A Dynamic Generalization of Becker’s Assortative Matching Result*

Axel Anderson†

Georgetown University
Economics Department
Washington, DC 20057

(March 2015)

Abstract

This paper considers a dynamic matching model in which each agent’s future productivity depends in part on their current match, as in labor markets, schooling, intergenerational marriage markets, and other environments. The Planner’s endogenous rankings of human distributions are characterized. These Planner rankings are then used to develop sufficient conditions for positive assortative matching to be dynamically efficient. One lesson that emerges is that complementarity assumptions alone are insufficient for a robust sorting theory — the curvature of the static production function is also critical to determine optimal sorting patterns. In addition, the Planner’s ranking of distributions over human capital yield characterizations of individual attitudes toward human capital gambles in an associated market equilibrium. Finally, the implied dynamics for (1) individual wages and (2) wage distributions across age cohorts are characterized.

---


†Email: aza@georgetown.edu
1 Introduction

In 1973, Becker introduced a pairwise matching model with type-dependent perfectly divisible match output. Becker showed that if productive types are complements, then equilibrium sorting is positive: the highest types match together, then the next highest, etc. In the decades since, Becker’s matching model has been embellished in many ways and fruitfully applied to study sorting in marriage markets, job markets, schools, and neighborhoods. His original theoretical prediction has proven to be quite robust: given sufficient complementarity assumptions on the match production function, positive sorting obtains across a wide spectrum of related matching models.\footnote{The definition of sorting is context-dependent. Chade (2005) provides a definition suitable for a noisy environment. See Legros and Newman (2002) for several alternative notions of positive sorting.}

Anderson and Smith (2010) consider an infinite horizon matching model with peer effects, allowing the evolution of individual agent characteristics to depend on their match partners. Intuitively, workers may learn how to produce more efficiently from star coworkers or pick up bad habits from less productive partners. Anderson and Smith (2010) establish existence and the welfare theorems, but they have little to say about sorting patterns. In fact, their main matching result is that sorting cannot obtain with sufficiently patient agents in an incomplete information special case of their general model. In contrast, one contribution of the current paper is characterizing when perfect sorting obtains with dynamic matching and peer effects.

In static matching models, the match value function is exogenous. Here the value to any match depends on both the static production function and continuation values. These continuation values depends on the human capital transition function and the equilibrium value of human capital. Thus, the equilibrium value of human capital must first be characterized before analyzing sorting incentives. Given the welfare theorems, we characterize values using the Planner’s problem. Specifically, the Planner chooses the distribution over matches in each period to maximize the present discounted value of aggregate output, taking the current distribution over human capital as given. In doing so, the Planner may trade off lower static production in the current period for a better distribution over human capital in the future. But what is a better distribution over human capital? Lemma 1 provides an answer. An immediate implication of the scaling of productive types (higher is better) is that first order shifts in the human cap-
ital distribution increase discounted aggregate output (Lemma 1: (a)). The Planner’s ranking of second order changes in the distribution of human capital is more nuanced and depends on both the curvature of the production function and stochastic transitions in human capital (Lemma 1: (b)). For example, the Lemma offers conditions under which mean-preserving spreads in the distribution over human capital reduce total discounted output: there need not be a conflict between efficiency and equity when it comes to human capital distributions.\footnote{There is a small literature on efficient macroeconomic inequality. In Atkeson and Lucas (1992) inequality makes private information revelation incentive compatible. Welch (1999) is mostly an empirical piece that draws a distinction between unequal outcomes and unequal opportunity, arguing that unequal outcomes can provide incentives for effort, education, etc. Eeckhout (2006) considers a matching framework with an observable but inessential parameter. Equilibria in which sorting is conditioned on this inessential parameter can payoff dominate equilibria in which sorting is blind to the parameter. Again, inequality is efficient.}

Having characterized values, we turn to characterizations of sorting. In static matching models, sorting follows from complementarity (supermodularity) of the production function. In the dynamic model, we must consider both static and dynamic complementarity. Theorem 1 provides the transition complementarity assumption sufficient to imply positive sorting for any static production function that is both increasing and supermodular. However, the subsequent discussion underscores that the required complementarity condition on the transition kernel is extremely strong, unlikely to be satisfied in most economic environments. Recognizing this, Theorem 2 incorporates curvature assumptions on static output, which allow for significant weakening of the transition complementarity assumptions. Taken together, these results suggest that when peer effects are important, sorting outcomes depend on the curvature of static production, not on complementarity alone.

