Positive Skill Clustering in Role-Assignment Matching Models[*]

Axel Anderson[†]

Economics Department
Georgetown University
Washington, DC 20057

February 21, 2022

Abstract

Kremer and Maskin (1996) explore optimal matching when production involves defined roles. Despite underlying Cobb-Douglas production functions, the induced maximum production function is not supermodular, and positive sorting does not arise. This paper introduces and solves a general class of role-assignment matching models with a continuum of types and general supermodular production functions.

The unique equilibrium entails a novel blend of positive sorting in the large, and locally negative sorting that I call *positive clustering*. I show how the equilibrium matching changes as the production function changes. In a dynamic extension, I show that sorting, mobility, and wage inequality positively covary with changes in production across matching markets (or over time).

[*]This paper benefitted from helpful feedback from Iourii Manovskii, Lones Smith, and three anonymous referees.

[†]Email: aza@georgetown.edu; web page: sites.google.com/view/axel-anderson

1 Introduction

In the Kremer and Maskin (1996) *role-assignment* model, workers are sorted into pairs, and in each pair one worker is assigned to the supervisor role and the other to the assistant role. Once these roles are assigned, output is supermodular in supervisor and assistant skill, which encourages sorting like types in equilibrium. However, output is more sensitive to supervisor skill, which yields a countervailing incentive to assign high types as supervisors and low types as assistants. This countervailing effect implies that the output function prior to role-assignment is not globally supermodular, and sorting like types is not an equilibrium (Theorem 0 in the current paper).

The core assumptions in the role-assignment model are broadly applicable. Roles could be hierarchical: a sales manager and a sales associate or a lead counsel and co-counsel. In such applications, it is natural to assume that output is more sensitive to the skills of those placed higher in the role hierarchy and that the assistant's marginal product rises in the skill of the supervisor (and vice versa). An alternative source of role differentiation is task specialization: output could be a software project that merges two separate modules. As long as the two tasks are complementary and the success of the project is more heavily reliant on one of the modules, the role-assignment model applies.

The model is quite streamlined, and yet little is known about the equilibrium matching except in extreme cases. For example, Kremer and Maskin (1996) assume a Cobb-Douglas production function in which a type x supervisor and a type y assistant produce output xy^2. They fully characterize the equilibrium matching assuming three skill types with middle types so prevalent that some must match together in equilibrium, pinning down the wages for middle types. The role-assignment models since Kremer and Maskin (1996), surveyed below, either make similarly extreme assumptions, limiting the applicability of the model, or have little to say about equilibrium matching patterns.

This paper gives assumptions on production that yield a flexible and tractable role-assignment model. In doing so, a novel matching pattern emerges: *positive clustering*. In this blend of positive sorting in the large and locally negative sorting, workers endogenously segment into skill intervals (or clusters). Within each cluster, *median matching* obtains: all workers below the median are matched as assistants to supervisors with skill above the median, and higher skill assistants are matched to higher skill supervisors.

Positive clustering is an intermediate matching pattern between median matching and perfect sorting that generalizes both. Median matching is the special case with only one skill cluster, and perfect sorting emerges in the limit as each cluster collapses to only one skill type. Not only are median matching and perfect sorting extreme matching patterns, they are inflexible: there is only one way to median match or perfectly sort a given distribution of types. In contrast, positive clustering is a flexible matching pattern: any sequence of clusters defines a feasible matching that obeys positive clustering.

Since production is supermodular once roles have been assigned, the equilibrium matching must involve positive sorting between sets of supervisors and assistants. Among all such matchings, median matching is the least sorted, as it maximizes the distance between the skills of matched partners. Since deviations from perfect sorting owe to output being more sensitive to supervisor skill, we should expect large deviations from perfect sorting when the wedge between supervisor and assistant marginal products is sufficiently large. Theorem 1 captures this intuition: if the supervisor marginal product is everywhere above the assistant marginal product, then median matching is the unique equilibrium. In contrast, median matching cannot be an equilibrium if the median type has a higher marginal product when matched as an assistant to a supervisor of the highest type than when matched as a supervisor to an assistant of the lowest type.

The *shifted marginal rate of technical substitution* is the marginal product of a skill y supervisor matched down to an assistant of type $z \leq y$ divided by the marginal product of this same skill y placed in the assistant role matched up to $x \geq y$. The *smooth pasting* condition is satisfied for a given cluster if the shifted MRTS is 1 when x is the highest type in the cluster, y is the median type, and z is the lowest type. When the shifted MRTS is non-decreasing in x (Assumption 1), then there exists a unique *smooth positive clustering solution* in which the smooth pasting condition is satisfied on all but (perhaps) the lowest skill cluster (Lemma 2). One can construct this solution by recursively equating two functions on a scalar domain: a *market clearing* curve, capturing feasibility of median matching within a cluster, and the smooth pasting condition.

Smooth positive clustering is not an equilibrium for all production functions obeying Assumption 1. Lemma 1 provides a necessary and sufficient condition for verifying that the recursively constructed smooth positive clustering solution is the unique equilibrium. Theorem 2 establishes that homogeneity and Assumption 1 together imply that smooth positive clustering is the unique equilibrium.[1] Corollary 3 extends Theorem 2 to models in which supervisor-assistant pairs buy capital on competitive markets after matches are formed. Theorem 4 relaxes homogeneity and establishes that smooth positive clustering is the unique equilibrium when the shifted marginal rate of technical substitution is

[1]Standard results in the optimal transport literature imply that the unique equilibrium is close to the smooth positive clustering solution when production is nearly homogenous.

log-convex and log-supermodular.

The equilibrium construction affords a straightforward and intuitive approach to comparative statics. Specifically, the equilibrium construction involves recursively equating the market clearing and smooth pasting locus. Thus, since the market clearing condition is independent of the production function, comparative statics follow from simply shifting the smooth pasting curve, i.e. from changes in the shifted MRTS. Any monotone change in the shifted MRTS implies a monotone shift in the smooth pasting curve. Monotonicity is not necessary for comparative statics. Instead, I consider an ordinal weakening: namely, production becomes *more biased towards supervisor skill* when the shifted MRTS satisfies a single crossing condition. If production becomes more biased towards supervisor skills *across* quantiles of the skill distribution, then the ratio of the top to bottom skill quantile in a cluster *falls* (i.e. clusters get smaller) as we move from lower to higher skill quantiles (Lemma 3). Corollary 1 applies the same logic to comparative statics across markets, establishing that markets with production that is more biased towards supervisor skills have a smaller ratio of top to bottom skill quantiles in all clusters. Intuitively, increasing the relative sensitivity of output to supervisor skill makes clusters smaller; and thus, the matching closer to perfect sorting.

I then explore the relationship between sorting and wage inequality given CES production. I consider both sorting across matches and sorting across clusters, measuring the similarity of skills in a match by the difference between the skill quantiles of the supervisor and assistant. *Sorting across matches rises* if the average distance between matched supervisor and assistant skill quantiles falls. *Sorting across clusters rises* if it takes more clusters to account for any fraction of the work force. *Sorting rises* if both types of sorting increase. Theorem 3 establishes that sorting increases in output elasticity and falls in the relative returns to supervisor skill and the elasticity of substitution between supervisors and assistants.

A common measure of wage inequality is the ratio of wage quantiles, e.g., the 90th percentile wage divided by the 10th percentile wage. Restricting attention to wage ratios across clusters, for workers at same relative position within their cluster, we find that wage inequality and sorting must positively covary across markets (Corollary 2). That is, in the CES model, any parametric change that increases sorting either increases wage inequality or leaves wage inequality unchanged. This positive covariance in sorting and inequality is both intuitive and consistent with recent empirical work (e.g., Song, Price, Guvenen, Bloom, and von Wachter (2019)).

Section 9 introduces a dynamic extension of the role-assignment model in which each individual's skill evolves over time. In steady state, the distribution over skills is fixed, but as long as individual workers remain in the market they move between roles and clusters as their skills evolve over time. Theorem 5 asserts that both types of mobility fall when production becomes more biased towards supervisor skills. Given CES production, mobility, sorting, and inequality all positively covary across markets. If we interpret clusters as coarse occupational (one-digit) groupings (e.g., managers and administrators vs. clerical) and roles as finer (three-digit) groupings (e.g., sales manager vs. sales clerk), then the positive link between mobility and wage inequality fits the stylized facts for US data uncovered by Kambourov and Manovskii (2009).

This is not the first role-assignment paper to follow Kremer and Maskin (1996). Legros and Newman (2002) establish that median matching is optimal with Cobb-Douglas production and a sufficiently tight skill distribution. Li and Suen (2001) assume production $x^{\alpha\beta}y^{(1-\alpha)\beta}$ for $\alpha \in (1/2, 1)$, and show that median matching is optimal for sufficiently tight skill distributions and not optimal for wide skill distributions with sufficient weight in the lower tail. They also show that if it is possible to match pairs such that *every* matched pair has supervisor to assistant skill ratio precisely $[\alpha/(1-\alpha)]^{1/\beta}$, then such a matching must be optimal.

Gavilan (2012) posits production function $\max_{k\geq 0} \sqrt{x}(\alpha y^{\sigma} + (1-\alpha)k^{\sigma})^{\frac{1}{2\sigma}} - pk$ with $\sigma \in (0, 1/2)$, $\alpha \in (0, 1)$. Gavilan assumes positive clustering[2] and numerically explores how sorting and inequality vary in discrete approximations to the continuum model. We capture Gavilan production as a special case in Corollary 3; thereby establishing that positive clustering is the unique equilibrium in Gavilan and providing a straightforward methodology for analytic comparative statics.

Mak and Siow (2018) consider a role-assignment model with two-dimensional types. These underlying types map into two separate scalar indices representing a worker's effective skill in the supervisor and assistant roles. As in the current work, the equilibrium follows from a market clearing curve and a curve that ensures that role-assignment choices are incentive compatible (the analogue of my smooth pasting condition). In this general model, they show that the wage functions inherit convexity from the production function.

[2]The text asserts that this is the unique equilibrium (his Lemma 3). However, the "proof" in the Appendix states that it is established numerically by solving the linear programming problem with N types. Specifically, he writes that, "it turns out that the optimal assignment of those N skill types obtained solving this maximization problem always coincides with the one established in Lemma 3." I infer that this means that he solved the discrete approximation numerically for a range of parameters and observed that the numerical solution approximated positive clustering in each case.

McCann and Trokhimtchouk (2010) provide rigorous mathematical underpinnings for the role-assignment model, allowing for multi-dimensional types. They establish the welfare theorems (duality) and uniqueness of the optimal matching with complementarity between supervisor and assistant skills. They also establish an equality that must hold on any positive measure of types matched as both assistants and supervisors. This equality is the smooth pasting condition that is satisfied at the median type within each cluster (and perhaps nowhere else) in the current work.[3]

Anderson and Smith (2021) consider the comparative statics of sorting in general matching models. Applying their theory to the role-assignment model, they show that sorting cannot fall as the relative return to assistant skill rises. However, they show by example that sorting need not rise. In particular, when the relative returns to assistant skill rises, the matching can shift from one that is (roughly) sorted for low types and mismatched for higher types to one that is sorted for high types and mismatched for lower types. Thus, any theory of sorting comparative statics in the role-assignment model must further restrict the production function, as done here.

Positive clustering is similar to *block segregation*, a matching pattern that arises in non-transferable utility, two-sided, search-and-matching models (see Burdett and Coles (1997) and Smith (2006)). Under block segregation agents also match within clusters, but realized matches within each cluster depend on the random matching process.

The next section presents the core model. The median matching characterization appears in Section 3. Section 4 introduces the market clearing and smooth pasting curves and shows how to use them to construct the equilibrium positive clustering solution. Section 5 uses these curves to deduce comparative statics across clusters and across markets as the skills distribution or production function changes. The comparative statics of the sorting and wage inequality with CES production appears in Section 6. Section 7 extends the model to allow for endogenous capital choice, while Section 8 provides alternative sufficient conditions for positive clustering for non-homogenous production. Section 9 introduces a dynamic extension of the model and derives the comparative statics of occupational mobility. Proofs follow results or appear in the Appendix.

[3]Since the set of such median types is measure zero in my positive clustering solution, I could not use their result directly, but this was my inspiration for exploring solutions satisfying this condition.

2 The Role-Assignment Model

This section introduces the *role-assignment model* studied in the current paper. As in Becker (1973), assume a pairwise "unisex" matching model with no defined sides. The skill of each individual is encoded by a scalar *type* x. The skill distribution is summarized by cdf $H : [\underline{x}, \bar{x}] \mapsto [0, 1]$ for $\underline{x} \geq 0$ and density h on the full support.

A match between a supervisor of type x and an assistant of type y yields perfectly divisible, strictly supermodular (SPM) output $g(x, y) \geq 0$ (> 0 for $(x, y) > 0$). Output is more sensitive to supervisor skill $g(x, y) \gtreqless g(y, x)$ as $x \gtreqless y$, and g is C^2, with derivatives $g_1(x, y)$ and $g_2(y, x)$ uniformly bounded in y on any open interval $x \in (a, b)$, and ordered $g_1(x, x) > g_2(x, x)$ for all $x > 0$. Taking into account that role-assignment is flexible, pair (x, y) produces output:

$$f(x, y) \equiv \max\{g(x, y), g(y, x)\} \tag{1}$$

Toward defining a market equilibrium, let $w(x)$ be the equilibrium wage of type x. In competitive equilibrium any agent of type x may hire another agent of type y by paying the equilibrium wage $w(y)$. Thus, wages obey the maximization:

$$w(x) = \max_{y} [f(x, y) - w(y)] \tag{2}$$

A *matching* is described by a symmetric bivariate cdf M on $[\underline{x}, \bar{x}]^2$. The set of *feasible* matchings $\mathscr{M}(H)$ is the space of symmetric cdfs on $[\underline{x}, \bar{x}]^2$ with marginals M_x and M_y, obeying $M_x(z) + M_y(z) = H(z)$ for all $z \in [\underline{x}, \bar{x}]$. A matching need not specify a unique match partner for each type: types x and y are *matched* if (x, y) lies in the *matching set* — the support of $M(x, y)$.

A *Competitive Equilibrium (CE)* is a pair (w, M) such that: the matching is feasible $M \in \mathscr{M}(H)$, and wages are individually rational $w \geq 0$, satisfy the optimization (2), and matches are incentive compatible given wages:

$$(\hat{x}, \hat{y}) \in \text{supp}(M) \quad \Rightarrow \quad \begin{array}{rcl} \hat{y} & \in & \arg\max_y [f(\hat{x}, y) - w(y)] \\ \hat{x} & \in & \arg\max_x [f(x, \hat{y}) - w(x)] \end{array} \tag{3}$$

The competitive equilibrium is *unique* if the disagreement between any two CE matching cdfs is measure zero.

The Planner's Problem is to choose a matching to maximize aggregate output:

$$V(H) = \max_{M \in \mathscr{M}(H)} \int_{[\underline{x},\bar{x}]^2} f(x,y) dM(x,y) \tag{4}$$

Gretsky, Ostroy, and Zame (1992) established the welfare theorems for pairwise matching problems with separate populations (e.g., buyers and sellers). For the current class of models McCann and Trokhimtchouk (2010) establish:

Lemma 0 (McCann and Trokhimtchouk (2010))**.** *The pair* (w^*, M^*) *is an equilibrium if and only if* M^* *solves the Planner's Problem* (4) *and* w^* *solves the dual Problem:*

$$\min_w \int w(x)h(x)dx \quad s.t. \;\; w(x) + w(y) \geq f(x,y) \tag{5}$$

For any matching, let $\mathscr{A} = \{x : (x,y) \in \text{supp}(M), x \leq y\}$ be the set of *assistants*, types matched to weakly higher skilled partners, and $\mathscr{S} = \{x : x \geq y, (x,y) \in \text{supp}(M)\}$ be the set of *supervisors*, types matched to weakly lower skilled partners. A type x can be both an assistant and a supervisor if some individuals of skill x match up to some $y > x$, while other individuals of type x match down to some $y' < x$. A matching with supervisors $\mathscr{S}$ and assistants $\mathscr{A}$ is a *pure pairing* if there exists a bijective *pairing function* $\mu_A : \mathscr{S} \mapsto \mathscr{A}$; such that the matching assigns measure zero to the set $\{(x,y) : y \in \mathscr{A}, y \neq \mu_A(x)\}$. In other words, a pure pairing is defined by a set of supervisors and assistants, and a function assigning assistants to supervisors. A pure pairing allows for overlapping sets of types assigned as supervisors and assistants. Following Legros and Newman (2002), a pure pairing is *increasing* if μ_A is increasing. Types may *self-match*, i.e. $(x,x) \in \text{supp}(M)$, then x is both an assistant and a supervisor. *Perfect sorting* obtains if all types self-match, and no other matches form.

Theorem 0. *The unique equilibrium is an increasing pure pairing, and marginal wages* w' *are uniquely defined almost everywhere. Perfect sorting is not an equilibrium.*

Monotonicity of the matching of supervisors to assistants follows from g SPM by standard reasoning (e.g., Becker (1973)). The next lemma provides a useful method to check whether a candidate pure pairing is an equilibrium.

