > w_M(v_M, \theta_1) \text{ and } w_S(v_S, \theta_2) > w_M(v_M, \theta_2)\}$.

[74] In fact, it is the change in the relative skill distribution that spreads out the wage in services in the first place.

function constant, the fall in the supply of relative skill has an ambiguous effect on variance, hence the overall effect of an increase in vertical differentiation of manufacturing workers on wage variance in services is ambiguous.

Both the range and the variance are measures of absolute wage inequality. In many contexts, however, relative inequality might be of more interest. In manufacturing, the ratio of wages earned by the highest skilled workers to wages earned by the lowest skilled workers strictly increases, because high-skilled workers earn strictly more, and workers of lowest skill strictly less, than previously. In services, however, the direction of the change in relative inequality depends on the workers' reservation wage. Thus far, the reservation wage has been normalized to 0, as it is of no import for the equilibrium, changes in sorting patterns, and changes in wage levels.[75] It does, however, matter for the change in relative inequality in services: A given increase in the difference in wages can result in a higher or lower ratio of wages, depending on what the original wage level was. In particular, if jobs are strictly scarce and the reservation wage is high enough, relative wage inequality increases in services, in that the ratio of wages earned by high-skilled workers to wages earned by low-skilled workers goes up.[76] Note that this could never happen in a model where workers are perfect substitutes.[77]

Proposition 6. If jobs are scarce and manufacturing workers become more vertically differentiated, then both total surplus and profits fall in services. If, further, the level of surplus in manufacturing increases universally, then total surplus produced in manufacturing increases as well, as does total surplus produced in the economy.

As was the case with wages, profits and total surplus in services are affected only by the fall in skill supply. Lower supply of skills means that firms are matched with less skilled workers, which decreases both their profits and the surplus they produce. In manufacturing, total surplus and profits are affected both by the increase in skill supply and the change in the surplus function. If surplus levels increase universally, then the surplus produced by any manufacturing firm increases and, as a result, total surplus in

[75] As long as every match produces more than the sum of the workers' reservation wage and the firms' reservation profit.

[76] We can write the ratio of wages earned by two different workers in services as

$$\frac{w_S(v_S'') + p_w}{w_S(v_S') + p_w} = 1 + \frac{w_S(v_S'') - w_S(v_S')}{p_w + w_S(v_S')},$$

where p_W denotes the workers' reservation wage and $w_S(\cdot)$ is the wage function under $p_W = 0$. If v_s'' is sufficiently close to 1 and v_S' is sufficiently close to $v_S^c(\theta_1)$, then it follows from Proposition 5 that $w_S(v_S'') - w_S(v_S')$ strictly increases, and thus we can write

$$p_w > \frac{(w_S(v_S''; \theta_1) - w_S(v_S'; \theta_1))w_S(v_S'; \theta_2) - (w_S(v_S''; \theta_2) - w_S(v_S'; \theta_2))w_S(v_S'; \theta_1)}{(w_S(v_S''; \theta_2) - w_S(v_S''; \theta_1)) - (w_S(v_S'; \theta_2) - w_S(v_S'; \theta_1))}.$$

Therefore, there exists a high enough p_w for which relative inequality increases.

[77] With perfect substitution, any change in skill supply in services affects wages of all workers proportionally, thus leaving relative inequality unchanged.

manufacturing rises. If vertical differentiation increases but surplus levels fall universally, the effect on production in manufacturing is ambiguous. An example is provided in Section 4.1.2.

The change in profits of manufacturing firms depends not only on the level of surplus and the supply of skill but also on the vertical differentiation of firms, that is, the change in $\frac{\partial}{\partial h}\pi_M$. This is explored in more detail in Sections 4.2 and 5.3.

4.1.1 Gaussian-exponential Specification

In the rest of my analysis of the scarce-jobs case, I provide three examples, each of them illustrating a different type of change that could cause an increase (or decrease) in vertical differentiation of workers. I do this using the following Gaussian-exponential (GE) specification of the model.

The vector $\mathbf{x}$ of basic skills is jointly normally distributed with mean $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Sigma}_x$. Productivity z, conditional on the firm operating in sector i, is distributed uniformly on $[\underline{\beta}_i, \bar{\beta}_i]$, with $\bar{\beta}_i > \underline{\beta}_i \geq 0$. The basic surplus function in sector i is given by $\Pi(\mathbf{x}, z, i) = \Big(A_i + \frac{1-e^{-\delta\boldsymbol{\alpha}_i\mathbf{x}^T}}{\delta}\Big)z^{\gamma_i}$, where $\boldsymbol{\alpha}_i = [\alpha_{i1}, \alpha_{i2}\ldots\alpha_{iN}]^T$ is an N-dimensional vector of (basic) skill requirements, $A_i > 0$ determines the extent to which productivity influences surplus *irrespective* of skill, δ determines the curvature of surplus as a function of skill, and $\gamma_i \geq 0$ determines the extent to which skills and productivity are supermodular. Note that with $\delta < 0$ this model is equivalent to a model in which $\ln(\mathbf{x})$ is jointly normally distributed and surplus is multiplicative: $(A_i + \boldsymbol{\alpha}_i\mathbf{x}^T)z_i^{\gamma_i}$.[78]

In the Gaussian-exponential specification, each sector uses some linear combination of the basic skill components in its production process. Denote the linear combination of skills required by sector i by $v_i'(\mathbf{x}) = \boldsymbol{\alpha}_i\mathbf{x}^T$, which gives us a vector of indices (v_M', v_S') that are $\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distributed, with

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_M \\ \mu_S, \end{bmatrix} = \boldsymbol{\alpha}\boldsymbol{\mu}_x,\ \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_M^2 & \rho\sigma_M\sigma_S \\ \rho\sigma_M\sigma_S & \sigma_S^2 \end{bmatrix} = \boldsymbol{\alpha}\boldsymbol{\Sigma}_x\boldsymbol{\alpha}^T, \text{ and } \boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_M \\ \boldsymbol{\alpha}_S \end{bmatrix}.$$

Normalizing v_i' and $z|i$ in such a way that their marginal distributions are standard normal, yields the following canonical formulation of the model:

$$\begin{aligned} \pi_i(v_i, h_i) &= \Big(A_i + \frac{1}{\delta}\big(1 - e^{-\delta\left(\Phi^{-1}(v_i)\sigma_i + \mu_i\right)}\big)\Big)\Big((\bar{\beta}_i - \underline{\beta}_i)h_i + \underline{\beta}_i\Big)^{\gamma_i}, \\ C(v_M, v_S) &= \boldsymbol{\Phi}_\rho\Big(\Phi^{-1}(v_M), \Phi^{-1}(v_S)\Big), \end{aligned}$$

where $\boldsymbol{\Phi}_\rho$ is the cdf of a standardized bivariate normal distribution with correlation ρ

[78]Note that if jobs are abundant in both sectors ($R_i > 1$), and if $\delta = -1$, $A_i = 1$, and $\gamma_i = 0$, then this specification reduces to the model in Roy (1951): The logarithm of skills is joint normally distributed, and the surplus function does not depend on firms' productivity.

and Φ is the cdf of the univariate standard normal distribution.[79] In the three following examples, the assumption that jobs are scarce is maintained; in general, however, the GE specification is applicable even if jobs are abundant.[80]

4.1.2 Public Investment in Education

Suppose the government would like to boost the total surplus produced in manufacturing, by investing in the training of a skill that is used more intensively in manufacturing than in services.[81] To make things really simple, I focus on the extreme case in which one of the basic skills (x_1) is (nearly) manufacturing specific, with $\alpha_{S1} \approx 0$.[82] Investment in the quality of x_1 training increases its mean μ_1. Table 1 provides a numerical example with three basic skills: x_1 is manufacturing specific, x_2 is services specific, and x_3 is a general purpose skill, used with equal weight in both sectors.

The investment in skill x_1 will have two effects: the *direct* effect and the *sorting* effect. The direct effect is the change that would occur if there were no re-sorting of workers; the sorting effect captures the impact of re-sorting. The direct effect is positive, because an improvement in μ_{M1} increases π_M for any possible match. The direction of the sorting effect, however, depends on whether the investment in x_1 increases or decreases the vertical differentiation of manufacturing workers. As a benchmark, note that for univariate normally distributed variables, a change in μ has no effect on the spread of their distribution. Therefore, for surplus that is linear in skill ($\delta \to 0$), an investment in μ_1 does not affect workers' vertical differentiation, hence there is no sorting effect and manufacturing expands solely via the positive direct effect.

The situation is quite different if δ is negative. In this case the basic surplus function is convex in skill, hence any improvement in the distribution of x_1 in the FOSD sense increases the differences in the surplus produced by workers of different relative skill. This causes a positive sorting effect in manufacturing, in line with Proposition 4. At the

[79]The Gaussian-exponential specification does not satisfy Assumption 1, as it is not defined for $v_i = 1$ and not differentiable for $v_i \in \{0, 1\}$. Formally, I solve this problem by working with a surplus function with an additional truncation parameter a_i, which approaches π_i as $a_i \to 0$ (see Online Appendix OA.5 for details). Then in simulations, I just set a_i close to 0. This procedure is equivalent to first aggregating $\mathbf{x}$ and then truncating the aggregate v_i' at values far removed from the mean. For expositional simplicity, I will ignore this problem in the main body of the paper.

[80]With no changes if $\delta < 0$; otherwise, v_i' needs to be truncated to ensure that $\pi_i > 0$, because otherwise π_M would admit negative values for $v_i \approx 0$, regardless of the values of other parameters. See Online Appendix OA.5 for details.

[81]There exists a plethora of reasons why this might be the government's goal, some of them economically justifiable, others less so. The government might be worried about deindustrialization for strategic reasons, or because the manufacturing sector is a powerful engine for productivity growth and experiences unconditional productivity convergence, unlike services (Rodrik, 2013). Alternatively, the government might be focused on manufacturing for purely political reasons: thinking manufacturing growth is more salient or having an electorate and donors that are concentrated in manufacturing.

[82]If the model was initially symmetric, this assumption is not needed; $\alpha_{M1} > \alpha_{S1}$ is sufficient. In the asymmetric case, however, the exact shape of the equilibrium separation function matters as regards how much more important x_1 needs to be in manufacturing than in services for Proposition 4 to apply.

Sector	Type of effect	% change in output for $\delta =$		
		−1 (convex)	0 (linear)	1 (concave)
Manufacturing	direct	0.5	0.24	0.26
	sorting	0.62	0	−0.32
	overall	**1.12**	**0.24**	**−0.07**
Services	**overall**	**−0.66**	**0**	**0.34**

Table 1: The effects of an increase in μ_1 from 0 to 1.
Computed for Gaussian-exponential specification: $N = 3$; $\mathbf{x} \approx \mathbf{N}(\mathbf{0}, \mathbf{I})$ ($\mathbf{I}$ is the identity matrix); $\boldsymbol{\alpha}_S = [0.1, 0, \sqrt{0.99}]^T$, $\boldsymbol{\alpha}_S = [0, 0.1, \sqrt{0.99}]^T$; $A_i = 41$; $\underline{\beta}_i = 2$, $\bar{\beta}_i = 3$, $\gamma_i = 2$, $R_i = 0.49$.

same time, the supply of skills decreases in services, causing its contraction.

If δ is positive, then surplus in services is concave in skills. By analogous reasoning, an increase in μ_1 makes manufacturing workers less vertically differentiated. This is plausible in the context of education, where often the greatest gains in productivity come from closing the gap between students that are already high achieving and those who are underperforming. However, if this is the case, then the sorting effect is negative: The investment in the manufacturing-specific skill makes firms less willing to hire workers of high relative skill and worsens sorting into that sector. In extreme cases, the negative sorting effect can dominate the positive direct effect and cause a decline in manufacturing (Table 1 provides an example). Thus this seemingly straightforward policy can easily backfire in this model. Note that this could never happen in a selection model with perfect substitution of workers.[83]

The results for the concave case have further counterintuitive implications. For example, if the government had to choose whether to invest in x_1 and x_2, then, wanting to boost manufacturing, its best option might be to invest in the *services-specific* skill. Furthermore, such an investment could lead to a contraction of services, implying that manufacturing expands by more than the economy as a whole. This puts in doubt the conclusions from Justman and Thisse (1997) and Poutvaara (2008), who argue that if workers can migrate between regions (sectors), then governments will necessarily under-invest in the training of skills, foreign-specific skills (here: services-specific skills) in particular.[84]

[83] In such models, the change in skill supply has two effects: It changes sorting and the relative price of the manufacturing task (which is analogous to the relative manufacturing skill in this model). Suppose manufacturing contracted following an increase in the supply of the manufacturing-specific skill. This is possible only if fewer efficiency units of the manufacturing task are supplied, which in turn implies that more efficiency units of the services task are supplied. But this would raise the relative price of the manufacturing task, which—together with the increase in the supply of x_1—would lead to an expansion of manufacturing. Contradiction!

[84] This literature models the strategic interactions much more carefully, which is beyond the scope of this paper; nevertheless, this example should make it clear that in my model it is not certain that there will be underinvestment in foreign-specific skills.

As an aside, this example is a good showcase for the advantages of the canonical formulation. In terms of the basic formulation, an increase in μ_1 affects only the (multivariate) sectoral supply of (basic) skills. In equilibrium, the sectoral supply of basic skills is driven partly by the increase in μ_1 and partly by sorting, generally with an ambiguous end effect. In the canonical formulation, this is decomposed into an improvement in the reduced surplus function, which captures the direct effect of the increase in μ_1, and a change in the sectoral supply of relative skill, which captures the effect on sorting. In particular, this allows for a meaningful comparison of the quality of workers who sort into manufacturing, even in cases when the distribution of basic skill has changed.

4.1.3 Productivity Distribution and Inequality Transmission

I will now address the impact of an improvement in the distribution of firms' productivity. The most natural interpretation of such an improvement is the introduction of a more efficient technology, although trade liberalization could have a similar effect (see Melitz, 2003; Sampson, 2014).[85] In the Gaussian-exponential specification, the distribution of firms' productivity z in sector i depends on two parameters: $\underline{\beta}_i$ and $\bar{\beta}_i$. An increase in either of them improves the distribution of productivity in the FOSD sense but their effect on the spread of productivity distribution differs. Specifically, an increase in $\underline{\beta}_i$ makes firms more similar, in the sense that $\frac{\partial}{\partial h}\pi_i$ falls, whereas $\bar{\beta}_i$ has the opposite effect. This difference matters as regards how firms' profits change but in the baseline model discussed here it has the same effect on sorting, qualitatively speaking.[86] This is because any improvement in the distribution of $z|i$ increases the basic productivity of a firm with relative productivity h_i. As basic surplus is strictly supermodular in the Gaussian-exponential specification (for $\gamma_i > 0$), this means that for any relative productivity h_i the difference in the surplus produced by workers of high and low relative skill increases. Thus surplus levels increase, and manufacturing workers become more vertically differentiated. This reasoning applies more generally than just for the Gaussian-exponential specification.

Lemma 4. If $\Pi(\mathbf{x}, z, i)$ is strictly increasing in productivity, then under Assumptions 1 and 3 an improvement in the distribution of $(Z|M)$ in the FOSD sense implies that manufacturing workers become more vertically differentiated.

Therefore, if jobs are scarce, Propositions 4–6 hold if either $\underline{\beta}_i$ or $\bar{\beta}_i$ increases. Consequently, the adoption of more efficient technologies makes the least skilled workers in that sector worse off (Proposition 5).[87] Crucially, the relation between the distribution

[85] Note, however, that in the Melitz model, trade liberalization also changes the mass of firms, which would then increase the number of workers. The effect of such expansion on wage inequality is discussed in Section 5.3.2.

[86] In the extended model from Section 5, where firms' entry decisions are endogenous, these two changes would have different effects on sorting.

[87] This is reminiscent of the effects in the task-based model of Acemoglu and Autor (2011), where, for example, high-skilled augmenting technology could result in a fall in wages for medium-skilled workers.

		% change in		
Sector	Type of effect	surplus	wage variance	wage range
Manufacturing	direct	19.35	82.32	39.53
	overall	20.57	−39.98	11.83
Services	overall	−1.035	67.24	11.42

Table 2: The effects of an increase in $\underline{\beta}_M$ from 2 to 2.5.
Computed for Gaussian-exponential specification: $N = 3$; $\mathbf{x} \approx \mathbf{N}(\mathbf{0}, \mathbf{I})$ ($\mathbf{I}$ is the identity matrix); $\boldsymbol{\alpha}_S = [0.1, 0, \sqrt{0.99}]^T$, $\boldsymbol{\alpha}_S = [0, 0.1, \sqrt{0.99}]^T$; $\delta = 1$; $A_i = 41$; $\underline{\beta}_i = 2$, $\bar{\beta}_i = 3$, $\gamma_i = 2$, $R_i = 0.49$.

of productivity and workers' differentiation implies that technological change which is restricted to just one sector can increase wage inequality in many industries. A numerical example using the concave case of the Gaussian-exponential specification is provided in Table 2. In this example, $\underline{\beta}_M$ increases and all manufacturing firms become more productive but also more similar. This affects wage inequality in manufacturing directly, increasing drastically both the wage range and the wage variance.[88] However, the sorting effect works in the opposite direction: Wage inequality falls in manufacturing and increases in services. Overall, and in line with Proposition 4, the wage range increases in both sectors. The increase is stronger in manufacturing (which is true in general, not just in this example) but the magnitude is very similar in the two sectors. The results for variance are even more striking, albeit less general: The variance *decreases* in manufacturing overall but increases in services. This suggests that, for example, Rosen's (1981) superstar effect could be driving the increases in wage inequality even in sectors that are not directly affected by the improvements in communication technology.

The effect of an improvement in $\underline{\beta}_i$ on total surplus is as implied by Proposition 6 and Lemma 4: There is an increase in manufacturing and a fall in services.

4.1.4 Inter-regional Competition for Skills as a Force for Inequality

Suppose that the two sectors are concentrated in two distinct regions: the manufacturing region and the services region. In the English context this could be thought of as the North of England and London. The manufacturing region considers an investment in regional broadband infrastructure. I will assume that such an investment would improve the surplus produced by all matches in manufacturing but particularly so for matches

[88] As explained in footnote 79, in the simulations I truncate the distribution of $\alpha_{1i}x_1 + \alpha_{2i}x_2 + \alpha_{3i}x_3$, and only for that reason there is any change in wage range—without this truncation wage range is always equal to infinity. For that reason, the object called "wage range" in Tables 2 and 3 should perhaps be more correctly thought of as an approximation of the change in the difference between the wages earned by workers at the 0.1 and 99.9 percentile in the within-sector wage distribution.

Sector	Type of effect	% change in surplus	wage variance	wage range
Manufacturing	direct	9.86	23.42	11.39
	overall	10.24	10.31	10.92
Services	overall	−0.66	−31.57	0.71

Table 3: The effects of an increase in γ_M from 2 to 2.1.
Computed for Gaussian-exponential specification: $N = 3$; $\mathbf{x} \approx \mathbf{N}(\mathbf{0}, \mathbf{I})$ ($\mathbf{I}$ is the identity matrix); $\boldsymbol{\alpha}_S = [0.1, 0, \sqrt{0.99}]^T$, $\boldsymbol{\alpha}_S = [0, 0.1, \sqrt{0.99}]^T$; $\delta = -1$; $A_i = 41$; $\underline{\beta}_i = 2$, $\bar{\beta}_i = 3$, $\gamma_i = 2$, $R_i = 0.49$.

involving high-skilled workers.[89] Formally, this will be captured by an increase in the exponent γ_M.[90]

An increase in γ_M increases both the level of surplus and workers' vertical differentiation in manufacturing. Thus its direct effect is an increase in total surplus and an increase in inequality. After workers re-sort, total surplus increases further, while the wage range decreases. In the numerical example provided in Table 3, the overall effect on the wage variance is positive in manufacturing and negative in services.[91] In services, total surplus falls and the wage range increases.

Suppose that the two sectors are symmetric and both regions need to decide whether to invest in broadband. In particular, suppose that the direct increase in total surplus is lower than the cost of the investment but the overall increase is well worth the cost, regardless of what the other region decides. The two regional governments have to simultaneously choose whether to invest in broadband infrastructure, and their payoff is the percentage change in average surplus net of the investment cost per capita (if any). The strategic form of this game is presented below for the specification from Table 3 and investment cost per capita equal to 10% of the pre-game average surplus.

		Manufacturing: $\gamma_M = 2$	Manufacturing: $\gamma_M = 2.1$
Services	$\gamma_S = 2$	0%, 0%	−0.66%, 0.24%
	$\gamma_S = 2.1$	0.24%, −0.66%	−0.14%, −0.14%

This is clearly a case of the Prisoner's Dilemma, with both governments investing in broadband in equilibrium. This means that neither of them improves the supply of skills in their region and their gains are limited to the direct effect of the investment, which is not worth its cost. As a consequence, wage inequality increases in both regions as well (see

[89] For suggestive evidence that broadband internet is indeed a complement of skill see Akerman, Gaarder, and Mogstad (2015).

[90] I will assume here that $\underline{\beta}_M \geq 1$, to ensure that an increase in γ_M raises the surplus.

[91] As we have seen in Section 4.1.3, this effect is not general.

Table 3).[92] In general, investments that increase workers' vertical differentiation impose a negative externality on other regions while also increasing overall wage inequality. This suggests that interregional or international competition for high-skilled workers is a force for greater inequality.[93]

4.2 Abundant Jobs

If jobs are abundant ($R^1+R^2 > 1$), the level of surplus plays a role in determining whether a firm hires any worker at all or exits the market, that is, in determining the extensive margin of a firm's hiring decision. In particular, if there is no change in workers' vertical differentiation in manufacturing but the level of surplus falls, then some low-productivity manufacturing firms will likely decide to leave the market, which will shift the demand for relative manufacturing skill downward. To address this, in this section I focus on changes in reduced surplus that both increase workers' vertical differentiation and increase the levels of surplus.

Proposition 7. If (a) jobs are abundant, (b) surplus levels in manufacturing increase universally, and (c) manufacturing workers become more vertically differentiated, then more relative skill is supplied to manufacturing and less to services in equilibrium ($S_M(v_M)$ increases and $S_S(v_S)$ falls for all v_i). If the increase in vertical differentiation is strict, then the changes in relative skill supply are strict for some v_i.

Let us again consider the change in manufacturing firms' hiring decisions after the reduced surplus function has changed but before wage functions have adjusted. By the logic outlined in Section 4.1, every firm will want to hire a more skilled worker than previously. Additionally, some firms that did not find it profitable to hire anyone previously will now decide to hire a low-skilled worker, because of the increase in surplus levels. Thus again the demand for relative skill in manufacturing shifts upward, which draws in additional workers from services, so that employment rises in manufacturing and falls in services.[94] Note that this time some of those additional workers could be of relatively low skill, as the increase in surplus levels implies that manufacturing generally becomes more productive relative to services.

The difference between the scarce- and abundant-jobs cases is that in the former only the intensive margin of demand (which worker is hired by firm h_M) matters, whereas in the latter the extensive margin (whether firm h_M hires any worker) matters as well.

[92]Note that because there is no re-sorting in the equilibrium of the investment game, both sectors will experience just the direct effect of the investment.

[93]In contrast to the example in Section 4.1.3, the insights from this example carry over virtually unchanged even if firms' entry decisions are endogenous, especially in Costrell–Loury specifications (defined in Section 5.1.1).

[94]The increase (fall) in manufacturing (services) employment follows immediately from the increase (fall) in skill supply.

Furthermore, if jobs are abundant in both sectors ($R_i \geq 1$), then in the important special case of Roy-like models the intensive margin does not matter at all for the equilibrium supply of relative skill.[95] In general, a demand shift at the intensive margin increases the relative market power of high-skilled workers, allowing them to receive a greater share of surplus. In Roy-like models, however, firms are not sufficiently heterogeneous to have any market power at all, as each firm can always be replaced by an identical, unmatched company. Therefore, workers always receive the entire surplus, regardless of how differentiated they are. More generally, vertical differentiation of workers has an impact on equilibrium sorting only if firms' heterogeneity is substantial enough to allow at least some of them a degree of market power.

Proposition 8. If (a) jobs are abundant, (b) surplus levels in manufacturing increase universally, and (c) manufacturing workers become more vertically differentiated, then all wages in services increase, as do wages of the most skilled manufacturing workers.

Less skill is supplied to services, and thus again wages rise.[96] In manufacturing, the wages of high-skilled workers rise, as they are in higher demand. However, the change in wages of the least skilled workers is ambiguous, because the demand for those workers might not have fallen. While it is true that the firms that employed them previously will now demand better workers, manufacturing firms that previously were not employing anyone might now want to hire those low-skilled workers.

The impact on the wage range is ambiguous in both sectors, as in certain cases the increase in surplus levels may be equality enhancing.

Proposition 9. If (a) jobs are abundant, (b) surplus levels in manufacturing increase universally, and (c) manufacturing workers become more vertically differentiated, then both total surplus and profits fall in services, whereas the total surplus produced in manufacturing rises.

In the abundant-jobs case, the results for profits and total surplus are exactly the same as in the scarce-jobs case, as is the intuition behind them. In the abundant-jobs case, however, it is particularly easy to see why an increase in workers' differentiation and surplus levels is not enough to ensure an increase in manufacturing profits. Suppose that the old surplus function $\pi_M(v_M, h_M; \theta_1) = v_M h_M$ changes to $\pi_M(v_M, h_M; \theta_2) = v_M$ and that $R_M > 1$. This increases both the surplus produced in any match and the marginal surplus of skill, in both cases because $h_M \leq 1$. Under the new surplus function, however, firms are identical and abundant, and thus realize no profit (regardless of the supply of skill in manufacturing).

[95] A formal result is provided in Online Appendix OA.3.

[96] It is worth noting that although the difference in the wages earned by high- and low-skilled workers increases, this does not necessarily result in an increase in the wage range. The reason is that critical skill in sector two may rise as the measure of active services firms falls.

5 Endogenous Entry of Firms

In this section, I extend the model to allow for endogenous firm entry. The equilibrium of the extended model exists and is unique in a sense stronger than in the baseline model: The zero-expected-profits condition pinpoints wages for all matched workers. I also show that the comparative statics results on the impact of (a) greater skill concordance on wage polarization (Proposition 3 from Section 3) and (b) increased worker differentiation on skill supply (Proposition 7 from Section 4) generalize naturally. All the analysis in this section focuses exclusively on the canonical formulation of the model.

5.1 The Extended Model

I will model firm entry by extending the approach from Melitz (2003) to the two-sector case. Suppose that there exists an unlimited supply of *potential* firms that are *ex ante* identical. If a potential firm decides to enter sector i, it pays cost $c_i > 0$ and then draws productivity h_i from a standard uniform distribution. This implies that the measure of firms in sector i (i.e., R_i) is determined endogenously but the *ex post* distribution of productivity in each sector is always standard uniform.

Recall that, for given sectoral wage functions w_M, w_S, the profit of firm h_i in sector i is given as

$$r_i(h_i) = \max_{v \in [0,1]} \ \pi_i(v, h_i) - w_i(v).$$

Then the expected profit in sector i can be defined as

$$\bar{r}_i = \int_0^1 \max\{r_i(h), 0\} \mathrm{d}h.$$

Firms enter the sector that maximizes their expected profits net of entry cost (if any). This implies that if entry is positive in sector i ($R_i > 0$), then the expected profit must be equal to the cost of entry: $\bar{r}_i = c_i$.

Definition 8 (Competitive Equilibrium)**.** An equilibrium consists of sectoral skill supply functions S_M, S_S, sectoral skill demand functions D_M, D_S, and sectoral wage functions w_M, w_S that satisfy conditions (i)–(iii) from Definition 1, as well as (iv) two *sectoral measures of firms*, $R_M, R_S \in \mathbf{R}_{\geq \mathbf{0}}$, such that $\bar{r}_i = c_i$ if $R_i > 0$ and $\bar{r}_i \leq c_i$ otherwise.

It follows that any equilibrium of the extended model must also be an equilibrium of the baseline model.[97] Formally, denote by E_B the set of all such worker and firm allocation quadruples $E = (S_M, S_S, R_M, R_S)$ such that the supply functions (S_M, S_S) hold in an

[97]Note that if $R_i = 0$ for some $i \in \{M, S\}$, then the baseline model reduces to the standard single-sector model from Sattinger (1979).

equilibrium of the baseline model if the sectoral firm measures are $R_M, R_S \in \mathbf{R}_{\geq \mathbf{0}}$.[98] For any $R_M, R_S > 0$, the corresponding S_M, S_S were characterized in Section 2. If $R_i = 0$, then $S_i(v) = 0$ and $S_j(v) = \min\{1 - v, R_j\}$ in any equilibrium of the baseline model.[99]

Analogously, denote by $\mathbf{E_E}$ the set of all quadruples $E = (S_M, S_S, R_M, R_S)$ such that S_M, S_S are the supply functions and R_M, R_S the sectoral firm measures that hold in an equilibrium of the extended model. Clearly, $E \in \mathbf{E_E}$ if and only if $E \in \mathbf{E_B}$ and satisfies condition (iv) from Definition 8.