The analysis focuses on steady states\footnote{In fact, the results in this paper hold in and out of steady state. I focus on steady state as this streamlines the notation and is a common restriction in applications.} in which the aggregate distribution over human capital is constant. Nonetheless, the model allows for rich dynamics at the level of the individual and across age cohorts. Corollary 1 characterizes the dynamics for individual wages implied by Theorem 2, while Theorem 3 characterizes the dynamics in the wage distribution by age cohort. If the noise in transition dynamics is type independent (a strong assumption often made in empirical applications), then Corollary 2 provides conditions such that wage inequality within a cohort rises in age.
There are many environments with endogenous group formation in which people are changed by their associates. In Section 7, I consider intergenerational marital sorting, internal and external labor markets, schooling, and peer effects in neighborhoods, in each case showing how the model assumptions translate to the particular environment.

**Additional Related Literature:** Jovanovic (1979) shares both matching and type evolution with the current work. The critical difference is that all heterogeneity in Jovanovic (1979) is match-specific: unemployed workers are homogenous. Jovanovic and Nyarko (1995) consider an overlapping generations model in which old and young match. The production process is uncertain, and older workers have private information about productivity from signals they have observed and any information revealed by older workers with whom they were matched when young. Contracts for information transmission are signed in the shadow of this asymmetric information problem. Although their model incorporates restrictions not assumed here, it allows for private information, and so is not a special case of the current model. Indeed, Jovanovic and Nyarko (1995) focuses on understanding the inefficiency caused by private information.

A more closely related OLG matching model is Jovanovic (2014). As in Jovanovic and Nyarko (1995), matching takes place between an old and young worker. However, as in the current paper, the productivity of an old worker depends (in part) on the type of their match partner when young. His aim is to characterize balanced growth paths, while here the focus is on steady state. In addition, for his general model he characterizes positive sorting equilibria assuming such an equilibrium exists, and constructs a positive sorting equilibrium for a particular case of his model. In contrast, my focus is on establishing general conditions that support positive sorting as an equilibrium. In Section 8, I compare the results in the current paper to Jovanovic (2014) in detail.

Moscarini (2005) incorporates Jovanovic’s match-specific productivity in a search framework. As in any model with match-specific quality and endogenous match dissolution, over time worker-firms sort into quality matches. Quality matches generate higher rents, which are shared by workers, and so ultimately wage inequality owes to ex post sorting. Moscarini (2005) inherits the ex ante homogeneity of Jovanovic (1979).

Eeckhout and Wang (2010) consider a dynamic continuum firm-worker matching model in which firms with known productivity match with workers with unknown
productivity. In particular, all workers are either high or low productivity, but all market participants merely know the probability that a worker is high productivity (the worker’s reputation). Matched firm-worker output follows a publicly observed Brownian motion with drift that depends on the firm-worker match obscured by independent noise. In this one-sided learning model, supermodularity of the drift term is necessary and sufficient to generate positive assortative matching on firm type and worker reputation. Papageorgiou (2014) also studies a matching model in which workers learn their own types by observing Brownian output with drift that depends on match type and noise that depends on firm (occupation) type. One salient difference from the current work is that Papageorgiou (2014) is a model of horizontal differentiation of workers: that is, output is not monotone in worker types across firms (occupations). A second substantial difference is that Papageorgiou (2014) considers search frictions.

Work on comparative dynamics in dynamic models with complementarities may be the most technically related to the current work. For example, Amir, Mirman, and Perkins (1991) characterize the planner’s problem in a one-sector growth model, including comparative dynamic results, while Amir (1996) provides comparative dynamics for a multi-sector growth model.