Lemma 1. *Let* μ_A *be any feasible increasing pure pairing. Then* μ_A *is an equilibrium* iff *the following inequality obtains for all supervisors* $x \geq y$*:*

$$\varphi(x,y) \equiv g(x,\mu_A(x)) + g(y,\mu_A(y)) - g(x,y) - g(\mu_A(x),\mu_A(y)) \geq 0 \tag{6}$$

3 Median Matching

A preliminary step in the analysis is characterizing median matching. In a *pure* matching every type is assigned a unique match, i.e. there exists a bijective *matching function* $\mu : [\underline{x}, \bar{x}] \mapsto [\underline{x}, \bar{x}]$, such that the matching assigns measure zero to the set $\{(x, y) : y \neq \mu(x)\}$.[4] Legros and Newman (2002) introduced median matching, a type of pure matching. Specifically, if $\bar{m}$ is the median skill in the economy, *median matching* obtains when all types $x \in (\bar{m}, \bar{x})$ match with a type $y \in (\underline{x}, \bar{m})$, and for all such x above the median, the slope of the matching function is $\mu'(x) = h(x)/h(\mu(x))$.

Legros and Newman (2002) show that for Cobb-Douglas production $g(x, y) = x^{1-\alpha}y^{\alpha}$, $\alpha \in (0, 1/2)$, median matching obtains *iff* $\bar{x}/\underline{x} \leq (\alpha/(1-\alpha))^{1/(1-\alpha)}$. That is, median matching obtains *iff* the skills domain is sufficiently tight.[5] The next result asserts that median matching obtains for sufficiently tight skills domains and implies that median matching does not obtain for sufficiently spread skills distribution quite generally.

Theorem 1. *Median matching is the unique equilibrium when* $g_1(\cdot, \underline{x}) \geq g_2(\bar{x}, \cdot)$ *and fails to obtain if* $g_1(\bar{m}, \underline{x}) < g_2(\bar{x}, \bar{m})$. *In particular, median matching is the unique equilibrium if the skill lower bound* $\underline{x}$ *is sufficiently close to the upper bound* $\bar{x}$.

Proof: For any pure pairing μ_A and any pair of supervisors x, y with $y \geq \mu_A(x)$, the fundamental theorem of calculus for (6) yields:

$$\varphi(x, y) = \int_{\mu_A(x)}^{y} [g_1(s, \mu_A(y)) - g_2(x, s)]\, ds \quad \forall\, y \geq \mu_A(x) \tag{7}$$

Since $\mu_A(x) \leq y$ for any pair of supervisors (x, y) given median matching, use $g_1(\cdot, \underline{x}) \geq g_2(\bar{x}, \cdot)$, followed by g SPM, and then (7) to discover:

$$0 \leq \int_{\mu_A(x)}^{y} [g_1(s, \underline{x}) - g_2(\bar{x}, s)]\, ds \leq \int_{\mu_A(x)}^{y} [g_1(s, \mu_A(y)) - g_2(x, s)]\, ds = \varphi(x, y)$$

Thus, median matching is the unique equilibrium by Lemma 1 and Theorem 0. Finally, since g_1 and g_2 are continuous and $g_1(x, x) > g_2(x, x)$, inequality $g_1(\cdot, \underline{x}) > g_2(\bar{x}, \cdot)$ obtains whenever $\underline{x}$ and $\bar{x}$ are sufficiently close together.

[4] A pure matching is more restrictive than a pure pairing, as the latter allows for a positive measure of types assigned as both assistants and managers.

[5] McCann and Trokhimtchouk (2010) generalize this result to Cobb-Douglas production with multi-dimensional types.

To see that median matching cannot obtain when $g_1(\bar{m}, \underline{x}) < g_2(\bar{x}, \bar{m})$, assume (on the contrary) that this inequality holds and median matching is an equilibrium. Given median matching, we have $\mu_A(\bar{x}) = \bar{m}$; and thus, by (6) we have $\varphi(\bar{x}, \bar{m}) = 0$. Then, since φ is continuously differentiable in x on $(\bar{m}, \bar{x})$ and $\mu' > 0$ on the same domain, we find:

$$\lim_{x \uparrow \bar{x}} \varphi_1(\bar{x}, \bar{m}) = [g_2(\bar{x}, \bar{m}) - g_1(\bar{m}, \underline{x})] \lim_{x \uparrow \bar{x}} \mu'(\bar{x}) > 0$$

Altogether, inequality (6) is violated at $(\bar{x} - \varepsilon, \bar{m})$ for small $\varepsilon > 0$ and median matching cannot be an equilibrium by Lemma 1. □

Theorem 1 does not directly assert that median matching fails for sufficiently spread skill distributions. However, an immediate corollary is that median matching cannot obtain for sufficiently small $\underline{x}$ when $g_1(\bar{m}, 0) = 0$ (e.g., Cobb-Douglas), and also cannot obtain for sufficiently larger $\bar{x}$ provided $\lim_{\bar{x} \to \infty} g_2(\bar{x}, \bar{m}) = \infty$ (e.g., CES). Thus, median matching fails to obtain for sufficiently spread skill distribution quite generally.

4 Positive Clustering

This section defines positive clustering, shows how to recursively construct the unique matching obeying *smooth positive clustering* (SPC), and establishes conditions under which this matching is the unique equilibrium.

4.1 Positive Clustering Defined

Fix any decreasing sequence $\bar{x} \equiv x_0 > x_1 > \ldots x_N = 0$, allowing $N = \infty$. Then, the median type m_n in *skill cluster* $[x_{n+1}, x_n]$ obeys the *market clearing* condition:

$$H(m_n) - H(x_{n+1}) = H(x_n) - H(m_n) \tag{8}$$

A matching displays *positive clustering* if no type matches outside of their skill cluster and median matching obtains within every skill cluster. Thus, positive clustering is a form of pure matching in which the matching function μ obeys:

$$\forall x \in (m_n, x_n): \quad H(x_n) - H(x) = H(m_n) - H(\mu(x)) \quad \Rightarrow \quad \frac{h(x)}{h(\mu(x))} = \mu'(x) \tag{9}$$

Figure 1 illustrates that positive clustering mixes elements of positive and negative

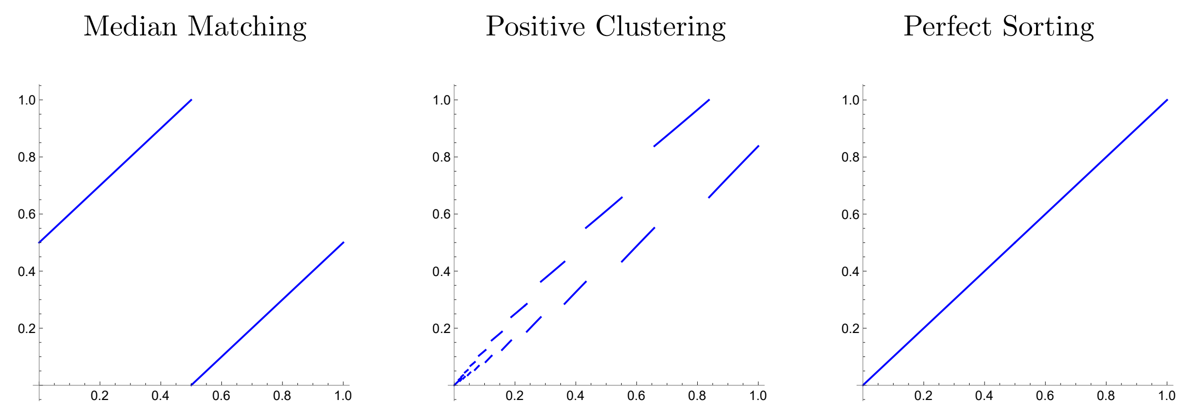

Figure 1: **Matching Patterns.** Positive clustering is an intermediate matching pattern that generalizes both median matching (left) and perfect sorting (right). The left and middle graphs assume uniformly distributed types.

sorting. *Conditional* on the set of assistants and supervisors, the matching is positively assortative, and there is *positive sorting across clusters*, since every type in a cluster is higher than all types in any lower cluster. However, there is *negative sorting within clusters*, since all assistants x_1 and x_2 match with supervisors y_1 and y_2, for which $\max(x_1, x_2) < \min(y_1, y_2)$. Positive clustering generalizes both median matching (left) and perfect sorting (right), since median matching is the special case of positive clustering with one skill cluster, and perfect sorting obtains in the limit as every type becomes its own cluster. One advantage of positive clustering for applied work is flexibility: there is only one way to perfectly sort or median match, while any sequence of clusters defines a feasible matching obeying positive clustering.

By Theorem 0, wages are differentiable almost everywhere; thus, if positive clustering is an equilibrium, we can apply the Envelope Theorem to wage maximization (2) to derive the associated marginal wages:

$$\begin{aligned} w'(x) &= g_2(\mu(x), x) \quad \forall x \in (x_{n+1}, m_n) \\ w'(x) &= g_1(x, \mu(x)) \quad \forall x \in (m_n, x_n) \end{aligned} \tag{10}$$

4.2 Smooth Positive Clustering Construction

In this subsection we construct a special case of positive clustering. The first step of the construction is to convert to quantile space. First, define the type function $X : [0, 1] \mapsto [\underline{x}, \bar{x}]$ by $H(X(q)) \equiv q$, i.e., $X(q)$ is the type of skill *quantile* q. The *quantile production function* is then $\mathrm{g}(p, q) \equiv g(X(p), X(q))$. We will assume henceforth that the skills density h is C^1. Under this assumption the quantile production function inherits all the properties imposed on g in Section 2.[6]

Toward constructing and illustrating the desired positive clustering solution, define the ratio of lowest to highest skill quantile $r_n \equiv q_{n+1}/q_n$ for each cluster n. Similarly, define the ratio of median quantile to top quantile $R_n \equiv (q_{n+1} + q_n)/(2q_n)$; i.e., $R_n q_n$ is the median quantile on cluster n. In quantile space the market clearing condition (8) relates these two ratios as $r_n = \rho_M(R_n)$, where

$$\rho_M(R_n) = 2R_n - 1 \quad \forall R_n \in \left[\frac{1}{2}, 1\right] \tag{11}$$

Toward the second key function, consider the following equation:

$$\mathrm{g}_1(R_n q_n, \rho_S q_n) - \mathrm{g}_2(q_n, R_n q_n) = 0 \tag{12}$$

Since $\mathrm{g}(p, q)$ is SPM, the LHS is strictly increasing in ρ_S. Thus, we may define the *function* $\rho_S(R|q) \in [0, 1]$ as $\rho_S(R|q) = 1$ when $\mathrm{g}_1(Rq, q) < \mathrm{g}_2(q, Rq)$, $\rho_S(R|q) = 0$ when $\lim_{p\to 0} \mathrm{g}_1(Rq, p) \geq \mathrm{g}_2(q, Rq)$, and the unique solution to (12) otherwise.

To gain some intuition for condition (12), consider the median type m_n in cluster $[x_{n+1}, x_n]$. If the median type is indifferent between matching down to x_{n+1} and up to x_n, then these two types should solve maximization (2) at $x = m_n$. In addition, if the envelope condition (10) is satisfied for both $y = x_{n+1}$ and $y = x_n$, then:

$$\begin{aligned} &w'(m_n) = g_1(m_n, x_{n+1}) \text{ and } w'(m_n) = g_2(x_n, m_n) \quad \Rightarrow \quad g_1(m_n, x_{n+1}) = g_2(x_n, m_n) \Leftrightarrow \\ &g_1(m_n, x_{n+1})X'(m_n) = g_2(x_n, m_n)X'(m_n) \quad \Leftrightarrow \quad \mathrm{g}_1(R_n q_n, r_n q_n) = \mathrm{g}_2(q_n, R_n q_n) \end{aligned} \tag{13}$$

In words, if (13) is satisfied with equality, then the rate of change in the median quantile's $(R_n q_n)$ value when matching up to q_n is equal to the rate of change in the median quantile's value when matching down to $q_{n+1} \equiv r_n q_n$. We refer to both (12) and (13) as *smooth pasting conditions* and ρ_S as the *smooth pasting locus*.[7]

[6]This smoothness assumption on the skills density is made for expository convenience. Without it the quantile production function inherits all properties from g, except that g may only be C^1 (rather than C^2). To see why, differentiate the identity $H(X(q)) \equiv q$, to get $X'(q) = h(X(q))^{-1}$. Thus, the density h must be C^1 for $g(X(p), X(q))$ to be C^2.

[7]I borrow the term *smooth pasting* from optimal stopping problems, in which the optimal harvest time equates the rate of change in the value of the planted tree to the rate of change in the value of the harvested wood. Analogously, equation (13) equates the rate of change in the value of the median type when matching up to the rate of change in the value of the medium type when matching down.

We will not assume *ex ante* that smooth pasting must be satisfied. Instead, we construct a positive clustering solution that obeys this condition on all but (perhaps) the lowest cluster and then verify that the constructed matching is indeed an equilibrium. Our construction makes use of the following assumption.

$$\mathrm{g}_1(b,c) = \mathrm{g}_2(a,b) \quad \Rightarrow \quad 2\mathrm{g}_{12}(b,c) \geq \mathrm{g}_{22}(a,b) - \mathrm{g}_{11}(b,c) \qquad \forall\; a \geq b \geq c \tag{14}$$

Lemma 2. *Given condition* (14)*, there is a unique* $R^* \in [1/2, 1)$ *s.t.* $\rho_M(R^*) = \rho_S(R^*|q) \equiv r^*(q) \in [0,1)$*, and there is a unique matching obeying positive clustering with breakpoints given by* $q_0 = 1$ *and* $q_{n+1} = r^*(q_n)q_n$*. If* $g_1(\bar{m}, \underline{x}) \geq g_2(\bar{x}, \bar{m})$*, then median matching obtains, else there are* $N \geq 2$ *clusters and* (13) *holds for all* $n < N$*.*

Since the smooth pasting condition (13) holds on all but (perhaps) the lowest skill cluster, we refer to this matching as the *smooth positive clustering solution* (SPCS). Lemma 2 provides a constructive algorithm: start with the top quantile $q_0 = 1$ and solve for the ratio $r^*(1)$ by equating the market clearing ρ_M and smooth pasting ρ_S functions.[8] The next cluster breakpoint is then $q_1 = r^*(1)$. Then recursively solve for each subsequent cluster by equating market clearing and smooth pasting evaluated at q_n to determine $r^*(q_n)$, and thus, $q_{n+1} = r^*(q_n)q_n$. This recursion may not end within a finite number of steps, but each step is computationally trivial and the exact matching on any domain $(\hat{q}, 1]$ with $\hat{q} > 0$ can be computed in a finite number of steps.

Lemmas 1 and 2 together provide a recipe for confirming when smooth positive clustering obtains in role-assignment models obeying (14). Specifically, since the smooth positive clustering solution is a feasible pure pairing (by construction), one can verify that this is an equilibrium by confirming inequality (6) for all pairs of supervisors (x, y) such that $x \geq y$. We explore exogenous characteristics of the production function that guarantee that the SPCS is the unique equilibrium below.

4.3 Positive Clustering with Homogeneous Production

As I later argue, the following monotone ratio property is satisfied by commonly used homogenous production functions.

Assumption 1. *The ratio* $\mathrm{g}_1(b,c)/\mathrm{g}_2(a,b)$ *is non-decreasing in* b *for all* $a \geq b \geq c$*.*

[8] Condition (14) ensures that ρ_S crosses ρ_M once, necessarily from above.

The marginal rate of technical substitution between supervisor and assistant skill quantiles is $\mathrm{g}_1(p,q)/\mathrm{g}_2(p,q)$. Assumption 1 considers the marginal product of skill quantile b in the supervisor role (matching down to c) and the marginal product of this *same* skill quantile in the assistant role (matching up to a). Assumption 1 demands that this *shifted MRTS* be monotone in b.

An immediate implication of Assumption 1 is that the smooth pasting locus ρ_S is non-increasing in R. To see this, rewrite the smooth pasting condition (12) as $\mathrm{g}_1(R_n q_n, \rho_S q_n)/\mathrm{g}_2(q_n, R_n q_n) = 1$, and notice that the left hand side is increasing in R; and thus, ρ_S must fall in R by g SPM.

Theorem 2. *Impose Assumption 1. Median matching is the unique equilibrium* iff $g_1(\bar{m}, \underline{x}) \geq g_2(\bar{x}, \bar{m})$. *If* $g_1(\bar{m}, \underline{x}) < g_2(\bar{x}, \bar{m})$ *and* g *is homogenous, then SPC is the unique equilibrium and the ratio* $r_n = q_{n+1}/q_n > 0$ *is constant across clusters.*

For some insight into the proof in Appendix A.4, note that Assumption 1 implies (14).[9] Consequently, there exists a unique smooth positive clustering solution by Lemma 2; feasible by construction. Since the equilibrium matching is unique by Theorem 0, we need only verify that the proposed matching is an equilibrium. The Appendix does this by showing that inequality (6) holds for all supervisor pairs $x \geq y$.