Existence and Uniqueness In order to show existence and uniqueness of the equilibrium, it will be useful to first show that results analogous to the two Fundamental Welfare Theorems hold in the extended model.

It has been known since at least Gretsky et al. (1992) that in an assignment model (such as the baseline model from Section 2) any equilibrium is efficient and any efficient assignment is an equilibrium. The extended model presented here, however, is not a special case of the model in Gretsky et al. (1992), and thus efficiency of equilibria needs to be established separately.

The total gross surplus is defined analogously to the total surplus in the baseline model (Equation (18)). That is, total gross surplus produced in sector i in equilibrium $E \in \mathbf{E_B}$ is equal to the sum of surpluses produced by all workers who joined sector i, taking into account that within-sector matching is positive and assortative:

$$T_i(E) = \begin{cases} 0 & \text{if } R_i = 0, \\ \int_1^0 \pi_i\left(v_i, 1 - \frac{S_i(v_i)}{R_i}\right) \mathrm{d}S_i(v_i) & \text{otherwise.} \end{cases} \tag{23}$$

The total net surplus produced in the economy is equal to the sum of the gross surplus produced in the two sectors net of entry costs: $V(E) = T_M(E) + T_S(E) - c_M R_M - c_S R_S$.

Proposition 10. A worker and firm allocation $E \in \mathbf{E_B}$ can hold in an equilibrium of the extended model if and only if it uniquely maximizes the total net surplus, that is,

$$E^* \in \mathbf{E_E} \Leftrightarrow V(E^*) - V(E') > 0 \text{ for all } E \in \mathbf{E_B} \setminus \{E^*\}.$$

This result can be interpreted as an analogue of the First and Second Welfare Theorems for this economy. First, it means that any equilibrium is efficient. Secondly, it means that any efficient allocation of workers and firms to sectors holds in some equilibrium.[100] Finally, it implies that any equilibrium allocation of workers and firms E

[98] Formally, the quadruples in E are such that if the sectoral firm measures are R_M, R_S, then there exist wage functions w_M, w_S that, together with the supply functions S_M, S_S and demand functions $D_M = S_M, D_S = S_S$, satisfy the conditions from Definition 1.

[99] The first part is trivial. The second part follows from the fact that all workers will be available to join the other sector but market clearing requires that at most R_j can actually be hired by sector j firms.

[100] This is because any efficient allocation of workers given R_M, R_S is an equilibrium of the baseline model, which follows from the results in Gretsky et al. (1992).

maximizes total net surplus *uniquely*. It follows trivially that if an equilibrium exists it must be (essentially) unique.

Theorem 2. An equilibrium exists and is essentially unique, in that the equilibrium measure of firms entering each sector, as well as relative skill supply and demand, are unique. Further, the equilibrium wage functions are uniquely determined for all matched workers (i.e., for $v_i \geq v_i^c$).

Both the existence results and the uniqueness results are new.[101] This is in contrast to the baseline model, where uniqueness of the equilibrium was a new result but its existence could have easily been shown from existing results for assignment models (Gretsky et al., 1992). Further, the uniqueness of equilibrium is stronger here than in the baseline model, as wages are *de facto* uniquely determined even if $R_M + R_S = 1$. This is because constant average profits pinpoint the split of surplus in the least productive match (see below).

5.1.1 Relation to Other Models

The extended model nests Roy's model in a manner similar to the baseline model. However, as far as single-sector assignment models are concerned, the extended model nests the model of Costrell and Loury (2004) rather than that of Sattinger (1979).[102]

Costrell and Loury (2004) In the hierarchical job assignment model of Costrell and Loury (2004), firms are homogeneous but consist of a hierarchy of heterogeneous jobs. The surplus produced by a firm is simply the sum of the surpluses produced by all the jobs (and all of them need to be filled to produce anything). Surplus produced in any job is supermodular in the job's rank and the skill of the worker assigned to it. The zero profit condition ensures that, in equilibrium, the measure of all jobs is equal to the measure of workers. Because of positive and assortative matching, a worker with skill v is assigned to a job of rank $h = G(v)$ (where $G(\cdot)$ denotes the cdf of skill). Using my notation, the wage paid to a worker with skill v_S in the Costrell and Loury model is

$$w^{\mathrm{CL}}(v) = \pi(v, G(v)) + \int_0^1 \int_{G(v)}^h \frac{\partial}{\partial h_i} \pi_M\Big(G^{-1}(t), t\Big)\, \mathrm{d}t\, \mathrm{d}h. \tag{24}$$

In equilibrium, the more productive and profitable jobs cross-subsidize the less productive jobs, leading to firm-wide profit of zero.[103]

[101] The only other paper I am aware of that allows for endogenous entry of firms in an assignment model is Costrell and Loury (2004), which is a single-sector, one-dimensional model.

[102] Sattinger (1979) could easily be nested as well if we allowed the measure of *potential* firms to be finite.

[103] Technically, Costrell and Loury (2004) allow only for multiplicative surplus functions, in the form $\mu(v)\beta(h)$. There is no problem, however, with generalizing their framework to supermodular surplus functions.

The wage function in the extended model is a generalization of Equation (24). To see this, denote the *positive and assortative matching function* in sector i by $P_i(v_i) = 1 - \frac{S_i(v_i)}{R_i}$, and its inverse by P_i^{-1}. Then the profit of firm h_i in sector i can be written as

$$r_i(h_i) = \int_{h_i^c}^{h_i} \frac{\partial}{\partial h_i} \pi_i\Big(P_i^{-1}(h), h\Big) \mathrm{d}h + r_i(h_i^c), \tag{25}$$

where $h_i^c = P(v_i^c)$ denotes the productivity of the least productive matched firm. The profit earned by said firm is pinned down by the zero-expected-profits condition:

$$r_i(h_i^c) + \int_{P_i(v_i^c)}^{1} \int_{P_i(v_i^c)}^{h} \frac{\partial}{\partial h_i} \pi_M\Big(P_M^{-1}(t), t\Big) \mathrm{d}t\mathrm{d}h = c_M. \tag{26}$$

The wage received by a worker of skill v_i is $w_i(v_i) = \pi(v_i, P_i(v_i)) - r_i(P_i(v_i))$. Substituting Equations (25) and (26) into this expression yields

$$w_i(v_i) = \pi(v_i, P_i(v_i)) + \int_{h_i^c}^{1} \int_{P_i(v_i)}^{h} \frac{\partial}{\partial h_i} \pi_i\Big(P_i^{-1}(t), t\Big) \mathrm{d}t\mathrm{d}h - h_i^c r_i(P_i(v_i)) - c_i. \tag{27}$$

This is similar to Equation (24) but the two wage functions differ if the skill and matching functions are not identical, that is, if $R_i \neq S_i(0)$. In addition to this, in the Costrell and Loury (2004) model all workers are employed and all tasks must be filled (by assumption) but this is not necessarily the case here. However, the extended model nests a two-sector version of the hierarchical job assignment model if the total measure of firms is necessarily equal to 1 in equilibrium.

Assumption 4 (Costrell–Loury Specification)**.** $\pi_i(1,1) - \pi_i(1,0) \leq c_i$ for both $i \in \{M, S\}$, and $\pi_i(0,0) > c_i$ for some $i \in \{M, S\}$.

This ensures that the total measure of firms, $R_M + R_S$, is equal to 1 in equilibrium.[104] Any specification of the extended model that meets Assumption 4 will be referred to as a *Costrell–Loury (CL) specification.* In Sections 5.2 and 5.3, I will focus on Costrell–Loury specifications, as they are much more tractable than the general model because of the property that the measures of workers and firms are equal. Nevertheless, in formal results Assumption 4 will not be imposed globally; it will be invoked explicitly wherever necessary.[105] In any CL specification, a worker with skill v_i matches a firm with productivity $h_i = G_i(v_i)$, and her wage is $w^{\mathrm{CL}}(v_i)$. Further, every firm hires a worker, and all

[104] If $R_M + R_S < 1$, then $\bar{r}_i \geq \pi_i(0,0) > c_i$ in one of the sectors, violating the zero-expected-profit condition. Similarly, if $R_M + R_S > 1$, then $r_i(0) = 0$ in at least one sector, implying that $c_i \geq \int_0^1 \frac{\partial}{\partial h_i} \pi_i\Big(1, h\Big) \mathrm{d}h \geq r_i(1) \geq \bar{r}$, and thus again violating the zero-expected-profit condition. Note, by the way, that Assumption 4 can be weakened significantly. For example, $\int_0^1 \int_0^h \frac{\partial}{\partial h_i} \pi_M\Big(1, t\Big) \mathrm{d}t\mathrm{d}h \leq c_M$ is sufficient, as the LHS must be greater than $\bar{r}$.

[105] That is, any result that does not invoke Assumption 4 holds generally.

workers are employed. Therefore, the Costrell–Loury specification of my model can be reinterpreted as a model in which firms are homogeneous within each sector but consist of a hierarchy of heterogeneous jobs.[106] I use this fact in Online Appendix OA.6 to develop a dynamic, overlapping generations version of the extended model[107]

Trivially, the fact that the extended model nests a two-sector version of the Costrell–Loury model implies that it can also reduce to its single-sector version. Firstly, suppose that services are an empty sector, which is ensured if $\pi_S(1,1) < c_S$. Then Assumption 4 implies that wages in manufacturing are given by $w^{\text{CL}}(v_M)$ if calculated for $G_M(v_M) = v_M$. Alternatively, the extended model reduces to the Costrell–Loury model for any symmetric Costrell–Loury specification.[108] In any such specification, wages are also given by $w_i^{\text{CL}}(v_i)$ with $G_i(v_i) = C(v_i, v_i)$. This is equivalent to a Costrell–Loury model in which workers' skill is $v = \max\{v_M, v_S\}$.

Roy (1951) The extended model can reduce to Roy's self-selection model in the same two ways as the baseline model. Firstly, suppose that the surplus function does not depend on a firm's type ($\frac{\partial}{\partial h}\pi_M(\bullet) = \frac{\partial}{\partial h}\pi_S(\bullet) = 0$). In such a case, firms can earn enough profit to warrant entry only if (weakly) fewer firms enter the market in equilibrium than there are workers ($R_M + R_S \leq 1$). Then all firms in sector i earn exactly the same profit, which—in equilibrium—is equal to the entry cost. It follows, then, that workers' wages are given exogenously and are equal to $\pi_i(v_i) - c_i$.[109]

The second case is when entry into each sector is truly free, that is, $c_i \to 0$ for all $i \in \{M, S\}$. Then the number of firms in each sector is unlimited ($R_i \to \infty$), and the equilibrium wage function becomes $w_i(v_i) = \pi_i(v_i, 1)$ by the same logic as in Section 2.2.3.

5.1.2 Complements and Substitutes

I will now show that in the extended model, workers of very high and very low skill levels are Hicks complements, that is, that the arrival of additional low-skilled workers in a sector increases the wages of high-skilled workers in that sector and *vice versa*. This complementarity will be of crucial importance for the comparative statics results

[106] Note that the extended model can be reinterpreted in this way more generally. If we adopt this interpretation, then the only difference between the extended model and a two-sector version of Costrell and Loury (2004) lies in the facts that in my model (a) the hierarchical firms are allowed to leave unprofitable jobs vacant and (b) the cost of entry can be positive.

[107] The fact that each sector consists of homogeneous, hierarchical firms means that the within-sector assignment is problem solved within firms within-firm, and thus is *de facto* decided centrally. This allows me to side-step the issue of how is the stable assignment reached in a decentralized dynamic process.

[108] The definition of the symmetric case in the extended model differs from Definition 2 only in that the condition $R_M = R_S$ is replaced by $c_M = c_S$, which—provided all other symmetry conditions are met—implies that $R_M = R_S$ in equilibrium.

[109] In fact, if the cost of entry is high enough, that is, if $c_i \geq \int_0^1 \pi_i(0, h)\mathrm{d}h$, then the same logic holds even if surplus is not supermodular but does depend on firm's type, that is, if $\pi_i(v_i, h_i) = \mu_i(v_i) + \beta_i(h_i)$, where $\mu(\cdot)$ and $\beta(\cdot)$ are strictly increasing. To see this, note that in such a case $h_i^c = 0$ and $r_i(0) = c_i - \int_0^1 \pi_i(0, h)\mathrm{d}h$, giving $w_i(v_i) = \pi_i(v_i, 0) + \int_0^1 \pi_i(0, h)\mathrm{d}h - c_i$.

presented in Sections 5.2 and 5.3, as it is in contrast to the baseline model, where (within each sector) all workers were imperfect substitutes. Nevertheless, the complementarity between workers of different skill levels should come as no surprise, as this is also the case in Costrell and Loury (2004).

Lemma 5. Suppose that the supply of skill changes in services for a reason exogenous to that sector (e.g., in response to a technological change in manufacturing). If there exists some $v'_S \in [\max v_i^c, 1]$ such that $w_S(v_S)$ changes strictly, then for $j \in \{1,2\}$ there exist $v_S^j \in [v_S^c(\theta_j), 1]$ such that $w_S(v_S^1)$ and $w_S(v_S^2)$ strictly increases and decreases, respectively.

To see why this must be the case, first note that a fall in the wage paid to any worker implies an increase in the profit of the firm that employed her originally.[110] Therefore, a fall in wages of all workers is inconsistent with the zero-expected-profits condition. Put differently, as average profits remain unchanged, some firms must gain and some must lose; however, as there is no change in the surplus function, firms' gains must come from some workers' losses (and *vice versa*).

Lemma 5 tells us that workers with *some* skill levels must be complements but is silent on *which* ones. To establish this, I follow the method introduced in Section III.B of Costrell and Loury (2004). Without loss of generality, I will focus on the single-sector version of my model (with $R_M = 0$) and consider an addition of a measure Δ_p of workers with skill $V_S = p$ to the labor force in services.[111] The notation is chosen so that Δ_{v_S} represents the additional measure of workers with skill v_S. I will examine the response of the wage of workers with skill $V_S = r$ to an infinitesimal Δ_p (i.e., $\frac{\partial}{\partial \Delta_p} w_S(r)$ evaluated at $\Delta_p = 0$).[112] I will also restrict attention to the case of strictly supermodular surplus functions.[113]

Proposition 11. Workers with similar skill levels are substitutes $(\frac{\partial}{\partial \Delta_p} w_S(p)) < 0)$, whereas workers with sufficiently different skill levels are complements (if $p = v_S^c$ and $r = 1$, then $\frac{\partial}{\partial \Delta_p} w_S(r)) = \frac{\partial}{\partial \Delta_r} w_S(p)) > 0$).

To see why workers of very different skill levels must be complements, note that with no change in the measure of firms, the addition of extra workers will increase average profits. Therefore, the number of firms must increase in equilibrium. If the newcomers are of low skill, this means that high-skilled workers are matched with more productive

[110]This follows from profit maximization: $r_S(P_S(v_S;\theta_1);\theta_2) \geq \pi_S(v_S, P_S(v_S;\theta_1)) - w_S(v_S;\theta_2) > \pi_S(v_S, P_S(v_S;\theta_1)) - w_S(v_S;\theta_1) = r_S(P_S(v_S;\theta_1);\theta_1)$.

[111]I focus on the single-sector version here to obtain simple expressions for the supply and matching functions. However, all results hold for changes in the supply of skill in services that are exogenous to that sector. To see this, note that we can always define a new variable u_S such that the supply in services at u_s is given by $S_S(u_S) = 1 - u_S$ and then write surplus as a function of u_S.

[112]I will suppress $|_{\Delta_p=0}$ from notation, so that $\frac{\partial}{\partial \Delta_p}|_{\Delta_p=0} w_S(r)$ will be denoted simply by $\frac{\partial}{\partial \Delta_p} w_S(r)$.

[113]Note that if $\frac{\partial^2}{\partial v_S \partial h_S} \pi_S(v_S, h_S) = 0$ for all matches, then the wage of each worker is equal to the surplus they produce and thus does not depend on the supply of skill.

firms than previously, which increases their wages. The effect is symmetric, and thus additional high-skilled workers increase wages of the least skilled workers. That workers of very similar skill levels must be substitutes follows from the fact that Lemma 5 implies that *some* workers must be substitutes and, of course, the closest substitute for a worker with skill p is another worker with skill p.

Proposition 11 generalizes Proposition 2 in Costrell and Loury (2004). In particular, if Assumption 4 holds, then taking the derivative of Equation (27) with respect to Δ_p gives, for $p < r$:

$$\frac{\partial}{\partial \Delta_p} w_S(r) = -\int_0^p h^2 \frac{\partial^2}{\partial v_S \partial h_S} \pi_S(h,h)\,\mathrm{d}h + \int_p^r h(1-h) \frac{\partial^2}{\partial v_S \partial h_S} \pi_S(h,h)\,\mathrm{d}h \\ - \int_r^1 (1-h)^2 \frac{\partial^2}{\partial v_S \partial h_S} \pi_S(h,h)\,\mathrm{d}h = \frac{\partial}{\partial \Delta_r} w_S(p), \tag{28}$$

which is the same as Equation (13) in Costrell and Loury (2004).[114] This expression makes it abundantly clear that workers of very similar skill levels must be substitutes, as for $p = r$ the middle term vanishes. For an excellent and detailed discussion of this expression and the intuition behind the complementarity between workers of very different skill levels, see Section III.B in Costrell and Loury (2004).[115]

The value added by my analysis is the insight that firms need not be homogeneous (which is the assumption in Costrell and Loury (2004)) for workers of different level of skill to become complements as long as the number of jobs in the economy is determined endogenously. However, endogenous entry is not a sufficient condition for workers of different skill levels to become complements. The sufficient condition, which is met in this model, is that expected profits be constant; this is required both for the wage function to be of the form derived in Equation (27) and for Lemma 5 to hold.[116]

[114]See Online Appendix OA.4 for details of the derivation.

[115]Note that Costrell and Loury (2004) restrict attention to multiplicative surplus functions of the form $\pi(v,h) = \mu(v)\beta(h)$, where $\mu(\cdot)$ and $\beta(\cdot)$ are strictly increasing functions. The more general surplus function allowed here changes slightly their insights with respect to the factors determining the number of workers whose wages decrease as a result of an arrival of additional low-skilled workers. With a multiplicative surplus function, this depends mostly on the convexity of the surplus function with respect to h. However, as we can see by inspection of Equation (28), what truly matters is the *gradient of the cross-partial* of the surplus function with respect to h. It just so happens that with multiplicative surplus this is captured by the convexity of surplus.

[116]Endogenous entry does not guarantee constant expected profits. For example, suppose that the measure of *potential* firms is limited (but very large) and that those potential firms differ *ex ante* in the cost of entry, according to some distribution. Then expected profits would need only to be greater than the entry cost of the *marginal* firm, which in turn would depend on the number of firms in the sector. In limiting cases, where some firms find it very cheap to enter but for others the cost of entry is prohibitively expensive, this even more general model would reduce to the standard assignment model, in which all skill levels are (imperfect) substitutes.

5.2 Skill Interdependence and Wage Polarization

In this section, I revisit the link between relative skill interdependence and wage polarization. I focus on Costrell–Loury specifications of the extended model, that is, I assume that Assumption 4 holds, and I find that increases in interdependence have an impact on the extended model which is qualitatively similar to their impact on the baseline model. Analogously to Section 3, I will first discuss the composition and wage effects in general, and then show that in the single-sector Costrell–Loury model the composition effect must dominate and increase wage polarization in absolute terms.

Note that the expressions for the distribution of wages, the inverse distribution of wages, and the decomposition into the wage and composition effects in the extended model are exactly the same as in the baseline model.

The composition effect is unchanged compared to the baseline model. The baseline and extended models differ only in how wages react to changes in the supply of skill. However, as the composition effect keeps wages constant, it must be the same (starting from the same wage function). Therefore, the composition effect of increased skill interdependence increases wage polarization in both absolute and relative terms.

The wage effect is still equal to the sum of changes in wages in each sector, weighted by the probability that a worker of given rank works in that sector. However, the change in wages is going to be different, because high-skilled workers and low-skilled workers are now complements.

In Costrell–Loury specifications, symmetry ensures that the wage effect increases wage inequality in both absolute and relative terms (see below). It follows, therefore, that only falls in the supply of skill that are biased toward one of the sectors could possibly create a wage effect that increases wage polarization.

The overall effect is ambiguous. As in the baseline model, if the extended model is symmetric, then the composition effect dominates the wage effect and increases wage polarization.

5.2.1 The Symmetric Costrell–Loury Specification

I defined the symmetric Costrell–Loury specification in Section 5.1.1 and showed that it is equivalent to the model of Costrell and Loury (2004), that is, the standard single-sector assignment model in which high-skilled workers and low-skilled workers are complements.[117] All results derived in this section also hold for any decrease in skill supply in the FOSD sense in the Costrell and Loury (2004) model.

In the symmetric case, the wage effect of an increase in skill concordance results in

[117] This is analogous to Section 3.4, which focuses on a specification equivalent to the Sattinger model, that is, the standard assignment model with imperfect substitution of skill.

a worse distribution of skill in each sector, because $G_i(v_i) = C(v_i, v_i)$. By Proposition 4 in Costrell and Loury (2004), this decreases the lowest wages and increases the highest wages. Additionally, it increases the difference in wages earned by workers of any two skill levels. Therefore, the wage effect clearly increases wage inequality in both absolute and relative terms, thus counteracting the increase in wage polarization caused by the composition effect. Nevertheless, similarly to the baseline model, the overall effect of an increase in relative skill interdependence raises wage polarization in absolute terms under fairly general conditions. However, to ensure that polarization increases also in relative terms, a slightly stronger notion regularity is needed than the one defined in Section 3.4. I will say that a change in interdependence is *strongly regular* if $\frac{\mathrm{d}}{\mathrm{d}v}\big(C_{v_i}(v,v,\theta_2) - C_{v_i}(v,v,\theta_1)\big)|_{v=0} > 0$. This implies that $C(v,v,\theta_2) - C(v,v,\theta_1) > 0$ for v close to 0.

Proposition 12 (Wage Polarization)**.** Suppose that the model is symmetric, Assumption 4 holds, the concordance of the relative skills distribution increases, and this increase is regular. Then wage polarization increases in absolute terms. If, further, the change in strongly regular and either (i) the lowest wage was sufficiently high originally or (ii) the supermodularity of the surplus function is sufficiently weak, then polarization increases in relative terms as well.

Note that in the Costrell–Loury model the difference between the wage received by a workers at rank t and the lowest wage is exactly the same as in the baseline model with $R_M = R_S = \frac{1}{2}$, that is,

$$W(t) - W(0) = \int_0^{G_i^{-1}(t)} \frac{\partial}{\partial v_i}\pi_i(s, G_i(s))\,\mathrm{d}s.$$

Therefore, the increase in wage polarization in absolute terms follows by exactly same reasoning as in Section 3.

The case of wage polarization in relative terms is more complicated, as

$$\log W(t) - \log W(0) = \int_0^{G_i^{-1}(t)} \frac{\frac{\partial}{\partial v_i}\pi_i(G_i^{-1}(s), s)}{g(G_i^{-1}(s))W(s)}\,\mathrm{d}s$$

and thus depends also on the level of wages. As noted above, a fall in skill supply decreases wages for the least skilled workers. This contributes to an increase in wage inequality in the left tail and counteracts the increase in polarization. However, it can be shown that if the change in interdependence is strongly regular, then the limit of the change in $\frac{\mathrm{d}}{\mathrm{d}t}W(t) = \frac{\frac{\partial}{\partial v_i}\pi_i(G_i^{-1}(t),t)}{g(G_i^{-1}(t))}$ is strictly positive at 0. Therefore, if the relative fall in wages is small for $t \to 0$, wage polarization has to increase in relative terms. This will happen either if the fall in $W(t)$ is small itself, or if wages were high originally. The former is ensured by the lack of supermodularity, in which case the wage effect is very weak. The latter is ensured if, for example, $\pi_i(0,0) - c_i$ is sufficiently large. Further, note that if the increase in interdependence is caused by a technological change that otherwise sufficiently decreases wage inequality, for example if $\pi_i(\bullet;\theta_2) = A + \pi_i(\bullet;\theta_1)$, then left-tail inequality has to fall in relative terms, by the same reasoning as in footnote 63.

Finally, note that because the Costrell and Loury model is a single-sector model in which workers are Hicks complements, it bears a certain qualitative resemblance to the canonical model of the labor market (Acemoglu and Autor, 2011), with two skill types and a CES production function. In fact, it is easy to see that in the canonical model a fall in the supply of high-skilled workers leads to an increase in wage polarization if the composition effect of such a change is taken into account.[118]

5.3 Changes to Reduced Surplus Functions

In this section I investigate the effects of changes to manufacturing's reduced surplus function. I again focus on Costrell–Loury specifications but the main result (Proposition 13) holds in general. I find that with endogenous entry, sorting depends not only on the vertical differentiation of workers but also on the vertical differentiation of firms. Perverse output effects are still possible: If manufacturing firms become less vertically differentiated, then manufacturing might contract even if surplus increases universally. Finally, I discuss the effect that changes in manufacturing's surplus have on wages in services.

As in Section 4, in the comparative statics results I will consider only specifications that result in non-degenerate equilibria, that is, those with $R_i(c_j) > 0$ for $i \in \{M, S\}$ and $j \in \{1, 2\}$.

Definition 9 (Firms' Vertical Differentiation)**.** Firms in manufacturing become (strictly) more *vertically differentiated* if, for any $v_M \in [0,1]$ and any $0 \le h'_M < h''_M \le 1$

$$\pi_M(v_M, h''_M;\theta_2) - \pi_M(v_M, h'_M;\theta_2)(>) \ge \pi_M(v_M, h''_M;\theta_1) - \pi_M(v_M, h'_M;\theta_1).$$

This is an exact analogue of an increase in workers' vertical differentiation (see Definition 6 in Section 4). To guarantee that the supply of skill increases in manufacturing, apart from an increase in workers' differentiation and a universal increase in surplus levels, we also need an increase in differentiation in manufacturing firms. To see this, let us consider what could happen otherwise. Keeping sorting constant, a fall in differentiation might reduce firms' profits, thus decreasing entry into manufacturing. This decreases the demand for skill in manufacturing—and, as a result, could lower the supply of skill

[118] A fall in the supply of high-skilled workers increases wages of the remaining high-skilled workers, decreases wages of low-skilled workers and strictly decreases left-tail inequality. To see the last point, note that there will exist a quantile t that was occupied by a high-skill worker before the change but is occupied by a low-skill worker afterwards.

in equilibrium.[119] To rule out this possibility, firms' differentiation needs to increase (weakly).

Proposition 13. If both workers and firms in manufacturing become more vertically differentiated and the surplus produced in manufacturing increases universally, then the measure of firms and supply of relative skill *increase in manufacturing* and *fall in services.*

As explained in Section 4, for a given measure of firms, an increase in workers' differentiation together with a universal increase in surplus attracts additional high-skilled workers to manufacturing. Combined with an increase in firms' differentiation, this increases profits in manufacturing, induces more firms to enter, and thus further increases demand for manufacturing skill. This results in an increase in equilibrium supply of skill in manufacturing.

Note that many plausible changes to the surplus function meet those three conditions. This includes multiplicative changes to surplus, which arise very naturally in many applications—for example as a consequence of price increases—and have been widely studied in the literature (e.g., Tervio, 2008; Dupuy, 2015). Further, two of the examples discussed in Section 4.1 also satisfy all three conditions: the increase in the quality of training in manufacturing that favors high-skilled workers (Section 4.1.2), the improvement in the quality of firms that favors high-skilled firms (Section 4.1.3), and the investment in infrastructure (Section 4.1.4).[120]

If firms' differentiation does not increase (weakly), not only is it possible that manufacturing will attract fewer skilled workers, but also that sector will produce less surplus overall, in both gross and net terms. This is best illustrated through an example.

5.3.1 Example: Perverse Output Effects

In this section, I consider a very stylized example, to demonstrate that if the conditions of Proposition 13 are not met, manufacturing's output can contract even if the surplus produced by all matches increases.