2 The Model

As in Anderson and Smith (2010), consider a discrete time, infinite horizon, dynamic partnership model with a continuum of heterogenous agents. At the start of every period $t$ each agent in the market is described by his scalar human capital, $x_t \in [0, 1]$, while the current state of the economy is defined by a cdf $H_t$ over human capital.

The distribution over matches in a given period is described by a measure $\mu_t$ on the product space $[0, 1]^2$. This allows for multivalued matching in which each type of agent does not have a unique partner type. We say that $x$ and $y$ are matched in period $t$ if $(x, y)$ lies in the joint support of $\mu_t$. Positive assortative matching (PAM) obtains iff $\mu_t(x, y) = H_t(x) \wedge H_t(y)$, i.e. the support of $\mu_t$ is the 45 degree line.

Of course, the matching must be feasible given $H_t$. Toward formally defining this feasibility constraint, let $\lambda_H(A)$ be the measure of any measurable set $A \subset [0, 1]$ asso-

\footnote{The basic insights can be extended to a mode with defined sides at some notational cost.
associated with the CDF $H$, let $Z$ be the space of cdfs on $[0, 1]$, and let $M$ be the space of bounded measures on $[0, 1]^2$. Then the feasibility constraint $\Phi : Z \Rightarrow M$ is the correspondence from distributions $H \in Z$ into matching measures $\mu \in M$ given by:

$$\Phi(H) = \{ \mu \in M : \lambda_H(A) = \mu(A \times [0, 1]) \forall A \text{ measurable} \}.$$  \hspace{1cm} (1)

This will be the only constraint placed on matching agents in a given period. For example, there are no costs associated with breaking up prior matches.

Within each period all matched pairs produce output that depends only on the human capital of each partner, $Q : [0, 1]^2 \mapsto R_+$. As agents are risk neutral, one may interpret $Q$ as deterministic or as expected output. I assume $Q$ is bounded, continuously increasing in each argument, and supermodular (SPM). Given matching $\mu_t$, total production in a given period is $\int_{[0,1]^2} Q(x, y) d\mu_t(x, y) \equiv \int Q d\mu_t$.

Each agent’s future human capital depends on her own human capital and the human capital of her partner. Specifically, if $x$ and $y$ match in a period, the transition cdf $T(s|x, y)$ gives the chance that $x$ updates to at most $s$ before the start of the next period. I assume $T(s|x, y)$ is monotonically decreasing in $x$ and in $y$. This captures two ideas: human capital is persistent, and being matched with better partners improves one’s future human capital.

Agents may discount the future for two separate reasons: they are mortal and they prefer immediate to delayed gratification. The discount rate $\delta > 0$ captures inherent time preferences, while $\sigma > 0$ is the probability that an agent survives between periods. I require $\sigma < 1$ to ensure a steady state exists, and $\delta < 1$ to ensure a bounded Planner value. Altogether, the implicit discount rate is $\beta \equiv \delta \sigma \in (0, 1)$.

At the start of every period a new cohort is born with cdf $H^0$ and mass $1 - \sigma$: Thus, the size of the market is constant and normalized to one. While the model allows for stochastic transitions at the micro level, the aggregate dynamics are deterministic. Specifically, given any matching $\mu_t$, $\mathcal{H}(\mu_t)$ gives the updated distribution over human capital among all survivors at the start of period $t + 1$:

$$\mathcal{H}(\mu_t)(s) \equiv \int_{[0,1]^2} T(s|x, y) d\mu_t(x, y).$$  \hspace{1cm} (2)
Altogether, for any matching $\mu_t$, the new distribution over human capital is:

$$H_{t+1} = (1 - \sigma)H^0 + \sigma\mathcal{H}(\mu_t). \quad (3)$$

To summarize the within-period timing: a new cohort enters, matches are formed, output is realized, human capital transitions, and deaths occur.