For supervisors in the same cluster, Assumption 1 and $g_1(m_n, x_n) \geq g_2(x_n, m_n)$ together guarantee that φ is increasing in y and weakly downcrossing in x.[10] Since $\varphi(x_n, m_n) = 0$ (trivially by (1)), this is sufficient for $\varphi \geq 0$ for supervisors in the same cluster. That is, under Assumption 1, the inequality $g_1(m_n, x_n) \geq g_2(x_n, m_n)$ is sufficient for *within cluster incentive compatibility.* The inequality is also necessary by $\varphi_y(x_n, m_n) \propto g_1(m_n, x_{n+1}) - g_2(x_n, m_n)$. Since the smooth positive clustering solution *is* median matching when $g_1(\bar{m}, \underline{x}) \geq g_2(\bar{x}, \bar{m})$ (Lemma 2), this inequality is necessary and sufficient for median matching. More generally, within cluster incentive compatibility obtains for all clusters since $g_1(m_n, x_n) \geq g_2(x_n, m_n)$ on all clusters for the SPCS.

When $g_1(\bar{m}, \underline{x}) < g_2(\bar{x}, \bar{m})$, the SPCS has more than one cluster (Lemma 2). In this case, we must also consider deviations across clusters. Easily, $\varphi(m_n, x_{n+1}) = 0$ and $\varphi_y(m_n, x_{n+1}) \propto g_1(x_{n+1}, m_{n+1}) - g_2(m_n, x_{n+1})$; and thus, $g_1(x_{n+1}, m_{n+1}) \leq g_2(m_n, x_{n+1})$ is necessary for $\varphi(m_n, y) \geq 0$ for supervisors y just below x_{n+1}. This inequality is also sufficient for $\varphi(x, y) \geq 0$ when x is in a higher cluster than y. In particular, Appendix A.4

[9] Indeed, since g is C^2, Assumption 1 is equivalent to $\mathrm{g}_{11}(b,c)/\mathrm{g}_1(b,c) \geq \mathrm{g}_{22}(a,b)/\mathrm{g}_2(a,b)$ for $a \geq b \geq c$. Thus, since g is increasing in both arguments, Assumption 1 implies $\mathrm{g}_{11}(b,c) \geq \mathrm{g}_{22}(a,b)$ whenever $\mathrm{g}_1(b,c) = \mathrm{g}_2(a,b)$, and so, since $\mathrm{g}_{12} \geq 0$, condition (14) holds.

[10] That is, $\varphi(x', y) < 0$ implies $\varphi(x'', y) \leq 0$ for all $x'' \geq x' \geq y$.

proves that Assumption 1 and $g_1(x_{n+1}, m_{n+1}) \leq g_2(m_n, x_{n+1})$ together imply that φ is decreasing in y and weakly upcrossing in x for all y in a lower cluster than x. When the quantile production function is homogenous we can "scale down" the smooth pasting condition in cluster n, and apply Assumption 1 to establish $g_1(x_{n+1}, m_{n+1}) \leq g_2(m_n, x_{n+1})$.

In fact, in the homogenous case, smooth positive clustering takes a particularly simple form. If $g_1(\bar{m}, \underline{x}) \geq g_2(\bar{x}, \bar{m})$ then median matching obtains. Otherwise, the matching is fully described by the unique pair (r^*, R^*) that satisfies $r^* = \rho_S(R^*) = \rho_M(R^*) > 0$, and there is an infinite sequence of clusters with $q_n = (r^*)^n$.

Appendix A.6 verifies that Theorem 2 applies to the following two examples.

Example 1 (CES)**.** *Smooth positive clustering is the unique equilibrium when* $\mathrm{g}(p, q) = (\alpha p^\sigma + (1-\alpha)q^\sigma)^{\frac{\beta}{\sigma}}$ *with* $\beta > \sigma \geq 0$ *and* $\alpha \in (1/2, 1)$*.*

The parametric restrictions are necessary. Quantile production g is strictly SPM *iff* $\beta > \sigma$, and output is more sensitive to supervisor skill *iff* $\alpha > 1 - \alpha$ and $\sigma \geq 0$. Given $\beta > \sigma$, Assumption 1 is satisfied *iff* $\alpha \in (1/2, 1)$ and $\sigma \geq 0$. Section 6 explores this example in more detail.

Example 2 (Gavilan (2012))**.** *Smooth positive clustering is the unique equilibrium for* $g(x, y) = \max_{k \geq 0} x^{\frac{1}{2}}(\alpha y^\sigma + (1-\alpha)\kappa^\sigma)^{\frac{1}{2\sigma}} - \varrho\kappa$ *with* $\sigma \in (0, 1/2)$ *and* $\alpha \in (0, 1)$ *is regular.*

This example, assumes that supervisor-assistant pairs purchase capital in a competitive market after matches have been formed. Section 7 extends the role-assignment model to accomodate endogenous capital choices and provides sufficient conditions for smooth positive clustering that subsume the Gavilan production function as a special case (Corollary 3). Roughly, Corollary 3 requires that capital be more productivity enhancing for supervisors than for assistants.

5 Comparative Statics Across Clusters and Markets

As shown in the last section, the unique smooth positive clustering solution follows from recursively equating the market clearing locus ρ_M and the smooth pasting condition ρ_S (Figure 2, left). This approach simplifies the equilibrium construction and affords powerful and intuitive tools for analyzing how the positive clustering solution changes with shifts in the production function.

First, consider variation across clusters. Easily, the ratios $r_n = q_{n+1}/q_n$ and $R_n = (q_n + q_{n+1})/(2q_n)$ must be co-monotone across clusters, since the market clearing locus is

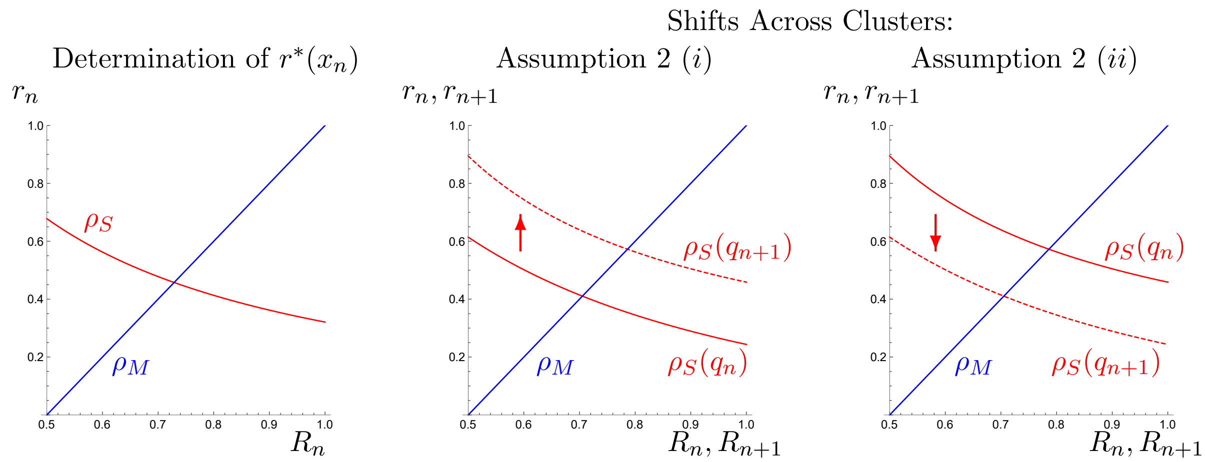

Figure 2: **Illustrations for Lemma 3.** The sequence of clusters follows from $q_0 = 1$ and difference equation $q_{n+1} = r^*(q_n)q_n$, where $r^*(q_n)$ is determined by the intersection of the market clearing locus ρ_M and the smooth pasting locus ρ_S (left). The market clearing locus ρ_M is independent of q_n, while ρ_S shifts down in q_n given Assumption 2 (i) (middle) and up in q_n when Assumption 2 (ii) holds.

independent of q and increasing in R. Thus, by Theorem 2, both ratios are constant in n when g is homogenous. This owes to the fact that the difference $g_1(kb, kc) - g_2(ka, kb)$ has a constant sign in k for homogeneous production functions; and so, the smooth pasting locus ρ_S defined by (12) is independent of q_n. Toward generalizing homogeneity, recall that the function $\eta : \mathbb{R} \mapsto \mathbb{R}$ is *upcrossing* if $\eta(z) \geq (>)0$ implies $\eta(z') \geq (>)0$ for all $z' \geq z$. A function η is *downcrossing* if $-\eta$ is upcrossing. We then have the following two natural generalizations of homogeneity:

Assumption 2. (i) $g_1(kb, kc) - g_2(ka, kb)$ *is upcrossing in* k *for* $a \geq b \geq c$ *and* (ii) $g_1(kb, kc) - g_2(ka, kb)$ *is downcrossing in* k *for* $a \geq b \geq c$.

These assumptions imply that the SPC solution is monotone across clusters.[11]

Lemma 3. *For any smooth positive clustering solution, the ratios* (r_n, R_n) *are non-decreasing in* n *under Assumption 2*(i) *and non-increasing in* n *under Assumption 2*(ii).

PROOF: The difference $g_1(Rq, \rho_S q) - g_2(q, Rq)$ is increasing in ρ_S by g SPM: thus, ρ_S shifts down in q under Assumption 2(i) and up in q under Assumption 2(ii). Monotonicity of the sequence then follows from the market clearing condition upward sloping, q_n falling in n, and (r_n, R_n) uniquely solving $r_n = \rho_S(R_n|q_n) = \rho_M(R_n)$ (Lemma 2). □

[11]Moreover, these upcrossing assumptions are robustly necessary for ρ_S to be monotone in q.

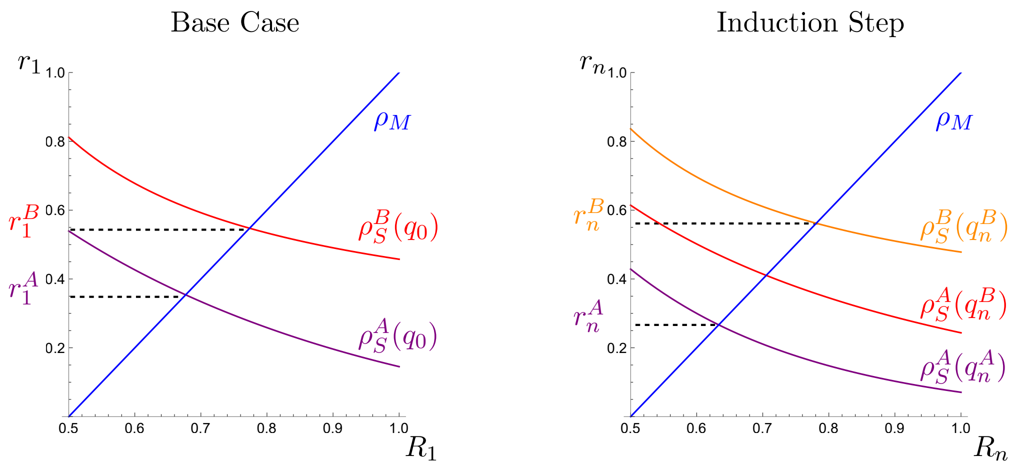

Figure 3: **Induction Proof of Lemma 4 Illustrated.** When production becomes more biased towards supervisor skill, $\rho_S(\cdot|q_n)$ shifts down for any fixed q_n; and thus, (r_1, R_1) falls as shown on the left. The change in all lower clusters also depends on how $\rho_S(\cdot|q_n)$ responds to changes in q_n. When Assumption 2 (i) holds the endogenous response reinforces the initial effect, as shown on the right.

By definition $r_{n+1} \gtreqless r_n$ as $q_{n-1}/q_n \gtreqless q_n/q_{n+1}$. Thus, cluster size, as measured by the ratio of the top quantile to bottom quantile in a cluster, increases moving from lower to higher skill clusters (i.e., as n falls) under Assumption 2(i). Conversely, clusters decrease in size when moving from lower to higher skill clusters under Assumption 2(ii). Both cases are illustrated in Figure 2.

Now consider comparisons across separate matching markets A and B with quantile production g^A and g^B. Recall that g_1 is the supervisor marginal product and g_2 is the assistant marginal product. Thus, an increase in the marginal rate of technical substitution between supervisors and assistants $\mathrm{g}_1/\mathrm{g}_2$, implies an increase in the relative returns to supervisor skill. We impose an ordinal version of this assumption; namely, production is *more biased towards supervisor skill* in market A than market B if:

$$\mathrm{g}_1^B(b,c) \geq \mathrm{g}_2^B(a,b) \quad \Rightarrow \quad \mathrm{g}_1^A(b,c) \geq \mathrm{g}_2^A(a,b) \qquad \forall\ a \geq b \geq c \tag{15}$$

Lemma 4. *If Assumptions 1 and 2(i) hold in markets A and B, then the smooth positive clustering solutions obey* $r_n^A \leq r_n^B$ *and* $q_n^A \leq q_n^B$ $\forall n$ *when production is more biased towards supervisor skill in market A.*

Proof: Let ρ_S^A and ρ_S^B be the smooth pasting conditions in markets A and B, and let $r^A(q), R^A(q)$ be the unique (by Lemma 2) ratios equating market clearing and smooth pasting in market A, i.e., $r^A(q) \equiv \rho_M(R^A) = \rho_S^A(R^A|q)$; and similarly define the unique ratios $r^B(q), R^B(q)$ in market B. Now, since the difference $g_1(Rq, \rho_S q) - g_2(q, Rq)$ is increasing in ρ_S by g strictly SPM, condition (15) implies $\rho_S^A(R|q) \leq \rho_S^B(R|q)$; and thus, $r^A(q) \leq r^B(q)$, since ρ_M is upward sloping (Figure 3, left).

We establish $q_n^A \leq q_n^B$ by induction. Easily, $q_1^A \equiv r^A(1) \leq r^B(1) \equiv q_1^B$. The proof of Lemma 3 showed that Assumption 2(i) implies that $\rho_S(R|q)$ shifts down in q. Figure 3 (right) illustrates the induction step: assume $q_n^A \leq q_n^B$, then apply $r^A(q)$ non-increasing in q, followed by $r^A(q) \leq r^B(q)$, and then $q_{n+1} = r(q_n)q_n$ in each market, to get:

$$q_n^A \leq q_n^B \;\Rightarrow\; r^A(q_n^A)q_n^A \leq r^A(q_n^B)q_n^B \;\Rightarrow\; r^A(q_n^A)q_n^A \leq r^B(q_n^B)q_n^B \;\Leftrightarrow\; q_{n+1}^A \leq q_{n+1}^B \qquad \square$$

Since homogeneous production satisfies Assumption 2(i), Theorem 2 and Lemma 4 provide the following immediate corollary.

Corollary 1. *Assume production is homogenous and obeys Assumption 1 in markets A and B with production more biased towards supervisor skill in market A. Then SPC obtains in each market with* $r^A \leq r^B$ *and* $q_n^A \leq q_n^B$ $\forall n$.

6 Sorting and Wage Inequality

In this section, I consider the interplay between sorting and wage inequality for a special case of the model. Specifically, assume CES quantile production $g(p, q) = (\alpha p^\sigma + (1 - \alpha)q^\sigma)^{\frac{\beta}{\sigma}}$ for $\alpha \in (1/2, 1)$ and $\beta > \sigma \geq 0$. Smooth positive clustering is the unique equilibrium by Example 1 where the critical threshold separating median matching from positive clustering with more than one cluster is $g_2(1, 0.5) = g_1(0.5, 0)$, i.e.:

$$\left(2^\sigma + \frac{1-\alpha}{\alpha}\right)^{\frac{\beta-\sigma}{\sigma}} = \frac{\alpha}{1-\alpha} \tag{16}$$

Given CES production, the smooth pasting condition (12) is independent of q_n:

$$\left(\frac{\alpha + (1-\alpha)R^\sigma}{\alpha R^\sigma + (1-\alpha)\rho_S(R)^\sigma}\right)^{\frac{\beta-\sigma}{\sigma}} \equiv \frac{\alpha}{1-\alpha} \tag{17}$$

These two equations allows for a sharp characterization of the equilibrium:

Lemma 5. *There exists a unique* $\beta^*(\alpha, \sigma) > \sigma$ *satisfying* (16)*, and this* β^* *is increasing in* α *and* σ*. Median matching obtains* iff $\beta \leq \beta^*$*. If* $\beta > \beta^*$*, then SPC obtains with* $q_n = r^n$ *for some* $r \in (0, 1)$*. This* r *rises in* β *and falls in* α *and* σ*. Perfect sorting obtains in the limit as* $\beta \uparrow \infty$ *or* $\alpha \downarrow \frac{1}{2}$.