Suppose that workers' skill $\mathbf{x}$ is bivariate and binomial, with $x_j \in \{0,1\}$ for $j \in \{1,2\}$. The joint distribution of skill is symmetric: One-quarter of workers have skill $(1,0)$, another quarter have skill $(0,1)$ and the other have skill $(0,0)$. Firms' productivity is also binomial: $z_i \in \{0,1\}$, and the fraction of sector i firms with productivity $z_i = 1$ is equal to p_i. I assume that highly productive firms are more common in manufacturing, with $p_M > p_S = \frac{1}{4}$. The surplus function in manufacturing is simply $\Pi_M(x_1, z_M) = \frac{1}{2}x_1 z_M + \frac{3}{2}$,

[119] Of course, the increase in workers' differentiation pushes in the opposite direction, as explained in Section 4. Nevertheless, the impact of lower differentiation of firms can easily be the dominant force.

[120] For details on why all three conditions are satisfied in the first example, see Section 4.1.2. In short, because of supermodularity, any increase in skill or productivity levels on one side of the market increases differentiation on the other side.

and in services it is $\Pi_S(x_2, z_S) = \frac{1}{2}x_2z_S + \frac{3}{2}$. Thus x_1 is used only in manufacturing, and x_2 is used only in services. Finally, the costs of entry are $c_M = c_S = 1$.

In this example, a measure of $R_M = \frac{1}{4p_M}$ firms enter manufacturing in equilibrium.[121] If there were more, then—because all high-skilled workers are already matched with high-productivity firms—it would always be strictly better to remove one.[122] If there were fewer, then an additional high-productivity firm in manufacturing would match with a high-skilled worker, and thus creation of more manufacturing firms would always be worth it.[123] Because it's always optimal to match high-skilled manufacturing workers with high-productivity manufacturing firms, it follows that the gross total surplus produced in manufacturing is

$$T_i = \frac{1}{2} + \frac{3}{2}\frac{1-p_M}{4p_M},$$

whereas the net total surplus is $V_i = \frac{1}{4} + \frac{1}{2}\frac{1-p_M}{4p_M}$.

This implies that any increase in the fraction of highly productive firms in manufacturing decreases both the gross and net total surplus in that sector. To see this, consider the case where all manufacturing firms become highly productive and identical ($p_M = 1$). Then T_i shrinks to $\frac{1}{2}$ and V_i to $\frac{1}{4}$. Thus manufacturing *contracts in both net and gross terms*, despite the fact that surplus increased for all matches. The increase in the share of productive firms to $p_M = 1$ could be the outcome, for example, of the most successful business models and technologies becoming common knowledge as manufacturing matures.

To understand the mechanism, consider the problem from the social planner's perspective. Because the surplus produced by low-skilled workers depends neither on their sector nor on the productivity of their firm, the social planner's only goal is to maximize the number of matches of high-productivity firms with high-skilled workers. As high-productivity firms are more common in manufacturing, it is more efficient to first produce such matches in that sector; thus all high-skilled manufacturing workers always find a match in equilibrium. However, the social planner wants to exactly match the number of high-productivity firms with the number of high-skilled workers in manufacturing, as any excess in either direction is a waste of resources. Thus, as highly productive firms become more common in manufacturing, fewer need to be created. However, as even matches involving low-skilled workers more than cover entry costs, a fall in entry

[121] Note that the measure of firms in services will be $1 - \frac{1}{4p_M}$, as this example satisfies Assumption 4.

[122] The marginal firm could either produce more surplus in services (if $R_M + R_S \leq 1$) or produce no surplus at all (otherwise). In the former case, it follows that $R_S p_S < 0.25$ and there are some high-skilled services workers not matched with high-productivity firms. Hence an additional firm produces $2p_S + (1 - p_S)\frac{3}{2} = \frac{3}{2} + (2 - \frac{3}{2})p_S$ in services and only $\frac{3}{2}$ in manufacturing. In the latter, a marginal firm is unmatched, produces nothing, and doesn't cover the entry cost.

[123] Either from the pool of potential firms (if $R_M + R_S < 1$) or from services (otherwise). The former case is trivial, as any matched firm produces enough surplus to cover the entry cost. In the latter case, an additional firm would produce $\frac{3}{2} + (2 - \frac{3}{2})p_M$ in manufacturing and just $\frac{3}{2} + (2 - \frac{3}{2})p_M$ in services.

decreases the total surplus, not just in gross terms but also net of entry costs. It is worth noting that while a contraction in gross terms would be possible even in a single-sector model, contraction in net terms requires the existence of a second sector.[124]

Of course, this example is special: The assumption that there are no workers who are highly skilled in both sectors is particularly strong. More realistically, an improvement in firms' productivity distribution should also attract some additional high-skilled workers to manufacturing, for the same reasons as in Section 4. However, this second effect can easily be dominated by the one described above. This can happen even if the two skills are positively interdependent (i.e., even if there are more workers who have high skill in both dimensions than workers with high skill in manufacturing only).

5.3.2 Wages in Services

In this section, I will discuss the impact that changes to manufacturing's surplus have on wages in services. In Section 4.1, I showed that in the baseline model, jobs' scarcity ensures that any increase in workers' vertical differentiation must increase wage inequality in services. This is no longer true with endogenous entry, especially if the increase in workers' differentiation is accompanied by a (weak) increase in surplus levels and firms' differentiation. Such a change will have two distinct effects on wages in services. The first is the *baseline effect*, which captures the change in wages caused by an increase in workers' differentiation when firms' entry is kept constant. In the baseline model, this is the only effect present. In the extended model, however, there is also the *entry effect*, which captures how wages are affected by changes to firm entry that are induced by an increase in differentiation—an expansion in manufacturing and a contraction of services—but keeps the differentiation itself constant.

Formally, Assumption 4 makes it possible to decompose the overall change in wage inequality by considering an intermediate specification of the model, denoted by θ_3. In this specification, the surplus function in manufacturing is shifted upwards by a constant compared to the original specification, $\pi_M(\bullet, \theta_3) = \pi_M(\bullet, \theta_1) + A$, in such a way that entry is the same as in the final specification, $R_i(\theta_3) = R_i(\theta_2)$.[125] To ensure comparability with the results in Section 4.1, I focus on the wage range as the measure of inequality, defined as $w_i^R = w_i(1) - w_i(v_i^c)$. We can write

$$w_S^R(\theta_2) - w_S^R(\theta_1) = \underbrace{w_S^R(\theta_2) - w_S^R(\theta_3)}_{\text{baseline effect}} + \underbrace{w_S^R(\theta_3) - w_S^R(\theta_1)}_{\text{entry effect}}.$$

[124] With a single sector, the only reason to stop producing matches with low-skilled workers is if they do not cover the entry cost, but then this increases net surplus in that sector. With two sectors, however, the opportunity cost of creating a manufacturing match is the creation of a services match.

[125] The size of A which is required for this to occur can easily be determined. Consider a baseline model with $R_i = R_i(\theta_2)$ and $\pi_i(\bullet) = \pi_i(\bullet; \theta_1)$. Assumption 4 implies that there exists a range of $w_M(v_M^c)$ such that $\bar{r}_M > c_M$ and $\bar{r}_S > c_S$ but $\bar{r}_M \neq \bar{r}_S$ in general. Hence A needs to be set so that $A = \bar{r}_S - \bar{r}_M$.

The baseline effect increases the wage range in services, just as in the baseline model. Effectively, if we consider just the baseline effect, workers of different skill levels still act as (imperfect) substitutes, and thus changes in (absolute) wage inequality in services depend only on changes to the supply of skill in that sector. Because the supply of skill falls in services, wage inequality increases in that sector unambiguously.

The entry effect is different. An expansion in manufacturing's size still causes a fall in the supply of skill in services but this does not necessarily increase wage inequality. The reason is that because of the change in entry, high-skilled workers and low-skilled workers are no longer substitutes; they are complements (see Section 5.1.2). Thus the change in wage inequality depends on whether the fall in skill supply in services is primarily caused by an exodus of high- or low-skilled workers. In general, the sign of the entry effect is ambiguous and depends on a number of factors, such as the shape of the surplus function, skill interdependence, and the features of the original equilibrium. Most importantly, however, it depends on the expansion's size. In particular, it is definitely possible that in the case of a small expansion the entry effect would increase wage inequality in services. It can be shown, however, that if the changes in entry are large enough, then the entry effect must *reduce* wage inequality in services.

Proposition 14. Suppose Assumption 4 holds for both specifications, that both workers and firms in manufacturing become more vertically differentiated, and that surplus levels increase universally in manufacturing. The entry effect of such a change has a negative impact on wage inequality in services ($w_S^R(\theta_3) - w_S^R(\theta_1) < 0$) as long as the contraction in the number of firms in services is sufficiently large ($\frac{R_S(\theta_2)}{R_S(\theta_1)}$ is small enough).

Self-selection implies that workers who are skilled only in manufacturing have already joined that sector. Therefore, any expansion of manufacturing must be fueled primarily by workers who are either skilled in both sectors or skilled in neither. If the expansion is small, it is not clear which of those two groups constitutes a higher *proportion* of leavers from services. If the expansion is sufficiently large, however, it must be the case that only workers who are high-skilled in services but low-skilled in manufacturing remain in services. Because of skill complementarity, this increases wages of low-skilled services workers and decreases wages of high-skilled services workers, thus decreasing inequality.

Thus if the change in entry of firms is large enough, the entry effect *must be negative.* Note that the change in entry levels might be large either because the increase in vertical differentiation is very large (so that even if the surplus produced by the least skilled workers does not change, other workers produce much more than before) or because the increase in differentiation is accompanied by a large increase in either surplus levels or firms' differentiation. In either case, the overall effect of, say, a multiplicative increase in manufacturing's surplus on wage inequality in services is ambiguous, as the positive baseline effect is counteracted by the (likely) negative entry effect.

Finally, note that Proposition 14 suggests a general link between sector sizes and within-sector wage inequality. In particular, if workers' differentiation is kept constant, a large reduction in entry into services decreases inequality in this sector. The flip side of this result is that a rapid expansion of services that starts at low levels must result in an increase in wage inequality in that sector. This is consistent with what happened in the second half of the 20th century.[126]

6 Related Literature

This paper builds on the work of Becker (1973), Sattinger (1979), and Costrell and Loury (2004) on one side, and Roy (1951) on the other, and combines their approaches to matching and self-selection, respectively. My model nests one-sector assignment models and two-sector comparative advantage models within a single framework. The sectors in Roy (1951) can be interpreted as Sattinger-like matching markets with homogeneous and abundant firms, implying that companies have no market power and workers earn the entire surplus. In Roy's model, therefore, sorting depends only on surplus levels but not on vertical differentiation. The introduction of firm heterogeneity gives firms some market power and is the reason why the vertical differentiation of workers matters for skill supply. Compared to Becker (1973), Sattinger (1979), and Costrell and Loury (2004), the addition of another sector allows the study of interactions between two matching markets as well as the determinants of sectoral skill supply.

There are a number of papers that provide comparative statics results for standard, one-sector differential rents models (e.g., Costrell and Loury, 2004; Gabaix and Landier, 2008; Tervio, 2008).[127] These results capture only the direct effect of exogenous changes, as the within-sector distribution of skill is fixed. As noted in Costrell and Loury (2004), this is a serious limitation—one that is addressed directly in this paper. In particular, I show that if entry is exogenous, then the two types of shocks which are of particular importance in this literature (i.e., multiplicative surplus shocks and first-order stochastic dominance improvements in the distribution of firms' productivity) result in a greater supply of skill in the affected sector, increasing, as they do, both the levels of surplus and workers' vertical differentiation (Section 4.2).[128] This undoes at least part of their positive direct effect on wage levels and inequality. In fact, with scarce jobs and exogenous entry, wages will certainly fall for the least skilled workers in the industry in which the change took place.

As both the self-selection model and the differential rents model are assignment models

[126]See, for example, Figures 1 and 2 in Gould (2002).

[127]See Sattinger (1993) for an overview of the different types of assignment models.

[128]If entry is endogenous, then the above still holds for multiplicative shocks but not necessarily for improvements in the distribution of firms' productivity, as sorting then depends also on firms' vertical differentiation (Section 5.3).

(Sattinger, 1993), this paper is also related to the wider literature on optimal assignment (Shapley and Shubik, 1971; Gretsky et al., 1992). In particular, in Section 5 I extend the basic result from this literature—that the core of an assignment game is efficient—to a setting where the measure of agents on one side of the market is endogenous. The literature on general assignment games is sparse on comparative statics results, whereas the recent, more specialized models of multivariate matching are focused on marriage markets (Anderson, 2003; Chiappori, Oreffice, and Quintana-Domeque, 2011, 2012).[129] A notable exception to this rule is a recent paper by Lindenlaub (2017) which investigates multivariate matching in labor markets. Lindenlaub defines positive and assortative matching in a general setting with bivariate skills and skill demands, and provides sufficient conditions for its existence. However, the model is solved—and comparative statics are provided—for only the very special Gaussian-quadratic case.[130] The comparative statics results focus on technological change, modeled as a multiplicative surplus shock, and only the knife-edge case of no unmatched firms and workers is considered. My paper studies the determinants of skill supply more generally, including changes to the distribution of skills, and all my results hold for general surplus and distribution functions.[131]

There exists a small but quickly growing literature on multi-sector matching. The models in McCann, Shi, Siow, and Wolthoff (2015) and Grossman, Helpman, and Kircher (2013), however, differ substantially from the one presented here, in that they both focus on one-to-many rather than one-to-one matching. McCann et al. (2015) have a complicated model, with three markets and schooling. This comes at the cost of using specific functional forms and not providing comparative statics results. In another vein, Grossman et al. (2013) focus on the impact of trade liberalization, rather than changes in skill and productivity distributions. Their skills are one dimensional, and they restrict attention to cases where re-sorting happens at the extensive margin only. The model in Dupuy (2015) is quite similar to the baseline model presented here (albeit less general), as it is a differential-rent matching model with two-dimensional skills.[132] Dupuy (2015) proves the existence (but not uniqueness) of an equilibrium and then proceeds to study

[129] To the best of my knowledge, the only papers that deal with general assignment games and provide comparative statics results are Kelso and Crawford (1982) and Mo (1988), both of which focus on the impact of the entry of new agents.

[130] The quadratic surplus function coupled with normal skill distribution implies a surplus function that is not monotonic in workers' skills, so that workers with extremely high and extremely low skills produce the same surplus and earn the same wages. Adding non-interaction skill terms does not resolve this problem in a satisfactory manner, as evidenced by the fact that the surplus function estimated in Lindenlaub (2017) (Table 10) is non-monotonic in manual skill.

[131] On the other hand, my model assumes that firms are divided into two sectors and that within each sector firms agree on the ranking of workers. The Gaussian-quadratic specification of Lindenlaub's setting could be reinterpreted as a model where firms are divided into a continuum of such sectors, and is thus more general in that dimension.

[132] Generally, the type of model considered in Dupuy differs from mine, as firms are one dimensional and the size of any sector can never be exogenously fixed. However, as Dupuy further assumes that the masses of workers and firms are equal, his specification is equivalent to the special case of my model for which the measure of firms is equal to 1 in each sector.

the impact of multiplicative shocks on self-selection and inequality. However, unlike this paper, Dupuy (2015) does not show the equilibrium effect of such shocks, providing only a first-order result.[133] Mak and Siow (2017) also propose a similar model, with the difference that workers self-select into different sides of the market, rather than into separate sectors. They then calibrate the model to Brazilian data to explain changes in within- and across-firm wage inequality but they provide no comparative statics results.

My model extends Roy (1951) in a different direction than the strand of "Roy-like" assignment models (Sattinger, 1975; Teulings, 1995, 2005), in which comparative advantage drives the assignment of workers to a continuum of tasks/sectors. However, because skills are one dimensional, firms are homogeneous within sectors, and surplus is log-supermodular, all firms in a given sector employ workers of the same, single type. For that reason, these models are ill-suited to study the determinants of the within-sector wage and skill distributions, which is the focus of Section 4 of this paper. They have, however, been widely and successfully applied to modeling of the impact of task-biased technological change on wage polarization (Costinot and Vogel, 2010; Acemoglu and Autor, 2011; Cortes, 2016). In Section 3, I argue that wage polarization can increase even if technological change is unbiased, as long as it causes sectors to use skill sets that are more similar. This effect could also be achieved in ''Roy-like" assignment models, through changes to the supply of skill but the existing literature either ignores such changes or, in the case of Costinot and Vogel (2010), ignores the fact that they have a composition effect. Furthermore, as has been argued forcefully by Boehm (2015) and Lindenlaub (2017), multi-dimensional models such as the one I employ are generally better suited to the study of changes in wage distributions.[134] That is because, empirically, changes in wage distributions usually involve workers switching ranks in the wage distribution, which is harder to achieve in one-dimensional models.[135]

Section 3 of my paper is closely related to the theoretical part of Gould (2002), which shows that in the Heckman and Sedlacek (1985) model of self-selection an increase in the

[133]The equilibrium adjustment process can be thought of as a following chain of events: the shock changes wages, which impacts sorting, which impacts wages, which impacts sorting etc. until a new equilibrium is reached. The results in Dupuy (2015) consider only the first change in sorting, not the subsequent ones. Proposition 7 in this paper does derive the full equilibrium effect for the type of shock considered by Dupuy.

[134]Boehm (2015) uses a three-sector Heckman and Sedlacek (1985) model to study wage polarization and explicitly assumes that skill content of sectors is unchanged, so the mechanism in his model is evidently different than mine. The relationship between the mechanism proposed here and that in Lindenlaub (2017) is more complex: My conjecture is that the change she considers would, if translated into the language employed here, both increase the interdependence of relative skills and increase vertical differentiation in cognitive-intensive "sectors", and thus combines all the effects discussed here. Lindenlaub links her results, however, to task-biased technological change only.

[135]In the models of Costinot and Vogel (2010) and Acemoglu and Autor (2011) in particular, workers' rank in the distribution of wages need not correspond to their rank in the distribution of skill, as productivity is not assumed to increase in skill. Thus rank switching is possible. Nevertheless, both papers focus exclusively on the changes to returns to skill, and thus ignore the impact of rank-switching on the distribution of wages.

correlation of sector-specific skills is likely to increase wage inequality. The (important) differences are that Gould (a) does not link this to changes in skill supply (in single-sector models or otherwise) and (b) does not study its impact on the polarization of the wage distribution. Further, the statistical literature on order statistics is concerned with related issues. For example, Owen and Steck (1962) study how do the first four moments of the max of a multivariate normally distributed random vector change as correlation increases, and find that the mean must fall and the change in variance is ambiguous. My paper extends their results in a number of ways. Firstly, I provide an economic interpretation for results of this type. Secondly, I show that the results wrt mean hold much more generally than just for multivariate normal variables. Thirdly, I consider the impact on the entire distribution of the max, rather than just the moments. And finally, this entire literature is concerned only with the composition effect, whereas I consider also the wage effect

7 Conclusions

This article developed a new model of workers' self-selection, which allows for imperfect substitution and complementarity of skills within sectors. This was accomplished by merging a standard model of self-selection (across sectors) in the vein of Roy (1951) with an assignment model (within sectors) in the vein of Becker (1973), Sattinger (1979) and Costrell and Loury (2004). The within-sector assignment was found to create imperfect substitution/complementarity of skills but also to cause wage functions in each sector to depend on the entire distribution of skill in that sector. Despite this difficulty, I was able to derive a series of sharp monotone comparative statics results without making functional form assumptions.

Firstly, I studied what happens if sectors start requiring different sets of skills than previously. This changes the interdependence between the sector-specific skill indices (*relative skills*). In particular, if sectors start using skill sets that are more similar, then the interdependence of relative skills increases. As a result, the overall supply of skill declines: There are fewer workers available that have a high relative skill in at least one sector. The composition effect of such a decline in the supply of skill increases wage polarization unambiguously, while its wage effect is ambiguous. I show, however, that (in contrast to the mechanism underpinning the task-biased technological change explanation for increased wage inequality) if the fall in supply of skill affects wages in both sectors symmetrically, then the composition effect dominates under mild conditions and wage polarization increases.

Secondly, I have shown that if the level of surplus produced by manufacturing increases for all matches and both manufacturing workers and firms become more vertically differentiated, then the supply of skills is guaranteed to increase in manufacturing. If firms' entry is exogenous and jobs are scarce, this increases wage inequality in both sectors. However, if the improvement in surplus levels decreases vertical differentiation of either manufacturing workers or firms, then the supply of skill might fall in that sector. As a result, manufacturing might contract.

The model developed in this paper can contribute to a variety of areas in economics. For example, in another paper (Burzyński and Gola, 2017), we embedded the extended model into a Krugman (1979) trade model, to study migration decisions and the distributional impact of migration on wages. As workers are complements in the model, we are able to shed light on who gains and who loses from migration. Calibrating the model to US and Mexican wage data, we find that although migration increases the average wage in the US, a significant proportion of (predominantly low-skilled) US residents earns lower wages than in a counterfactual in which there is no migration.

A Demand: Formal Definition and Shifts

The definition of sectoral demand for relative skill (Section 2) holds for a given hiring function and under the assumption that profit is strictly increasing. However, if, for example, surplus does not depend on firm productivity, then (a) firms will be indifferent between many different workers and there will exist many different hiring functions and (b) all firms will make the same profits. Here, I amend the definition of sectoral demand to allow for such possibilities. Accordingly, the economy will be in equilibrium if there exists at least one demand function consistent with firms maximization problem for which the market clears.

Definition 10. A mapping $v_i^* : [0,1] \to [0,1] \cup \{-1\}$ is a *hiring function* in sector i for wage function w_i, if (a) for $v^*(h) \in [0,1]$, $v_i^*(h) \in \arg\max_{v_i} \pi_i(v_i, h) - w_i(v_i)$ and $\pi_i(v_i^*(h), h) - w_i(v_i^*) \geq 0$ and (b) for $v_i^*(h) = -1$, $\pi_i(v_i, h) - w_i(v_i) \leq 0$ for all $v_i \in [0,1]$.

Given a talent level v_i and an input function v_i^*, define the set $B(v_M, v_i^*) = \{h \in [0,1], v_i^*(h) \geq v_M\}$.

Definition 11. A mapping $D_i : [0,1] \to [0,R]$ is a *sector i demand function for relative skill* given wage function w_i, if there exists a hiring function such that $R_M \int_{B(v_i,v_i^*)} 1 \mathrm{d}v_i = D_i(v_i)$, for all $v_i \in [0,1]$.

For any matching problem, I will denote as $DC(\theta)$ the set of all possible cumulative demand functions and as $DC(v_M, \theta)$ the set of their values for talent v_M.

Definition 12. Demand for skill *shifts up* if—given the old equilibrium wage function $w_M(\cdot;\theta_1)$ and for all $v_M \in [0,1]$—for any $h'' \in DC(v_M;\theta_2)$ and $h' \in DC(v_M;\theta_2)$ we have that $\max\{h'', h'\} \in DC(v_M;\theta_2)$ and $\min\{h'', h'\} \in DC(v_M;\theta_1)$.

Proposition 15. If both manufacturing workers become more vertically differentiated and surplus levels increase universally in manufacturing, the demand for relative skill shifts up in manufacturing. If jobs are scarce, an increase in workers' vertical differentiation alone suffices for an upward shift of skill demand.

Proof. The partial order $([0,1],\geq)$ is clearly a lattice and the function $\pi_i(v,h)-w_i(v)$ is supermodular in v (for any h). Thus, as an increase in vertical differentiation implies that $\pi_i(v,h)-w_i(v)$ has increasing differences in c it follows from the results in Topkis (1978) and Milgrom and Shannon (1994) that the set $V^*(c_i)=\{v\in[0,1]: v\in \arg\max \pi_i(v,h,(c)-w_i(v)\}$ increases in the strong set order sense with a change from θ_1 to θ_2. This proves the second statement, as $v_i^*(h)\in[0,1]$ for all firms in that case. As for the first claim, note that the increase in surplus levels means that each firm's profit increases for the old choice of inputs, and hence, by profit maximization, also for the new choice. Thus, no firms leave the market and the result follows. $\square$

B Equilibrium Characterization

Proof of Proposition 1

This proof will refer to the formal definition of a sectoral demand function introduced in Appendix A rather than the simplified definition from Section 2.1.2. First, I propose the following hiring function $v^*:[0,1]\to[v_i^c,1]\cup\{-1\}$, where

$$v^*(h)=\begin{cases}-1 & \text{for } h\in[0,1-\frac{S_i(0)}{R_i})\\ \max\{v\in[0,1]: S(v)=R_i(1-h)\} & \text{otherwise.}\end{cases}$$

Clearly, the set $B(v,v^*)=\{h\in[0,1]: h\geq 1-\frac{S_i(v)}{R_i}\}$ for this v^*. Thus, the corresponding demand schedule is simply $D_i(v,v^*)=R_i\int_{1-\frac{S_i(v)}{R_i}}^{1} 1\mathrm{d}t = S_i(v)$, as required. In other words, hiring function v^* ensures that demand equals supply.

Secondly, I show that the hiring function is consistent with firms' maximization problem for the wage functions proposed. Consider any firm with $h\in[1-\frac{S_i(0)}{R_i},1]$; for v^* to be consistent with that firm's profit maximization it must be the case that

$$\underset{v\in[0,1]}{\forall}\quad \pi_M(v_M^*(h),h)-\pi_M(v,h)\geq w_M(v_M^*(h))-w_M(v).$$

This is met as long as

$$\int_v^{v^*(h)}\int_{1-\frac{S_i(s)}{R_i}}^{h}\frac{\partial^2}{\partial v\partial h}\pi_i(s,t)\mathrm{d}t.$$

Note that $h=1-\frac{S_i(v^*(h))}{R_i}$ and that $1-\frac{S_i(v)}{R_i}$ is an increasing function of v. Thus, if $v^*(h)>v$ then $1-\frac{S_i(s)}{R_i}\leq h$ for all $s\in[v,v^*(h)]$. Similarly, if $v^*(h)\leq v$ then $1-\frac{S_i(s)}{R_i}\geq h$

$s \in [v, v^*(h)]$. This and the supermodularity of reduced surplus ensure that the above condition is always met.

Note that the set $[0, 1 - \frac{S_i(0)}{R_i})$ is non-empty only if $S_i(0) < R_i$. If that's the case, for any firm with $h \in [0, 1 - \frac{S_i(0)}{R_i})$ it must be the case that

$$\underset{v \in [0,1]}{\forall} \quad \pi_M(v, h) - w_M(v) \leq 0,$$

which for $v \in [0, v_i^c)$ gives:

$$\pi_i(v, h) - \pi_i(v, 1 - \frac{S_i(0)}{R_i}) \leq w_i(v_i^c) - \pi_i(v_i^c, 1 - \frac{S_i(0)}{R_i}) \geq 0.$$

The LHS is greatest for $h \approx 1 - \frac{S_i(0)}{R_i}$, in which case the LHS is arbitrarily close to 0. Thus, this condition is ensured to be met if and only if $w_i(v_i^c) = \pi_i(v_i^c, 1 - \frac{S_i(0)}{R_i})$, as required. For $v \in [v_i^c, 1]$ we need

$$\pi_i(v, h) - \pi_i(v, 1 - \frac{S_i(0)}{R_i}) + \pi_i(v, 1 - \frac{S_i(0)}{R_i}) - w_i(v) \leq 0. \tag{29}$$

Note that $\pi_i(v, 1 - \frac{S_i(0)}{R_i}) - w_i(v)$ is the profit firm $1 - \frac{S_i(0)}{R_i}$ would make from hiring worker v. This is weakly smaller than $1 - \frac{S_i(0)}{R_i}$ makes by hiring $v^*(1 - \frac{S_i(0)}{R_i}) = v_i^c$, which is equal to 0, because $w_i(v_i^c) = \pi_i(v_i^c, 1 - \frac{S_i(0)}{R_i})$. Thus, the LHS of Equation (29) is negative, as required.

Given a supply function S_i sector i of my model is trivially a special case of the assignment model specified in Chiappori, McCann, and Nesheim (2010). Label the firms as buyers and the workers as sellers. Then this model meets the conditions of the semi-convex buyer setting from Chiappori et al. (2010).[136] Hence, their Proposition 3 holds. This implies that in any equilibrium the marginal profits are equal to $\frac{\partial}{\partial h} r_M(h)$ almost everywhere for $h \in [1 - \frac{S_i(0)}{R_i}, 1]$, where $r_M(h) = \pi_M(v^*(h), h) - w_M(v^*(h))$. This in turn implies that the wage function for any stable matching is of the proposed form.[137]

Proof of Lemma 1

Denote $\sup\{v_M \in [0, 1] : w_M(v_M) \leq \max\{w_S(0), 0\} \text{ or } v_M = 0\}$ as v'_M.