The Planner maximizes discounted production: $\sum_{t=0}^{\infty} \delta^t \int Qd\mu_t$, subject to feasibility constraint (1), and human capital dynamics (3). Anderson and Smith (2010) establish that this dynamic problem is well-defined: The planner’s discounted optimization problem can be replaced with the static Bellman representation:

$$V(H) = \max_{\mu \in \Phi(H)} \left[ (1 - \beta) \int Qd\mu + \delta V \left( (1 - \sigma)H^0 + \sigma\mathcal{H}(\mu) \right) \right]. \quad (4)$$

More precisely, defining $v(x)$ as the shadow value to the planner of relaxing the feasibility constraint at type $x$, so that $V(H) = \int v(x)dH(x)$. Then a Steady State Pareto Optimum is a 4-tuple $(V, v, \mu, H)$, such that:

1. $(V, v)$ satisfies the Bellman Equation (4),
2. $\mu$ solves the Bellman Equation (4) given $(V, H)$, and
3. Steady state obtains: $H = (1 - \sigma)H^0 + \mathcal{H}(\mu)$.

The argument of the Planner’s value function is a distribution. Thus, to characterize this value function we must consider partial orders on distributions. In particular, the following commonly used partial orders on cdfs are used throughout the paper:

- **First Order Stochastic Dominance**: $\hat{H} \succ_{FSD} H$ iff $H(s) \geq \hat{H}(s) \forall s$.
- **Increasing Concave Order**: $\hat{H} \succ_{ICV} H$ iff $\int_0^z H(s)ds \geq \int_0^z \hat{H}(s)ds \forall z$.
- **Increasing Convex Order**: $\hat{H} \succ_{ICX} H$ iff $\int_z^1 (1 - \hat{H}(s))ds \geq \int_z^1 (1 - H(s))ds \forall z$.
- **Convex Order**: $\hat{H} \succ_{CX} H$ iff $\hat{H} \succ_{ICX} H$ and $\int_0^1 sd\hat{H}(s)ds = \int_0^1 sdH(s)ds$.
  (i.e. $\hat{H}$ is a Mean-Preserving Spread of $H$)
3 Motivational Two Period Model

To help motivate the more general theory, first consider a two period model. Given SPM static output \( Q \), PAM is optimal in the final period; thus, the Planner’s continuation value given any first period matching \( \mu \) is:

\[
V^0(\mu) = \frac{1}{2} \int Q(x, x) d\mathcal{H}(\mu)(x).
\]  

(5)

For sufficient relative weight on second period output, PAM is optimal in the first period if and only if it maximizes (5) across all feasible matchings. PAM maximizes (5) for all increasing \( Q(x, x) \) if and only if the continuation distribution \( \mathcal{H} \) under PAM first order stochastically dominates the continuation distribution for all other feasible matchings. Equivalently, PAM solves \( \min_{\mu \in \Phi(H^0)} \mathcal{H}(\mu)(s) \) for all \( s \). Since the continuation distribution (2) is an average of \( \mathcal{T}(s|\cdot) \) over \( \mu \), PAM minimizes \( \mathcal{H}(\mu)(s) \) for all initial distributions \( H^0 \) if and only if \( \mathcal{T}(s|x, y) \) is SBM in \( (x, y) \) for all \( s \) by Lorentz (1953). Altogether, the robust condition for PAM in the first period given \( Q(x, x) \) increasing is \( \mathcal{T}(s|x, y) \) SBM in \( (x, y) \) for all \( s \).

I argue in Section 5 that SBM of \( \mathcal{T}(s|\cdot) \) for all \( s \) is a very strong condition, unlikely to be met in many practical applications. In order to loosen the restriction on transition dynamics, one must further restrict the static production function. For example, PAM will maximize the continuation value (5) for all increasing and convex \( Q(x, x) \), if and only if it maximizes the continuation distribution \( \mathcal{H}(\mu) \) in the increasing convex order. Equivalently, PAM must maximize:

\[
\int_z^1 (1 - \mathcal{H}(\mu)(s)) ds = \int_z^1 \left[ 1 - \int \mathcal{T}(s|x, y) d\mu(x, y) \right] ds = 1 - z - \int \left[ \int_z^1 \mathcal{T}(s|x, y) ds \right] d\mu(x, y)
\]

for all \( z \) across all feasible matchings. For this to hold for all \( H^0 \) we need \( \int_z^1 \mathcal{T}(s|x, y) ds \) SBM in \( (x, y) \) for all \( z \). If, instead, \( Q(x, x) \) were increasing and concave, then we would require that PAM maximize \( \mathcal{H}(\mu) \) in the increasing concave order.