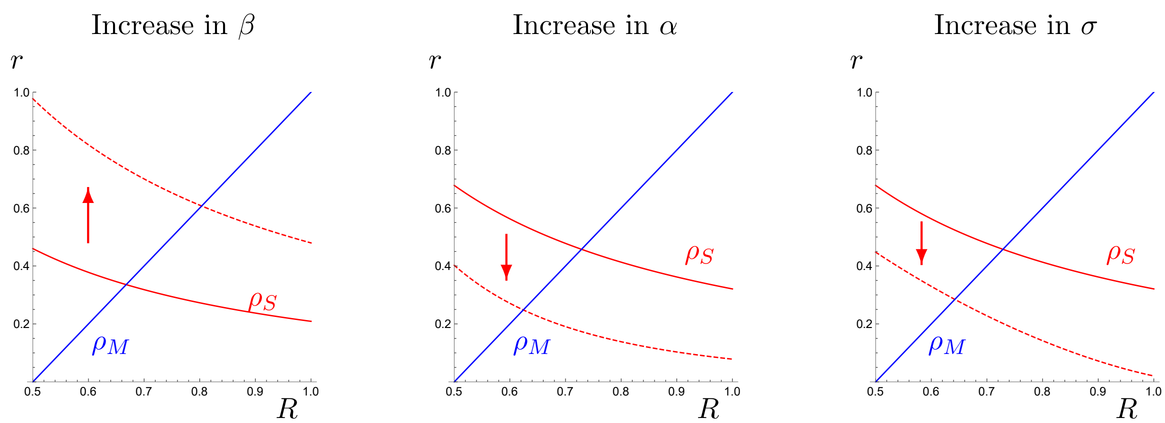

Figure 4: **The Comparative Statics of Segregation.** Left: A shift up in the output elasticity β. Middle: An increase in returns to supervisor skill α. Right: An increase in the elasticity of substitution between supervisors and assistants σ.

Consider *the elasticity of output* β. When β is low, median matching obtains. Once β increases above β^*, equilibrium involves an infinite sequence of clusters. As β increases further, the ratio $r = q_{n+1}/q_n$ rises; and thus, each q_n shifts up. Since β^* is increasing in all arguments, the opposite occurs for α and σ: median matching obtains for high values of the parameter, and once β^* exceeds β, r rises as the parameter falls (Figure 4).

Next consider the implications of these comparative statics for sorting and inequality. First consider sorting *across matches*. In order to ensure that sorting is independent of the skill scaling, consider the *quantile matching function* $\hat{\mu}(p) \equiv H(\mu(H^{-1}(p)))$. *Sorting across matches increases* in some parameter if the average distance between partners, $\int_0^1 |\hat{\mu}(p) - p|dp$, falls in the parameter.[12]

Next consider sorting *across clusters*. For any matching μ obeying positive clustering, let $n(k)$ be the index of the cluster with the kth largest measure of workers $q_{n(k)-1} - q_{n(k)}$, and $Q_K(\mu)$ be the sum $\sum_{k=1}^{k} q_{n(k)-1} - x_{n(k)}$ across the K clusters with the largest measure of workers. Trivially, perfect sorting has $Q_K = 0$ for all finite K, while for median matching $Q_K = 1$ for all $K \geq 1$. Let μ and μ' be two matchings obeying positive clustering. We say that μ is *more sorted across clusters* than μ' *iff* $Q_K(\mu) \leq Q_K(\mu')$.

Finally, *sorting increases* in some parameter *iff* sorting increases across matches and across clusters. The comparative statics of sorting are summarized in Theorem 3.

Theorem 3. *Sorting falls in the relative returns to supervisor skill (α) and the elasticity of substitution between supervisors and assistants (σ), and rises in output elasticity (β).*

[12] All derived comparative static predictions hold for $\int_0^1 |\hat{\mu}(p)/p - 1|dp$.

PROOF: The measure of each cluster $q_n - q_{n+1} = q_n(1-r)$ falls in n. Thus, $Q_K(\mu) = \sum_{n=1}^{K}(q_{n-1} - q_n) = 1 - r^n$ falls in r; or equivalently, sorting across clusters rises in r, implying that sorting across clusters rises in β and falls in α and σ by Lemma 5. Appendix A.8 establishes that sorting across matches rises in r. □

These changes in sorting are intuitive. Sorting should increase as the complementarity between supervisor and assistant skills increases. Thus, since the cross partial of the production function g is proportional to $(\beta - \sigma)$, sorting rises in β and falls in σ. Now, recall that deviations from perfect sorting are driven by the asymmetric roles of supervisors and assistants, and $\alpha \in (1/2, 1)$ parameterizes this asymmetry. In the limit as $\alpha \downarrow 1/2$, asymmetries vanish, and the matching converges to perfect sorting. More generally, sorting is falling in the supervisor-assistant production asymmetry.

Now consider a standard measure of inequality, the ratio of wages at different quantiles of the wage distribution. Since wages are increasing in skill, we can define the quantile wage function $W(p) = w(X(p))$. Using homogeneity of the quantile production function and the quantile analogue of marginal wages (10), Appendix A.9 proves:

$$W(r^n p) = r^{\beta n} W(p) \qquad \forall\ n = 1, 2, \ldots \tag{18}$$

Consequently, if $q \in (r^{n+1}, r^n)$ and $q' \in (r^{n'+1}, r^{n'})$ for integers $n > n'$ obey $q/r^n = q'/r^{n'}$, then $W(p) = r^{(n'-n)\beta} W(p')$. That is, moving up k clusters increases wages by a factor of $r^{-\beta k}$. Equation (18) affords a simple measure of *inequality across clusters*. Specifically, if $p'/p = r^{-k}$ for some integer k, then by (18) wage inequality between quantiles p' and $p < p'$ is $W(p')/W(p) = (r^{-k})^\beta = (p'/p)^\beta$.[13] Trivially, inequality across clusters is rising in the output elasticity (β). Combining this insight with the comparative statics of sorting in Theorem 3 we have:

Corollary 2. *Sorting and inequality positively covary across matching markets.*

There is empirical support for this prediction. Song, et. al. (2019) find positive co-movement of wage inequality and measures of *wage sorting* within firms. Håkanson, Lindqvist, and Vlachos (2021) measure sorting by skills directly (vs. inferring sorting by skills from sorting by wages) in Sweden from 1986 to 2008 and document that sorting increased substantially along with wage inequality. Changes in search frictions are an alternative mechanism linking sorting and inequality. However, there is strong prima facie evidence that changes in search technology are not the primary driver of the changes

[13] The function $(p'/p)^\beta$ will closely approximate wage inequality for all quantiles that are sufficiently far apart, or for *any* fixed pair of quantiles when r is sufficiently low.

in Swedish data. In particular, most of the increase in sorting took place from 1986 to 1995 (see their Figure 1), predating widespread use of the internet. Thus, for search technology to be the main protagonist of the story requires a change in search technology from 1986 to 1995 that swamps the impact of the internet from 1996 to 2008.[14]

7 Endogenous Capital

Now allow for endogenous capital choice after matches have been formed. Specifically, supervisor-assistant pair (p, q) with *capital* $\kappa \geq 0$ produces C^3 and strictly SPM gross production $\hat{g}(p, q, \kappa) > 0$ for $(p, q, \kappa) > 0$, with $\hat{g}_1(p, q, \kappa)$ and $\hat{g}_2(p, q, \kappa)$ uniformly bounded on open balls. After matches form, any $\kappa \geq 0$ can be purchased at per unit price $\varrho > 0$. The quantile production function is then $g(p, q) = \max_{k \geq 0} \hat{g}(p, q, \kappa) - \varrho\kappa$.

To ensure that g is more sensitive to supervisor skill, assume $\hat{g}(p, q, \kappa) \gtreqless \hat{g}(q, p, \kappa)$ as $p \gtreqless q$ for all $\kappa > 0$ and $\hat{g}_1(p, p, \kappa) > \hat{g}_2(p, p, \kappa)$ for all $(p, \kappa) > 0$. To guarantee that the optimal capital stock is well-defined and strictly positive, assume $\hat{g}_{33} < 0$ with $\lim_{\kappa \to 0} \hat{g}_3(p, 0, \kappa) > \varrho$ for all $p > 0$ and $\lim_{\kappa \to \infty} \hat{g}_3(1, 1, \kappa) < \varrho$.

The *role-assignment model with capital* is a role-assignment model with g as described in the preceding two paragraphs. Appendix A.6 verifies that such a $\hat{g}$ obeys all of the assumptions of the model in Section 2. But to extend Theorem 2 we need assumptions on $\hat{g}$ that imply a monotone shifted MRTS as in Assumption 1. The first is simply an analogous condition on gross production:

Assumption 3. *The ratio* $\frac{\hat{g}_1(b,c,\kappa)}{\hat{g}_2(a,b,\kappa')}$ *is non-decreasing in* b *for all* $a \geq b \geq c$ *and* $\kappa' \geq \kappa$.

This new assumption is not sufficient, as it ignores changes in the marginal products $\hat{g}_1$ and $\hat{g}_2$ induced by changes in the optimal capital stock. To ensure that such changes do not overturn the desired monotone ratio property, also consider:

$$-\frac{\hat{g}_{13}(b, c, \kappa)^2}{\hat{g}_1(b, c, \kappa)\hat{g}_{33}(b, c, \kappa)} \geq -\frac{\hat{g}_{23}(a, b, \kappa')^2}{\hat{g}_2(a, b, \kappa')\hat{g}_{33}(a, b, \kappa')} \quad \forall a \geq b \geq c \text{ and } \kappa' \geq \kappa \tag{19}$$

Since $\hat{g}$ is strictly SPM and strictly concave in κ, this inequality demands that capital be more productivity enhancing for supervisors than for assistants. To see how this relates to the monotone shifted MRTS, note that the optimal capital stock rises in both the supervisor and assistant skill quantiles by $\hat{g}$ SPM (Topkis (1998)). Thus, if

[14] I do not claim that there is no role for search frictions in explaining the observed changes in sorting and inequality, only that it seems unlikely that changes in search frictions are the *full* story.

inequality (19) did not hold, then the endogenous response to an increase in b would work to counteract the direct effect imposed in Assumption 3.

The Appendix shows that Assumption 3 and inequality (19) together guarantee that Assumption 1 is satisfied in the role-assignment model with capital and establishes the following corollary of Theorem 2:

Corollary 3. *Smooth positive clustering is the unique equilibrium in the role-assignment model with capital if* $\hat{\mathrm{g}}$ *is h.d. 1 and obeys Assumption 3 and inequality* (19).

Altogether, positive clustering is robust to endogenous capital choice, when capital is predominantly productivity enhancing for supervisors.

8 Robust Smooth Positive Clustering

It is well known that optimal matchings are continuous in production; and thus, the SPC solution will well-approximate the optimal matching when production is "close to" meeting the premise of Theorem 2.[15] In fact, if "non-decreasing" is replaced with "increasing" in Assumption 1, then the unique equilibrium will be exactly the smooth positive clustering solution for production functions that are nearly homogeneous. Rather than make this last statement precise, I consider a general class of non-homogenous production functions for which SPC is the unique equilibrium.

Absent homogeneity we must discipline the second order properties of the smooth pasting locus, which in turn follow from the second order properties of the shifted marginal rate of technical substitution.

Assumption 4. *The ratio* $\mathrm{g}_1(Rq, rq)/\mathrm{g}_2(q, Rq)$ *is individually log-convex in* r, R, *and* q, *and log-SPM in* (r, R, q^{-1})

Theorem 4. *If Assumptions 1, 2(i), and 4 hold, then smooth positive clustering is the unique equilibrium,* (r_n, R_n) *is non-decreasing in* n, *and all cluster breakpoints* q_n *fall as production becomes more biased towards supervisor skill.*

Appendix A.10 establishes that SPC is the unique equilibrium. Monotonicity of the ratios (r_n, R_n) then follows from Lemma 3, and the comparative statics of the cluster breakpoints q_n across matching markets follows from Lemma 4.

[15]Precisely, fix a sequence of production functions $\{g^k\}$ meeting the assumptions of the role-assignment model in Section 2, and converging uniformly to a production function g that meets the premise of Theorem 2. Then a subsequence $\{M_k\}$ of the optimal matching sequence weakly converges to the optimal matching for g, which obeys SPC by Theorem 2. See Theorem 5.20 in Villani (2008).

To gain some insight into the role of Assumptions 2(i) and 4, recall that homogeneity is only used in the proof of Theorem 2 to show that

$$\frac{g_1(q_{n+1}, R_{n+1}q_{n+1})}{g_2(R_n q_n, q_{n+1})} \leq 1 \tag{20}$$

which in turn rules out profitable deviations across clusters given Assumption 1. Since, r_n is non-decreasing in n under Assumption 2(i), we have an infinite sequence of clusters in Theorem 4. Consequently, the smooth pasting condition (13) holds on all clusters. In particular, on clusters n and $n+1$, i.e.:

$$\frac{g_1(R_n q_n, r_n q_n)}{g_2(q_n, R_n q_n)} = \frac{g_1(R_{n+1}q_{n+1}, r_{n+1}q_{n+1})}{g_2(q_{n+1}, R_{n+1}q_{n+1})} = 1 \tag{21}$$

Now notice that the arguments on the LHS of inequality (20) are pointwise between the arguments in the two ratios in (21); namely,

$$R_{n+1}q_{n+1} \leq q_{n+1} \leq R_n q_n, \quad r_{n+1}q_{n+1} \leq R_{n+1}q_{n+1} \leq r_n q_n, \quad q_{n+1} \leq R_n q_n \leq q_n$$

The proof in Appendix A.10 shows that (21) and Assumption 4 together imply that $g_1(\bar{R}\bar{q}, \bar{r}\bar{q})/g_2(\bar{q}, \bar{R}\bar{q}) \leq 1$ for the averages $\bar{R} = (R_n + R_{n+1})/2$, $\bar{r} = (r_n + r_{n+1})/2$, and $\bar{q} = (q_n + q_{n+1})/2$. Of course, these averages are not precisely the arguments in (20). But the additional structure afforded by Assumptions 1 and 2(i), along with g SPM, establishes that inequality (20) follows from $g_1(\bar{R}\bar{q}, \bar{r}\bar{q})/g_2(\bar{q}, \bar{R}\bar{q}) \leq 1$.

9 Mobility Across Roles and Clusters

This section considers mobility across roles and clusters over time for individual agents in a dynamic extension of the role-assignment model.

9.1 Mobility in a Dynamic Matching Model

Assume that the role-assignment matching model is repeated in periods $1, 2, \ldots, \infty$ with static production function $g(x, y)$. At the start of each period the distribution over skill types on $[0, \infty)$ is given by H_t. At the end of each period, fraction δ of agents leave the market and are replaced by a new cohort of mass δ with cdf H_0. The evolution of skills is independent of the matching. Specifically, if an individual with skill type x_t in period t survives, then $T(x_{t+1}|x_t)$ is the cdf over her skill types in period $t+1$, which is first order increasing in x_t. Altogether, the skills distribution evolves as:

$$H_{t+1}(x) = \delta H_0(x) + (1-\delta) \int_0^\infty T(x|s) dH_t(s) \tag{22}$$

Let H be the unique steady state distribution over skills implied by this contraction. Assume the Planner maximizes aggregate steady state output, i.e., solves (4) given H.[16] Since skill transitions are independent of partner skill, the Planner solution can be decentralized as a market equilibrium in which all agents choose partners to maximize their present discounted value of wages (as in Lemma 0).[17]

9.2 Occupational Mobility in a Stylized Example

Kambourov and Manovskii (2008) studies mobility across occupations and industries in the US from 1968-1997, finding that both forms of mobility are substantial and increase over this time period. Kambourov and Manovskii (2009) study the link between occupational mobility and wage inequality. They document that occupational mobility and wage inequality positively covary in US data, a finding that is robust to alternative specifications of occupational mobility and wages. They then posit a model of occupation specific human capital, calibrate the model to the mobility data, and show that the mobility calibration well-approximates the wage inequality data.

The frictionless role-assignment model is also consistent with a positive covariance between mobility (occupation or industry) and wage inequality.[18] In order to make this link precise, interpret roles as three-digit occupations (e.g., sales manager vs. sales clerk), and clusters as larger occupational (one-digit) groupings (e.g., managers and administrators vs. clerical).[19] We can then analyze occupational mobility in the steady state of the dynamic matching model.

To simplify the analysis, consider the following continuum time limit of the dynamic matching model. All workers enter the market as a type $x = 0$. Poisson death occurs

[16] If an impatient Planner maximizes the pdv of aggregate output starting from any H_t, then the distribution will converge to H and the sequence of optimal matchings M_t will converge to the matching that maximizes steady state flow output.

[17] In fact, Anderson and Smith (2010) show that the welfare theorems extend to the case where the transition cdf T depends on own type *and* current partner type.

[18] Intuitively, frictions make it easier to support positive clustering as an equilibrium, since frictions introduce switching costs across roles or clusters. In a role-assignment model with frictions, changes in sorting, mobility, and inequality would flow from changes in production and search technology.