I will first show that either $w_M(v_M^c) = \max\{w_S(0), 0\}$ or $v_M^c \in \{0, 1\}$. First, suppose that $w_M(v_M^c) < \max\{w_S(0), 0\}$. This is possible only if $v_M^c = 1$. Otherwise, as w_M is continuous for $v_M \geq v_M^c$, there must exist some $\epsilon > 0$ such that $w_M(v_M) < \max\{w_S(0), 0\}$

[136] Surplus function is twice differentiable, $[0, 1]$ and $[v_M^c, 1]$ are smooth manifolds and standard uniform distribution puts zero mass on any $h \in [0, 1]$. See Definition 4 in Chiappori et al. (2010) and its discussion.

[137] If for any $v_i \in [0, v_i^c]$ we had $w_i(v_i) < w_i(v_i^c) + \pi_i\Big(v_i, 1 - \frac{S_i(0)}{R_i}\Big) - \pi_i\Big(v_i^c, 1 - \frac{S_i(0)}{R_i}\Big)$, then this worker would be demanded by firm $h = 1 - \frac{S_i(0)}{R_i}$. If this was the case for a set of points of positive Lebesgue measure, then $D_i(0) \neq S_i(0)$.

for all workers with $v_M \in [v_M^c, v_M^c + \epsilon]$. Suppose that $v_S^c > 0$, then $w_S(0) \leq 0$.[138] In such a case, all workers with $v_M \in [v_M^c, v_M^c + \epsilon]$ prefer to remain unemployed than join manufacturing and $S_M(v_M^c) = S_M(v_M')$, which contradicts the definition of v_M^c. Similarly, if $v_S^c = 0$, then all workers in $v_M \in [v_M^c, v_M^c + \epsilon]$ prefer to join services than manufacturing, as w_S is strictly increasing for $v_S \geq v_S^c$, also contradicting the definition of v_M^c.

Suppose that $w_M(v_M^c) > \max\{w_S(0), 0\}$. Firstly, consider $v_S^c > 0$. In such a case, all (but possibly a zero measure of) workers with $(v_M, v_S) \in [0, v_M^c] \times [0, v_S^c]$ would join manufacturing. As C has full support, if $v_M^c > 0$ then a strictly positive measure of workers lives in this rectangle, which contradicts the definition of v_M^c. Secondly, v_S^c could be equal to 0. By continuity of the reduced surplus function, there exists some $v_M'' < v_M^c$, such that

$$w_M(v_M^c) + \pi_M\Big(v_M'', 1 - \frac{S_M(0)}{R_M}\Big) - \pi_M\Big(v_M^c, 1 - \frac{S_M(0)}{R_i}\Big) > w_S(0).$$

Further, w_S is continuous for all $v_S \geq v_S^c$: thus, there must exist some $v_S'' > 0$, such that

$$w_S(v_S) < w_M(v_M^c) + \pi_M\Big(v_M'', 1 - \frac{S_M(0)}{R_M}\Big) - \pi_M\Big(v_M^c, 1 - \frac{S_M(0)}{R_i}\Big) \leq w_M(v_M'')$$

for all $v_S \in [0, v_S'']$. Thus, all workers with $(v_M, v_S) \in (v_M'', v_M^c) \times [0, v_S'']$ would join manufacturing, which contradicts the definition of v_M^c. The proof for v_S^c is analogous.

It follows that $v_M^c = v_M'$. First, suppose that $v_M' < v_M^c$ then by the definition of v_M' follows that $w_M(v_M) > \max\{w_S(0), 0\}$; contradiction. Now suppose that $v_M' > v_M^c$, which is possible only if $v_M^c < 1$. This implies that there exists some $v_M \in (v_M^c, v_M')$ such that $w_M(v_M) \leq \max\{w_S(0), 0\}$. By Proposition1 w_M is strictly increasing for $v_M \geq v_M^c$ and hence $w_M(v_M^c) < \max\{w_S(0), 0\}$ which was shown to be impossible.

Finally, let me prove the last statement. First, I will consider the case of $v_M^c, v_S^c \in (0, 1)$. This implies that (a) some workers are unemployed (because workers with $(v_M, v_S) < (v_M^c, v_S^c)$ cannot join either sector by definition of critical skills and (b) that $w_M(v_M^c) = \max\{w_S(0), 0\}$ and $w_S(v_S^c) = \max\{w_M(0), 0\}$. Suppose $w_S(0) > 0$; then by Proposition 1 there exists an $\epsilon_3 > 0$, such that all workers with $(v_M, v_S) \in (v_M^c - \epsilon_3, v_M^c) \times (0, v_S^c)$ prefer to join one of the sectors than to remain unemployed, which contradicts the definition of either v_S^c or v_M^c; thus $w_M(v_M^c) = 0$. An analogous reasoning holds for $w_S(v_S^c)$.

Now suppose that $v_M^c = 0$. It follows immediately that $w_S(v_S^c) \neq w_M(v_M^c)$ only if $w_S(v_S^c) > w_M(v_M^c)$. By Proposition 1 there must then exist an $\epsilon_4 > 0$ such that $w_M(v_M) < w_S(v_S)$ for all $(v_M, v_S) \in [0, \epsilon_4] \times [v_S^c - \epsilon_4, v_S^c]$, so that all workers with such skill vectors prefer to work in services over manufacturing, which contradicts the definitions of v_M^c and v_S^c.

[138]Otherwise, there exists some $\epsilon_1 > 0$ such that for all $(v_S, v_M) \in (0, \epsilon_1) \times (v_M^c, v_M^c + \epsilon)$, $w_S(v_S) > \max\{w_M(v_M), 0)\}$, which contradicts the definition of v_S^c.

Proof of Lemma 2

I start with manufacturing. The probability that a worker with relative skill $v_M \geq v_M^c$ chooses services is $\Pr(\psi(V_S) < v_M | V_M = v_M)$. Note that because ψ is weakly increasing, it follows that if $\psi(v_S') < v_M$ then $\psi(v_S'') < v_M$ for any $v_S' \geq v_S'' \geq v_S^C$. Thus:

$$\Pr(\psi(V_S) < v_M | V_M = v_M) = \frac{\partial}{\partial v_M} C(v_M, \phi(v_M)) \qquad \text{for } v_M \geq v_M^c,$$

where $\phi(v_M) = \sup\{v_S \in [v_S^c, 1] : \psi(v_S) < v_M\}$. Because $S_M(1) = 0$, this gives us the required expression for $S(v_M)$ if $v_M \geq v_M^c$. And, of course, for any $v_M < v_M^c$, $S_M(v_M) = S_M(0)$ by the definition of critical relative skill.

The proof for $S_S(\cdot)$ is analogous.

Proof of Theorem 1

Define the *extended separating function* $\psi^e : [v_S^c, 1] \to [v_M^c, 1 + B]$ as

$$\psi^e(v_S) = v_M^c + \int_{v_S^c}^{v_S} \frac{\frac{\partial}{\partial v_S} \pi_S^e\left(t, 1 - \frac{\int_t^1 \frac{\partial}{\partial v_S} C^e(\psi(r), r) \mathrm{d}r}{R_S}\right)}{\frac{\partial}{\partial v_M} \pi_M^e\left(\psi(t), 1 - \frac{\int_t^1 \frac{\partial}{\partial v_M} C^e(r, \phi(r)) \mathrm{d}r}{R_M}\right)} \mathrm{d}t, \tag{30}$$

where he extended functions $C^e(\bullet)$, $\pi_M^e(\bullet)$ and $\pi_S^e(\bullet)$ are defined as follows (1) $C^e : [0, 1 + B] \times [0, 1] \to [0, 1]$

$$C^e(v_M, v_S) = \begin{cases} C(v_M, v_S) & \text{for } (v_M, v_S) \in [0, 1] \times [0, 1] \\ v_S & \text{for } (v_M, v_S) \in (1, 1 + B] \times [0, 1], \end{cases}$$

(2): $\pi_M^e(v_M, h) : [0, 1 + B] \times [0, \frac{1+R_M}{R_M}] \to \mathbf{R}^+$

$$\pi_M^e(v_M, h) = \begin{cases} \pi_M(v_M, h) & \text{for } (v_M, h) \in [0, 1]^2 \\ \pi_M(1, h) + (v_M - 1)\frac{\partial}{\partial v_M} \pi_M(1, h) & \text{for } (v_M, h) \in (1, B] \times [0, 1], \\ \pi_M(v_M, 1) & \text{for } (v_M, h) \in [0, 1] \times (1, \frac{1+R_M}{R_M}], \\ \pi_M(1, 1) + (v_M - 1)\frac{\partial}{\partial v_M} \pi_M(1, 1) & \text{for } (v_M, h) \in (1, B] \times (1, \frac{1+R_M}{R_M}], \end{cases}$$

(3): $\pi_S^e(v_S, h) : [0, 1] \times [0, 1 + \frac{1}{R_S}] \to \mathbf{R}^+$

$$\pi_S^e(v_S, h) = \begin{cases} \pi_S(v_S, h) & \text{for } (v_M, h) \in [0, 1]^2 \\ \pi_S(v_S, 1) & \text{for } (v_M, h) \in [0, 1] \times (1, 1 + \frac{1}{R_S}], \end{cases}$$

and $B = \frac{\max \frac{\partial}{\partial v_S}\pi_S}{\min \frac{\partial}{\partial v_M}\pi_M}$. Note that $C^e(\cdot, v_S)$, $\frac{\partial}{\partial v_S}C^e(\cdot, v_S)$, $\frac{\partial}{\partial v_M}\pi^e_M(\cdot,\cdot)$ and $\frac{\partial}{\partial v_S}\pi^e_S(v_S,\cdot)$ are Lipschitz continuous[139]; denote their Lipschitz-constants as L^1, L^2 ,L^3, L^4 and L^5 respectively.

Clearly, given v^c_M and v^c_S the separating function ψ uniquely determines the extended separation function ψ^e. Similarly, it should be clear that

$$\psi(v_S) = \begin{cases} \psi^e(v_S) & \text{if } \psi^e(v_S) \leq 1, \\ 1 & \text{otherwise.} \end{cases}$$

The result for $\psi^e(v_S) \leq 1$ follows from noting that ψ^e is strictly increasing and then substituting Equation (9) into Equation (14), differentiating wrt v_S, dividing both sides by $\frac{\partial}{\partial v_M}\pi_M\Big(\psi(v_S), \frac{1}{R_M}\int_{v^c_M}^{\psi(v_S)} \frac{\partial}{\partial v_M}C(r,\psi^{-1}(r))\mathrm{d}t\Big)$ and then integrating from v^c_S to v_S (and remembering that $\psi(v^c_S) = v^c_M$).[140] The other part follows from the fact that for v_S's such that $w_S(v_S) \leq w_M(1)$ we have $\psi(v_S) = 1$ and $\psi^e(v_S) > 1$ (because ψ^e is strictly increasing).

Thus, it is sufficient to prove that ψ^e, v^c_M, v^c_S exist and are unique. Let me make a few observations that will prove useful.

Relation Between Supply Functions By differentiating $C(\psi(r), r)$ rearranging and integrating from v^c_S to v_S, we arrive at

$$S_M(0) - S_M(\psi(v_S)) + S_S(0) - S_S(v_S) = C(\psi(v_S), v_S) - C(v^c_M, v^c_S). \tag{31}$$

Determining the Critical Skills As the critical skills v^c_M, v^c_S are also unknown, we need to find conditions that will pin them down. Let me start by denoting the measure of employed workers as $M = S_M(0) + S_S(0)$. Clearly, $M = \min\{R_M + R_S, 1\}$ in equilibrium: otherwise we have $S_i(0) < R_i$ in some sector i, implying that a positive measure of

[139] I will do this in detail for $\frac{\partial}{\partial v_S}C^e(v_M, v_S)$—the reasoning for the other two is analogous. $\frac{\partial}{\partial v_S}C^e(v_M, v_S) : [0, 1+B] \times [0,1] \to [0,1]$:

$$\frac{\partial}{\partial v_S}C^e(v_M, v_S) = \begin{cases} \frac{\partial}{\partial v_S}C(v_M, v_S) & \text{for } (v_M, v_S) \in [0,1] \times [0,1] \\ 1 & \text{for } (v_M, v_S) \in (1, 1+B] \times [0,1], \end{cases}$$

is clearly continuous in u. It is equally easy to see that the function $\frac{\partial}{\partial v_S}C^e(\cdot, v_S)$ is differentiable almost everywhere and its derivative is Lebesque integrable. It is also the case that for any $(v_M, v_S) \in (1, 1+B] \times [0,1]$ we have:

$$\frac{\partial}{\partial v_S}C^e(a, v_S) + \int_a^1 C^e_{uv}(r, v_S)\mathrm{d}r + \int_1^{v_M} 0\mathrm{d}r = 1,$$

which means that $\frac{\partial}{\partial v_S}C^e(\cdot, v_S)$ is absolutely continuous. Moreover, as $C^e(\bullet)$ is twice continuously differentiable and any continuous function defined on a compact set is bounded it follows that $\frac{\partial}{\partial v_S}C^e(\cdot, v_S)$ is essentially bounded; and a differentiable almost everywhere, absolutely continuous function with an essentially bounded derivative is Lipschitz-continuous.

[140] This gives us Equation (30), but with ψ rather than ψ^e on the right hand side.

workers with relative skill below (v_M^c, v_S^c) would strictly prefer to join sector i than remain unemployed. By Equation (31) this gives $1 - M = C^e(v_M^c, v_S^c)$, determining one of the critical skills as a function of the other. Furthermore, note that Assumption 3 implies that $v_M^c, v_S^c < 1$ and thus $S_M(0), S_S(0) > 0$.[141] Therefore, from Proposition 1 and Lemma 1 it follows that if $S_M(0) < R_M$ then:

$$\pi_M(v_M^c, 1 - \frac{S_M(0)}{R_M}) = w_M(v_M^c) = w_S(v_S^c) \leq \pi_S(v_S^c, 1 - \frac{M - S_M(0)}{R_S}),$$

and analogously for services. This determines the other critical skill if $R_M + R_S > 1$. Finally, recall that market clearing implies that $S_i(0) \leq R_i$, implying that if $R_M + R_S \leq 1$ we have $S_M(0) = R_M$ and $S_S(0) = R_S$.

The Set of Equations and Inequalities By substituting $S_i(v_i) = S_i(0) - S_i(v_i)$ and Equation (31) into Equation (30) we arrive at

$$\psi^e(v_S) = v_M^c + \int_{v_S^c}^{v_S} \frac{\frac{\partial}{\partial v_S}\pi_S^e\left(t, \frac{R_S - S_S(0) + \int_{v_S^c}^t \frac{\partial}{\partial v_S} C^e(\psi(r), r)\mathrm{d}r}{R_S}\right)}{\frac{\partial}{\partial v_M}\pi_M^e\left(\psi^e(t), \frac{R_M - 1 + S_S(0) + C^e(\psi^e(t), t) - \int_{v_S^c}^t \frac{\partial}{\partial v_S} C^e(\psi^e(r), r)\mathrm{d}r}{R_M}\right)} \mathrm{d}t. \tag{32}$$

This, together with

$$\begin{aligned}
M &= \min\{R_M + R_S, 1\} && (33)\\
1 - M &= C^e(v_M^c, v_S^c), && (34)\\
S_S(0) &= \int_{v_S^c}^1 \frac{\partial}{\partial v_S} C^e(\psi(r), r)\mathrm{d}r, && (35)\\
S_S(0) &\in \Theta(M) = [\max\{0, M - R_M\}, \min\{1, R_S\}] && (36)\\
S_M(0) < R_M &\Rightarrow \pi_M^e(v_M^c, 1 - \frac{S_M(0)}{R_M}) \leq \pi_S^e(v_S^c, 1 - \frac{M - S_M(0)}{R_S}), && (37)\\
S_S(0) < R_S &\Rightarrow \pi_S^e(v_S^c, 1 - \frac{M - S_M(0)}{R_S}) \leq \pi_M^e(v_M^c, 1 - \frac{S_M(0)}{R_M}). && (38)
\end{aligned}$$

constitutes the set of Equations and Inequalities that determines ψ^e, v_M^c, v_S^c.

The remainder of the proof shows that there exists a unique solution to Equations (32)-(38). Define the set

$$K = \{d \in C[0, 1] : d(v_S) \in [0, 1 + B]\},$$

where $C[0, 1]$ is the set of all continuous functions that map from $[0, 1]$. The constant function $d(v_S) = 1$ lies in K and hence the set is non-empty. Define a (Bielecki) norm,

[141] If $R_i < 1$ this follows immediately from $1 - M = C^e(v_M^c, v_S^c)$. Otherwise, suppose that $v_M^c = 1$; then $S_M(0) = 0 < R_M$ and $w_M(1) = \pi_M(1, 1) > \pi_S(0, 1 - \frac{1}{R_S}) \geq w_S(0)$. But then, by continuity of π_M and Proposition 1 follows that there must exist some $\epsilon > 0$ such that all workers with $(v_M, v_S) \in [0, \epsilon] \times [1 - \epsilon, 1]$ would prefer to join manufacturing, contradicting $v_M^c = 1$.

$||\cdot||_\lambda$ on $C[0,1]$:

$$||h||_\lambda = \sup_{[0,1]} e^{-\lambda v_S}|h(v_S)|,$$

where λ is some weakly positive number. K is a complete metric space for the metric implied by this norm.[142]

Endow the sets $[0,1]^2$ and $\Theta(M)$ with the Euclidean norm and define a mapping $\mathscr{T} : K \times [0,1]^2 \times \Theta(M) \to K$

$$(\mathscr{T}d)(v_S, v_S^c, v_M^c, S_S(0)) = v_M^c + \begin{cases} 0 & \text{for } v_S < v_S^c \\ \int_{v_S^c}^{v_S} \frac{\frac{\partial}{\partial v_S}\pi_S^e\left(t, \frac{R_S - S_S(0) + \int_{v_S^c}^t \frac{\partial}{\partial v_S} C^e(d(r),r)}{R_S}\mathrm{d}r\right)}{\frac{\partial}{\partial v_M}\pi_M^e\left(d(t), \frac{R_M - 1 + S_S(0) + C^e(d(t),t) - \int_{v_S^c}^t \frac{\partial}{\partial v_S} C^e(d(r),r)\mathrm{d}r}{R_M}\mathrm{d}r\right)}\mathrm{d}t & \text{for } v_S \geq v_S^c. \end{cases}$$

Note that this map is well-defined, as for any $v_S^c \in [0,1]$ and $d \in K$:

$$\frac{R_S - S_S(0) + \int_{v_S^c}^t \frac{\partial}{\partial v_S} C^e(d(r),r)}{R_S}\mathrm{d}r \leq 1 + \int_{v_S^c}^t \frac{1}{R_S}\mathrm{d}r \leq \frac{1}{R_S} + 1$$

$$\frac{R_M - 1 + S_S(0) + C(d(t),t) - \int_{v_S^c}^t \frac{\partial}{\partial v_S} C^e(d(r),r)\mathrm{d}r}{R_M} \leq \frac{R_M + C(d(t),t)}{R_M} \leq \frac{1}{R_M} + 1;$$

and that it is continuous in v, v_S^c, v_M^c and $S_S(0)$. Clearly, $(\mathscr{T}d)(v_S, v_S^c, v_M^c, S_S(0)) \geq v_M^c \geq 0$. Further, for $v_S \geq v_S^c$:

$$(\mathscr{T}d)(v_S, v_S^c, v_M^c, S_S(0)) \leq \int_{v_S^c}^{v_S} B\mathrm{d}t + v_M^c \leq 1 + B,$$

and for $v_S < v_S^c$:

$$(\mathscr{T}d)(v_S, v_S^c, v_M^c, S_S(0)) \leq v_M^c - 1 \leq 1 + B,$$

so indeed $\mathscr{T}(K) \subset K$. Finally, it should be clear that for any $(v_S^c, v_M^c, S_S(0))$ the restriction of any fixed point of $(\mathscr{T}d)(\bullet)$ to $[v_S^c, 1]$ gives us the solution to (32) and that any solution to (32) can be easily extended into a fixed point of $(\mathscr{T}d)(\bullet)$. Therefore, it suffices to show that there exists such a λ that for any $(v_S^c, v_M^c, S_S(0)) \in [0,1]^2 \times \Theta(M)$, $\mathscr{T}d(\bullet)$ is a contraction wrt to the norm $||\cdot||_\lambda$ to show that (32) has a unique solution for any feasible $(v_M^c, v_S^c, S_S(0))$.

Let us drop $(v_S^c, v_M^c, S_S(0))$ from the definition of the map (remembering that we are keeping them constant) and enhance our notation by new maps: $S_S : [v_S^c, 1] \times K \to$

[142] If we endowed K with the sup-norm, then K would be a closed subset of $C[0,1]$; since $C[0,1]$ is complete in the sup-norm, so is K. It was shown by Bielecki (1956) that the $||\cdot||_\lambda$ norm is equivalent to the sup-norm for any $C[a,b]$. As K is a closed subset of $C[a,b]$ under the metric implied by Bielecki norm, it is also complete and thus K endowed with the Bielecki metric is a complete metric space for $||\cdot||_\lambda$.

$[0,1], P_S : [v_S^c, 1] \times K \to [0, 1+\frac{1}{R_S}]$ and $P_M : [0, B] \times K \to [0, 1 + \frac{1}{R_M}]$

$$\begin{aligned}
(S_S d)(v_S) &= S_S(0) - \int_{v_S^c}^{v_S} \frac{\partial}{\partial v_S} C^e(d(r), r)\mathrm{d}r, \\
(P_S d)(v_S) &= \frac{R_S - (S_S d)(v_S)}{R_S}, \\
(P_M d)(d(v_S)) &= \frac{R_M - 1 + C^e(d(v_S), v_S) + (S_S d)(v_S)}{R_M}.
\end{aligned}$$

Take any $t \geq v_S^c$ and any $d_1, d_2 \in S$ and for any map $(fd)(t)$ denote $(fd_1)(t) - (fd_2)(t)$ as $\Delta_d(fd)(t)$. Then we have:

$$\begin{aligned}
|\Delta_d(S_S(0)d)(t)| &= |\int_{v^c}^{t} C_v^e(d_1(r), r) - C_v^e(d_2(r), r)\mathrm{d}r| \qquad (39) \\
&\leq \int_{v^c}^{t} |C_v^e(d_1(r), r) - C_v^e(d_2(r), r)|\mathrm{d}r \leq \int_{v^c}^{t} L_2|d_1(r) - d_2(r)|\mathrm{d}r| \\
&= L_2 \int_{v^c}^{t} e^{\lambda r} e^{-\lambda r}|d_1(r) - d_2(r)|\mathrm{d}r \leq L_2||d_1 - d_2||_\lambda \int_{v^c}^{t} e^{\lambda r}\mathrm{d}r \\
&= \frac{L_2}{\lambda}||d_1 - d_2||_\lambda (e^{\lambda t} - e^{\lambda v^c}) \leq \frac{L_2}{\lambda}||d_1 - d_2||_\lambda e^{\lambda t},
\end{aligned}$$

which can be used to establish

$$|\Delta_d(P_S d)(t)| \leq \frac{L_2}{\lambda R_S}||d_1 - d_2||_\lambda e^{\lambda t} \qquad (40)$$

$$\begin{aligned}
|(P_M d_1)(d_1(t)) - (P_M d_2)(d_2(t))| &= |\frac{C^e(d_1(v), v) - C^e(d_2(v), v) - \Delta_d(S_S(0)d)(v)}{R_M}| \qquad (41) \\
&\leq \frac{1}{R_M}(|C^e(d_1(v), v) - C^e(d_2(v), v)| + |\Delta_d(S_S(0)d)(v)|) \\
&\leq \frac{L_2}{\lambda_M}||d_1 - d_2||_\lambda e^{\lambda t} + \frac{L^1}{R_M}|d_1(t) - d_2(t)|.
\end{aligned}$$

Denote $L_6 = \sup \frac{\partial}{\partial v_S}\pi_S(v_S, h)$, $L_7 = \inf \frac{\partial}{\partial v_M}\pi_M(v_M, h)$ and note that continuity of $\frac{\partial}{\partial v_M}\pi_M$ and $\frac{\partial}{\partial v_S}\pi_S$ and the fact that $\frac{\partial}{\partial v_M}\pi_M > 0$ imply that both L^6 and L^7 are finite. Using all this, we can write, for any $v_S \geq v_S^c$ and any $d_1, d_2 \in S$:

$$\begin{aligned}
|\Delta_d(\mathscr{T}d)(v)| = |\int_{v^c}^{v} & \frac{\frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_1)(t))}{\frac{\partial}{\partial v_M}\pi_M^e(d_1(r), (P_M d_1)(d_1(t)))} - \frac{\frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_2)(t))}{\frac{\partial}{\partial v_M}\pi_M^e(d_2(r), (P_M d_2)(d_2(t)))}\mathrm{d}t| \\
\leq \int_{v^c}^{v} & |\frac{\frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_1)(t))}{\frac{\partial}{\partial v_M}\pi_M^e(d_1(r), (P_M d_1)(d_1(t)))} - \frac{\frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_2)(v,))}{\frac{\partial}{\partial v_M}\pi_M^e(d_1(r), (P_M d_1)(d_1(t)))} \\
& + \frac{\frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_2)(t))}{\frac{\partial}{\partial v_M}\pi_M^e(d_1(r), (P_M d_1)(d_1(t)))} - \frac{\frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_2)(t))}{\frac{\partial}{\partial v_M}\pi_M^e(d_2(r), (P_M d_2)(d_2(t)))}|\mathrm{d}t \\
\leq \int_{v^c}^{v} & \frac{|\frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_1)(t)) - \frac{\partial}{\partial v_S}\pi_S^e(t, (P_S d_2)(t))|}{L_7}
\end{aligned}$$

$$
\begin{aligned}
&+L_6(|\frac{\frac{\partial}{\partial v_M}\pi^e_M(d_1(r),(P_M d_1)(d_1(t)))-\frac{\partial}{\partial v_M}\pi^e_M(d_2(r),(P_M d_2)(d_2(t)))}{\frac{\partial}{\partial v_M}\pi^e_M(d_1(r),(P_M d_1)(d_2(t)))\frac{\partial}{\partial v_M}\pi^e_M(d_2(r),(P_M d_2)(d_2(t)))}|\mathrm{d}t\\
\leq &\int_{v^c}^{v}\frac{L_5}{L_7}|\Delta_d(P_S d)(t)|\\
&+\frac{L_6}{L_7^2}[|\frac{\partial}{\partial v_M}\pi^e_M(d_1(r),(P_M d_1)(d_1(t)))-\frac{\partial}{\partial v_M}\pi^e_M(d_2(t),(P_M d_1)(d_1(t))|]\\
&+\frac{L_6}{L_7^2}[|\frac{\partial}{\partial v_M}\pi^e_M(d_2(t),(P_M d_1)(d_1(t))-\frac{\partial}{\partial v_M}\pi^e_M(d_2(r),(P_M d_2)(d_2(t)))|]\mathrm{d}t\\
\leq &\int_{v^c}^{v}\frac{L_5L_2}{\lambda L_7 R_S}||d_1-d_2||_\lambda e^{\lambda(t-v^c)}+\frac{L_3L_6}{L_7^2}|d_1(t)-d_2(t)|\\
&+\frac{L_4L_6}{L_7^2}|(P_M d_1)(d_1(t))-P_M d_2)(d_2(t))|\mathrm{d}t\\
\leq &\frac{L_5L_2}{\lambda^2 L_7 R_S}||d_1-d_2||_\lambda e^{\lambda v}+\frac{L_3L_6}{\lambda L_7^2}||d_1-d_2||_\lambda e^{\lambda v}\\
&+\int_{v^c}^{v}\frac{L_4L_6}{L_7^2}(\frac{L_2}{\lambda_M}||d_1-d_2||_\lambda e^{\lambda(t-v^c)}+\frac{L_1}{R_M}|d_1(t)-d_2(t)|)\mathrm{d}t\\
\leq &\frac{1}{\lambda}||d_1-d_2||_\lambda e^{\lambda v}\Big[\frac{L_5L_2}{\lambda L_7 R_S}+\frac{L_3L_6}{L_7^2}+\frac{L_4L_6}{L_7^2}\Big(\frac{L_2}{\lambda_M}+\frac{L_1}{R_M}\Big)\Big]
\end{aligned}
$$

Now, for $v_S < v_S^c$ this has to hold as well, as then $|(\mathscr{T}d_1)(v_S)-\mathscr{T}(d_2)(v_S)|=0$; therefore, for any $v_S \in [0,1]$ we have that

$$
|\Delta_d(\mathscr{T}d)(v_S)| \leq \frac{1}{\lambda}||d_1-d_2||_\lambda e^{\lambda v_S}\Big[\frac{L_5L_2}{\lambda L_7 R_S}+\frac{L_3L_6}{L_7^2}+\frac{L_4L_6}{L_7^2}\Big(\frac{L_2}{\lambda_M}+\frac{L_1}{R_M}\Big)\Big].
$$

Dividing both sides of that by $e^{\lambda v_S}$ and then taking sup on both sides we get

$$
||(\mathscr{T}d_1)(t)-\mathscr{T}(d_2)(t)||_\lambda \leq \frac{1}{\lambda}||d_1-d_2||_\lambda\Big[\frac{L_5L_2}{\lambda L_7 R_S}+\frac{L_3L_6}{L_7^2}+\frac{L_4L_6}{L_7^2}\Big(\frac{L_2}{\lambda_M}+\frac{L_1}{R_M}\Big)\Big]. \tag{42}
$$

Therefore, there has to exist a high enough λ for which our map $(\mathscr{T}d)(v_S)$ is a contraction in the metric space $(S, ||\cdot||_\lambda)$—which, by Banach's Fixed-Point Theorem means that $(\mathscr{T}d)(v_S)$ has a unique fixed point, which in turn means that Equation (32) has a single solution for any given $(v_S^c, v_M^c, S_S(0)) \in [0,1]^2 \times \Theta(M)$. Note that Equation (42) does not depend on $(v_S^c, v_M^c, S_S(0))$—and thus, by standard results (see e.g. Hasselblatt and Katok, 2003, p. 68) it follows that as $(\mathscr{T}d)(v_S, v_S^c, v_M^c, S_S(0))$ is continuous in v_S^c, v_M^c and $S_S(0)$ the fixed point—and thus the solution of (32)— is continuous in them as well.