In the two period model, the final period value function \( V^0 \) is exogenous. In order to develop robust sufficient conditions for PAM in the infinite horizon model, we now characterize the Planner’s endogenous value \( V \) of human capital distributions.
4 Value Characterization

The Planner chooses a matching to maximize a linear combination of flow payoffs and continuation values. Lorentz (1953) showed that $Q$ SPM, implies that PAM is \textit{statically efficient}, i.e. solves $\max_{\mu \in \Phi(H)} \int Qd\mu$, while Becker (1973) also showed that PAM is a static market outcome with transferable utility and $Q$ SPM. To this static problem dynamics in human capital have been added, and we seek to understand \textit{dynamic efficiency}, i.e. what maximizes continuation values:

$$\max_{\mu \in \Phi(H)} V \left( (1 - \sigma)H^0 + \sigma\mathcal{H}(\mu) \right).$$

Unlike the static problem, the answer depends on an endogenous value function $V$ which ranks distributions $H$, as well as the transition function $T$, which maps from matching choices $\mu$ to continuation distributions $\mathcal{H}(\mu)$. Thus, $V$ and $T$ jointly determine an endogenous ranking over matchings $\mu \in \Phi(H)$ via the composite map $V(\mathcal{H}(\mu))$.

Call a function of two variables bi-convex (bi-concave) if it is convex (concave) in both variables individually. The following Lemma characterizes $V$.

\textbf{Lemma 1 (Value Characterization)} We have:

(a) $V(\hat{H}) \geq V(H)$ for all $\hat{H} \succ_{FSD} H$ and $V$ is increasing; and

(b) If $Q$ is bi-concave and $\int_0^z T(s|x, y)ds$ is bi-convex in $x$ and $y \forall z$, then $V(\hat{H}) \geq V(H)$ for all $\hat{H} \succ_{ICV} H$ and $V$ is concave. If $Q$ is bi-convex and $\int_0^1 Tds$ is bi-concave, then $V(\hat{H}) \geq V(H)$ for all $\hat{H} \succ_{ICX} H$ and $V$ is convex.

An immediate corollary to Lemma 1 is that the Planner dislikes (likes) mean-preserving spreads in $H$ when $Q$ bi-concave (bi-convex) and $\int_0^z Tds$ is bi-convex (bi-concave).\footnote{Anderson and Smith (2010) establish $V$ increasing in mean-preserving spreads given stronger assumptions: $Q$ bi-linear, with $\int_0^z T(s|x, y)ds$ bi-concave and convex along the diagonal $y = x$.}

The proof (in the Appendix) exploits the fact that the Planner’s problem can be written as a recursive contraction mapping $B$ on the space $V$ of bounded, continuous, and homogenous functions defined on cdfs. Thus, the unique fixed point $BV = V$ is the limit of any sequence $U^{N+1} = BU^N$, starting from any arbitrary $U^0 \in V$. Thus, if $BU$ inherits a given property from any arbitrary $U$ (ex. increasing in FSD shifts in $H$), we can conclude that the fixed point $V$ also has this property. The key to the
verification step is showing that if $\hat{H}$ dominates $H$ in the given order, then for any matching $\mu$ feasible for $H$, we can construct a matching $\hat{\mu}$ feasible for $\hat{H}$ that yields both a higher static payoff and a higher continuation distribution in the assumed order.

Having characterized $V$, properties of the Planner’s shadow value function $v$ follow from the identity $V(H) = \int v(x)dH(x)$ (from the Planner’s dual minimization problem). Intuitively, the shadow value function $v$ must be increasing and concave (convex) when $V$ rises in the ICV order (increasing convex order).

Both $Q$ and $T$ are monotonic. That these assumptions imply $V$ increasing in FSD shifts is not surprising. That the Planner’s attitude toward risk follows from curvature assumptions on static output, and that concave static output is associated with a concave $v$ and an aversion to mean-preserving spreads, is also not surprising. But notice that I only require bi-concavity/convexity, and not the usual strong requirement that these bi-variate functions be convex or concave.