[19] For alternative applications, clusters can be interpreted as firms with matched pairs assigned to tasks within the firm (e.g., attorneys assigned as lead and co-counsel to cases within law firms).

within t units of entry with chance $1 - e^{-\lambda t}$. While in the market, each worker's skill increases linearly over time: A worker surviving for t units of time becomes skill type γt. Thus, the steady state cdf evaluated at x is the chance that an individual dies before making it to skill x, i.e., $H(x) = 1 - e^{-\frac{\lambda}{\gamma}x}$. Inverting this cdf, and using $x_t = \gamma t$, we find:

$$p = 1 - e^{-\frac{\lambda}{\gamma}X(p)} \quad \Rightarrow \quad p_t = 1 - e^{-\lambda t} \quad \Rightarrow \quad \tau(p|q) = \lambda^{-1}\left[\log(1-p) - \log(1-q)\right] \tag{23}$$

where p_t is the quantile of a worker who survives until time t, and $\tau(p|q)$ is the time it takes a worker to move from skill quantile p to skill quantile $q > p$.

Given any steady state, define *τ-mobility within clusters* as the fraction of workers that start as assistants and transition to supervisors at least once during an interval of length τ. Similarly, Define *τ-mobility across clusters* as the fraction of workers that move between clusters at least once over time horizon τ. Now consider comparing mobility across separate matching markets, A and B. We say that *mobility within (across) clusters is higher* in market B than A if τ-mobility within (across) clusters is higher in B than A for all τ. We say that *mobility is higher* in B than A if both forms of mobility are higher in B. For markets with a constant ratio $r = q_n/q_{n+1}$ (e.g., homogeneous production), Appendix A.11 establishes that mobility rises in r, which combined with Lemma 4 yields the following result:

Theorem 5. *If steady state production is homogenous and obeys Assumption 1 in markets A and B, with production more biased towards supervisor skill in market A, then mobility is higher in market B than A.*

An immediate implication is that sorting and mobility positively covary across such matching markets. In particular, with CES production $\mathrm{g}(p,q) = (\alpha p^{\sigma} + (1-\alpha)q^{\sigma})^{\frac{\beta}{\sigma}}$ as in Section 6,[20] mobility falls in the relative returns to supervisor skill (α) and the elasticity of substitution between supervisors and assistants (σ), and rises in output elasticity (β). We can then embellish Corollary 2 as follows:

Corollary 4. *Sorting, mobility, and inequality positively covary across markets.*

10 Conclusion

Kremer and Maskin (1996) is perhaps the most famous early example of a transferable utility partnership model without Becker's supermodularity premise. The core assump-

[20] That is, $g(x,y) = (\alpha H(x)^{\sigma} + (1-\alpha)H(y)^{\sigma})^{\frac{\beta}{\sigma}}$ for $H(x) = 1 - e^{-\frac{\lambda}{\gamma}x}$.

tions of the model are natural, but characterizing the solution and determining comparative statics has proven difficult. Intuitively, absent any restrictions, the role-assignment model is too flexible, and thus not much can be said about equilibrium sorting patterns or comparative statics. On the other hand, models with globally supermodular payoffs are too rigid, allowing no deviations from perfect sorting.

This paper makes progress by introducing smooth positive clustering, identifying sufficient conditions for positive clustering, and showing how the solution can be obtained by solving a sequence of straightforward equations in scalar unknowns. Comparative statics follow from shifting the smooth pasting curve. The methodology was illustrated for CES production, determining the comparative statics of sorting and wage inequality.

A dynamic extension assumed an inflow of workers each period, exogenous death, and type dynamics for all surviving workers. While the type distribution and equilibrium matching remain constant by assumption, this extension allows for interesting career paths for individual workers. For example, under the natural assumption that (conditional on survival) worker types move up over time, the model can rationalize transitions from a supervisory role in a low productivity match (firm) to an assistant role in a high productivity match (firm). In a stylized version of the model, sorting, mobility, and wage inequality positively covary across matching markets, including over time, as in US data (see Håkanson, Lindqvist, and Vlachos (2021) and Kambourov and Manovskii (2009)).[21]

A second extension would be to allow for more than two roles per match. For example, assume production requires an assistant, manager, and supervisor. Posit a sequence of skill clusters. Within each cluster, assume all workers below the 1/3rd quantile are assistants, and all above the 2/3 are supervisors, and all others assigned as managers, and that there is positive sorting conditional on assigned roles. Then generalize the unique matching in Lemma 2 by imposing smooth pasting at the highest assistant and highest manager in each cluster.[22] Verifying that this unique matching is immune to tri-lateral deviations is less straightforward.

[21] A further extension posits learning from match partners as in Anderson and Smith (2010).

[22] Alternatively, assume firms with one supervisor and two assistants, where all workers below the 2/3rd quantile within a cluster are assistants. A more ambitious extension allows for scaling up the size of firms (as in Eeckhout and Kircher (2018)), with an endogenous assistant to supervisor ratio.

A Omitted Proofs

A.1 Proof of Theorem 0

Step 0. *The unique equilibrium is an increasing pure pairing.*

Corollary 1, Remark 4 in McCann and Trokhimtchouk (2010) establishes uniqueness of the optimal matching. Now notice that the production function f satisfies the weak increasing differences condition in Legros and Newman (2002), since $x_1 > x_2 \geq x_3 > x_4$ and g SPM implies:

$$g(x_2, x_3) - g(x_3, x_4) \leq g(x_1, x_2) - g(x_1, x_4) \;\Rightarrow\; f(x_2, x_3) - f(x_3, x_4) \leq f(x_1, x_2) - f(x_1, x_4)$$

Also, the type distribution admits a density. Altogether, the matching is an increasing pure pairing by Propositions 3 and 5 in Legros and Newman (2002).

Step 1. *Locally Lipschitz and locally semiconvex functions.*

The function $f(x, y)$ is *locally Lipschitz*, if it is Lipschitz in x on any open interval (a, b) with Lipschitz constant independent of y (not necessarily independent of (a, b)). The function $f(x, y)$ is *locally semiconvex* if for all open intervals (a, b), there exists an *error function* $\chi : \mathbb{R} \to \mathbb{R}$ obeying $\lim_{z \to 0} \chi(z)/z = 0$, such that for all x, x' in (a, b), and $t \in [0, 1]$,

$$f((1-t)x + tx', y) \leq (1-t)f(x, y) + tf(x', y) + t(1-t)\chi(|x - x'|) \tag{24}$$

Step 2. *The production function is locally Lipschitz.*

By assumption, the derivatives of g are uniformly bounded on any interval (a, b), i.e., we have $g_1(x, y) \leq L_1$ and $g_2(y, x) \leq L_2$ for all y, for finite L_1, L_2. Also, the pointwise maximum of Lipschitz functions with common Lipschitz constant L is Lipschitz with constant L by Lemma 2.1 in Heinonen (2005). Altogether, $f(x, y)$ is Lipschitz with constant $L = \max\{L_1, L_2\}$ independent of y for $x \in (a, b)$ as required.

Step 3. *The production function is locally semiconvex.*[23]

[23] This may be a known result. There are results establishing that the pointwise maximum of "semiconvex functions" is "semiconvex" (ex. Mifflin (1977)). However, the definition of semiconvexity varies across references, and I am not aware of a result for the form of semiconvexity needed in Step 4.

The production function g is C^2; and thus, for any open interval (a, b), there exists bounded constant $c > 0$ such that $\min\{g_{11}(x, y), g_{22}(y, x)\} \geq -c$, which implies that the functions $g(x, y)$ and $g(y, x)$ are locally semiconvex with common error function $\chi(z) = cz^2/2$ (Villani (2008) Example 10.11). Thus, for all $(x, x') \in (a, b)$:

$$\begin{aligned} g((1-t)x + tx', y) &\leq (1-t)g(x, y) + tg(x', y) + ct(1-t)(x - x')^2 \quad \text{and} \\ g(y, (1-t)x + tx') &\leq (1-t)g(y, x) + tg(y, x') + ct(1-t)(x - x')^2 \end{aligned}$$

Given these inequalities and convexity of the max operator, $f((1-t)x + tx', y)$ equals

$$\begin{aligned} & \max\{g((1-t)x + tx', y), g(y, (1-t)x + tx')\} \\ \leq\ & \max\{(1-t)g(x, y) + tg(x', y), (1-t)g(y, x) + tg(y, x')\} + ct(1-t)(x - x')^2 \\ \leq\ & (1-t)\max\{g(x, y), g(y, x)\} + t\max\{g(x', y), g(y, x')\} + ct(1-t)(x - x')^2 \\ =\ & (1-t)f(x, y) + tf(x', y) + t(1-t)c(x - x')^2 \end{aligned}$$

Step 4. *Wages are a.e. differentiable and returns to skill, w' are a.e. unique.*

Since the distribution over types admits a density, the premise of Proposition 3 in Chiappori, McCann, and Nesheim (2010) is met by Steps 2 and 3.[24] Thus, any equilibrium wage function must be almost everywhere differentiable, and any two equilibrium wage functions w and $\hat{w}$ obey $w' = \hat{w}'$ a.e..

Step 5. *Perfect sorting cannot be an equilibrium.*

Use g C^2 and $g_1(x, x) > g_2(x, x)$ to discover the following for small $\varepsilon > 0$:

$$g(x, x) + g(x + \varepsilon, x + \varepsilon) - 2g(x + \varepsilon, x) \approx \varepsilon\,[g_2(x, x) - g_1(x, x)] < 0 \quad \forall x > 0$$

Thus, the payoff for two mismatched pairs $(x + \varepsilon, x)$ beats positively sorting (x, x) and $(x + \varepsilon, x + \varepsilon)$, and perfect sorting cannot be an equilibrium.

[24] The definition of semiconvexity in Chiappori, McCann, and Nesheim (2010) *appears* to differ from the one I use here, but the two notions are equivalent (see Villani (2008) Proposition 10.12). Chiappori et al (2010) applies for fixed sets of assistants and supervisor. But Step 0 established the uniqueness of the allocation into sets of supervisors and assistants.

A.2 Proof of Lemma 1

Any feasible pure pairing matching μ_A is an equilibrium, *iff* output cannot be increased by rematching any pairs $(x, \mu_A(x))$ and $(y, \mu_A(y))$, i.e.:

$$f(x,\mu_A(x))+f(y,\mu_A(y))-\max\{f(x,y)+f(\mu_A(x),\mu_A(y)),f(x,\mu_A(y))+f(y,\mu_A(x))\} \geq 0$$

Since μ_A is an increasing pure pairing, we may WLOG assume $x \geq y$ are supervisors; and thus, $x \geq \mu_A(x)$, $y \geq \mu_A(y)$, and $\mu_A(x) \geq \mu_A(y)$. So, the preceding inequality is:

$$g(x,\mu_A(x))+g(y,\mu_A(y))-\max\{g(x,y)+g(\mu_A(x),\mu_A(y)),g(x,\mu_A(y))+f(y,\mu_A(x))\} \geq 0 \tag{25}$$

Now, consider mutually exclusive and exhaustive cases: $y > \mu_A(x)$ and $y \leq \mu_A(x)$.

Case 1: $y > \mu_A(x)$. In this case $f(y, \mu_A(x)) = g(y, \mu_A(x))$. Making this substitution and using g SPM we find:

$$\begin{aligned} & g(x,\mu_A(x)) + g(y,\mu_A(y)) - g(x,\mu_A(y)) - f(y,\mu_A(x)) \\ = \quad & g(x,\mu_A(x)) + g(y,\mu_A(y)) - g(x,\mu_A(y)) - g(y,\mu_A(x)) \quad \geq 0 \end{aligned}$$

Thus, $\varphi(x, y) \geq 0$ (i.e., inequality (6)), is equivalent to inequality (25).

Case 2: $y \leq \mu_A(x)$. In this case $f(y, \mu_A(x)) = g(\mu_A(x), y)$. Making this substitution and using g SPM we find:

$$\begin{aligned} & \max\{g(x,y) + g(\mu_A(x),\mu_A(y)), g(x,\mu_A(y)) + f(y,\mu_A(x))\} \\ = \quad & \max\{g(x,y) + g(\mu_A(x),\mu_A(y)), g(x,\mu_A(y)) + g(\mu_A(x),y)\} \quad = g(x,y) + g(\mu_A(x),\mu_A(y)) \end{aligned}$$

Again we see $\varphi(x, y) \geq 0$ (i.e., inequality (6)) is equivalent to inequality (25).

A.3 Proof of Lemma 2

Step 1. *The function ρ_S.*

Since g is C^2, $\rho_S(R|q)$ is continuous in R with $\rho'_S(R)$ well-defined for all $\rho_S(R) \in (0, 1)$. Thus, we totally differentiate (12), rearrange and use (14) with $1 \geq R \geq \rho_S$ to find:

$$\rho_S(R|q) \in (0, R) \quad \Rightarrow \quad \rho'_S(R|q) = \frac{\mathrm{g}_{22}(q, Rq) - \mathrm{g}_{11}(Rq, \rho_S(R)q)}{\mathrm{g}_{12}(Rq, \rho_S(R)q)} \leq 2$$

Step 2. *There is a unique $R^* \in [1/2, 1)$ s.t. $\rho_M(R^*) = \rho_S(R^*) \equiv r^*$ and $r^* \in [0, 1)$.*

Trivially, $\rho'_M(R) = 2$ with endpoints $\rho_M(1/2) = 0$ and $\rho_M(1) = 1$, while ρ_S endpoints obey $\rho_S(1|q) < 1$ (by $\mathrm{g}_1(q,q) > \mathrm{g}_2(q,q)$) and by definition of ρ_S:

$$\rho_S(1/2|q) > 0 \quad \Leftrightarrow \quad \mathrm{g}_2(q, q/2) > \lim_{p\downarrow 0} \mathrm{g}_1(q/2, p) \tag{26}$$

Thus, since ρ_M and ρ_S are both continuous in R, there exists a unique $R^* \in [1/2, 1)$ such that $\rho_M(R^*) = \rho_S(R^*) \equiv r^*(q) \in [0,1)$ for all $q \in (0,1]$. Further, by (26), $r^*(q) > 0$ *iff* $\mathrm{g}_2(q, q/2) > \lim_{p\downarrow 0} \mathrm{g}_1(q/2, p)$. In particular, median matching obtains (equivalently, $r^*(1) = 0$) when $\mathrm{g}_2(1, 1/2) \leq \lim_{p\downarrow 0} \mathrm{g}_1(1/2, p)$, i.e., when $g_2(\bar{x}, \bar{m}) \leq g_1(\bar{m}, \underline{x})$.

Step 3. *On all clusters we must have* $\mathrm{g}_1(R_n q_n, r_n q_n) \geq \mathrm{g}_2(q_n, R_n q_n)$.

Toward a contradiction, assume some cluster with $\mathrm{g}_1(R_n q_n, r_n q_n) < \mathrm{g}_2(q_n, R_n q_n)$, i.e., $\mathrm{g}_1(R^*(q_n)q_n, \rho_S(R^*(q_n)|q_n)q_n) < \mathrm{g}_2(q_n, R^*(q_n)q_n)$. Then by construction of ρ_S, we must have $r^*(q_n) \equiv \rho_S(R^*(q_n)|x_n) = 1$, contradicting Step 2.

Step 4. *The sequence of clusters.*

As shown in Step 2, $r^*(q) \in [0,1)$; and thus the sequence defined by $q_{n+1} = r^*(q_n)q_n$ must be strictly decreasing in n. There are two possibilities: $r^*(q_k) = 0$ for some $k < \infty$, in which case the sequence is finite, or $r^*(q_n) > 0$ for all n in which case the sequence is infinite. As shown in Step 2, $\mathrm{g}_2(q, q/2) > \lim_{p\downarrow 0} \mathrm{g}_1(q/2, p)$ implies $q_1 = r^*(1) > 0$; and thus at least two clusters, with (13) satisfied on all but the lowest cluster by construction.

Now consider the infinite sequence case. Since the sequence is decreasing and contained in the closed interval $[0,1]$, it must converge. To verify that the sequence $\{q_n\}$ converges to 0, assume not, i.e., $q_n \to q_\infty > 0$; and thus, $q_{n+1} \to q_\infty$ and the median on any interval $R_n q_n \to q_\infty$. Then by continuity of $\mathrm{g}_1, \mathrm{g}_2$ we must have $\mathrm{g}_1(q_n, R_n q_n) \to \mathrm{g}_1(q_\infty, q_\infty)$ and $\mathrm{g}_2(R_n q_n, q_{n+1}) \to g_2(q_\infty, q_\infty)$, but since $\mathrm{g}_1(q,q) > \mathrm{g}_2(q,q)$ for all $q \geq 0$, these limits contradict (13) and $q_\infty > 0$. $\square$

A.4 Proof of Theorem 2

The result follows from Lemmas 6 and 7, stated and proven below. We say that smooth positive clustering obeys *across cluster incentive compatibility* if inequality (20) obtains on all but (perhaps) the top cluster (i.e., for all $n > 0$).

Lemma 6. *If* g *is homogenous and Assumption 1 holds, then the smooth positive clustering solution uniquely exists and obeys across cluster incentive compatibility.*

Proof: The text established that the SPCS uniquely exists by Lemma 2 given Assumption 1. And since g is homogenous, the difference $g_1(kb, kc) - g_2(ka, kb)$ has a constant sign in k, implying that $\rho_S(R|q)$ is constant in q (by (12)): thus, R_n is constant in n. Further, Assumption 1 implies $g_1(b, c) - g_2(a, b)$ is upcrossing in b for $a \geq b \geq c$.