Denote the fixed point of $(\mathscr{T}d)(\cdot, v_S^c, v_M^c, S_S(0))$ as $d^*(\cdot, v_S^c, v_M^c, S_S(0))$—then the following result holds

Lemma 6. The function $d^*(\cdot, v_S^c, v_M^c, S_S(0))$ is weakly decreasing in v_S^c and $S_S(0)$ and weakly increasing in v_M^c for all v_S's. Moreover, for some v_S's, $d^*(\cdot, v_S^c, v_M^c, S_S(0))$ is strictly decreasing in v_S^c and $S_S(0)$ (strictly increasing in v_M^c).

Proof. I start with the claims regarding $d(v_S, \cdot, v_M^c, S_S(0))$ and suppress v_M^c and $S_S(0)$

from notation for that part of the proof. Take any $v^c_{S2} > v^c_{S1} \in [0,1]$, denote $d^*(v_S, v^c_{S2}) - d^*(v_S, v^c_{S1})$ as $\Delta_{v^c_S} d^*(v_S, v^c_S)$ and for $v_S \geq v^c_S$ define

$$\begin{aligned} S_S(v_S, v^c_S) &= S_S(0) - \int_{v^c_S}^{v_S} \frac{\partial}{\partial v_S} C(d^*(r, v^c_S), r)\mathrm{d}r, \\ P_S(v_S, v^c_S) &= \frac{R_S - S_S(v_S, v^c_S)}{R_S}, \\ P_M(d^*(v_S, v^c_S), v^c_S) &= \frac{R_M - 1 + C(d^*(v_S, v^c_S), r) + S_S(v_S, v^c_S)}{R_M}. \end{aligned}$$

Then for any $v_S \geq v^c_{S2}$ we have

$$\begin{aligned} \Delta_{v^c} d^*(v, v^c) &= v^c_{S2} - v^c_{S1} \\ &+ \int_{v^c_S}^{v} \frac{\frac{\partial}{\partial v_S}\pi^e_S(t, P_S(t, v^c_{S2}))}{\frac{\partial}{\partial v_M}\pi^e_M(d^*(t, v^c_{S2}), P_M(d^*(t, v^c_{S2}), v^c_{S2}))} - \frac{\frac{\partial}{\partial v_S}\pi^e_S(t, P_S(t, v^c_{S1}))}{\frac{\partial}{\partial v_M}\pi^e_M(d^*(t, v^c_{S1}), P_M(d^*(t, v^c_{S1}), v^c_{S1}))}\mathrm{d}t. \end{aligned}$$

It is trivial that for any $v_S \in [v^c_{S1}, v^c_{S2})$ we have $\Delta_{v^c_S} d^*(v_S, v^c_S) < 0$, which proves the second (strict) part of this claim. Thus, we only need to show now that $\Delta_{v^c_S} d^*(v_S, v^c_S) \leq 0$ for all $v_S \in [v^c_{S2}, 1]$. Suppose not. Then the set $\Omega^{gen} = \{v_S \in [v^c_{S2}, 1] : \Delta_{v^c_S} d^*(v_S, v^c_S) > 0\}$ has to be non-empty. Then we have that for ${v_S}^g = \inf \Omega^{gen}$, $\Delta_{v^c_S} d^*({v_S}^g, v^c_S) = 0$ and $\Delta_{v^c_S} \frac{\partial}{\partial v_S} d^*({v_S}^g, v^c_S) \geq 0$. The sign of $\Delta_{v^c_S} \frac{\partial}{\partial v_S} d^*({v_S}^g, v^c_S)$ depends only on the signs of

$$\frac{\partial}{\partial v_S}\pi^e_S({v_S}^g, P_S(v^c_{S2}, {v_S}^g)) - \frac{\partial}{\partial v_S}\pi^e_S({v_S}^g, P_S(v^c_{S1}, {v_S}^g))$$

and

$$\frac{\partial}{\partial v_M}\pi^e_M(d^*({v_S}^g, v^c_{S1}), P_M(d^*({v_S}^g, v^c_{S1}), v^c_{S1})) - \frac{\partial}{\partial v_M}\pi^e_M(d^*({v_S}^g, v^c_{S2}), P_M(d^*({v_S}^g, v^c_{S2}), v^c_{S2})).$$

This can be easily seen by differentiation the expression that gives $\Delta_{v^c} d^*(v, v^c)$.[143] However, as $\Delta_{v^c_S} d^*({v_S}^g, v^c_S) = 0$ and both surplus functions are weakly supermodular, these in turn depend only on the sign of $S_S(v^c_{S2}, {v_S}^g) - S_S(v^c_{S1}, {v_S}^g)$. As for any $v_S \leq {v_S}^g$ it was the case that $\Delta_{v^c_S} d^*({v_S}^g, v^c_S) \leq 0$ and $v^c_{S2} > v^c_{S1}$, it follows that: $S_S(v^c_{S2}, {v_S}^g) - S_S(v^c_{S1}, {v_S}^g) < 0$ and thus:

$$\Delta_{v^c_S} \frac{\partial}{\partial v_S} d^*(v_S, v^c_S) < 0,$$

[143] To see this, note that

$$\begin{aligned} &\Delta_{v^c} \frac{\partial}{\partial v_S} d^*(v^g_S, v^c_S) = \frac{\frac{\partial}{\partial v_S}\pi^e_S(v^g_S, P_S(v^g_S, v^c_{S2}))}{\frac{\partial}{\partial v_M}\pi^e_M(d^*(v^g_S, v^c_{S2}), P_M(d^*(v^g_S, v^c_{S1}), v^c_{S1}))} - \frac{\frac{\partial}{\partial v_S}\pi^e_S(v^g_S, P_S(v^g_S, v^c_{S1}))}{\frac{\partial}{\partial v_M}\pi^e_M(d^*(v^g_S, v^c_{S1}), P_M(d^*(v^g_S, v^c_{S1}), v^c_{S1}))} \\ &= \frac{\frac{\partial}{\partial v_S}\pi^e_S(v^g_S, P_S(v^g_S, v^c_{S2})) - \frac{\partial}{\partial v_S}\pi^e_S(v^g_S, P_S(v^g_S, v^c_{S1}))}{\frac{\partial}{\partial v_M}\pi^e_M(d^*(v^g_S, v^c_{S2}), P_M(d^*(v^g_S, v^c_{S1}), v^c_{S1}))} \\ &+ \frac{\partial}{\partial v_S}\pi^e_S(v^g_S, P_S(v^g_S, v^c_{S1}))\frac{\frac{\partial}{\partial v_M}\pi^e_M(d^*(v^g_S, v^c_{S1}), P_M(d^*(v^g_S, v^c_{S1}), v^c_{S1})) - \frac{\partial}{\partial v_M}\pi^e_M(d^*(v^g_S, v^c_{S2}), P_M(d^*(v^g_S, v^c_{S1}), v^c_{S1}))}{\frac{\partial}{\partial v_M}\pi^e_M(d^*(v^g_S, v^c_{S1}), P_M(d^*(v^g_S, v^c_{S1}), v^c_{S1}))\frac{\partial}{\partial v_M}\pi^e_M(d^*(v^g_S, v^c_{S2}), P_M(d^*(v^g_S, v^c_{S1}), v^c_{S1}))}. \end{aligned}$$

which means that Ω^{gen} has to be empty and proves our first claim.

The proof for $S_S(0)$ is analogous.[144] For v_M^c, note that for a change in v_M^c, $\Delta_{v_M^c} d^*(v_S^c, v_M^c)$ is positive. The subsequent reasoning is analogous, but with opposite signs (the strict decreasingness follows from $\Delta_{v_M^c} d^*(v_S^c, v_M^c) < 0$ and continuity). $\square$

Everything I derived so far applies both for cases with abundant and scarce jobs. From now on, however, I will consider those cases separately.

Scarce jobs If $R_M + R_S < 1$, then $M = R_M + R_S$, which reduces (38) to $S_S(0) = R_S$ and gives $C(v_M^c, v_S^c) = 1 - R_M - R_S > 0$. For $(v_M, v_S) > 0$, $C(\bullet)$ is strictly increasing in both parameters, which allows us to define v_M^c as a strictly decreasing, continuous function of v_S^c. Define $\underline{v_S}$ as $v_M^c(\underline{v_S}) = 1$ and note that, as $v_M^c \in [0,1]$, Equation (34) shrinks the range of feasible v_S^c's to $[\underline{v_S}, 1]$. Hence, $d^*(v_S, v_S^c, v_M^c, S_S(0))$ depends only on v_S and v_S^c and is decreasing and continuous in v_S^c—I will denote it as $d^*(v_S, v_S^c)$ from now on. Thus, the modified system of equations reduces to

$$R_S = \int_{v_S^c}^{1} \frac{\partial}{\partial v_S} C^e(d^*(r, v_S^c), r)\mathrm{d}r.$$

The RHS is continuous in v_S^c, as $d^*(v_S, v_S^c))$ is continuous in v_S^c. For $v_S^c = \underline{v_S}$, we have $d^*(v_S, v_S^c) \geq 1$ regardless of v_S and therefore $\int_0^1 \frac{\partial}{\partial v_S} C^e(d^*(r, v_S^c), r)\mathrm{d}r = 1$, whereas for $v_S^c = 1$, $\int_1^1 \frac{\partial}{\partial v_S} C^e(d^*(r, v_S^c), r)\mathrm{d}r = 0$; thus, a solution to (35) (given $R_S \in (0,1)$) exists. It is unique, as $d^*(v_S, \cdot)$ is weakly decreasing for all v_S and strictly decreasing for some v_S and thus the RHS crosses R_S only once from above.

Abundant jobs If $R_M + R_S \geq 1$, then $M = 1$ and thus $C(v_M^c, v_S^c) = 0$. Hence, $\min\{v_M^c, v_S^c\} = 0$ and I cannot define v_M^c as a function of v_S^c, as there is a continuum of v_S^c's for which $C(0, v_S^c) = 0$. I address this by defining the set $\Gamma^c = \{(v_M^c, v_S^c) : \min\{v_M^c, v_S^c\} = 0\}$, a new variable $a \in [-1,1]$ and writing v_M^c and v_S^c as

$$v_M^c(a) = \begin{cases} -a & \text{for } a \leq 0, \\ 0 & \text{for } a > 0, \end{cases} \qquad v_S^c(a) = \begin{cases} 0 & \text{for } a \leq 0, \\ a & \text{for } a > 0. \end{cases}$$

For any a, $(v_M^c(a), v_S^c(a)) \in \Gamma^c$ and for any $(v_M^c, v_S^c) \in \Gamma^c$ there exists a unique a, such that $(v_M^c(a), v_S^c(a)) = (v_M^c, v_S^c)$. Thus, if there exists a unique a that solves Equation (35), there also exists a unique (v_M^c, v_S^c) that solves it. Moreover, $v_S^c(a)$ is continuous and increasing, and $v_M^c(a)$ is continuous and decreasing. Therefore the function $d^*(v_S, a, S_S(0)) = d^*(v_S, v_S^c(a), v_M^c(a), S_S(0))$ is continuous and decreasing (strictly for

[144] For $v_S = v_S^c$ we have $\Delta_{S_S(0)} d^*(v_S, S_S(0)) = 0$ and $\Delta_{S_S(0)} \frac{\partial}{\partial v_S} d^*(v_S, S_S(0)) < 0$. The sign of $\Delta_{S_S(0)} \frac{\partial}{\partial v_S} d^*(v_S{}^g, S_S(0))$ depends on $S_{S1}(0) - S_{S2}(0) < 0$ and the difference in $S_S(v_S, S_S(0))$, which is weakly negative for the same reasons as above. Thus, $\Delta_{S_S(0)} \frac{\partial}{\partial v_S} d^*(v_S{}^e, S_S(0)) \leq 0$, which implies that $d^*(v_S, \cdot)$ will never strictly increase.

some v_S's) in a. Thus, I can write Equation (35) as

$$S_S(0) = \begin{cases} \int_0^1 \frac{\partial}{\partial v_S} C^e(d^*(r,a,S_S(0)),r)\mathrm{d}r & \text{for } a < 0, \\ \int_a^1 \frac{\partial}{\partial v_S} C^e(d^*(r,a,S_S(0)),r)\mathrm{d}r & \text{for } a \geq 0. \end{cases}$$

The RHS is continuous in a, as $d^*(v_S,a,S_S(0))$ is continuous in a. For $a=-1$, we have $\int_0^1 \frac{\partial}{\partial v_S} C^e(d^*(r,a,S_S(0)),r)\mathrm{d}r = 1$; for $a=1$, we have $\int_a^1 \frac{\partial}{\partial v_S} C^e(d^*(r,a,S_S(0)),r)\mathrm{d}r = 0$; thus, a solution to (35) (given $S_S(0) \in \Theta(1)$) exists. It is unique, as $d^*(v_S,\cdot,S_S(0))$ is weakly decreasing for all and strictly decreasing for some v_S and thus the RHS crosses $S_S(0)$ only once from above.

As $d^*(v_S,\cdot,\cdot)$ is continuous, $a(S_S(0))$ is continuous as well. It is strictly decreasing in $S_S(0)$, as the LHS is strictly increasing in $S_S(0)$ and the RHS is weakly decreasing in $S_S(0)$ and strictly decreasing in a; thus, if $S_S(0)$ increases, Equation (35) is met only if a decreases. As $a(S_S(0))$ is unique and a defines uniquely (v_M^c, v_S^c), there exist unique $v_M^c(S_S(0))$ and $v_S^c(S_S(0))$; the former is non-decreasing and the latter non-increasing; and for any $S_{S2}(0) > S_{S1}(0)$ we have that $v_M^c(S_{S2}(0)) > v_M^c(S_{S1}(0))$ or $v_S^c(S_{S2}(0)) < v_S^c(S_{S1}(0))$.

The modified set reduces to

$$S_S(0) > 1 - R_M \Rightarrow \pi_M\Big(u^c(S_S(0)), \frac{R_M - 1 + S_S(0)}{R_M}\Big) \leq \pi_S\Big(v^c(S_S(0)), \frac{R_S - S_S(0)}{R_S}\Big) \tag{43}$$

$$S_S(0) < R_S \Rightarrow \pi_M\Big(u^c(S_S(0)), \frac{R_M - 1 + S_S(0)}{R_M}\Big) \geq \pi_S\Big(v^c(S_S(0)), \frac{R_S - S_S(0)}{R_S}\Big) \tag{44}$$

$$S_S(0) \in \Theta(1). \tag{45}$$

Note that $v_M^c(0) = 0$, $v_S^c(0) = 1$, $v_M^c(1) = 1$ and $v_S^c(1) = 0$. Condition (43)—(44) will be trivially met if there exists some $S_S(0) \in \Theta(1)$ such that

$$\pi_M\Big(v_M^c(S_S(0)), \frac{R_M - 1 + S_S(0)}{R_M}\Big) = \pi_S\Big(v_S^c(S_S(0)), \frac{R_S - S_S(0)}{R_S}\Big).$$

If there is no such $S_S(0)$, then it has to be the case that either (a) $LHS > RHS$ for all $S_S(0) \in \Theta(1)$ or (b) $RHS > LHS$ for all $S_S(0) \in \Theta(1)$. However, (a) is possible only if $\max\{0, 1 - R_M\} = 1 - R_M$, as $LHS > RHS$ for $S_S(0) = 0$ violates condition (d). And for $S_S(0) = 1 - R_M$, $LHS > RHS$ meets (43)—(44), as the first inequality doesn't have to hold. For similar reasons, (b) is possible only if $\min\{1, R_S\} = R_S$, in which case $RHS > LHS$ meets (43)—(44). Thus, existence of a solution to (43)—(44) follows. Hence, there exists a solution to the modified and original sets.

For uniqueness, remember that $d^*(v_S, a(S_S(0)), S_S(0))$ is unique and, thus, it suffices to show that the solution to (43)—(44) is unique. Denote the set of all $S_S(0) \in \Theta(1)$ that meet (43)—(44) as Ω^M. Consider $\min \Omega^M = S_{S1}(0)$. Note that $S_{S1}(0)$ exists as Ω^M

is non-empty and $\pi_M(\cdot,\cdot)$, $\pi_S(\cdot,\cdot)$, $v_S^c(\cdot)$ and $v_M^c(\cdot)$ are continuous. Suppose $S_{S1}(0) = \min\{1, R_S\}$—then the solution is unique. Now suppose that $S_{S1}(0) < \min\{1, R_S\}$, which implies that for any $S_{S2}(0) \in \Omega^M$ such that $S_{S2}(0) > S_{S1}(0)$ we need to have

$$\pi_M\Big(v_M^c(S_{S2}(0)), \frac{R_M - 1 + S_{S2}(0)}{R_M}\Big) \leq \pi_S\Big(v_S^c(S_{S2}(0)), \frac{R_S - S_{S2}(0)}{R_S}\Big)$$

and for $S_{S1}(0)$ we have:

$$\pi_M\Big(v_M^c(S_{S1}(0)), \frac{R_M - 1 + S_{S1}(0)}{R_M}\Big) \geq \pi_S\Big(v_S^c(S_{S1}(0)), \frac{R_S - S_{S1}(0)}{R_S}\Big).$$

This is a contradiction, as $\frac{\partial}{\partial v_M}\pi_M > 0$, $\frac{\partial}{\partial h}\pi_M \geq 0$, $\frac{\partial}{\partial v_S}\pi_S > 0$, $\frac{\partial}{\partial h}\pi_S \geq 0$, $v_M^c(\cdot)$ is weakly increasing, $v_S^c(\cdot)$ is weakly decreasing and $v_M^c(S_{S2}(0)) > v_M^c(S_{S1}(0)) \vee v_S^c(S_{S2}(0)) < v_S^c(S_{S1}(0))$. Thus $S_{S2}(0)$ does not exist and $S_{S1}(0)$ is the only element in Ω^M, which completes the proof.

Proof of Lemma 3

First note that if $S_M(0) + S_S(0) < 1$ then $v_M^c, v_S^c \in (0,1)$, and the reasoning from the proof of Lemma 1 implies then that $w_i(v_i^c) = 0$. As $S_i(0) \leq R_i$ this proves the last statement immediately. The first claim follows by Proposition 1, because $S_i(0) < R_i$ implies that both $w_i(v_i^c) > 0$ and $S_M(0)+S_S(0) < 1$—contradiction. Finally, $G_i(v_i) = 1 - \frac{S_i(v_i)}{R_i}$ follows from $S_i(0) = R_i$ and the definitions of the distribution of relative skills in sector i and the supply of relative skills in sector i.

C Comparative Statics

To simplify what follows, I first introduce new notation. The difference between the new and old values of any object O is denoted as $\Delta_\theta O$. The greater of the old and new values of O is denoted as $\max O$. Thus, for instance, the measure of manufacturing workers is denoted by $\Delta_\theta S_M(0)$ and the greater critical skill in services is denoted by $\max v_S^c$. Additionally, I define the *star skill in services* as

$$\bar{v}_S = \sup\{v_S \in [0,1] : \psi(v_S) < 1\}.$$

This definition implies that all workers with $v_S > \bar{v}_S$ join services. The manufacturing analogue can be defined as $\bar{v}_M = \psi(\bar{v}_S)$. Finally, the positive assortative matching function is defined as $P_i(v_i) = 1 - \frac{S_i(v_i)}{R_i}$.

C.1 Wage Polarization

Proof of Proposition 2

I start with the 'only if' part. Suppose that there exists some (v'_M, v'_S) such that $C(v'_M, v'_S, \theta_1) > C(v'_M, v'_S, \theta_2)$. Then there exists a quadruple (π_M, π_S, R_M, R_S) that meets Assumptions 1 and 3 for which $\max_{S\in\mathbf{S}(\theta_2)} T(S) > \max_{S\in\mathbf{S}(\theta_1)} T(S)$. Consider $R_M = R_S = 1$ and following surplus functions: $\pi_M(v_M) = 0$ if $v_M \leq v'_M$ and $\pi_M(v_M) = 1$ otherwise, whereas $\pi_S(v_S) = 0.5$ if $v_S \leq v'_S$ and $\pi_S(v_S) = 1.5$ otherwise. Then the efficient assignment of workers to sectors is such that any worker with $v_M > v'_M$ and $v_S < v'_S$ works in manufacturing and all other workers work in services. The measure of workers in manufacturing is, thus, $v'_S - C(v'_M, v'_S)$ and the maximal total surplus produced in the economy is $1.5(1-v'_S)+0.5C(v'_M, v'_S)+1(v'_S-C(v'_M, v'_S))$ giving $1.5-0.5v'_S-0.5C(v'_M, v'_S)$ which is then lower for $C(v'_M, v'_S, \theta_1)$ than $C(v'_M, v'_S, \theta_2)$, as required.

The proof of the 'only if' part is not complete yet, as these surplus functions do not meet the differentiability assumption. However, they can be approximated by the following pair of surplus functions that meet Assumptions 1 and 3: $\pi_M(v_M, h_M) = \frac{1}{1+\exp(-2k(v_M - v'_M))}$ and $\pi_S(v_S, h_S) = \frac{1}{1+\exp(-2k(v_S - v'_S))} + 0.5$. As $k \to \infty$ these two functions approach the functions outlined above pointwise. It has been shown by Gretsky et al. (1992) that the equilibrium of an assignment game is efficient and it follows from the proof of Theorem 3 that the equilibrium is continuous in any parameters in which the surplus functions are continuous. Thus, it follows by the definition of a limit and by Assumption 2 that for any difference in copulas $C(v'_M, v'_S, \theta_1) - C(v'_M, v'_S, \theta_2) > 0$ there exists k large enough that $\max_{S\in\mathbf{S}(\theta_2)} T(S) > \max_{S\in\mathbf{S}(\theta_1)} T(S)$.

As for the 'if' part it suffices to show that an increase in concordance implies that $\mathbf{S}(\theta_2) \subset \mathbf{S}(\theta_1)$. Define a function $PPF(S; \theta_j) = S_S(v_S) + S_M(\psi(v_S)) - 1 + C(\psi(v_S), v_M)$. It follows that $S \in \mathbf{S}(\theta_j)$ if and only if $PPF(S; \theta_j) \leq 0$. I will now show that if $PPF(S; \theta_2) \leq 0$ then $PPF(S; \theta_1) \leq 0$ as well, which will prove that the set of feasible pairs of supply functions expands as concordance falls.

$$\begin{aligned} PPF(S;\theta_1) - PPF(S;\theta_2) = {} & C(\psi(v_S;\theta_1), v_S;\theta_1) - C(\psi(v_S;\theta_1), v_S;\theta_2) \\ & + \int_{\psi(v,\theta_2)}^{\psi(v,\theta_1)} \frac{\partial}{\partial v_M} C\big(r, v_S, \theta_2)\big) - \frac{\partial}{\partial v_M} C\big(r, \phi(r,\theta_2), \theta_2\big) \mathrm{d}r. \end{aligned}$$

If $\psi(v_S, \theta_1) \geq \psi(v_S, \theta_2)$ then for any $r \in [\psi(v_S, \theta_2), \psi(v_S, \theta_1)]$, $\phi(r, \theta_2) \geq v_S$ and the expression is negative, as required. If $\psi(v_S, \theta_1) < \psi(v_S, \theta_2)$ then for any $r \in [\psi(v_S, \theta_1), \psi(v_S, \theta_2)]$, $\phi(r, \theta_2) < v_S$ and the expression is negative again.

Proof of Proposition 3

The increase in absolute terms can be proved by just slightly generalizing the reasoning from the main body. Formally, this follows from the following lemma, which is proved in Online Appendix OA.1. Given the old equilibrium separation function the *generalized regularity condition* is defined as follows: there exists some $v' > v_S^c$, such that $\frac{\partial}{\partial v_S}C(\psi(v';\theta_1), v';\theta_2) - \frac{\partial}{\partial v_S}C(\psi(v';\theta_1), v';\theta_1) \neq 0$ and $\mathrm{sgn}\big(\frac{\partial}{\partial v_S}C(\psi(v;\theta_1), v;\theta_2) - \frac{\partial}{\partial v_S}C(\psi(v;\theta_1), v;\theta_1)\big) = \mathrm{sgn}\big(\frac{\partial}{\partial v_S}C(\psi(v';\theta_1), v';\theta_2) - \frac{\partial}{\partial v_S}C(\psi(v';\theta_1), v';\theta_1)\big)$ for all $v \in (v_S^c, v')$.

Lemma 7. Suppose the concordance of the distribution of relative skills increases and the change satisfies the generalized regularity condition. As long as (i) the change in concordance is unbiased, i.e. $\Delta_\theta w_S(v_S) = \Delta_\theta w_M(\psi(v_S;\theta_1))$ for all $v_S \geq v_S^c$ and (ii) the increase in concordance leaves the critical skill levels (v_M^c, v_S^c) unchanged, then wage polarization increases in absolute terms. Further, the lowest wage remains unchanged ($\Delta_\theta w_i(v_i^c) = 0$) and the highest wage increases ($\Delta_\theta w_i(1) \geq 0$).

Clearly, under symmetry $w_M(v_M) = w_S(v_S)$ and $\psi(v_S) = v_S$ for all $v_S \geq v_S^c$ and so the Lemma applies. The increase in log-wage range follows trivially from $\Delta_\theta w_i(v_i^c) = 0$ and $\Delta_\theta w_i(1) \geq 0$. To see that left-tail log-wage inequality must fall, note that

$$\Delta_\theta\Big(\ln W(t) - \ln W(0)\Big) = \Delta_\theta \ln W(t) = \ln\Big(\frac{\Delta_\theta W(t)}{W(t;\theta_1)} + 1\Big).$$

The RHS has the same sign as $\Delta_\theta W(t)$ and the fall in left-tail inequality in relative terms follows from the fact that it falls in absolute terms.

C.2 Vertical Differentiation

Definition 13. *Vertical differentiation* in manufacturing increases by (strictly) more than in services if, for all (v_M, h):

$$\psi_{v_S}(v_S;\theta_1)\Delta_c \frac{\partial}{\partial v_M}\pi_M(\psi(v_M;\theta_1), P_M(\psi(v_s);\theta_1)) \geq (>) \Delta_c \frac{\partial}{\partial v_S}\pi_s(v_s, P_S(v_S;\theta_1)).$$

Note that a manufacturing-specific increase in vertical differentiation (Definition 6) implies trivially that vertical differentiation increased by more in manufacturing than in services.

Definition 14. The matching problems $(Q(\theta_1), Q(\theta_2))$ have *(strong) impossibility property* if it is impossible that $v_S^c(\theta_2) < (\leq) v_S^c(\theta_1)$ and $\Delta_\theta S_S(0) > (\geq) 0$.

Theorem 3. Suppose $(Q(\theta_1), Q(\theta_2))$ exhibit the impossibility property, R_S is unchanged and R_M weakly increases. If vertical differentiation increases by more in manufacturing than services then (i) $S_S(v_S;\theta_2) \leq S_S(v_S;\theta_1)$ for all v_S and (ii) $S_M(v_M;\theta_2) \geq S_M(v_M;\theta_1)$ for all v_M. If the impossibility property is strong, then (i) holds strictly for a positive measure of v_S and (ii) for a positive measure of v_M. If the increase in differentiation is strict, then (iii) $S_S(v_S;\theta_2) < S_S(v_S;\theta_1)$ for all $v_S \in [\max v_S^c, \max \bar{v}_S)$ and (iv) $S_M(v_M;\theta_2) > S_M(v_M;\theta_1)$ for all $v_M \in [\max v_M^c, \max \bar{v}_M)$.