To show that (20) holds for all $n > 0$, impose the smooth pasting condition (13) (valid by Lemma 2 and $n > 0$), apply $g_1(kb, kc) - g_2(ka, kb)$ constant in $k = R_n$, then $g_1(b, c) - g_2(a, b)$ is upcrossing in b with $R_n^2 > r_n$ (trivially true for conditional quantiles), followed by $R_n = R_{n+1}$, and then $q_{n+1} \equiv r_n q_n$ to recover inequality (20):

$$\begin{aligned} g_1(R_n q_n, r_n q_n) = g_2(q_n, R_n q_n) \;\; &\Leftrightarrow \;\; g_1(R_n^2 q_n, R_n r_n q_n) = g_2(R_n q_n, R_n^2 q_n) \\ &\Rightarrow \;\; g_1(r_n q_n, R_n r_n q_n) \leq g_2(R_n q_n, r_n q_n) \\ &\Rightarrow \;\; g_1(r_n q_n, R_{n+1} r_n q_n) \leq g_2(R_n q_n, r_n q_n) \\ &\Leftrightarrow \;\; g_1(q_{n+1}, R_{n+1} q_{n+1}) \leq g_2(R_n q_n, q_{n+1}) \qquad \square \end{aligned}$$

Lemma 7. *Impose Assumption 1. If* $g_1(\bar{m}, \underline{x}) \geq g_2(\bar{x}, \bar{m})$, *then median matching is the unique equilibrium, if* $g_1(\bar{m}, \underline{x}) < g_2(\bar{x}, \bar{m})$, *then the SPCS is the unique equilibrium if it obeys across cluster incentive compatibility (i.e., inequality* (20) *for all* n*).*

Proof: Assume the quantile matching function $\hat{\mu}(p) \equiv H(\mu(H^{-1}(p))$ obeys smooth positive clustering, which uniquely exists by Lemma 2, and let $\mathscr{S}$ and $\mathscr{A}$ be the associated sets of supervisor and assistant skill quantiles with subsets $\mathscr{S}_n$ and $\mathscr{A}_n$ on clusters n. Recall φ defined in (6) and define the associated quantile function $\psi(p, q) \equiv \varphi(X(p), X(q))$ for all supervisors p, q. All steps impose Assumption 1, while inequality (20) will only be imposed when explicitly stated in the step header.

The function $\eta : \mathbb{R} \mapsto \mathbb{R}$ is *strictly upcrossing* if $\eta(z) \geq 0 \Rightarrow \eta(z') > 0$ for all $z' > z$.

Step 1. *The function* $g_1(q, \hat{\mu}(q)) - g_2(a, q)$ *is strictly upcrossing in* q *on* $\mathscr{S} \bigcap \{q : q \leq a\}$.

This follows since we can write the given function as the product of a positive function and an increasing function (by Assumption 1 and $g_{12} > 0$):

$$g_1(q, \hat{\mu}(q)) - g_2(a, q) = g_2(a, q) \left[\frac{g_1(q, \hat{\mu}(q))}{g_2(a, q)} - 1 \right]$$

Step 2. $\psi_2(p, q) > 0$ *on* $[R_n q_n, q_n] \times (R_n q_n, p]$.

By definition of ψ, the quantile analogue of (7) holds for $q \geq \hat{\mu}(p)$:

$$\psi(p, q) = \int_{\hat{\mu}(p)}^{q} [g_1(s, \hat{\mu}(q)) - g_2(p, s)] \, ds \tag{27}$$

Now, apply Step 3 in the proof of Lemma 2, followed by g SPM and $p \leq q_n$, and then Step 1 with $q > R_n q_n$ and $\hat{\mu}(R_n q_n) = q_{n+1}$, and finally $\mathrm{g}_{12} \geq 0$ to get:

$$\mathrm{g}_1(R_n q_n, q_{n+1}) - \mathrm{g}_2(q_n, R_n q_n) \geq 0 \;\;\Rightarrow\;\; \mathrm{g}_1(R_n q_n, q_{n+1}) - \mathrm{g}_2(p, R_n q_n) \geq 0 \;\;\Rightarrow$$
$$\mathrm{g}_1(q, \hat{\mu}(q)) - \mathrm{g}_2(p, q) > 0 \;\;\Rightarrow\;\; \psi_2(p,q) = \mathrm{g}_1(q, \hat{\mu}(q)) - \mathrm{g}_2(p,q) + \int_{\hat{\mu}(p)}^{q} \mathrm{g}_{12}(s, \hat{\mu}(q)) ds > 0$$

Step 3. *Given $q \in [R_n q_n, p]$, $\psi(p,q)$ is weakly downcrossing in p on $[R_n q_n, q_n]$.*

Let $\sigma : \mathbb{R}^2 \mapsto \mathbb{R}$, denoting by $\sigma^+ \equiv \max(\sigma, 0)$ and $\sigma^- \equiv -\min(\sigma, 0)$ the positive and negative parts of σ. The function $\Sigma(t) = \int_{[a,b]} \sigma(s,t) ds$ is *weakly upcrossing* in t if $\Sigma(t) > 0 \Rightarrow \Sigma(t') \geq 0$ for all $t' \geq t$. Theorem 1 in Anderson and Smith (2021) asserts that $\Sigma(t)$ is weakly upcrossing in t when $\forall s, s' \in S$ and $t' \geq t$:[25]

$$\sigma^-(s \wedge s', t)\sigma^+(s \vee s', t') \geq \sigma^-(s, t')\sigma^+(s', t) \tag{28}$$

Notice that this inequality is satisfied for any function σ that is non-decreasing in s and t, since $(s \vee s', t') \geq (s', s) \Rightarrow \sigma^+(s \vee s', t') \geq \sigma^+(s', t)$, and $(s, t') \geq (s \wedge s', t) \Rightarrow \sigma^-(s \wedge s', t) \geq \sigma^-(s, t')$. In particular, the function $\hat{\sigma}(s,t) \equiv 1 - \mathrm{g}_2(1-t, s)/\mathrm{g}_1(s, \hat{\mu}(q))$ obeys (28) since g is SPM and the ratio $\mathrm{g}_2(1-t, s)/\mathrm{g}_1(s, \hat{\mu}(q))$ is non-increasing in s (Assumption 1). Likewise the function $\tilde{\sigma}(s,t) \equiv \mathrm{g}_1(s, \hat{\mu}(q))\mathbb{1}_{s \geq \hat{\mu}(1-t)}$ obeys (28) since the indicator function $\mathbb{1}_{s \geq \hat{\mu}(1-t)}$ is non-decreasing in s and t (the latter by $\hat{\mu}$ increasing) and $g_1(s, \hat{\mu}(q)) > 0$ is independent of t. Easily, the product $\sigma = \hat{\sigma} \cdot \tilde{\sigma}$ also satisfies (28). Now we can rewrite (27) as:

$$\psi(1-t, q) = \int_0^q \left[1 - \frac{\mathrm{g}_2(1-t, s)}{\mathrm{g}_1(s, \hat{\mu}(q))}\right] \mathrm{g}_1(s, \hat{\mu}(q))\mathbb{1}_{s \geq \hat{\mu}(1-t)} ds \equiv \int_0^q \sigma(s,t) ds \tag{29}$$

Altogether, $\psi(1-t, q)$ is weakly upcrossing in t, i.e., $\psi(p,q)$ is weakly downcrossing in p.

Step 4. *Given inequality* (20)*, $\psi_2(p,q) < 0$ on $[R_n q_n, q_n] \times \mathcal{S} \bigcap \{q : q < q_{n+1}\}$.*

Similarly to Step 2, the following holds for all supervisors $q \leq \hat{\mu}(p)$:

$$\psi(p,q) = \int_q^{\hat{\mu}(p)} \left[\mathrm{g}_2(p,s) - \mathrm{g}_1(s, \hat{\mu}(q))\right] ds \tag{30}$$

Now, apply inequality (20), followed by g SPM and $p \geq R_n q_n$, and then Step 1 with

[25] This Theorem applies to any function $\sigma : \mathbb{R}^N \times T$ for any sub-lattice $S \subseteq \mathbb{R}^N$ and poset $(T, \succeq)$.

$q < q_{n+1}$ and $\hat{\mu}(q_{n+1}) = R_{n+1}q_{n+1}$, and finally $\mathrm{g}_{12} > 0$ to get:

$$\mathrm{g}_1(q_{n+1}, R_{n+1}q_{n+1}) - \mathrm{g}_2(R_nq_n, q_{n+1}) \le 0 \quad\Rightarrow\quad \mathrm{g}_1(q_{n+1}, R_{n+1}q_{n+1}) - \mathrm{g}_2(p, q_{n+1}) \le 0 \;\Rightarrow$$
$$\mathrm{g}_1(q, \hat{\mu}(q)) - \mathrm{g}_2(p, q) < 0 \quad\Rightarrow\quad \psi_2(p, q) = \mathrm{g}_1(q, \hat{\mu}(q)) - \mathrm{g}_2(p, q) - \int_q^{\hat{\mu}(p)} \mathrm{g}_{12}(s, \hat{\mu}(q))ds < 0$$

Step 5. *Given* $q \in \mathcal{S} \bigcap \{q : q \le q_{n+1}\}$, $\psi(p, q)$ *is weakly upcrossing in* p *on* $[R_nq_n, q_n]$.

By g SPM and Assumption 1, $\hat{\sigma}(z, t) \equiv \mathrm{g}_2(t, 1-z)/\mathrm{g}_1(1-z, \hat{\mu}(q)) - 1$ is non-decreasing in t and z, and thus, obeys (28) (with $s = z$). Likewise, $\tilde{\sigma}(z, t) \equiv \mathrm{g}_1(1-z, \hat{\mu}(q))\mathbb{1}_{z\ge 1-\hat{\mu}(t)}$ obeys (28) since $\mathbb{1}_{z\ge 1-\hat{\mu}(t)}$ is non-decreasing in z and t and $g_1(1-z, \hat{\mu}(q)) > 0$ is independent of t. Thus, the product $\sigma = \hat{\sigma} \cdot \tilde{\sigma}$ also satisfies (28). Now, rewrite (30) using change of variable $z = 1 - s$ as:

$$\psi(t, q) = \int_0^{1-q} \left[\frac{\mathrm{g}_2(t, 1-z)}{\mathrm{g}_1(1-z, \hat{\mu}(q))} - 1\right] \mathrm{g}_1(1-z, \hat{\mu}(q))\mathbb{1}_{z\ge 1-\hat{\mu}(t)} \equiv \int_0^{1-q} \sigma(z, t)dz$$

which must be weakly upcrossing in t by Theorem 1 in Anderson and Smith (2021).

Step 6. $\psi(p, q) \ge 0$ *for all* $(p, q) \in [R_nq_n, q_n] \times [R_nq_n, p]$.

Trivially $\psi(q_n, R_nq_n) = 0$, with Step 2, then Step 3, followed by ψ continuous reveals:

$$\begin{aligned}\psi(q_n, R_nq_n) = 0 \quad&\Rightarrow\quad \psi(q_n, q) > 0 \quad \forall\, q \in (R_nq_n, p]\\ &\Rightarrow\quad \psi(p, q) \ge 0 \quad \forall\, (p, q) \in [R_nq_n, q_n] \times (R_nq_n, p]\\ &\Rightarrow\quad \psi(p, q) \ge 0 \quad \forall\, (p, q) \in [R_nq_n, q_n] \times [R_nq_n, p]\end{aligned}$$

Step 7. *Given* (20), $\psi(p, q) \ge 0$ *for all* $(p, q) \in [R_nq_n, q_n] \times [R_kq_k, q_k]$ *with* $q_k \le q_{n+1}$.

Assume $q_k \le q_{n+1}$. Then use $\psi(R_nq_n, q_{n+1}) = 0$ (trivial), along with Step 4, then Step 5, followed by ψ continuous to get:

$$\begin{aligned}\psi(R_nq_n, q_{n+1}) = 0 \quad&\Rightarrow\quad \psi(R_nq_n, q) > 0 \quad \forall\, q \in [R_kq_k, q_k)\\ &\Rightarrow\quad \psi(p, q) \ge 0 \qquad \forall (p, q) \in [R_nq_n, q_n] \times [R_kq_k, q_k)\\ &\Rightarrow\quad \psi(p, q) \ge 0 \qquad \forall (p, q) \in [R_nq_n, q_n] \times [R_kq_k, q_k]\end{aligned}$$

Step 8. *Median matching is the unique equilibrium given* $g_1(\bar{m}, \underline{x}) \ge g_2(\bar{x}, \bar{m})$. *SPC is the unique equilibrium given inequality* (20).

Since the equilibrium matching is unique by Theorem 0, we need only verify that smooth positive clustering is an equilibrium. SPC is an increasing pure pairing, and is feasible by construction. Thus, by Lemma 1 we complete the proof by showing $\varphi(x, y) \ge 0$ for all supervisors $x \geq y$. Trivially, $\varphi(x,y) \geq 0$ *iff* $\psi(p,q) \equiv \varphi(X(p),X(q)) \geq 0$. Step 6 establishes this inequality for all supervisors in the same cluster. This is sufficient when $g_1(\bar{m},\underline{x}) \geq g_2(\bar{x},\bar{m})$, since SPC *is* median matching in this case (by Lemma 2). When $g_1(\bar{m},\underline{x}) < g_2(\bar{x},\bar{m})$, Lemma 2 asserts that SPC involves at least two clusters, in which case Step 7 is also required.

A.5 Proof of Corollary 3

We prove Corollary 3 by verifying the premise of Theorem 2.

Step 1. *Characterization of the optimal capital stock.*

Define the optimal capital stock $\kappa^*(p,q) \equiv \arg\max_{k\geq 0}[\hat{\mathrm{g}}(p,q,\kappa) - \varrho\kappa]$. By $\hat{\mathrm{g}}$ strictly concave in κ with a continuous derivative $\hat{\mathrm{g}}_3$ obeying $\lim_{\kappa\to 0}\hat{\mathrm{g}}_3(p,0,\kappa) = \infty$ for $p > 0$ and $\lim_{\kappa\to\infty}\hat{\mathrm{g}}_3(1,1,\kappa) = 0$ with $\hat{\mathrm{g}}_3(p,q,\kappa)$ non-decreasing in (p,q) (by $\hat{\mathrm{g}}$ SPM), there is a unique optimal $\kappa^*(p,q)$ which necessarily satisfies the FOC:

$$\hat{\mathrm{g}}_3(p,q,\kappa^*(p,q)) = \varrho \qquad \forall\ p > 0 \tag{31}$$

Thus, since $\hat{\mathrm{g}}$ is C^3, the *function* κ^* is C^2 and strictly positive for all $p > 0$. And we may differentiate (31) in p and in q, and then apply $\hat{\mathrm{g}}$ strictly SPM and strictly concave in κ to get:[26]

$$\kappa_i^*(p,q) = -\frac{\hat{\mathrm{g}}_{i3}(p,q,\kappa^*(p,q))}{\hat{\mathrm{g}}_{33}(p,q,\kappa^*(p,q))} > 0 \quad \text{ for } i \in \{1,2\} \tag{32}$$

Step 2. *Induced production* g *meets all assumptions of the role-assignment model.*

By assumption, $\hat{\mathrm{g}}(p,q,\kappa) > 0$ for $(p,q,\kappa) > 0$, while $\kappa^*(p,q) > 0$ by Step 1. Thus, $\mathrm{g}(p,q) = \hat{\mathrm{g}}(p,q,\kappa^*(p,q)) > 0$ for all $(p,q) > 0$. Since the FOC holds with equality and $\hat{\mathrm{g}}$ is C^3, we apply the Envelope Theorem (ET) and Step 1 to discover g strictly SPM:

$$\mathrm{g}_i(p,q) = \hat{\mathrm{g}}_i(p,q,\kappa^*(p,q)) \ \Rightarrow\ \mathrm{g}_{12}(p,q) = \hat{\mathrm{g}}_{12}(p,q,\kappa^*(p,q)) + \hat{\mathrm{g}}_{i3}(p,q,\kappa^*(p,q))\kappa_i^*(p,q) > 0$$

By the same ET application, g must be C^2, since κ is C^2 and $\hat{\mathrm{g}}$ is C^3, and g_i are uniformly bounded in q on open intervals in p by κ^* and $\hat{\mathrm{g}}_1$ and $\hat{\mathrm{g}}_2$ uniformly bounded on open balls.

[26]Capital increasing in (p,q) and g SPM follows from Topkis (1998) without assuming differentiability of $\hat{\mathrm{g}}$. But we use the explicit expression for κ_i^* in later arguments.