Proof of Theorem 3. The results for services are proved in a series of lemmas and the result for manufacturing follow easily (details at the end of the proof). But first, I define the following three sets of services talent levels

$$\begin{aligned}
\Xi^0 &= \{v_S \in [\max v_S^c, \min \bar{v}_S] : \; S_S(v_S;\theta_2) \geq S_S(v_S;\theta_1)\} \\
\Xi^1 &= \{v_S \in [\max v_S^c, \min \bar{v}_S] : \; \psi(v_S;\theta_2) < \psi(v_S;\theta_1) \wedge S_S(v_S;\theta_2) > S_S(v_S;\theta_1)\} \\
\Xi^2 &= \{v_S \in [\max v_S^c, \min \bar{v}_S) : \; \psi(v_S;\theta_2) \leq \psi(v_S;\theta_1) \wedge S_S(v_S;\theta_2) \geq S_S(v_S;\theta_1)\}
\end{aligned}$$

as well as the function $\kappa : [\max v_S^c, \min \bar{v}_S] \to \mathbf{R}$:

$$\kappa(v_s) = \Delta_\theta w_S(v_s) - \Delta_\theta w_M(\psi(v_S;\theta_1)).$$

Lemma 8. Suppose that vertical differentiation increases by (strictly) more in manufacturing than services. Then $\frac{\partial}{\partial v_S}\kappa(v_s) \leq (<) 0$ for all $v_S \in \Xi^0$.

Proof of Lemma 8. Remember that $\frac{\partial}{\partial v_S} P_S(v_S) = \frac{\psi_{v_S}(v_S)\frac{\partial^2}{\partial v_M \partial v_S} C(\psi(v_S), v_S)}{R_S}$. Take any $v_{S0} \in \Xi^0$ Note that by Equation (31) we have $\Delta_\theta P_M(\psi(v_{S0};\theta_1)) \geq 0$. Then we have

$$\begin{aligned}
\Delta_\theta \frac{\partial}{\partial v_M} w_M(\psi(v_0;\theta_1)) &= \Delta_\theta \frac{\partial}{\partial v_M} \pi_M(\psi(v_0;\theta_1), P_M(\psi(v_0;\theta_1);\theta_2)) \\
&\quad + \int_{P_M(\psi(v_0;\theta_1);\theta_1)}^{P_M(\psi(v_0;\theta_1);\theta_2)} \frac{\partial^2}{\partial v_M \partial h} \pi_M(\psi(v_0;\theta_1), r;\theta_1) \mathrm{d}r \geq (>) 0,
\end{aligned}$$

as $\Delta_\theta \frac{\partial}{\partial v_M}\pi_M(v_M, h) \geq (>) 0$ for any (v_M, h), $\pi_M(\bullet)$ is supermodular and $\Delta_\theta P_M(\psi(v_{S0};\theta_1)) \geq 0$. Whereas for v_{S0} we have:

$$\Delta_\theta \frac{\partial}{\partial v_S} w_S(v_{S0}) = \Delta_c \frac{\partial}{\partial v_S} \pi_s(v_s, P_S(v_S;\theta_1)) + \int_{P_S(v_{S0};\theta_1)}^{P_S(v_{S0};\theta_2)} \frac{\partial^2}{\partial v_S \partial h} \pi_S(v_{S0}, r) \mathrm{d}r \leq 0,$$

as $\pi_S(\bullet)$ is supermodular and $\Delta_\theta P_S(v_{S0}) \leq 0$. By differentiating Equation (**??**) wrt to v for both θ_2 and θ_1, taking differences and rearranging, we arrive at

$$\frac{\partial}{\partial v_S}\kappa(v_s) = \left[\Delta_\theta \frac{\partial}{\partial v_S} w_S(v_{S0}) - \psi_{v_s}(v_{S0};\theta_1)\Delta_\theta \frac{\partial}{\partial v_M} w_M(\psi(v_{S0};\theta_1)\right],$$

from which follows trivially that $\frac{\partial}{\partial v_S}\kappa(v_s) \leq (<) 0$. □

Lemma 9. Suppose that $\frac{\partial}{\partial v_S}\kappa(v_s) \leq (<)0$ for all $v_S \in \Xi^0$. Then for any $v_1 \in \Xi^1(\Xi^2)$ it is the case that $(v_1, \min \bar{v}_S] \subset \Xi^1$.

Proof of Lemma 9. First, note that $\kappa(v_s) = \int_{\psi(v_S;\theta_1)}^{\psi(v_S;\theta_1)} \frac{\partial}{\partial v_S}\pi_M(r, P(r;\theta_2);\theta_2)\mathrm{d}r$. Because $\frac{\partial}{\partial v_S}\pi_M > 0$ it follows that $\operatorname{sgn}(\Delta_\theta\psi(v_S)) = \operatorname{sgn}(\kappa(v_s))$. In particular, this means that $\kappa(v_1) < (\leq)0$. Second, define the set $\Xi^3 = \{v_S \in [v_1, \min \bar{v}_S] : v_S \notin \Xi^1\}$.

I will first show the result for $v_1 \in \Xi^1$. Suppose Ξ^3 is non-empty—then continuity of ψ and S_S implies that $\min \Xi^3$ exists; clearly $\min \Xi^3 > v_1$. Further, $[v_1, \min \Xi^3] \subset \Xi^0$. Therefore, the following is true for all $v \in [v_1, \min \Xi^3]$: (a) $\frac{\partial}{\partial v_S}\kappa(v_s) \leq 0$, (b) $\kappa(v_1) < 0$ and (a) imply that $\kappa(v_s) < 0$, which further implies that (c) $\Delta_\theta\psi(v_S) < 0$. However, the last fact implies that

$$\Delta_\theta S_S(\min X_3) = \Delta_\theta S_S(v_1) + \int_{\min v_1}^{\min \Xi_3} \int_{\psi(r;\theta_1)}^{\psi(r;\theta_2)} \frac{\partial^2}{\partial v_M \partial v_S} C(s, r)\mathrm{d}s\mathrm{d}r < 0,$$

and $\min \Xi_3 \in \Xi_1$; contradiction!

Now suppose that $v_1 \in \Xi^2$. By continuity of κ and the fact that $\frac{\partial}{\partial v_S}\kappa(v_1) < 0$, there must exist some $v_2 > v_1$ such that for all $v_S \in [v_1, v_2]$ we have $\frac{\partial}{\partial v_S}\kappa(v_1) < 0$. It follows that $\kappa(v_S) < 0$ and $\Delta_\theta\psi(v_S) < 0$ for all $v_S \in (v_1, v_2]$, from which follows that $\Delta_{\theta_3} S_S(v_S) < 0$ for all $v_S \in (v_1, v_2]$. Therefore, $(v_1, v_2] \subset \Xi^1$; by the reasoning above follows that $[v_2, \min \bar{v}_S \subset \Xi^1$; combining these two completes the proof. □

Lemma 10. Suppose that for any $v_1 \in \Xi^2(\Xi^1)$ it is the case that $[v_1, \min \bar{v}_S]\Xi^1$. Then Ξ^2 (Ξ^1) is empty.

Proof of Lemma 10. Take any $v_{S1} \in \Xi^2(X_1)$. This implies that $\Delta_\theta\psi(v_S) < 0$ for all $v_S \in [v_{S1}, \min \bar{v}_S]$, which implies that $\bar{v}_S(\theta_2) > \bar{v}_S(\theta_1)$. $\Delta_\theta S_S(v_{S1})$ can be expanded into:

$$\begin{aligned}
\Delta_\theta S_S(v_{S1}) &= \int_{v_{S1}}^{\bar{v}_S(\theta_2)} \frac{\partial}{\partial v_S} C(\psi(v_S;\theta_2), v_S)\mathrm{d}v_S - \int_{v_{S1}}^{\bar{v}_S(\theta_1)} \frac{\partial}{\partial v_S} C(\psi(v_S;\theta_1), v_S)\mathrm{d}v_S - \Delta_\theta \bar{v}_S \\
&= \int_{v_{S1}}^{\bar{v}_S(\theta_1)} \int_{\psi(v_S;\theta_1)}^{\psi(v_S;\theta_2)} \frac{\partial^2}{\partial v_M \partial v_S} C(s, v_S)\mathrm{d}s\mathrm{d}v_S - \int_{\bar{v}_S(\theta_1)}^{\bar{v}_S(\theta_2)} 1 - \frac{\partial}{\partial v_S} C(\psi(v_S;\theta_2), v_S)\mathrm{d}v_S.
\end{aligned}$$

The LHS is (strictly) positive, whereas the RHS is strictly negative—contradiction. Thus $\Xi^2(\Xi^1)$ must be empty, as required. □

Lemma 11. Suppose Ξ^1 is empty. Consider some $v_{Se} \in [\max v_S^c, \min \bar{v}_S]$. Then $\Delta_\theta S_S(v_{Se}) \leq 0$ implies $\Delta_\theta S_S(v_{Se}) \leq 0$ for all $v_S \in [v_{Se}, \min \bar{v}_S]$. If Ξ^2 is empty, then additionally $\Delta_\theta S_S(v_{Se}) < 0$ implies $\Delta_\theta S_S(v_{Se}) < 0$ for all $v_S \in [v_{Se}, \min \bar{v}_S)$.

Proof. I will start with the first claim. Suppose it is false. Then the set $\Upsilon^1 = \{v_S \in [v_{Se}, \min \bar{v}_S] : \Delta_\theta S_S(v_S) > 0\}$ has to be non-empty. Take some ${v_S}^1 \in \Upsilon^1$ and define $\Upsilon^2 = \{v_S \in [v_{Se}, v_{S1}] : \Delta_\theta S_S(v_S) \leq 0\}$. By continuity of $\Delta_\theta S_S(v_S)$ the point ${v_S}^2 = \max \Upsilon^2$

exists and is $< {v_S}^1$. Therefore, for any $v_S \in ({v_S}^2, {v_S}^1]$ we have $\Delta_\theta S_S(v_S) > 0$. However, as:

$$\Delta_\theta S_S({v_S}^1) = \Delta_\theta S_S({v_S}^2) - \int_{{v_S}^2}^{{v_S}^1} \int_{\psi(r;\theta_1)}^{\psi(r;\theta_2)} \frac{\partial^2}{\partial v_M \partial v_S} C(s,r) \mathrm{d}s\mathrm{d}r,$$

this implies that there exists some $v_{S1} \in ({v_S}^2, {v_S}^1]$ such that $\Delta_\theta \psi(v_{S1}) < 0$ and thus $v_{S1} \in \Xi^1$—contradiction.

Let us move to the second claim. Again, suppose it is false. Then the set $\Upsilon^3 = \{v_S \in [v_{Se}, \min \bar{v}_S] : \Delta_\theta S_S(v_S) \geq 0\}$ has to be non-empty; but as $\Delta_\theta S_S(v_S)$ is continuous in v, the non-emptiness implies that ${v_S}^3 = \min \Upsilon^3$ exists. Additionally, ${v_S}^3 > v_{Se}$, as $\Delta_\theta S_S(v_{Se}) < 0$. Define a new set $\Upsilon^4 = \{v_S \in [v_{Se}, {v_S}^3] : \Delta_\theta \psi(v_S) \leq 0\}$ and ${v_S}^4 = \max \Upsilon^4$; by definition of ${v_S}^3$, for any $v_S < {v_S}^3 \wedge \in \Upsilon^4$ we have that $\Delta_\theta S_S(v_S) < 0$. As $[v_{Se}, {v_S}^3]$ is a compact set and $\Delta_\theta \psi(v_S)$ is continuous ${v_S}^4$ won't exist only if Υ^4 is empty; but an empty Υ^4 implies that $\Delta_\theta \psi(v_S) > 0$ for any $v_S \in [v_{Se}, {v_S}^3]$, which in turn means that $\Delta_\theta S_S({v_S}^3) < 0$, which contradicts the definition of ${v_S}^3$. Therefore ${v_S}^4$ needs to exist. Now suppose that ${v_S}^4 < {v_S}^3$; then we have $\Delta_\theta S_S({v_S}^4) < 0$ and for any $v_S \in ({v_S}^4, {v_S}^3], \Delta_\theta \psi(v_S) > 0$, which implies that $\Delta_\theta S_S({v_S}^3) < 0$ and also contradicts the definition of ${v_S}^3$. Therefore it has to be the case that ${v_S}^3 = {v_S}^4$; but this implies that $\Delta_\theta(\psi({v_S}^3)) \leq 0$ and $\Delta_\theta S_S({v_S}^3) \geq 0$, which contradicts emptiness of Ξ^2 □

Lemma 12. $\Delta_\theta S_S(\min \bar{v}_S) \leq 0$ implies that (i) for any $v_S > \max \bar{v}_S$ we have $\Delta_\theta S_S(v_S) \leq 0$ and (ii) for all $v_S \in [\min \bar{v}_S, \max \bar{v}_S)$ we have $\Delta_\theta S_S(v_S) < 0$.

Proof. Note that $\Delta_\theta S_S(\min \bar{v}_S) < (\leq) 0$ implies that $\bar{v}_S(\theta_2) > (\geq) \bar{v}_S(\theta_1)$[145]. Thus, if $\Delta_\theta S_S(\min \bar{v}_S) = 0$ then $\min \bar{v}_S = \max v_S^c$ and the second claim follows trivially. Whereas if $\Delta_\theta S_S(\min \bar{v}_S) < 0$ then $\bar{v}_S(\theta_2) > \bar{v}_S(\theta_1)$ and by the fact that all agents with $v_S \in (\bar{v}_S, 1]$ join services for sure it follows that for $v_S \in (\bar{v}_S(\theta_1), \bar{v}_S(\theta_2))$ we also have $\Delta_\theta S_S(v_S) < 0$. Claim (i) for $v_S > \max \bar{v}_S$ follows easily from the aforementioned property of $\bar{v}_S$. □

Lemma 13. The (strong) impossibility property implies that if $v_S^c(\theta_2) < (\leq) v_S^c(\theta_1)$ then $\Delta_\theta S_S(v_S^c(\theta_1)) < 0$.

Proof. This follows from the fact that $\Delta_\theta v_S^c < (\leq) 0$ implies that $G_S(v_S^c(\theta_1)) > (\geq) 0$, the fact that:

$$\Delta_\theta S_S(v_S) = \big(1 - (G_S(v_S;\theta_2)\big) \Delta_\theta S_S(0) - S_S(0)(\theta_1) \Delta_\theta G_S(v_S) \tag{46}$$

and the fact that $v_S^c(\theta_1) < 1$ and thus $1 - G_S(v_S;\theta_2) > 0$. □

145 To see this, denote the θ_j for which $\bar{v}_S(c_i) = \max \bar{v}_S$ as θ_m; then we have

$$\Delta_\theta S_S(\min \bar{v}_S;\theta_1) = \int_{\bar{v}_S(\theta_2)}^{\bar{v}_S(\theta_1)} 1 - \frac{\partial}{\partial v_S} C(\psi(v_S, c_m), v_S) \mathrm{d}v_S.$$

As $1 - \frac{\partial}{\partial v_S} C(\psi(v_S, c_m), v_S) \geq 0$, the fact that $\Delta_\theta S_S(\min \bar{v}_S) < (\leq) 0$ implies that for this to hold we need $\bar{v}_S(\theta_2) > (\geq) \bar{v}_S(\theta_1)$.

Lemma 14. Empty Ξ^1 and impossibility property jointly imply $\Delta_\theta S_S(\max v_S^c) \leq 0$. If either the increase in vertical differentiation is strict or the property is strong then this inequality holds strictly.

Proof. Suppose (strong) impossibility property holds. Define a set $\Xi^5 = \{v_S \in [\max v_S^c, \max \bar{v}_S) : \Delta_\theta \psi(v_{S5}) < 0 \wedge \Delta_\theta S_S(v_{S5}) \geq 0\}$. By continuity, there has to exist some arbitrarily small $\epsilon > 0$ such that $v_{S5} + \epsilon \in \Xi^1$; thus, by Lemma 10, an increase in vertical differentiation implies that Ξ^5 has to be empty.

If $v_S^c(\theta_2) < (\leq) v_S^c(\theta_1)$, then by Lemma 13 we have $\Delta_\theta S_S(\max v_S^c) < 0$. If $v_S^c(\theta_2) \geq v_S^c(\theta_1)$ and $\max v_S^c \geq \min \bar{v}_S$, then—as $\bar{v}_S > v_S^c$—it has to be that $\bar{v}_S(\theta_2) > v_S^c(\theta_2) > \bar{v}_S(\theta_1)$. But as all agents with $v_S > \bar{v}_S$ join services, this implies $\Delta_\theta S_S(v_S^c(\theta_2)) < 0$.

Thus, we only need to show the result for $\max v_S^c < \min \bar{v}_S$ and $v_S^c(\theta_2) \geq (>) v_S^c(\theta_1)$. As $\Delta_\theta R_M \geq 0$ we have $C(v_M^c(\theta_1), v_S^c(\theta_1)) \geq C(v_M^c(\theta_2), v_S^c(\theta_2))$ and thus $\Delta_\theta v_S^c \geq (>) 0$ implies $\Delta_\theta v_M^c \leq 0$. As $\psi(v_S^c) = v_M^c$ and $\psi(v_S)$ is strictly increasing for any c we have: $\psi(v_S^c(\theta_2); \theta_1) \geq (>) v_M^c(\theta_1)$, $v_M^c(\theta_1) \geq v_M^c(\theta_2)$ and $v_M^c(\theta_2) = \psi(v_S^c(\theta_2); \theta_2)$, which trivially implies that

$$\Delta_\theta \psi(v_S^c(\theta_2)) \leq (<) 0.$$

If the impossibility property holds, then this inequality holds weakly, which together with empty Ξ^1 implies $\Delta_\theta S_S(v_S^c(\theta_1)) \leq 0$. If the impossibility property is strong, then $\Delta_\theta \psi(v_S^c(\theta_2)) < 0$, which—as Ξ^5 is empty—implies $\Delta_\theta S_S(\max v_S^c) < 0$. If Ξ^2 is empty, then we have that $\Delta_\theta \psi(v_S^c(\theta_2)) \leq 0$ implies $\Delta_\theta S_S(\max v_S^c) < 0$, which concludes the proof. □

Lemma 15. Empty Ξ^1 and impossibility properties imply jointly that for any $v_S < \max v_S^c$, $\Delta_\theta S_S(v_S) \leq 0$.

Proof. Suppose $\Delta_\theta v_S^c < 0$—then for all $v_S < \max v_S^c$ we have that $G_S(v_S^c(\theta_1)) \geq 0$ and by impossibility property that $\Delta_\theta S_S(0) \leq 0$. Thus, the claim follows from Equation (46). Now suppose that $\Delta_\theta v_S^c \geq 0$. This implies that for any $v_S \leq v_S^c(\theta_2)$ it is the case that $\Delta_\theta G_S(v_S^c(\theta_2)) = 0 - G_S(v_S; \theta_2) \leq 0$ and this expression is decreasing in v. As by Lemma 14 $\Delta_\theta S_S(v_S^c(\theta_2)) \leq 0$ it follows from Equation (46) that $\Delta_\theta S_S(v_S) \leq 0$ for all $v_S < \max v_S^c$, as required. Note that this implies also that $\Delta_\theta S_S(0, c) = \Delta S_S(0) \leq 0$. □

Lemma 16. For all $v_S \in [\max v_S^c, \min \bar{v}_S]$, if $\Delta_\theta S_S(v_S) \leq (<) 0$ then $\Delta_\theta S_M(\psi(v_S; \theta_2)) \geq (>) 0$.

Proof. From Equation (31) and Lemma 2 follows that

$$\begin{aligned}\Delta_\theta S_M(\psi(v; \theta_2)) = &-\Delta_\theta S_S(v) \\ &- \Big[\int_{\psi(v;\theta_1)}^{\psi(v;\theta_2)} \frac{\partial}{\partial v_M} C\big(r, v)\big) \mathrm{d}r - \int_{\psi(v;\theta_1)}^{\psi(v;\theta_2)} \frac{\partial}{\partial v_M} C\big(r, \phi(r; \theta_1)\big) \mathrm{d}r \Big].\end{aligned}$$

If $\psi(v_S;\theta_2) \geq \psi(v_S;\theta_1)$ then for any $r \in [\psi(v_S;\theta_1), \psi(v_S;\theta_2)]$, $\phi(r;\theta_1) \geq v_S$ and my claim follows. If $\psi(v_S;\theta_2) < \psi(v_S;\theta_1)$ then for any $r \in [\psi(v_S;\theta_2), \psi(v_S;\theta_1)]$, $\phi(r;\theta_1) < v_S$ and my claim follows as well. □

All results for services follow easily from Lemmas 8, 9, 10, 11, 12, 14 and 15 as well as continuity of $\Delta_\theta S_S(\cdot)$. As Lemma 12 has an exact manufacturing analogue, the manufacturing results for $v_M \geq \max v_M^c$ follow from services results and Lemma 16. The results for $v_M < \max v_M^c$ follow from reasoning analogous to that in proof of Lemma 15 once we note that $\Delta_\theta S_S(0) \leq 0$ implies $\Delta_\theta S_M(0) \geq 0$. □

Proof of Proposition 4

The impossibility property is met, as $\Delta_\theta S_S(0) = 0$; the result follows from Theorem 3.

Proof of Proposition 7

Suppose the impossibility property does not hold, then $\Delta_\theta S_S(0) > 0$ and $\Delta_\theta v_S^c < 0$, which implies that $\Delta_\theta S_M(0) < 0$, $\Delta_\theta v_M^c \geq 0$ and trivially $\Delta_\theta h_S^c < 0$ and $\Delta_\theta h_M^c > 0$. $\Delta_\theta S_S(0;\theta_1) > 0$ implies $S_S(0;\theta_2) < R_S$; $\Delta_\theta S_M(0) < 0$ implies $R_M > S_M(0;\theta_2)$, and thus from (37)—(38) in the proof of Theorem 1 follows that:

$$\pi_M(v_M^c(\theta_2), h_M^c(\theta_2); \theta_2) \leq \pi_S(v_M^c(\theta_2), h_S^c(\theta_2)) \tag{47}$$

$$\pi_S(v_S^c(\theta_1), h_S^c(\theta_1)) \leq \pi_M(v_M^c(\theta_1), h_M^c(\theta_1); \theta_1). \tag{48}$$

Given that $\frac{\partial}{\partial v_S}\pi_S > 0$ and $\frac{\partial}{\partial h}\pi_S \geq 0$, we have that RHS of (47) is strictly less than the LHS of (48) and therefore $\pi_M(v_M^c(\theta_2), h_S^c(\theta_2); \theta_2) < \pi_M(v_M^c(\theta_1), h_S^c(\theta_1); \theta_1)$. However, as $\Delta_\theta \pi_M(v_M^c(\theta_1), h_M^c(\theta_1)) \geq 0$, $\frac{\partial}{\partial v_M}\pi_M > 0$ and $\frac{\partial}{\partial h}\pi_M \geq 0$ this is impossible and impossibility property holds; the result follows from Theorem 3.

Lemma 17. Scarce jobs and a strict increase in vertical differentiation imply that $\Delta \bar{v}_M \leq 0$, $\Delta \bar{v}_S \geq 0$, with at least one of these holding strictly.

Proof. The first part follows trivially from Theorem 3 and Lemma 12. Suppose $\Delta \bar{v}_S = 0$; consider the set $\Omega^T = \{v_S \in [\max v_S^c, \min \bar{v}_S] : \Delta\psi(v_S) < 0\}$ and its minimum ${v_S}^5$. Suppose ${v_S}^5 \neq \min \bar{v}_S$. By Theorem 3 we have then that $\Delta S_S(v_S^5) < 0$, which implies $\Delta S_S(\min \bar{v}_S) < 0$, and thus $\Delta \bar{v}_S > 0$—contradiction. Therefore, if $\Delta \bar{v}_S = 0$, then $\Delta\psi(\min \bar{v}_S) < 0$, which implies $\Delta \bar{v}_M < 0$ and concludes the proof. □

Proof of Proposition 5

(i) From Theorem 3 in Appendix C and Lemma 3 follows that for (strictly) scarce jobs a strict increase in vertical differentiation results in a (strict) decrease in v_S^c. As $w_S(v_S^c)$

remains unchanged and surplus function is supermodular, the (strict) increase in lowest wage follows from inspection of Equation (**??**). Note that for any $v_S'' > v_S' \geq v_S^c$ we have

$$w_S(v_S'') = \int_{v_S'}^{v_S''} \frac{\partial}{\partial v_S} \pi_S(r, P_S(r)) \mathrm{d}r + w_S(v_S'). \tag{49}$$

As $P_S(r)$ increases and surplus is supermodular, it follows that $w_S(v_S'')$ increases by more than $w_S(v_S')$.

(ii) Proposition 4 and of Lemma 17 (in Appendix C) imply that $\bar{v}_M(\theta_2) \leq \bar{v}_M(\theta_1)$ and $\bar{v}_S(\theta_2) \geq \bar{v}_S(\theta_1)$ and at least one of these inequalities is strict. This gives

$$\begin{aligned} w_M(\bar{v}_M(\theta_1); \theta_2) &\geq w_M(\bar{v}_M(\theta_2); \theta_2) = w_S(\bar{v}_S(\theta_2); \theta_2) \geq w_S(\bar{v}_S(\theta_1); \theta_2) \\ w_S(\bar{v}_S(\theta_2); \theta_1) &\geq w_S(\bar{v}_S(\theta_1); \theta_1) = w_M(\bar{v}_M(\theta_1); \theta_1) \geq w_M(\bar{v}_M(\theta_2); \theta_1) \end{aligned}$$

with at least one inequality holding strictly, which trivially implies

$$w_M(\bar{v}_M(\theta_1); \theta_2) - w_M(\bar{v}_M(\theta_1); \theta_1) > w_S(\bar{v}_S(\theta_2); \theta_2) - w_S(\bar{v}_S(\theta_2); \theta_1). \tag{50}$$

Thus, $w_M(\bar{v}_M(\theta_1))$ increases strictly. For any $v_M > \bar{v}_M$ we have that

$$w_M(v_M) = \int_{\bar{v}_M}^{v_M} \frac{\partial}{\partial v_M} \pi_M(r, G_M(r)) \mathrm{d}r + w_M(\bar{v}_M(\theta_1)). \tag{51}$$

For $v_M > \bar{v}_M(\theta_1)$, $G_M(v_M)$ does not change; and as surplus' spread implies that $\frac{\partial}{\partial v_M} \pi_M(v_M, h)$ strictly increases, it follows that $w_M(v_M; \theta_2) > w_M(v_M; \theta_1)$ for any $v_M \in [\bar{v}_M(\theta_1), 1]$.

(iii) Follows from (i), (ii) and the fact that with scarce jobs $w_i(v_i^c(\theta_2); \theta_2) = w_i(v_i^c(\theta_1); \theta_1)$.

I will turn now to wages of the least skilled agents with strictly scarce jobs. By Theorem 3 we have that $v_M^c(\theta_2) > v_M^c(\theta_1)$. As wages strictly increase in talent, it follows from definition of critical skill that $w_M(v_M^c(\theta_2); \theta_1) > w_M(v_M^c(\theta_1); \theta_1) = 0$. Note that existence of a positive mass of agents for whom wages decrease (increase) follows from continuity of wage functions.

Proof of Proposition 6

The profit earned by firm h_i is equal to the difference between the surplus produced by its match and the wage it pays its worker. After some rearrangements this yields

$$r_i(h_i) = \int_{h_i^c}^{h_i} \frac{\partial}{\partial h_i} \pi_M(P_i^{-1}(s), s) \mathrm{d}s + r_i(h_i^c). \tag{52}$$

The first claim follows from inspection of Equation (52): Each firm is matched with a less productive agent, so $(P_S)^{-1}$ decreases for all h; and as $r_i(h_i^c) = \pi_S(v_S^c, 0)$ in the scarce jobs case, it falls as well (v_S^c falls by Proposition 4). The second follows trivially from the fact that the pool of services firms is unchanged, the supply of talent falls and surplus is increasing in talent.

As for the last two claims, note that, fixing sectoral supply functions, an increase in surplus' levels increases the total surplus in the economy. As the stable assignment is surplus maximising in this model, the change from the old to new stable assignment has to further improve total surplus.[146] Finally, as services's total surplus falls, it has to increase in manufacturing.

Proof of Lemma 4

If $\Pi(\mathbf{x}, z, i)$ is strictly increasing in productivity, then $h_i(\cdot)$ must be strictly increasing in z.[147] Denoting the distribution of Z conditional on $i = M$ as H_{Z_M} it follows from the fact that h_M has standard uniform distribution that $H_{Z_M}(z) = h_M(z)$. Thus, a FOSD improvement in the distribution of $Z|M$ implies that $h_M(z;\theta_2) \leq h_M(z;\theta_1)$. Consider such z and z' that $h_M(z';\theta_2) = h_M(z;\theta_1)$; it follows that $z' \geq z$. Thus:

$$\frac{\partial}{\partial v_M}\pi_M(v_M, h_M(z';\theta_2);\theta_2) = \frac{\partial}{\partial v_M}\pi_M(v_M, h_M(z';\theta_1);\theta_1) \geq \frac{\partial}{\partial v_M}\pi_M(v_M, h_M(z;\theta_1);\theta_1),$$

as required, where the final inequality follows from the supermodularity of reduced surplus function.

Lemma 18. If jobs are abundant, surplus levels increase universally in manufacturing and workers vertical differentiation increases in manufacturing, then the lowest profit rises in manufacturing and decreases in services ($r_M(h^c_M(\theta_2);\theta_2) \geq r_M(h^c_M(\theta_1);\theta_1)$ and $r_S(h^c_S(\theta_2);\theta_2) \leq r_S(h^c_S(\theta_1);\theta_1)$), where $h^c_i = 1 - \frac{S_i(0)}{R_i}$.

Proof. In manufacturing, there are two possibilities: $S_M(0;\theta_1) < R_M$ and $S_M(0;\theta_1) = R_M$. If the former is the case, then $r_M(h^c_M(\theta_1);\theta_1) = 0$ and the result follows trivially. If the latter is true, then $S_M(0;\theta_2) = R_M$ and by Proposition 7 we have $v^c_M(\theta_2) \geq v^c_M(\theta_1)$, $v^c_S(\theta_2) \leq v^c_S(\theta_1)$, $h^c_M(\theta_1) = h^c_S(\theta_2)$ and $h^c_S(\theta_2) = h^c_S(\theta_1)$.[148] This implies that $r_M(h^c_M(\theta_2);\theta_2) \geq r_M(h^c_M(\theta_1);\theta_1)$, as by market clearing and Equations (37)-(38)

$$r_M(h^c_M(c_i);\theta_j) = \pi_M(v^c_M(\theta_j), h^c_M(\theta_j)) - \pi_S(v^c_S(\theta_j), h^c_S(\theta_j)),$$

for $i = 1, 2$. The result for services follows from analogous reasoning, but the two cases are $S_S(0;\theta_2) < R_S$ and $S_S(0;\theta_2) = R_S$. □

[146] This is the case, as my model can be rewritten as a special case of the assignment model described in Gretsky et al. (1992) and thus the equivalence of stable and efficient matching showed by them holds for my model as well.

[147] For any $z' > z$ we have $\pi_i(v_i, h_i(z')) > \pi_i(v_i, h_i(z))$. Suppose that $h_i(z') \leq h_i(z)$. Then by Assumption A1.3 $\pi_i(v_i, h_i(z')) \leq \pi_i(v_i, h_i(z))$; contradiction.

[148] The change in critical levels follows from the fact that with fixed sector sizes an improvement in skill supply is equivalent to a FOSD improvement in the distribution of skill.

Proof of Proposition 8

Lemma 19. Suppose that (a) the surplus function in sector i weakly increases for all matches, (b) that the share of surplus produced by the worse match and received by the firm ($r_i(h_i^c)$) decreases (strictly) and (c) that $P_i(v_i;\theta_2) \geq P_i(v_i;\theta_2)$ for all $v_S \in [\max v_i^c, 1]$. Then it must be the case that $w_i(v_S;\theta_2) \geq (>) w_i(v_S;\theta_1)$ for all $v_S \in [\max v_i^c, 1]$.

Proof. Consider the agent with skill $\max_\theta v_i^c = \max\{v_i^c(\theta_1), v_i^c(\theta_2)\}$; that is a worker with the lowest skill level that will be observed in both matching problems in sector i. Her wage depends on two factors: positively on the surplus she produces and negatively on its share received by the firm she is matched with. The first factor always increases, as she is matched with a more productive firms. The change in the second factor can be both positive (for $\max_\theta v_S^c = v_S^c(\theta_1)$) and negative (for $\max_\theta v_S^c = v_S^c(\theta_2)$). If the former is the case, however, then the increase in surplus received by her firm

$$\Delta_\theta r_S(P_S(v_S^c(\theta_1)) = \int_{v_S^c(\theta_2)}^{v_S^c(\theta_1)} \frac{\partial}{\partial v_S} P_S(r;\theta_2) \frac{\partial}{\partial h} \pi_S(r, P_S(r, c_2)) \mathrm{d}r + \Delta_\theta r_i(h_i^c),$$

is always (strictly) less than the increase in the surplus she produces (as $r_i(h_i^c) \leq (<) 0$):

$$\pi_S(v_S^c(\theta_1), P_S(v_S^c(\theta_1)) - \pi_S(v_S^c(\theta_1), h_S^c) = \int_{v_S^c(\theta_2)}^{v_S^c(\theta_1)} \frac{\partial}{\partial v_S} P_S(r;\theta_2) \frac{\partial}{\partial h} \pi_S(v_S^c(\theta_1), P_S(r, c_2)) \mathrm{d}r.$$

Thus, $w_S(\max_\theta v_S^c;\theta_2) - w_S(\max_\theta v_S^c;\theta_1) \geq 0$ and by inspection of Equation (49) we have that $w_S(v_S'';\theta_2) - w_S(v_S'';\theta_1) \geq w_S(v_S^c(\theta_2);\theta_1) - w_S(v_S^c(\theta_1);\theta_1)$, for any $v_S'' \in [v_S^c(\theta_1), v_S']$. It follows that wages increase (strictly) for all $v_S \in [v_S^c(\theta_1), v_S']$. $\square$

Consider $T = A_A^1(\theta_2) \cap A_A^1(\theta_1)$, the set of agents who work in services in both matching problems. Clearly, the skill of the least skilled of those workers –$\inf_y T$—is $\max_\theta v_S^c$. Because $P_i(v_i;\theta_2) \geq P_i(v_i;\theta_2)$ for all $v_S \in [\max v_i^c, 1]$ by Proposition 7 it follows from Lemmas 18 and 19 that wages increase for all $v_S \in T$. This and revealed preference imply that all agents who used to work in services are better off[149]. As the top wages increase by more in manufacturing (by the same reasoning as in the proof of Proposition 5) it follows that wages increase for most skilled manufacturing workers.

Proof of Proposition 9

The result wrt services profits follows from Proposition 7, the definition of a decrease in supply, the definition of PAM ($P_i(\cdot)$), Lemma 18 in Appendix C and inspection of

[149] This is trivial if $v_S^c(\theta_2) \leq v_S^c(\theta_1)$. If $v_S^c(\theta_2) \geq v_S^c(\theta_1)$ then the agents with $v_S \in [v_S^c(\theta_1), v_S^c(\theta_2))$ will move to manufacturing; but as the lowest wages are the same in both sectors, they earn more than $w_S(v_S^c(\theta_2);\theta_2)$, which in turn is greater than their old wage.

Equation (52). The increases in total output follow from analogous reasoning as in the proof of Proposition 6.

D Endogenous Entry

Proof of Proposition 10

The proof will consist of two steps. First, I will prove that

$$E^* \in \mathbf{E_E} \Leftarrow V(E^*) - V(E') \geq 0 \text{ for all } E \in \mathbf{E_B}, \tag{53}$$

which trivially implies the "only if" part of the lemma. In the second step I will show the "if" part.

'Only if' Denote as $S_i(\cdot, R_M, R_S)$ the equilibrium supply of skill in sector i in the baseline model if the measures of firms are R_M, R_S.[150] Define

$$\begin{aligned} T_i(R_i) &= T_i(S_i(R_M, R_S), R_i), \\ T(R_M, R_S) &= T_M(S_M(R_M, R_S), R_M) + T_S(S_S(R_M, R_S), R_S), \\ V(R_M, R_S) &= V(S_M(R_M, R_S), S_S(R_M, R_S), R_M, R_S), \end{aligned}$$

so the gross and net surpluses holding in an equilibrium of the baseline model if the measures of firms are R_M, R_S. In a direct analogy, we can also denote the average profits in sector i holding in equilibrium for R_M, R_S as $\bar{r}_i(R_M, R_S)$. Note that these are defined uniquely if $R_i > 0$ and $R_M + R_S \neq 0$, otherwise they can take a range of values.

Lemma 20. Consider $(R_M, R_S), (R'_M, R'_S) \in \mathbf{R}^2_{\geq 0}$. For any $t \in [0, 1]$ define $R_i(t) = R_i + t(R'_i - R_i)$ and $V(t) = V(R_M(t), R_S(t))$. The following is true: (a) $V(\cdot)$ is absolutely continuous; (b) for any $t \in (0, 1)$ for which V is differentiable, we have

$$V_t(t) = (R'_M - R_M)(\bar{r}_M(R_M(t), R_S(t)) - c_M) + (R'_S - R_S)(\bar{r}_S(R_M(t), R_S(t)) - c_S),$$

giving

$$V(t) = V(0) + \int_0^t (R'_M - R_M)\bar{r}_M(R_M(s), R_S(s)) + (R'_S - R_S)\bar{r}_S(R_M(s), R_S(s))\mathrm{d}s. \tag{54}$$

Proof. As the baseline model is an assignment game, it follows from the results in Gretsky et al. (1992) that the equilibrium of the baseline model is efficient and, thus

$$V(R_S, R_M) = \max_{(S_M, S_S) \in \mathbf{S}_B} V(S_M, S_S, R_M, R_S), \tag{55}$$

[150] Formally, $S_M(\cdot, R_M, R_S) = S_i(\cdot)$ if and only if there exists some function $S_S(\cdot)$ such that $(S_M, S_S, R_M, R_S) \in \mathbf{E_B}$.

where $\mathbf{S}_B = \{(S_M, S_S) : \exists_{(R_M,R_S)\in\mathbf{R}^2_{\geq 0}} (S_M, S_S, R_M, R_S) \in \mathbf{E_B}\}$. For any $(R_M, R_S), (R'_M, R'_S) \in \mathbf{R}^2_{\geq 0}$ we can define $V(S_M, S_S, t) = V(S_M, S_S, R_M(t), R_S(t))$. Note that $V_t(S_M, S_S, t)$ exists as long as $R_i(t) \neq S_i(0)$, and $R_i(t) \neq 0$ so for all $t \in [0,1]$ but at most four. Further, whenever $V_t(S_M, S_S, t)$ does exist we have that

$$V_t(S_M, S_S, t) = (R'_M - R_M)(\bar{r}_S(S_M, R_M(t)) - c_M) + (R'_S - R_S)(\bar{r}(S_S, R_S(t)) - c_S),$$

where

$$\bar{r}_i(S_i, R_i) = \begin{cases} \int_0^1 \int_0^h \frac{\partial}{\partial h}\pi_i(S_i^{-1}((1-p)R_i), p)\mathrm{d}p + \pi_i(S_i^{-1}(R_i), 0)\,\mathrm{d}h & \text{for } R_i \in (0, S_i(0)), \\ \int_{1-\frac{S_i(0)}{R_i}}^1 \int_{1-\frac{S_i(0)}{R_i}}^h \frac{\partial}{\partial h}\pi_i(S_i^{-1}((1-p)R_i), p)\mathrm{d}p\,\mathrm{d}h & \text{for } R_i > S_i(0). \end{cases} \tag{56}$$

Thus,

$$V(S_M, S_S, t) = V(S_M, S_S, t_1) + \int_{t_1}^t (R'_M - R_M)\bar{r}_M(S_M, R_S(s)) + (R'_S - R_S)\bar{r}_S(S_S, R_S(s))\mathrm{d}s,$$

proving that $V(S_M, S_S, \cdot)$ is absolutely continuous for any $(S_M, S_S) \in \mathbf{S}_B$ and any choice of $(R_M, R_S), (R'_M, R'_S)$. Clearly, $\bar{r}(S_i, R_i(t)) - c_i \in [-c_i, \pi_i(1,1) - c_i]$, implying

$$|V_t(S_M, S_S, t)| \leq (R'_M - R_M)\max\{c_M, \pi_M(1,1)\} + (R'_S - R_S)\max\{c_S, \pi_S(1,1)\}$$

which proves $V(\cdot)$ is absolutely continuous by Theorem 2 in Milgrom and Segal (2002).

Let us turn to point (b). Define $T(t) = T(R_M(t), R_S(t))$ and pick any $t \in [0,1]$ for which $T(t)$ is differentiable. Consider two $c'_M, c'_S \in \mathbf{R}_{\geq 0}$ such that $c'_i = \bar{r}_i(R_M(t), R_S(t))$. For entry costs c'_M, c'_S, the quadruple $(S_M(R_M(t), R_S(t)), S_S(R_M(t), R_S(t)), R_M(t), R_S(t))$ is an equilibrium of the extended model, implying that it maximizes the function $V'(t) = T(t) - c'_M R_M(t) - c'_M R_M(t)$. Clearly, both $V(\cdot)$ and $V'(\cdot)$ are differentiable at t as well. It follows from first order conditions that $V'_t(t) = 0$ implying that

$$\begin{aligned} T_t(t) &= (R'_M - R_M)c'_M + (R'_S - R_S)c'_S \\ &= (R'_M - R_M)\bar{r}_M(R_M(t), R_S(t)) + (R'_S - R_S)\bar{r}_S(R_M(t), R_S(t)). \end{aligned}$$

This proves that

$$V_t(t) = (R'_M - R_M)(\bar{r}_M(R_M(t), R_S(t)) - c_M) + (R'_S - R_S)(\bar{r}_S(R_M(t), R_S(t)) - c_S),$$

which, together with the absolute continuity of $V(t)$ proves Equation (54) as well. □

Consider $R^M_M, R^M_S \geq 0$ for which $V(E)$ is maximized. I will show that R^M_M, R^M_S must satisfy condition (iv) of the equilibrium definition and, together with the corresponding supply functions, constitute an equilibrium.

First, I will show that if $R_M^M > 0$ then $\bar{r}_M - c_M \geq 0$. First, pick some $R'_M < R_M^M$ and define $V(t)$ for (R_M^M, R_S^M) and (R'_M, R_S^M). From Lemma 20 and the definition of maximum follows that there exists some $t' \in (0,1)$ such that for any $t < t'$ we have $\bar{r}_M(R_M(t), R_S^M(t)) > 0$. If $R_M^M + R_S^M \neq 1$ then this immediately implies $\bar{r}_M(R_M^M, R_S^M) \geq 0$ by continuity. If $R_M^M + R_S^M = 1$, there exist wage functions for which $\bar{r}_M(R_M^M, R_S^M) \geq 0$—and condition (iv) is satisfied as well. It remains to show that if $R_M^M \in\geq 0$ then $\bar{r}_M - c_M \leq 0$. The proof is analogous: pick some $R''_M > R_M$ and the result follows from an analogous reasoning as for $R_M^M > 0$. The proof for services is analogous.

'If' Assume that $\mathbf{E_E}$ is non-empty and consider some E^*, E' such that $E^* \in \mathbf{E_E}, E' \in \mathbf{E_B}$ and $E^* \neq E'$. Denote by $\mathbf{W}$ the set of pairs of sectoral functions $w = (w_M, w_S)$ that are of the form prescribed by Proposition 1 given E^* and consider an arbitrary $w^* \in \mathbf{W}$.

Denote the total wage bill in sector i under wage function w_i and supply function S_i as

$$\bar{w}_i(w_i, S_i) = -\int_0^1 w_i(t) s_i(t) \mathrm{d}t.$$

where $\frac{\partial}{\partial v} S_i(v) = s_i(v)$. Note that for $R_i > 0$ we have $\bar{w}_i(w_i, S_i) = R_i \int_{1-\frac{S_i(0)}{R_i}}^1 w_i(S_i^{-1}((1-h)R_i))\mathrm{d}h$. We can now denote the average wage in the economy i under wage schedule $w = (w_M, w_S)$ and supply functions $S = (S_M, S_S)$ as

$$\bar{w}(w, S) = \bar{w}_M(w_M, S_M) + \bar{w}_S(w_S, S_S).$$

Note that by the definition of a sectoral supply function $\bar{w}(w^*, S^*) \geq \bar{w}(w^*, S')$. Further, if $S^* \neq S'$, then this inequality holds strictly, because the measure of workers who are indifferent between joining manufacturing or services is equal to 0.

Profit maximization implies that, if $R'_i > 0$, then

$$\bar{r}_i^* - c_i = \int_0^1 \max\{\pi_i(v_i^*(h), h) - w_i^*(v_i^*(h)), 0\}\mathrm{d}h - c_i \geq \frac{T_i(S'_i, R'_i) - \bar{w}_i(w_i^*, S'_i)}{R'_i} - c_i, \quad (57)$$

where v_i^* is the hiring function defined in Section 2.1.2.

I will prove the result by first assuming that $R_i^*, R_S^* > 0$ and only later considering the alternative. Note that if $R_M^*, R_S^* > 0$, then $v_i^*(h) = (S_i^*)^{-1}((1-h)R_i^*)$ for $h \in [0, 1 - \frac{S_i^*(0)}{R_i^*}]$, whereas for $h \in [0, 1 - \frac{S_i(0)}{R_i}]$ we have $\pi_i(v, h) - w_i^*(v) \leq 0$ for all $v \in [0,1]$. This gives

$$\bar{r}_i^* - c_i = \frac{T_i(S_i^*, R_i^*) - \bar{w}_i(w_i^*, S_i^*)}{R_i^*} - c_i. \quad (58)$$

Note also that $R'_M(\bar{r}_M^* - c_M) + R'_S(\bar{r}_S^* - c_S) \geq V(E') - \bar{w}(w^*, S')$. If $R'_M, R'_S > 0$ this follows directly from Equation (57). If $R'_i = 0$, then it follows as $T_i(S'_i, R'_i) - R_i c_i - \bar{w}_i(w_i^*, S'_i) \leq$

$0 = R_i'(\bar{r}_i^* - c_i)$. Thus we can write

$$
\begin{aligned}
V(E^*) - \bar{w}(w^*, S^*) &= R_M^*(\bar{r}_M^* - c_M) + R_S^*(\bar{r}_S^* - c_S) = R_M'(\bar{r}_M^* - c_M) + R_S'(\bar{r}_S^* - c_S) \\
&\geq V(E') - \bar{w}(w^*, S'). \qquad (59)
\end{aligned}
$$

Now suppose that $R_i^* = 0$. By definition of equilibrium follows that $r_i - c_i \leq 0$. If $R_i' > 0$ we have that

$$0 = T_i(S_i^*, R_i^*) - \bar{w}_i(w_i^*, S_i^*) - R_i^* c_i \geq R_i'(\bar{r}_i^* - c_i) \geq T_i(S_i', R_i') - \bar{w}_i(w_i^*, S_i') - R_i' c_i. \quad (60)$$

Also, trivially, if $R_i' = 0$, then

$$0 = T_i(S_i^*, R_i^*) - \bar{w}_i(w_i^*, S_i^*) - R_i^* c_i = T_i(S_i', R_i') - \bar{w}_i(w_i^*, S_i') - R_i' c_i.$$

Thus, it follows that $V(E^*) - \bar{w}(w^*, S^*) \geq V(E') - \bar{w}(w^*, S')$.

$V(E^*) - \bar{w}(w^*, S^*) \geq V(E') - \bar{w}(w^*, S')$ and the fact that $\bar{w}(w^*, S^*) > \bar{w}(w^*, S')$ imply that

$$V(E^*) - V(E') \geq \bar{w}(w^*, S^*) - \bar{w}(w^*, S') \geq 0.$$

Suppose that $V(E^*) = V(E')$. From (53) follows that this is possible only if $E' \in \mathbf{E_E}$. Further, if $S^* \neq S'$, then $\bar{w}(w^*, S^*) - \bar{w}(w^*, S') > 0$, and thus $V(E^*) = V(E')$ is possible only if $S^* = S'$ and $R_i^* \neq R_i'$ for some $i \in \{i, j\}$. Finally, it follows from Equation (56) and Assumption 1 that if $R_i^* \neq R_i'$ then $\bar{r}(S_i^*, R_i^*) \neq \bar{r}(S_i^*, R_i') = c_i$, implying that $E^* \notin \mathbf{E_E}$; contradiction! Therefore, $V(E^*) > V(E')$.

Proof of Theorem 2

Existence. Denote as $S_i(\cdot, R_M, R_S)$ the equilibrium supply of skill in sector i in the baseline model, holding for R_M, R_S. It follows from the proof of Theorem 1 that S_S is continuous in R_M, R_S for any $R_M, R_S > 0$. Thus because $\int_{1-\frac{S_i(0)}{R_i}}^1 \pi_i(S_i^{-1}((1-h)R_i), h)\mathrm{d}h \leq \pi_i(1,1)$ for any $R_i > 0$, it follows that

$$V(R_M, R_S) = V(S_M(R_M, R_S), S_S(R_M, R_S), R_M, R_S)$$

is continuous in R_M, R_S.[151]

Lemma 21. If $R_M > \bar{R}_M = \frac{\pi_M(1,1)+\pi_S(1,1)}{c_M}$ then $V(R_M, R_S) < 0 = V(0,0)$.

Proof. It follows from Equation (23) that, trivially, $T_i(R_M, R_S) \leq \pi_i(1,1)$, where $T_i(R_M, R_S)$ is defined as in the proof of Proposition 10. Thus it follows from the definition of net

[151] This is because $\lim_{R_i \to 0} V_i(S_i, R_i) = 0 \cdot \lim_{R_i \to 0} \int_{1-\frac{S_i(0)}{R_i}}^1 \pi_i(S_i^{-1}((1-h)R_i), h)\mathrm{d}h \leq 0 \cdot \pi_M(0,0) = 0$.

total surplus that

$$V(R_M, R_S) \leq \pi_M(1,1) - R_M c_M + \pi_S(1,1).$$

Note that $\pi_M(1,1) - R_M c_M + \pi_S(1,1) < 0$ for any $R_M > \frac{\pi_M(1,1)+\pi_S(1,1)}{c_M}$, implying that $V(R_M, R_S) < 0 = V(0,0)$, as required. $\square$

Of course, an analogous result holds for services. Define the set $\bar{R} = \{(R_M, R_S) \in \mathbf{R}^{\mathbf{2}}_{\geq \mathbf{0}} : R_M \leq \bar{R}_M, R_S \leq \bar{R}_S$. Because the net total surplus for $(R_M, R_S) = (0,0)$ is zero, it follows from Lemma 21 that

$$\max_{(R_M,R_S)\in \mathbf{R}^{\mathbf{2}}_{\geq \mathbf{0}}} V(R_M, R_S) = \max_{(R_M,R_S)\in \bar{R}} V(R_M, R_S).$$

As $\bar{R}$ is closed and bounded, it follows from Weierstrass' Theorem that $V(R_M, R_S)$ admits a global maximum on $\mathbf{R}^2_{\geq 0}$. As by Proposition 10 any global maximum must be en equilibrium, existence follows.

Uniqueness. Follows immediately from Proposition 10.[152]

Wages First, if $R_M + R_S \neq 1$ in equilibrium, then this follows from the Proposition 1 and Lemma 1. Otherwise the constant of integration C_i is not uniquely determined in the baseline model; here, however, if $C_i' > C_i$, then $\bar{r}_i' > \bar{r}_i$ contradicting the requirement that both have to be equal to c_i.

Proof of Lemma 5

Of course, if $w_S(v_S')$ increases then the existence of v_S^1 follows trivially; analogously if $w_S(v_S')$ decreases. Therefore it suffices to show that if $w_S(v_S')$ increases (decreases) then v_S^2 (v_S^1) exists. I will prove the case of an increase in $w_S(v_S')$; the proof for the other case is analogous.

The proof will be by contradiction. Suppose that there exists a $v_S' \in [\max v_i^c, 1]$ such that $w_S(v_S')$ increases strictly, yet for all $v_S \geq v_S^c(\theta_2)$ we have that $w_S(v_S)$ increases weakly. Note that—because the surplus function is unchanged—the fact that $w_S(v_S;\theta_2) \geq (>) w_S(v_S;\theta_1)$, implies that $r_S(P_S(v_S;\theta_2);\theta_2) \leq (<) r_S(P_S(v_S;\theta_2);\theta_1)$. This follows directly from profit maximization, as

$$\begin{aligned} r_S(P_S(v_S;\theta_2);\theta_1) \quad &\geq \quad \pi_S(v_S, P_S(v_S;\theta_2)) - w_S(v_S;\theta_1) \\ &\geq (>) \pi_S(v_S, P_S(v_S;\theta_2)) - w_S(v_S;\theta_2) \\ &= \quad r_S(P_S(v_S;\theta_2);\theta_2). \end{aligned} \tag{61}$$

Therefore $r_S(h)$ falls weakly for all $h \in [h_S^c(\theta_2), 1]$, where $h_S^c(\theta_2) = P_S(v_S(\theta_2);\theta_2)$. This

[152] Consider a pair $E^*, E' \in \mathbf{E_E}$ and $E^* \neq E'$. Then (53) implies that $V(E^*) > V(E)$ and $V(E^*) < V(E)$ which is a contradiction.

further implies that $h_S^c(\theta_2) \geq h_S^c(\theta_1)$.[153] Continuity implies that there exists some ϵ such that $r_S(P_S(v_S;\theta_2);\theta_2) < r_S(P_S(v_S;\theta_2);\theta_1)$ for all $v_S \in (v_S' - \epsilon, v_S' + \epsilon)$. Altogether, this implies that $\bar{r}_i(\theta_2) < \bar{r}_i(\theta_1) = c_i$, which contradicts the zero-expected-profits condition.

Proof of Proposition 11

All derivatives wrt Δ_p will be evaluated at $\Delta_p = 0$. I will, thus, suppress $|_{\Delta_p=0}$ from notation, so that $\frac{\partial}{\partial \Delta_p}|_{\Delta_p=0}$ will be denoted just as $\frac{\partial}{\partial \Delta_p}$.

The addition of a measure Δ_p of workers with skill p, results in the following equilibrium matching function

$$P_S(v_S, \Delta_p) = \begin{cases} 1 - \frac{1 - v_S + \Delta_p}{R_S(\Delta_p)} & \text{for } v_S \in [v_i^c, p) \\ 1 - \frac{1 - v_S}{R_S(\Delta_p)} & \text{for } v_S \in (p, 1], \end{cases} \tag{62}$$

where $R_S(\Delta_p)$ is the equilibrium entry into services as a function of Δ_p. Note that the zero-expected-profit condition implies that (for strictly supermodular surplus functions) $\frac{\partial}{\partial \Delta_p} R_S > 0$. The change in wages in response to an infinitesimal Δ_p is, therefore:

$$\begin{aligned} \frac{\partial}{\partial \Delta_p} w_S(r) = & \frac{\mathrm{d}}{\mathrm{d}\Delta_p} w_S(v_S^c) - \frac{\partial}{\partial \Delta_p} v_S^c \frac{\partial}{\partial v_S} \pi_M(v_S^c, h_S^c) \\ & + \int_{v_S^c}^{r} \frac{\partial}{\partial \Delta_p} P_S(v) \frac{\partial^2}{\partial v_S h_S} \pi_M(v, P_S(v)) \mathrm{d}v. \end{aligned} \tag{63}$$

Trivially, for $p \neq r$, the first two derivatives of the above expression wrt v_S are

$$\frac{\partial}{\partial v_S} \frac{\partial}{\partial \Delta_p} w_S(r) = \frac{\partial}{\partial \Delta_p} P_S(r) \frac{\partial^2}{\partial v_S h_S} \pi_M(r, P_S(r)), \tag{64}$$

$$\begin{aligned} \frac{\partial^2}{\partial v_S^2} \frac{\partial}{\partial \Delta_p} w_S(r) = & \frac{\partial}{\partial v_S} \frac{\partial}{\partial \Delta_p} P_S(r) \frac{\partial^2}{\partial v_S h_S} \pi_M(r, P_S(r)) \\ & + \Big(\frac{\partial}{\partial \Delta_p} P_S(r) \Big)^2 \frac{\partial^3}{\partial v_S h_S^2} \pi_M(r, P_S(r)) \end{aligned} \tag{65}$$

I will now show that $\arg\min_{r \in [v_S^c, 1]} \frac{\partial}{\partial \Delta_p} w_S(r) = p$. It follows from Weierstrass' Theorem that $\frac{\partial}{\partial \Delta_p} w_S(\cdot)$ must attain a global minimum on $[v_S^c, 1]$. This minimum can be found only at (a) the boundaries, (b) stationary points and (c) at $p = r$, where $\frac{\partial}{\partial \Delta_p} w_S(r)$ is not differentiable. It follows immediately from Equations (64) and (65) that at any stationary point $\frac{\partial^2}{\partial v_S^2} \frac{\partial}{\partial \Delta_p} w_S(r) < 0$, so that no stationary point can be global min. The same is the case for $r = 1$, which leaves us with $r = v_S^c$ as the only possibility. This, of course, is possible only if $\frac{\partial}{\partial \Delta_p} P_S(v_S^c) > 0$. However, if $v_S^c \neq p$, then $\frac{\partial}{\partial \Delta_p} P_S(v_S^c) > 0$ implies that $\frac{\partial}{\partial \Delta_p} w_S(v_S^c) > 0$.[154] Thus it follows by Lemma 5 that $\arg\min_{r \in [v_S^c, 1]} \frac{\partial}{\partial \Delta_p} w_S(r) \neq v_S^c$.