Inequality $\mathrm{g}(p,q) \gtrless \mathrm{g}(q,p)$ as $p \gtrless q$ for all $\kappa > 0$ follows easily from $\hat{\mathrm{g}}(p,q,\kappa) \gtrless \hat{\mathrm{g}}(q,p,\kappa)$ as $p \gtrless q$ for all $\kappa > 0$ and $\kappa^* > 0$ for all $p > 0$. Finally, use $\kappa^* > 0$ for $p > 0$ and the assumed ordering on the derivatives of $\hat{\mathrm{g}}$, then the Envelope Theorem to find:

$$p > 0 \;\Rightarrow\; \hat{\mathrm{g}}_1(p,p,\kappa^*(p,p)) > \hat{\mathrm{g}}_2(p,p,\kappa^*(p,p)) \;\Leftrightarrow\; \mathrm{g}_1(p,p) > \mathrm{g}_2(p,p)$$

Step 3. *Induced production* g *is homogeneous of degree 1.*

Any differentiable h.d. 1 function has h.d. 0 first derivatives, since (for example):

$$\hat{\mathrm{g}}(tp,tq,t\kappa) = t\hat{\mathrm{g}}(p,q,\kappa) \quad \Rightarrow \quad t\hat{\mathrm{g}}_i(tp,tq,t\kappa) = t\hat{\mathrm{g}}_i(p,q,\kappa)$$

Combine this equality with the fact that $\kappa^*(p,q)$ uniquely solves the FOC (31) to get:

$$\hat{\mathrm{g}}_3(p,q,\kappa^*(p,q)) = \varrho \quad \Rightarrow \quad \hat{\mathrm{g}}_3(tp,tq,t\kappa^*(p,q)) = \varrho \quad \Rightarrow \quad \kappa^*(tp,tq) = t\kappa^*(p,q)$$

Step 4. *Assumption 3 and inequality* (19) *jointly imply Assumption 1.*

Fix any $a \geq b \geq c$. Then since $\kappa^*(a,b) \geq \kappa^*(b,c)$ (shown in Step 1), we combine Assumption 3 and inequality (19), followed by (32), to discover:

$$\begin{aligned}
& \frac{\hat{\mathrm{g}}_{11}(b,c,\kappa^*(b,c)) - \frac{\hat{\mathrm{g}}_{13}(b,c,\kappa^*(b,c))^2}{\hat{\mathrm{g}}_{33}(b,c,\kappa^*(b,c))}}{\hat{\mathrm{g}}_1(b,c,\kappa^*(b,c))} \geq \frac{\hat{\mathrm{g}}_{22}(a,b,\kappa^*(a,b)) - \frac{\hat{\mathrm{g}}_{23}(a,b,\kappa^*(a,b))^2}{\hat{\mathrm{g}}_{33}(a,b,\kappa^*(a,b))}}{\hat{\mathrm{g}}_2(a,b,\kappa^*(a,b))} \\
\Rightarrow \quad & \frac{\hat{\mathrm{g}}_{11}(b,c,\kappa^*(b,c)) - \hat{\mathrm{g}}_{13}(b,c,\kappa^*(b,c))\kappa_1^*(b,c)}{\hat{\mathrm{g}}_1(b,c,\kappa^*(b,c))} \geq \frac{\hat{\mathrm{g}}_{22}(a,b,\kappa^*(a,b)) - \hat{\mathrm{g}}_{23}(a,b,\kappa^*(a,b))\kappa_2^*(a,b)}{\hat{\mathrm{g}}_2(a,b,\kappa^*(a,b))} \\
\Rightarrow \quad & \frac{\frac{\partial \hat{\mathrm{g}}_1(b,c,\kappa^*(b,c))}{\partial b}}{\hat{\mathrm{g}}_1(b,c,\kappa^*(b,c))} \geq \frac{\frac{\partial \hat{\mathrm{g}}_2(a,b,\kappa^*(a,b))}{\partial b}}{\hat{\mathrm{g}}_2(a,b,\kappa^*(a,b))} \quad \Rightarrow \quad \frac{\hat{\mathrm{g}}_1(b,c,\kappa^*(b,c))}{\hat{\mathrm{g}}_2(a,b,\kappa^*(a,b))} \text{ non-decreasing in } b
\end{aligned}$$

Thus, by the Envelope Theorem $\mathrm{g}_1(b,c)/\mathrm{g}_2(a,b)$ is non-decreasing in b.

A.6 Verifying Examples for Theorem 2

Example 1: CES. The function g is homogeneous and strictly SPM, since

$$\mathrm{g}_{12}(p,q) = \alpha(1-\alpha)\beta(\beta-\sigma)(pq)^{\sigma-1}\mathrm{g}(p,q)^{\frac{\beta-2\sigma}{\beta}} > 0$$

Trivially, $\mathrm{g}(p,q) \gtreqless \mathrm{g}(q,p)$ as $p \gtreqless q$ by $\sigma > 0$ and $\alpha \in (1/2, 1)$, and $\mathrm{g}_1(p,p)/\mathrm{g}_2(p,p) = \alpha/(1-\alpha) > 1$. Assumption 1 is satisfied since:

$$\frac{\partial \log(\mathrm{g}_1(b,c)/\mathrm{g}_2(a,b))}{\partial b} = \frac{(\beta-\sigma)b^{\sigma-1}(\alpha^2 a^\sigma - (1-\alpha)^2 c^\sigma)}{(\alpha a^\sigma + (1-\alpha)b^\sigma)(\alpha b^\sigma + (1-\alpha)c^\sigma)} > 0 \quad \forall a \geq c$$

Example 2: Gavilan Production. We show that $\hat{\mathrm{g}}$ fits the role-assignment model with capital and the extra restrictions imposed by Corollary 3. Trivially, $\hat{\mathrm{g}}$ is h.d. 1.

Step 1. $\hat{\mathrm{g}}$ *meets the premise of the role-assignment model with capital.*

Trivially, $\hat{\mathrm{g}}$ is strictly SPM and strictly positive for $(p,q,\kappa) > 0$ and C^3 with uniformly bounded first derivatives on any open ball. It is straightforward to verify that $\hat{\mathrm{g}}$ is strictly concave in κ with a continuous derivative $\hat{\mathrm{g}}_3$ that obeys $\lim_{\kappa\to 0} \hat{\mathrm{g}}_3(p,0,\kappa) = \infty$ and $\lim_{\kappa\to\infty} \hat{\mathrm{g}}_3(1,1,\kappa) = 0$. Gavilan (2012) establishes that $\hat{\mathrm{g}}(p,q,\kappa) \gtreqless \hat{\mathrm{g}}(q,p,\kappa)$ as $p \gtreqless q$. Finally, using $\alpha \in (0,1)$, we have:

$$\frac{\hat{\mathrm{g}}_1(p,p,\kappa)}{\hat{\mathrm{g}}_2(p,p,\kappa)} = 1 + \frac{(1-\alpha)\kappa^\sigma}{\alpha p^\sigma} > 1 \quad \Rightarrow \quad \hat{\mathrm{g}}_1(p,p,\kappa) > \hat{\mathrm{g}}_2(p,p,\kappa)$$

Step 2. *Gross production* $\hat{\mathrm{g}}$ *satisfies Assumption 3.*

Using $\sigma \in (0, 1/2)$ and $\alpha \in (0,1)$, and computing the requisite derivatives of $\hat{\mathrm{g}}$, we find:

$$\begin{aligned} 1 < 2(1-\sigma) \;&\Rightarrow\; 1 \leq \frac{2(1-\sigma)(1-\alpha)(\kappa')^\sigma + \alpha b^\sigma}{(1-\alpha)(\kappa')^\sigma + \alpha b^\sigma} \\ \Rightarrow \quad \frac{\hat{\mathrm{g}}_{22}(a,b,\kappa')}{\hat{\mathrm{g}}_2(a,b,\kappa')} &= -\frac{2(1-\sigma)(1-\alpha)(\kappa')^\sigma + \alpha b^\sigma}{2b\left[(1-\alpha)(\kappa')^\sigma + \alpha b^\sigma\right]} \leq -\frac{1}{2b} = \frac{\hat{\mathrm{g}}_{11}(b,c,\kappa')}{\hat{\mathrm{g}}_1(b,c,\kappa)} \end{aligned}$$

And thus, the ratio $\hat{\mathrm{g}}_1(b,c,\kappa)/\hat{\mathrm{g}}_2(a,b,\kappa')$ is non-decreasing in b as required.

Step 3. *Gross production* $\hat{\mathrm{g}}$ *obeys inequality* (19).

Differentiating the given gross production function, we discover:

$$\begin{aligned} -\frac{\hat{\mathrm{g}}_{13}(b,c,\kappa)^2}{\hat{\mathrm{g}}_1(b,c,\kappa)\hat{\mathrm{g}}_{33}(b,c,\kappa)} &= \frac{(1-\alpha)\kappa^\sigma}{b^2\left[2(1-\alpha)\kappa^\sigma + 4\alpha(1-\sigma)c^\sigma\right]} \qquad \text{(falling in } \kappa \text{ and } c\text{)} \\ -\frac{\hat{\mathrm{g}}_{23}(a,b,\kappa')^2}{\hat{\mathrm{g}}_2(a,b,\kappa')\hat{\mathrm{g}}_{33}(a,b,\kappa')} &= \frac{\alpha(1-\alpha)(1-2\sigma)^2(\kappa')^\sigma b^{\sigma-1}}{2\left[(1-\alpha)(\kappa')^\sigma + \alpha b^\sigma\right]\left[(1-\alpha)(\kappa')^\sigma + 2\alpha(1-\sigma)b^\sigma\right]} \end{aligned}$$

Use monotonicity of the top function in κ and c, along with $\kappa' \geq \kappa$ and $b \geq c$ to get:

$$
\begin{aligned}
& -\frac{\hat{\mathrm{g}}_{13}(b,c,\kappa)^2}{\hat{\mathrm{g}}_1(b,c,\kappa)\hat{\mathrm{g}}_{33}(b,c,\kappa)} - \left(-\frac{\hat{\mathrm{g}}_{23}(a,b,\kappa')^2}{\hat{\mathrm{g}}_2(a,b,\kappa')\hat{\mathrm{g}}_{33}(a,b,\kappa')}\right) \\
\geq \quad & -\frac{\hat{\mathrm{g}}_{13}(b,b,\kappa')^2}{\hat{\mathrm{g}}_1(b,b,\kappa')\hat{\mathrm{g}}_{33}(b,b,\kappa')} - \left(-\frac{\hat{\mathrm{g}}_{23}(a,b,\kappa')^2}{\hat{\mathrm{g}}_2(a,b,\kappa')\hat{\mathrm{g}}_{33}(a,b,\kappa')}\right) \\
= \quad & \frac{(1-\alpha)(\kappa')^\sigma\left[(1-\alpha)(\kappa')^\sigma + 4\alpha(1-\sigma)\sigma b^\sigma\right]}{2b\left[(1-\alpha)(\kappa')^\sigma + \alpha b^\sigma\right]\left[(1-\alpha)(\kappa')^\sigma + 2\alpha(1-\sigma)b^\sigma\right]} \geq 0
\end{aligned}
$$

A.7 Proof of Lemma 5

Define the the LHS of (17), as:

$$\xi(\alpha,\sigma,\rho,R) \equiv \left(\frac{\alpha + (1-\alpha)R^\sigma}{\alpha R^\sigma + (1-\alpha)\rho^\sigma}\right)^{\frac{\beta-\sigma}{\sigma}}$$

Step 1. *ξ falls in σ for all $\rho \leq R^2$.*

Differentiating ξ in σ we discover:

$$\frac{\xi_\sigma(\cdot)\sigma^2}{\xi(\cdot)} = (\beta-\sigma)\sigma\left(\frac{(1-\alpha)\rho^\sigma(\log(R)-\log(\rho))}{\alpha R^\sigma + (1-\alpha)\rho^\sigma} - \frac{\alpha\log(R)}{\alpha + (1-\alpha)R^\sigma}\right) - \beta\log(\xi(\cdot))$$

Evaluating at $\rho = R^2$ we find:

$$\frac{\xi_\sigma(\alpha,\sigma,R^2,R)\sigma^2}{\xi(\alpha,\sigma,R^2,R)} \propto R^{\sigma-\beta}\log(R) < 0 \tag{33}$$

Differentiating a second time, we find:

$$\frac{\partial\left[\frac{\xi_\sigma(\alpha,\sigma,\rho,R)\sigma^2}{\xi(\alpha,\sigma,\rho,R)}\right]}{\partial\rho} \propto \frac{\rho^{\sigma-1}\left[(1-\alpha)\rho^\sigma + (1+(\beta-\sigma)(\log(R)-\log(\rho)))\,R^\sigma\right]}{((1-\alpha)\rho^\sigma + \alpha R^\sigma)^2} \geq 0 \tag{34}$$

Combining (33) and (34) we see that $\xi_\sigma\sigma^2/\xi$ is increasing in ρ, negative at $\rho = R^2$; and thus, $\xi_\sigma(\alpha,\sigma,\rho,R) < 0$ for all $\rho \leq R^2$.

Step 2. *ξ falls in α for all $\rho \leq R^2$.*

Differentiating, we find:

$$\frac{\partial \log(\xi(\alpha,\sigma,\rho,R))}{\partial\alpha} \propto \left(\frac{\alpha+(1-\alpha)R^\sigma}{\alpha R^\sigma+(1-\alpha)\rho^\sigma}\right)^{\frac{\beta}{\sigma}} \frac{\rho^\sigma - R^{2\sigma}}{(\alpha+(1-\alpha)R^\sigma)^2} \leq 0$$

Step 3. *There exists a unique $\beta^*(\alpha,\sigma) > \sigma$ satisfying* (16)*, and β^* is increasing in α and σ. Median matching obtains* iff $\beta \leq \beta^*$.

The LHS of (16) is continuous and strictly increasing in β, converging to ∞ as $\beta \to \infty$. And since $\alpha \in (1/2, 1)$, the LHS falls short of the RHS at $\beta = \sigma$. Altogether, there exists a unique $\beta^* > \sigma$ satisfying (16). To see that β^* must be increasing in σ and α note that the LHS of (16) is $\xi(\alpha,\sigma,0,2^{-1/\gamma})$, which decreases in α and σ by Steps 1 and 2. Since the LHS of (16) increases in β, median matching obtains *iff* $\beta \leq \beta^*$ by Theorem 2.

Step 4. *The locus ρ_S shifts down in α and σ at R^*, i.e., $\partial\rho_S(R^*)/\partial\sigma \leq 0$.*

Since $R^2 \geq \rho_M(R)$, $\xi_\alpha(\alpha,\sigma,\rho_S(R^*),R^*) < 0$ and $\xi_\sigma(\alpha,\sigma,\rho_S(R^*),R^*) < 0$ by Steps 1 and 2, while $\xi_\rho < 0$ (by $\beta > \sigma$). Further, the RHS of (17) is independent of σ and increasing in α. Altogether, $\partial\rho_S(R^*)/\partial\alpha \leq 0$ and $\partial\rho_S(R^*)/\partial\sigma \leq 0$.

Step 5. *Perfect sorting obtains in the limit as $\beta \uparrow \infty$ or $\alpha \downarrow \frac{1}{2}$.*

Rewrite the smooth pasting condition (17) as:

$$\frac{\alpha+(1-\alpha)R^\sigma}{\alpha R^\sigma+(1-\alpha)\rho_S(R)^\sigma} \equiv \left(\frac{\alpha}{1-\alpha}\right)^{\frac{\sigma}{\beta-\sigma}} \tag{35}$$

Notice that the RHS converges to 1 as $\beta \to \infty$ or $\alpha \to 1/2$. Thus, the LHS must also converge to 1, which requires $\rho_S(R) \to R \to 1$; and thus $r^* \to 1$. By definition, this implies q_{n+1}/q_n converges to 1, and the matching converges to perfect sorting.

A.8 Monotone Sorting: Proof of Theorem 3

By definition, the quantile matching function for smooth positive clustering is:

$$\hat{\mu}(p) = \begin{array}{ll} p + \frac{1}{2}(1-r)r^n & \forall\, p \in (r^{n+1}, Rr^n) \\ p - \frac{1}{2}(1-r)r^n & \forall\, p \in (Rr^n, r^n) \end{array}$$

Thus, for $p \in [r^{n+1}, r^n]$ we have $|\hat{\mu}(p) - p| = (1-r)r^n/2$; and so, the average distance between matched partners is

$$\frac{1}{2}\sum_{n=0}^{\infty}\left(r^n - r^{n+1}\right)(1-r)r^n = \frac{1}{2}(1-r)^2\sum_{n=0}^{\infty} r^{2n} = \frac{(1-r)^2}{2(1-r^2)}$$

which falls in r. The comparative statics in α, β, and σ then follow from Lemma 5.