[153] Note that if $r_S(h_S^c(\theta_j)) > 0$ then $h_S^c(\theta_j) = 0$. We have that $r_S(h;\theta_1) \geq 0$ for all $h_S \geq h_S^c(\theta_2)$ and, hence, $h_S^c(\theta_1) \leq h_S^c(\theta_2)$.

[154] Assume $v_S^c \neq p$. First, note that $\frac{\partial}{\partial \Delta_p} P_S(v_S^c) = \frac{\partial}{\partial \Delta_p} h_S^c - \frac{1}{R_S} \frac{\partial}{\partial \Delta_p} v_S^c$. Note that in the single sector

Therefore, $\arg\min_{r\in[v_S^c,1]} \frac{\partial}{\partial\Delta_p} w_S(r) = p$ as required. This proves that $\frac{\partial}{\partial v_S}\frac{\partial}{\partial\Delta_p} w_S(r) < 0$ by Lemma 5.

Note that the total surplus (net of entry costs) in services is equal to:

$$V_S = R_S \int_{h_S^c}^{1} \pi_S(P_S^{-1}(t), t)\mathrm{d}t - R_S c_S.$$

Lemma 22. In the single sector model the wage of worker r is equal to the response in total net surplus to the addition of an infinitesimal Δ_r measure of workers with skill r, with $\frac{\partial}{\partial\Delta_p} V_S = w(r)$.

This Lemma is a simple generalization of the result from Section III.B in Costrell and Loury (2004); a formal proof is provided in Online Appendix OA.4. By Schwartzmann's Theorem:

$$\frac{\partial}{\partial\Delta_p} w_S(r) = \frac{\partial}{\partial\Delta_p}\frac{\partial}{\partial\Delta_r} V_S = \frac{\partial}{\partial\Delta_r}\frac{\partial}{\partial\Delta_p} V_S = \frac{\partial}{\partial\Delta_r} w_S(p),$$

as required. If $p = v_S^c$ then $P_S(v_S, \Delta_p) > 0$ for all $v_S > v_S^c$. It follows by an analogous reasoning as above that in such a case $\arg\max_{r\in[v_S^c,1]} \frac{\partial}{\partial\Delta_p} w_S(r) = 1$, and thus $\frac{\partial}{\partial\Delta_p} w_S(1) > 0$ by Lemma 5. It follows trivially that $\frac{\partial}{\partial\Delta_r} w_S(p) = \frac{\partial}{\partial\Delta_p} w_S(r) > 0$ for $p = v_S^c$ and $r = 1$.

Proof of Proposition 12

As $R_M = R_S = \frac{1}{2}$ in the symmetric Costrell-Loury specification, it follows that $W(t) - W(0) = \int_0^{G_i^{-1}(t)} \frac{\partial}{\partial v_i}\pi_i(s, G_i(s))\mathrm{d}s$ and the increase in wage polarization in absolute terms follows from the proof of Proposition 3.

The proof of the second statement is more involved. Define two functions $A(t;\theta_j) = \frac{g(G_i^{-1}(t;\theta_1);\theta_1) - g(G_i^{-1}(t;\theta_2);\theta_2)}{g(G_i^{-1}(t;\theta_j);\theta_j)}$ and $B(t;\theta_j) = \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(t;\theta_1);\theta_1) - \frac{\partial}{\partial v_i} g(G_i^{-1}(t;\theta_2);\theta_2)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(t;\theta_j);\theta_j)}$, with $j \in \{1, 2\}$. Note that $B(t;\theta_j)$ is well defined for $t = 0$, because $\frac{\partial}{\partial v_i} g(0) = C_{v_M v_S}(0,0) + C_{v_M v_M}(0,0) = C_{v_M v_S}(0,0) > 0$, as $\frac{\partial}{\partial v_M} C(v_M, 0) = 0$ for all v_M. Further

$$\frac{\frac{\mathrm{d}}{\mathrm{d}t}\Big[g(G_i^{-1}(t;\theta_2);\theta_2) - g(G_i^{-1}(t;\theta_1);\theta_1)\Big]}{\frac{\mathrm{d}}{\mathrm{d}t} g(G_i^{-1}(t;\theta_j);\theta_j)} = B(0;\theta_j) - \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_k);\theta_k)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_j);\theta_j)} A(t;\theta_k), \quad (66)$$

where $k \neq j$. Therefore, it follows from the generalized l'Hopital rule (Theorem 2 in

model it is always the case that either $v_S^c = 0$ (if $R_S \geq 1$) or $h_S^c = 0$ (if $R_S \leq 1$). This implies, in particular, that it is impossible to have $\frac{\partial}{\partial\Delta_p} h_S^c > 0$ and $\frac{\partial}{\partial\Delta_p} v_S^c > 0$ or $\frac{\partial}{\partial\Delta_p} h_S^c < 0$ and $\frac{\partial}{\partial\Delta_p} v_S^c < 0$. Therefore, $\frac{\partial}{\partial\Delta_p} P_S(v_S^c) > 0$ implies that $\frac{\partial}{\partial\Delta_p} h_S^c \geq 0$ or $\frac{\partial}{\partial\Delta_p} v_S^c \leq 0$, with at least one of them holding strictly. There are three possibilities. Firstly, $R_S < 1$, in which case $w_S(v_S^c) = 0$ and $\frac{\partial}{\partial\Delta_p} v_S^c < 0$, giving $\frac{\partial}{\partial\Delta_p} w_S(v_S^c) > 0$. Secondly, $R_S = 1$, in which case $h_S^c = v_S^c = 0$. This implies that $\frac{\partial}{\partial\Delta_p} v_S^c = 0$ and, therefore, $\frac{\partial}{\partial\Delta_p} h_S^c > 0$. From Equation (62) we have then $\frac{\partial}{\partial\Delta_p} R_S > 1$, implying that $\frac{\mathrm{d}}{\mathrm{d}\Delta_p} w_S(v_S^c) > 0$ and thus also $\frac{\partial}{\partial\Delta_p} w_S(v_S^c) > 0$. And finally, if $R_S > 1$, then $w_S(v_S^c) = \pi_S(0, h_S^c)$ again giving $\frac{\partial}{\partial\Delta_p} w_S(v_S^c) > 0$.

Taylor (1952)) that

$$\limsup_{t\to 0} A(t;\theta_j) \leq B(0;\theta_j) - \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_k);\theta_k)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_j);\theta_j)} \liminf_{t\to 0} A(t;\theta_k), \tag{67}$$

$$\liminf_{t\to 0} A(t;\theta_j) \geq B(0;\theta_j) - \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_k);\theta_k)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_j);\theta_j)} \limsup_{t\to 0} A(t;\theta_k), \tag{68}$$

$$\liminf_{t\to 0} A(t;\theta_j) \leq B(0;\theta_j) - \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_k);\theta_k)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_j);\theta_j)} \liminf_{t\to 0} A(t;\theta_k). \tag{69}$$

Equations (67) and (68) give:

$$\limsup_{t\to 0} A(t;\theta_j) \leq B(0;\theta_j) - \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(t;\theta_k);\theta_k)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(t;\theta_j);\theta_j)} \liminf_{t\to 0} A(t;\theta_k) \leq \limsup_{t\to 0} A(t;\theta_j)$$

whereas Equations (67) and (69) give:

$$\limsup_{t\to 0} A(t;\theta_j) \leq B(0;\theta_j) - \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_k);\theta_k)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_j);\theta_j)} \liminf_{t\to 0} A(t;\theta_k) \geq \liminf_{t\to 0} A(t;\theta_j).$$

The latter shows that $\lim_{t\to 0} A(t;\theta_j)$ exists and the former that

$$\begin{aligned}\lim_{t\to 0} A(t;\theta_1) + \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_2);\theta_2)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_1);\theta_1)} \lim_{t\to 0} A(t;\theta_2) = &-\frac{\Delta_\theta \frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_2)))}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_1);\theta_1)} \\ &+ \frac{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_2);\theta_1) - \frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_1);\theta_1)}{\frac{\partial}{\partial v_i} g(G_i^{-1}(0;\theta_1);\theta_1)} < 0.\end{aligned}$$

It follows immediately from the strict increase in left-tail that $\lim_{t\to 0} A(t;\theta_1) > 0$ or $\lim_{t\to 0} A(t;\theta_2) > 0$. Denote $\frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t;\theta_2)}{W(t;\theta_2)} - \frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t;\theta_1)}{W(t;\theta_1)}$ as $\Delta_\theta \frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t)}{W(t)}$. Crucially, $\lim_{t\to 0} \Delta_\theta \frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t)}{W(t)} = \lim_{t\to 0} \frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t;\theta_2)}{W(t;\theta_2)} - \frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t;\theta_1)}{W(t;\theta_1)}$ can be written as

$$\begin{aligned}\lim_{t\to 0} \Delta_\theta \frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t)}{W(t)} = \lim_{t\to 0} \frac{1}{g(G_j^{-1}(t;\theta_k);\theta_k)} \Big[&\frac{\frac{\partial}{\partial v_i}\pi(G_j^{-1}(0;\theta_j),t)}{W(0;\theta_j)} \lim_{t\to 0} A(t;\theta_j) \\ &- \frac{\Delta_\theta W(0) \frac{\partial}{\partial v_i}\pi(G_j^{-1}(0;\theta_j),t)}{W(0;\theta_1)W(0;\theta_2)} \Big]\end{aligned} \tag{70}$$

for all $i \in \{M,S\}$, $k,j \in \{1,2\}$, and $k \neq j$. If $W(t;\theta_2)W(t;\theta_1)$ is large enough or $\Delta_\theta W(t)$ is sufficiently small, then the term in brackets in Equation (70) must be negative for some $j \in \{1,2\}$.[155] The first is ensured by a high enough reservation value, whereas the latter by weak enough supermodularity of the surplus function. This proves the second statement by the definition of a limit and the fact that $\ln W(t) = \int_0^t \frac{\frac{\mathrm{d}}{\mathrm{d}t}W(t)}{W(t)} + W(0)$.

[155]Note that if $\frac{\partial^2}{\partial h_i \partial v_i}\pi_i(v_i,h_i)$ is arbitrarily close to 0 for all (v_i,h_i) then $\Delta_\theta W(t)$ is arbitrarily close to 0 as well.

Proof of Proposition 13

Endow the space $\mathbf{R}^2_{\geq 0}$ with the following partial order $\succeq$: if $R'_M \geq R_M$ and $R'_S \leq R_S$ then $(R'_M, R'_S) \geq (R_M, R_S)$. Clearly, $(\mathbf{R}^2_{\geq 0}, \succeq)$ is a lattice.

Recall the function $V : \mathbf{R^2_{\geq 0}} \to \mathbf{R_{\geq 0}}$ defined in the proof of Theorem 2. I will argue that V is supermodular under order $\succeq$. Consider two points R''_M, R''_S and R_M, R_S, such that $R''_i \geq R_i$. $V(\bullet)$ is supermodular if and only if for any such pair of points it is the case that

$$V(R''_M, R_S) + V(R_M, R''_S) \geq V(R''_M, R''_S) + V(R_M, R_S).$$

We can rewrite the above as

$$V(R_M, R''_S) - V(R_M, R_S) \geq V(R''_M, R''_S) - V(R''_M, R_S).$$

By Lemma 20 this can be rewritten as

$$\int_{R_S}^{R''_S} \bar{r}_S(R_M, s)\mathrm{d}s \geq \int_{R_S}^{R''_S} \bar{r}_S(R''_M, s)\mathrm{d}s. \tag{71}$$

It follows immediately from Equation (56) and the fact that, by Theorem 3, $S^{\mathrm{e}}_S(R''_M, R_S) \leq S^{\mathrm{e}}_S(R_M, R_S)$, that for any R_S we have $\bar{r}_S(R_M, R_S) \geq \bar{r}_S(R''_M, R_S)$, which proves that Equation (71) must hold.

Further, consider some $(R'_M, R'_S) \succeq (R_M, R_S)$, then by Lemma 20 follows that

$$\begin{aligned} V(R'_M, R'_S) - V(R_M, R_S) &= V(R'_M, R'_S) - V(R'_M, R_S) + V(R'_M, R_S) - V(R_M, R_S) \\ &= -\int_{R'_S}^{R_S} \bar{r}_S(R'_M, s) - c_S \mathrm{d}s + \int_{R_M}^{R'_M} \bar{r}_S(m, R_S) - c_M \mathrm{d}m \end{aligned} \tag{72}$$

Note that, by Propositions 9 and **??** we have that $\bar{r}_S(R'_M, R_S, \theta_2) \leq \bar{r}_S(R'_M, R_S, \theta_1)$ and $\bar{r}_M(R_M, R_S, \theta_2) \geq \bar{r}_S(R_M, R_S, \theta_1)$. Denote the net surplus holding in the equilibrium of the baseline model in specification θ_j, with firm measures R_M, R_S as $V(R_M, R_S, \theta_j)$. Then it follows that

$$V(R'_M, R'_S, \theta_2) - V(R_M, R_S, \theta_1) \geq V(R'_M, R'_S, \theta_2) - V(R_M, R_S, \theta_1).$$

In other words $V(R_M, R_S, \theta)$ has increasing differences in $\theta \in \{\theta_1, \theta_2\}$. Finally, note that the equilibrium sectoral firm measures $R^*_M(\theta_j), R^*_S(\theta_j)$ are given by

$$(R^*_M(\theta_j), R^*_S(\theta_j)) = \underset{(R_M, R_S) \in \mathbf{R}^2_{\geq 0}}{\arg\max} \; V(R_M, R_S, \theta_j).$$

From the facts that $\mathbf{R}^2_{\geq 0}$ endowed with the $\succeq$ is a lattice, $\theta \in \{\theta_1, \theta_2\}$ endowed with the increasing order is a partially ordered set, $V(R_M, R_S, \theta_j)$ is supermodular in (R_M, R_S)

and satisfies the increasing differences property in $(R_M, R_S; \theta_j)$ it follows from Theorem 6.1 in Topkis (1978) (or, alternatively, Theorem 4 in Milgrom and Shannon (1994)) that $R_M(\theta_2) \geq R_M(\theta_1)$ and $R_S(\theta_2) \leq R_S(\theta_2)$.

Finally, note that a change from specification θ_1 to specification θ_2 of the extended model constitutes a change from specification $(R_M(\theta_1), R_S(\theta_1), \theta_1)$ to specification $(R_M(\theta_2), R_S(\theta_2), \theta_2)$, i.e. there is a simultaneous increase in R_M, fall in R_S, universal increase in surplus level, and an increase in vertical differentiation of both workers and firms. This can be broken down as a change from $(R_M(\theta_1), R_S(\theta_1), \theta_1)$ to $(R_M(\theta_1), R_S(\theta_1), \theta_1)$ and only then $(R_M(\theta_2), R_S(\theta_1), \theta_2)$; applying Theorem 3 to both changes proves the result.

Proof of Proposition 14

First, note that it follows from differentiating Equation (14) that $\psi_{v_S}(v_S; \theta_3) \geq \frac{\min \frac{\partial}{\partial v_S}\pi_S}{\max \frac{\partial}{\partial v_M}\pi_M}$, for any $v_S \in (v_S^c(\theta_3), \bar{v}(\theta_3))$. Note that $\frac{\min \frac{\partial}{\partial v_S}\pi_S}{\max \frac{\partial}{\partial v_M}\pi_M} > 0$ by Assumptions A1.2 and A1.3.

For any $a_R \in \mathbf{R}_{>\mathbf{0}}$ we can always define a function $\psi(\cdot, a_R) : [v_S^c(\theta_1), \bar{v}(\theta_1)] \to [0, 1]$ such that $\frac{\frac{\partial}{\partial v_S}C(\psi(v_s, a_R), v_S)}{a_R} = \frac{\partial}{\partial v_S}C(\psi(v_s; \theta_1), v_S)$. I will show that there must exist some $a_R^* > 0$ such that if $a_R < a_R^*$ then $\psi_{v_S}(v_S, a_R) < \frac{\min \frac{\partial}{\partial v_S}\pi_S}{\max \frac{\partial}{\partial v_M}\pi_M}$ for all $v_S \in [v_S^c(\theta_1), \bar{v}(\theta_1)]$. By differentiating the definition of $\psi(\cdot, a_R)$ we get $\psi_{v_S}(v_S, a_R) = \frac{-s_S'(v_S;\theta_1)a_R + C_{v_S v_S}(\psi(v_s, a_R), v_S)}{C_{v_M v_S}(\psi(v_s, a_R), v_S)}$, where $-s_S'(v_S; \theta_1) = \frac{\mathrm{d}}{\mathrm{d}v_S}[\frac{\partial}{\partial v_S}C(\psi(v_s; \theta_1), v_S)]$. Therefore, it is sufficient to show that

$$- s_S'(v_S; \theta_1)a_R + C_{v_S v_S}(\psi(v_s, a_R), v_S) < \underline{c}\frac{\min \frac{\partial}{\partial v_S}\pi_S}{\max \frac{\partial}{\partial v_M}\pi_M}, \tag{73}$$

where $\underline{c} = \min_{(v_M, v_S)\in[0,1]^2} C_{v_M v_S}(v_M, v_S) > 0$ by Assumption A2.2. Note that $\frac{\partial}{\partial v_S}C(\cdot, v_S)$ is continuously increasing, and that $\frac{\partial}{\partial v_S}C(0, v_S) = 0$ for all $v_S \in [0, 1]$. From the definition of $\psi(\cdot, a_R)$ this implies that for small enough a_R, $\psi(v_s, a_R)$ must be arbitrarily small as well and, hence, so does $C_{v_S v_S}(\psi(v_s, a_R), v_S)$. Altogether, these facts imply that for small enough a_R, Equation (73) must be met, as required.

Consider $\frac{R_S(\theta_2)}{R_S(\theta_1)} < a_R^*$ and set $a_R = \frac{R_S(\theta_2)}{R_S(\theta_1)}$.[156] This implies $\psi_{v_S}(v_S, a_R) < \psi_{v_S}(v_S; \theta_3)$ for all $v_S \in E = [\max\{v_S^c(\theta_3), v_S^c(a_R)\}, \min\{\bar{v}_S(\theta_3), \bar{v}_S(a_R)\}]$. Let me adapt the definitions of the sets Ξ^1, Ξ^2 in the proof of Theorem 3 in Appendix C as $\Xi^1 = \{v_S \in E : \psi(v_S; a_R) < \psi(v_S; \theta_3) \wedge S_S(v_S; a_R) > S_S(v_S; \theta_3)\}$, $\Xi^2 = \{v_S \in E : \psi(v_S; a_R) \leq \psi(v_S; \theta_3) \wedge S_S(v_S; a_R) \geq S_S(v_S; \theta_3)\}$. It follows trivially that if $v \in \Xi^2$ then $[v, \min\{\bar{v}_S(\theta_3), \bar{v}_S(a_R)\}] \subset \Xi_1$. Comparing the triples $(R_S(\theta_3), R_S(\theta_1), \psi(\cdot; R_S(\theta_3))$ and $(R_S(\theta_3), R_S(\theta_1), \psi(\cdot; \theta_3)$ it becomes clear that they meet the strong impossibility property (see Definition 14).[157] Therefore, Lemmas 10—15 can be applied directly. From the definition of $\psi(\cdot, a_R)$ follows that $P_S(\cdot, a_R) =$

[156] This can be always achieved by setting A to be sufficiently high, because $R_S(\theta_3) = 0$ if $\pi_(0, 0; \theta_1) + A - c_M \geq \pi_S(1, 1) - c_S$ by the efficiency of equilibrium.

[157] Recall that Assumption 4 implies that $S_S(0; R_S(\theta_3)) = S_S(0; \theta_3)) = R_S(\theta_3)$.

$P_S(\cdot;\theta_1)$, because $a_R = \frac{R_S(\theta_2)}{R_S(\theta_1)}$; thus it must be the case that $P_S(v_S;\theta_3) \leq P_S(v_S;\theta_1)$ for all $v_S \in [\max\{v_S^c(\theta_3), v_S^c(a_R)\}, 1]$. Assumption 4 implies that $P_S(v_S^c) = 0$, and thus $v_S^c(\theta_3) \geq v_S^c(\theta_1)$. Inspecting Equation (9) gives $w_S^R(\theta_3) = w_S(1;\theta_3) - w_S(v_S(\theta_3);\theta_3) < w_S(1;\theta_1) - w_S(v_S(\theta_3);\theta_1) + w_S(v_S(\theta_3);\theta_1) - w_S(v_S(\theta_1);\theta_1) = w_S^R(\theta_1)$, as required.

References

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.

Akerman, A., I. Gaarder, and M. Mogstad (2015). The Skill Complementarity of Broadband Internet. *The Quarterly Journal of Economics 130*(4), 1781–1824.

Anderson, S. (2003). Why Dowry Payments Declined with Modernization in Europe but Are Rising in India. *Journal of Political Economy 111*(2), pp. 269–310.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics 118*(4), 1279–1333.

Bakija, J., A. Cole, and B. Heim (2010, March). Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data. Department of Economics Working Papers 2010-22, Department of Economics, Williams College.

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

Bickel, P. and E. Lehmann (1979). Descriptive Statistics for Nonparametric Models IV. Spread. In J. Jureckova (Ed.), *Contributions to Statistics : Jaroslav Hajek Memorial Volume*. D. Reidel Pub. Co Dordrecht, Holland ; Boston.

Bielecki, A. (1956). Une remarque sur la methode de Banach-Cacciopoli-Tikhonov dans la theorie des equations differentielles ordinaires. *Bull. Acad. Polon. Sci. Cl. III. 4*, 261–264.

Boehm, M. (2015, 03). The Price of Polarization: Estimating Task Prices Under Routine-Biased Technical Change.

Burzyński, M. and P. Gola (2017). Mexican Immigration to the US: Selection, Sorting and Matching.

Card, D. and J. E. DiNardo (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of labor economics 20*(4), 733–783.

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 (2011). Matching with a Handicap: The Case of Smoking in the Marriage Market. (1011-07).

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012). Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market. *Journal of Political Economy 120*(4), 659 – 695.

Cortes, G. M. (2016). Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data. *Journal of Labor Economics 34*(1), 63–105.

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

Costrell, R. M. and G. C. Loury (2004, December). Distribution of Ability and Earnings in a Hierarchical Job Assignment Model. *Journal of Political Economy 112*(6), 1322–1363.

Dupuy, A. (2015, January). The Assignment of Workers to Tasks with Endogenous Supply of Skills. *Economica*.

Gabaix, X. and A. Landier (2008, February). Why Has CEO Pay Increased so Much? *The Quarterly Journal of Economics 123*(1), 49–100.

Gould, E. D. (2002). Rising Wage Inequality, Comparative Advantage, and the Growing Importance of General Skills in the United States. *Journal of Labor Economics 20*(1), 105–147.

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

Grossman, G. M., E. Helpman, and P. Kircher (2013, October). Matching and Sorting in a Global Economy. NBER Working Papers 19513, National Bureau of Economic Research, Inc.

Hasselblatt, B. and A. Katok (2003). *A First Course in Dynamics: With a Panorama of Recent Developments*. Cambridge University Press.

Heckman, J. and G. Sedlacek (1985). Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market. *Journal of Political Economy 93*(6), 1077–1125.

Hopenhayn, H. A. (1992). Entry, Exit, and firm Dynamics in Long Run Equilibrium. *Econometrica 60*(5), 1127–1150.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability. Chapman & Hall.

Justman, M. and J.-F. Thisse (1997, September). Implications of the mobility of skilled labor for local public funding of higher education. *Economics Letters 55*(3), 409–412.

Kaplan, S. and J. Rauh (2010). Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes? *Review of Financial Studies 23*(3), 1004–1050.

Kaplan, S. N. and J. Rauh (2013). It's the Market: The Broad-Based Rise in the Return to Top Talent. *Journal of Economic Perspectives 27*(3), 35–56.

Katz, L. and K. M. Murphy (1992). Changes in relative wages, 19631987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kelso, A. S. and V. P. Crawford (1982). Job matching, coalition formation, and gross substitutes. *Econometrica: Journal of the Econometric Society*, 1483–1504.

Krugman, P. (1979). Increasing Returns, Monopolistic Competition, and International Trade. *Journal of International Economics*.

Legros, P. and A. F. Newman (2002). Monotone Matching in Perfect and Imperfect Worlds. *Review of Economic Studies 69*, 925–942.

Lindenlaub, I. (2017). Sorting Multidimensional Types: Theory and Application. *Review of Economic Studies*.

Mak, E. and A. Siow (2017). Occupational Choice and Matching in the Labor Market. *IZA Discussion Paper No. 10584.*.

McCann, R. J., X. Shi, A. Siow, and R. Wolthoff (2015). Becker meets ricardo: Multisector matching with communication and cognitive skills. *Journal of Law, Economics, and Organization 31*(4), 690–720.

Melitz, M. J. (2003, November). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Milgrom, P. and I. Segal (2002). Envelope Theorems for Arbitrary Choice Sets. *Econometrica*.

Milgrom, P. and C. Shannon (1994, January). Monotone Comparative Statics. *Econometrica 62*.

Mo, J.-P. (1988). Entry and structures of interest groups in assignment games. *Journal of Economic Theory 46*(1), 66–96.

Owen, D. B. and G. P. Steck (1962). Moments of Order Statistics from the Equicorrelated Multivariate Normal Distribution. *The Annals of Mathematical Statistics*.

Poutvaara, P. (2008, 09). Public and Private Education in an Integrated Europe: Studying to Migrate and Teaching to Stay? *Scandinavian Journal of Economics 110*(3), 591–608.

Rodrik, D. (2013). Unconditional Convergence in Manufacturing. *Quarterly Journal of Economics 128*(1), 165–204. This is a substantially revised version of “Unconditional Convergence” below.

Rodrik, D. (2015). *Economic Rules: Why Economics Works, When it Fails and How to Tell the Difference*. Oxford: Oxford University Press.

Rosen, S. (1981). The Economics of Superstars. *American Economic Review 71*(5), 845–58.

Roy, A. D. (1951). Some Thoughts On The Distribution Of Earnings. *Oxford Economic Papers 3*(2), 135–146.

Ryan, C. L. and K. Bauman (2016). Educational Attainment in the United States: 2015. Technical report, United States Census Bureau.

Sampson, T. (2014). Selection into Trade and Wage Inequality. *American Economic Journal: Microeconomics 6*(3), 157–202.

Sattinger, M. (1975). Comparative Advantage and the Distributions of Earnings and Abilities. *Econometrica 43*(3), 455–468.

Sattinger, M. (1979, March). Differential Rents and the Distribution of Earnings. *Oxford Economic Papers 31*(1), 60–71.

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings. *Journal of Economic Literature 31*(2), pp. 831–880.

Scarsini, M. (1984). On measures of concordance. *Stochastica 8*(3), 201–218.

Shaked, M. and G. Shanthikumar (2007). *Stochastic Orders* (1 ed.). Springer Series in Statistics. Springer-Verlag New York.

Shapley, L. and M. Shubik (1971). The Assignment Game I: The Core. *International Journal of Game Theory 1*(1), 111–130.

Sklar, A. (1959). *Fonctions de Repartition a n Dimensions et Leurs Marges.*

Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics 24*(2), 235–270.

Taylor, A. (1952). L'Hospital's Rule. *The American Mathematical Monthly*.

Tervio, M. (2008, June). The Difference That CEOs Make: An Assignment Model Approach. *The American Economic Review 98*(3), 642–668.

Teulings, C. (2005). Comparative Advantage, Relative Wages, and the Accumulation of Human Capital. *Journal of Political Economy 113*(2), 425–461.

Teulings, C. N. (1995, April). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy 103*(2), 280–315.

Topkis, D. M. (1978). Minimizing a Submodular Function on a Lattice. *Operations Research 26*(2), 305–321.