A.9 Verification of Equation (18)

Easily, the quantile analogue of (10) holds; namely,

$$\begin{aligned} W'(p) &= \mathrm{g}_2(\hat{\mu}(p), p) \quad \forall p \in (q_{n+1}, Rq_n) \\ W'(p) &= \mathrm{g}_1(p, \hat{\mu}(p)) \quad \forall p \in (Rq_n, q_n) \end{aligned} \tag{36}$$

and since r is constant across clusters, the quantile matching function obeys $\hat{\mu}(rp) = r\hat{\mu}(p)$ for all p. Using these facts and homogeneity of the production function, we have the following equality for all assistants p:

$$W'(rp) = \mathrm{g}_2(\hat{\mu}(rp), rp) = \mathrm{g}_2(r\hat{\mu}(p), rp) = r^{\beta-1}\mathrm{g}_2(\hat{\mu}(p), p) = r^{\beta-1}W'(p)$$

Now use $W'(rp) = r^{\beta-1}W'(p)$, the fact that marginal wages are defined a.e., and change of variable $s = rt$, to get:

$$W(rp) = W(0) + \int_0^{rp} W'(s)ds = \int_0^{p} rW'(rt)dt = r^{\beta}\int_0^{p} W'(t)dt = W(0) + r^{\beta}W(p)$$

Then by $W(0) = \mathrm{g}(0,0)/2 = 0$, we discover $W(rp) = r^{\beta}W(p)$; and so (18) holds.

A.10 Proof of Theorem 4

Median matching is the unique equilibrium when $g_1(\bar{m}, \underline{x}) \geq g_2(\bar{x}, \bar{m})$ by Theorem 2. The following proof verifies that the premise of Lemma 7 is met in the remaining case, $g_1(\bar{m}, \underline{x}) < g_2(\bar{x}, \bar{m})$ for which median matching cannot obtain by Theorem 1.

Step 1. *The order of ratios.*

By Lemma 3, (r_n, R_n) is non-decreasing in n; and thus, the averages $\bar{R} = (R_n + R_{n+1})/2$,

$\bar{r} = (r_n + r_{n+1})/2$, and $\bar{q} = (q_n + q_{n+1})/2$ obey:

$$(R_{n+1}, r_{n+1}) \geq (\bar{R}, \bar{r}) \geq (R_n, r_n) \quad \text{and} \quad q_{n+1} \leq \bar{q} \leq q_n \tag{37}$$

Step 2. *Inequalities* $\frac{r_n}{R_n} < \bar{R}$ *and* $r_n \frac{R_{n+1}}{R_n} \leq \bar{r}$ *hold for all* n.

For the first inequality note that $R_n \in (1/2, 1)$ implies $R_n^2 > 2R_n - 1 \equiv \rho_M(R_n) = r_n$ and $\bar{R} \geq R_n$ by (37). For the second inequality use (37), followed by $R_n < 1$ to find:

$$\begin{aligned}
& R_{n+1} \geq R_n \;\Rightarrow\; R_{n+1}(R_n - 1) \leq R_n(R_n - 1) \\
\Leftrightarrow\;& R_{n+1}(R_n - 1) + R_{n+1}R_n \leq R_n(R_n - 1) + R_{n+1}R_n \\
\Leftrightarrow\;& (2R_n - 1)R_{n+1} \leq (R_{n+1} + R_n - 1)R_n \\
\Leftrightarrow\;& (2R_n - 1)R_{n+1} \leq (2\bar{R} - 1)R_n \;\Leftrightarrow\; r_n R_{n+1} \leq \bar{r} R_n
\end{aligned}$$

where the last implications uses $r_n = \rho_M(R_n) = 2R_n - 1$ and $\bar{r} = \rho_M(\bar{R}) \equiv 2\bar{R} - 1$ (by ρ_M linear and $r_n = \rho_M(R_n)$ and $r_{n+1} = \rho_M(R_{n+1})$).

Step 3. *We have* $g_1(\bar{R}\bar{q}, \bar{r}\bar{q})/g_2(\bar{q}, \bar{R}\bar{q}) \leq 1$ *for all* n.

To conserve space, define the function $\phi(r, R, q) \equiv g_1(Rq, rq)/g_2(q, Rq)$. Since (r_n, R_n) is non-decreasing (Lemma 3), the SPC solution has an infinite sequence of clusters whenever median matching does not obtain. Thus, by Lemma 2, equation (13) holds for all n. Equivalently, $\phi(r_n, R_n, q_n) = 1$ for all n.

Now, by assumption ϕ is separately log-convex in each argument and log-SPM in (r, R, q^{-1}) (i.e., with q reverse ordered); and thus, we can use in order: (i) log-convexity in r, (ii) log-convexity in R, (iii) log-convexity in q, (iv) log-SPM in (r, R, q^{-1}) with (37), and (v) $\phi(r_n, R_n, q_n) = 1$ for all n, to discover:

$$\begin{aligned}
& \phi(\bar{r}, \bar{R}, \bar{q})^8 \leq \phi(r_n, \bar{R}, \bar{q})^4 \phi(r_{n+1}, \bar{R}, \bar{q})^4 \\
\leq\;& \phi(r_n, R_n, \bar{q})^2 \phi(r_n, R_{n+1}, \bar{q})^2 \phi(r_{n+1}, R_n, \bar{q})^2 \phi(r_{n+1}, R_{n+1}, \bar{q})^2 \\
\leq\;& \phi(r_n, R_n, q_n)\phi(r_n, R_n, q_{n+1})\phi(r_n, R_{n+1}, q_n)\phi(r_n, R_{n+1}, q_{n+1}) \\
& \times \phi(r_{n+1}, R_n, q_n)\phi(r_{n+1}, R_n, q_{n+1})\phi(r_{n+1}, R_{n+1}, q_n)\phi(r_{n+1}, R_{n+1}, q_{n+1}) \\
\leq\;& \phi(r_n, R_n, q_n)^4 \phi(r_{n+1}, R_{n+1}, q_{n+1})^4 = 1
\end{aligned}$$

Step 4. *The SPCS obeys across cluster incentive compatibility, i.e.,* (20) *for all* n.

Use Step 3 along with the definition of ϕ, followed by $r_n \frac{R_{n+1}}{R_n} \leq \bar{r}$ (Step 2) with g SPM, then $\frac{r_n}{R_n} \leq \bar{R}$ (Step 2) with Assumption 1, and then $\bar{q} = (q_n + r_n q_n)/2 = R_n q_n$ to get:

$$1 \geq \frac{\mathrm{g}_1(\bar{R}\bar{q}, \bar{r}\bar{q})}{\mathrm{g}_2(\bar{q}, \bar{R}\bar{q})} \geq \frac{\mathrm{g}_1(\bar{R}\bar{q}, r_n R_{n+1} R_n^{-1}\bar{q})}{\mathrm{g}_2(\bar{q}, \bar{R}\bar{q})} \geq \frac{\mathrm{g}_1(r_n R_n^{-1}\bar{q}, r_n R_{n+1} R_n^{-1}\bar{q})}{\mathrm{g}_2(\bar{q}, r_n R_n^{-1}\bar{q})} = \frac{\mathrm{g}_1(r_n q_n, r_n R_{n+1} q_n)}{\mathrm{g}_2(R_n q_n, r_n q_n)}$$
$$\Rightarrow \quad \mathrm{g}_1(r_n q_n, r_n R_{n+1} q_n) \leq \mathrm{g}_2(R_n q_n, r_n q_n) \quad \Leftrightarrow \quad \mathrm{g}_1(q_{n+1}, R_{n+1} q_{n+1}) \leq \mathrm{g}_2(R_n q_n, q_{n+1})$$

A.11 Mobility increases in r

Lemma 8. *Assume SPC in two markets with constant ratios $r = q_{n+1}/q_n$ and $r' = q'_{n+1}/q'_n > r$. Mobility is higher under r' than r.*

Consider mobility over some time interval $\tau > 0$. Define $p_n^*(R) = r^{n+1}$ if $\tau(p|Rr^n) \leq \tau$ for all $p \in [r^{n+1}, Rr^n]$ and let $p_n^*(R) \in [r^{n+1}, Rr^n)$ solve $\tau(p_n^*(R)|Rr^n) = \tau$ otherwise.

Step 1. *If $\alpha r \leq 1$ then $\alpha R r^{n+1} - p_n^*(R)$ is upcrossing in n.*

$$\alpha R r^{n+1} - p_n^*(R) \geq (>) 0 \quad \Leftrightarrow \quad \tau \geq (>) \tau(\alpha R r^{n+1}|Rr^n) \quad \Leftrightarrow \quad e^{\lambda\tau} \geq (>) \frac{1 - \alpha R r^{n+1}}{1 - Rr^n}$$

Now note that the fraction on the right is non-increasing in n when $\alpha r \leq 1$.

Step 2. *Within cluster mobility increases in r.*

By definition within cluster mobility under r is:

$$m_W(r) = \sum_{n=0}^{\infty} \int_{r^{n+1}}^{Rr^n} \mathbb{1}_{\tau(p|Rr^n) \leq \tau} dp$$

Define $k \equiv r'/r > 1$ and use change of variable $p = (R/R')k^{-n}q$ to get:

$$m_W(r') = \sum_{n=0}^{\infty} \int_{(r')^{n+1}}^{R'(r')^n} \mathbb{1}_{\tau(q|R'(r')^n) \leq \tau} dq = \sum_{n=0}^{\infty} \int_{\frac{R}{R'}kr^{n+1}}^{Rr^n} \mathbb{1}_{\tau(p|Rr^n) \leq \tau} dp$$

Then we have:

$$m_W(r') - m_W(r) = \sum_{n=0}^{\infty} \Delta_n(R', R), \text{ where:} \tag{38}$$

$$\Delta_n(R', R) = \left\{ \begin{array}{lll} \Delta_n^2(R', R) & \equiv & R'k^n r^n - k^{n+1} r^{n+1} - (Rr^n - p_n^*(R)) \quad \forall\ p_n^*(R) \in \left[r^{n+1}, \frac{R}{R'}kr^{n+1}\right] \\ \Delta_n^3(R', R) & \equiv & \left(\frac{R'}{R}k^n - 1\right)(Rr^n - p_n^*(R)) \quad \forall\ p_n^*(R) \in \left(\frac{R}{R'}kr^{n+1}, Rr^n\right] \end{array} \right\}$$

Using $k > 1$, $R' > R$, and $p_n^*(R) \in [r^{n+1}, Rr^n]$ we have:

$$\Delta_n^3 \geq 0 \quad \text{and} \quad \Delta_n^2 \geq R'k^n r^n - k^{n+1} r^{n+1} - \left(Rr^n - r^{n+1}\right) \equiv \Delta_n^1(R', R) \tag{39}$$

Since $R' \geq r'$, we have $k/R' = [r'/(rR')]r \leq 1$; and thus, by Step 1 the difference $(R/R')kr^{n+1} - p_n^*(R)$ is upcrossing in n. Consequently, there are two possibilities, either $(R/R')kr^{n+1} \geq p_n^*(R)$ for all n, in which case:

$$\sum_{n=0}^{\infty} \Delta_n(R', R) = \sum_{n=0}^{\infty} \Delta_n^3(R', R) \geq 0 \quad \text{by} \;\; (39)$$

Alternatively, $\exists$ finite n^* s.t. $(R/R')kr^{n+1} \gtreqless p_n^*(R)$ as $n \gtreqless n^*$; and thus, by (39):

$$\begin{aligned}
\sum_{n=0}^{\infty} \Delta_n(R', R) &= \sum_{n=0}^{n^*-1} \Delta_n^3(R', R) + \sum_{n=n^*}^{\infty} \Delta_n^2(R', R) \geq \sum_{n=n^*}^{\infty} \left[R'k^n r^n - k^{n+1}r^{n+1} - Rr^n + r^{n+1}\right] \\
&= \sum_{n=n^*}^{\infty} \left[R'(r')^n - Rr^n + r^{n+1} - (r')^{n+1}\right] = \sum_{n=n^*}^{\infty} \left[(r')^n(R' - r') - r^n(R - r)\right] \\
&= \frac{1}{2} \sum_{n=n^*}^{\infty} \left[(r')^n(1 - r') - r^n(1 - r)\right] = \frac{1}{2} \left[\sum_{n=n^*}^{\infty} ((r')^n - r^n) - \sum_{n=n^*+1}^{\infty} ((r')^n - r^n)\right] \\
&= \frac{1}{2} \left[(r')^{n^*} - r^{n^*}\right] \geq 0 \quad \text{by } r' \geq r
\end{aligned}$$

where we have used $R' = (1 + r')/2$ and $R = (1 + r)/2$ in the third line.

Step 3. *Between cluster mobility increases in r.*

By definition between cluster mobility under r is:

$$m_B(r) = \sum_{n=0}^{\infty} \int_{r^{n+1}}^{r^n} \mathbb{1}_{\tau(p|r^n) \leq \tau} dp$$

Use $k \equiv r'/r$ and change of variable $p = k^{-n}q$ to get:

$$m_B(r') = \sum_{n=0}^{\infty} \int_{(r')^{n+1}}^{(r')^n} \mathbb{1}_{\tau(q|(r')^n) \leq \tau} dq = \sum_{n=0}^{\infty} \int_{kr^{n+1}}^{r^n} \mathbb{1}_{\tau(p|r^n) \leq \tau} dp$$

Then we have:

$$m_B(r') - m_B(r) = \sum_{n=0}^{\infty} \Delta_n(1, 1) \tag{40}$$

Since $kr = r' \leq 1$, Step 1 implies that $kr^{n+1} - p_n^*(1)$ is upcrossing in n. Thus, either $kr^{n+1} \geq p_n^*(1)$ for all n, in which case $\sum_{n=0}^{\infty} \Delta_n(1, 1) = \sum_{n=0}^{\infty} \Delta_n^3(1, 1) \geq 0$, or there exists finite n^* s.t. $kr^{n+1} \gtreqless p_n^*(1)$ as $n \gtreqless n^*$; and thus, by (39):

$$\begin{aligned}\sum_{n=0}^{\infty} \Delta_n(1,1) &= \sum_{n=0}^{n^*-1} \Delta_n^3(1,1) + \sum_{n=n^*}^{\infty} \Delta_n^2(1,1) \geq \sum_{n=n^*}^{\infty} \Delta_n^1(1,1) \\ &= \sum_{n=n^*}^{\infty} \left[k^n(r^n - kr^{n+1}) - (r^n - r^{n+1})\right] = \sum_{n=n^*}^{\infty} [k^n r^n - r^n] - \sum_{n=n^*+1}^{\infty} [k^n r^n - r^n] \\ &= k^{n^*} r^{n^*} - r^{n^*} \geq 0 \quad \text{by } k \geq 1\end{aligned}$$

References

Anderson, A., and L. Smith (2010): "Dynamic Matching and Evolving Reputations," *Review of Economic Studies*, 77(5), 3–29.

——— (2021): "The Comparative Statics of Sorting," working paper.

Becker, G. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–846.

Burdett, K., and M. G. Coles (1997): "Marriage and Class," *Quarterly Journal of Economics*, 112(1), 141–168.

Chiappori, P.-A., R. McCann, and L. Nesheim (2010): "Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness," *Economic Theory*, 42, 317–354.

Eeckhout, J., and P. Kircher (2018): "Assortative Matching with Large Firms," *Econometrica*, 86(1), 85–132.

Gavilan, A. (2012): "Wage Inequality, Segregation by Skill and the Price of Capital in an Assignment Model," *European Economic Review*, 56(1), 116–137.

Gretsky, N., J. Ostroy, and W. Zame (1992): "The Nonatomic Assignment Model," *Economic Theory*, 2, 102–127.

Håkanson, C., E. Lindqvist, and J. Vlachos (2021): "Firms and Skills: The Evolution of Worker Sorting," *Journal of Human Resources*, 56(2), 512–538.

Heinonen, J. (2005): "Lectures on Lipschitz Analysis," No. 100. University of Jyväskylä.

Kambourov, G., and I. Manovskii (2008): "Rising Occupational and Industry Mobility in the United States: 1968–97," *International Economic Review*, 49(1), 41–79.

——— (2009): "Occupational Mobility and Wage Inequality," *The Review of Economic Studies*, 76(2), 731–759.

Kremer, M., and E. Maskin (1996): "Wage Inequality and Segregation by Skill," NBER Working Paper.

Legros, P., and A. Newman (2002): "Monotone Matching in Perfect and Imperfect Worlds," *The Review of Economic Studies*, 69(4), 925–942.

Li, H., and W. Suen (2001): "Technological Change and Labor Market Segregation," Working Paper.

Mak, E., and A. Siow (2018): "Occupational Choice, Matching and Earnings Inequality," Working Paper.

McCann, R., and M. Trokhimtchouk (2010): "Optimal Partition of a Large Labor Force into Working Pairs," *Economic theory*, 42(2), 375–395.

Mifflin, R. (1977): "Semismooth and Semiconvex Functions in Constrained Optimization," *Siam Journal of Control and Optimization*, 15(6), 959–972.

Smith, L. (2006): "The Marriage Model with Search Frictions," *Journal of Political Economy*, 114(6), 1124–1144.

Song, J., D. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019): "Firming up Inequality," *The Quarterly Journal of Economics*, 134(1), 1–50.

Topkis, D. (1998): *Supermodularity and Complementarity.* Princeton University Press, Princeton, NJ.

Villani, C. (2008): *Optimal Transport: Old and New.* Springer Science & Business Media, New York, NY, first edn.