As the ICV (ICX) order compares integrals of human capital cdfs, it is the integral of the continuation cdf that is relevant. At first glance it may appear odd that the convexity (concavity) assumptions on $\int T ds$ are associated with concavity (convexity) assumptions on $v$. However, note that the shadow value $v(x)$ captures the marginal value to the Planner of relaxing the feasibility constraint, i.e. adding more of type $x$. Thus, if this shadow value function is concave, the Planner’s value will fall if we spread out the mass at $x$, keeping the mean constant. When $\int_0^x T(s)\cdot ds$ is bi-convex, spreading $x$ causes this integral to rise, which is associated with a lower continuation distribution precisely when continuation distributions are ordered by ICV.

For further intuition, define the expected shadow continuation value for a type $x$ matched to a type $y$:

$$\psi(x|y) \equiv E[v(x')|x, y] = \int v(s)T(ds|x, y)$$

and the sum $\Psi(x, y) \equiv \psi(x|y) + \psi(y|x)$, then the solution to (4) must obey the complementary slackness conditions:

$$v(x) + v(y) \geq (1 - \beta)Q(x, y) + \beta\Psi(x, y) \quad (6)$$

$$(x, y) \in \text{supp}(\mu) \Rightarrow v(x) + v(y) = (1 - \beta)Q(x, y) + \beta\Psi(x, y). \quad (7)$$
By inspection, these complementary slackness conditions require:

$$v(x) = \max_y [(1 - \beta)Q(x, y) + \beta \Psi(x, y) - v(y)]. \quad (8)$$

If this objective function is convex in $x$, then shadow values so defined will be convex.

For example, consider deterministic transitions. Specifically, if $x$ and $y$ match, then $x$ transitions to $f(x, y)$ and $y$ transitions to $f(y, x)$, where $f$ is assumed weakly increasing in both arguments. Given this specification for transitions, we have $E[v(x')|x, y] \equiv \Psi(x, y) = v(f(x, y)) + v(f(y, x))$. Twice differentiating the objective function in (8), we find: $$(1 - \beta)Q_{11}(x, y) + \Psi_{11}(x|y),$$

where:

$$\Psi_{11}(x, y) = v''(f(x, y))f_2(x, y)^2 + v''(f(y, x))f_1(y, x)^2 + v'(f(x, y))f_22(x, y) + v'(f(y, x))f_{11}(y, x).$$

Thus, if $Q$ and $f$ are bi-convex, the objective function in (8) will be convex in $x$ when $v$ is convex.\footnote{While maximization (8) is useful for providing intuition, using it to characterize $v$ is problematic. In particular, the maximization (8) does not define a contraction when $\beta > 1/2.$}

To map back to the results in Lemma 1, note that in the deterministic case:

$$\int_z^1 T(s|x, y)ds = 1 - \max\{z, f(x, y)\}. \quad (9)$$

As convexity is preserved by the max operator, $f$ bi-convex implies $\int_z^1 T(s|\cdot)ds$ bi-concave. Altogether, $Q$ and $f$ bi-convex imply $v$ convex and $V$ increasing in the ICX order. For the lower integral, we have $\int_0^z T(s|x, y)ds = \max\{0, z - f(x, y)\}$ and $f$ bi-concave implies $\int_0^z Tds$ is bi-convex.

### 5 Sufficient Conditions for Sorting

I now consider sufficient conditions for positive assortative matching (PAM) without making curvature assumptions on the convexity or concavity of static output.

**Theorem 1** If $T(s|x, y)$ is submodular (SBM) in $(x, y)$ for all $s$, then PAM is optimal.

The logic echoes that in Section 3. The mass in any lower tail of the continuation distribution, $\mathcal{H}(\mu)(s)$ is $\int T(s|x, y)d\mu(x, y)$. Thus, $T(s|\cdot)$ SBM for all $s$ implies that
PAM minimizes $\mathcal{H}(\mu)(s)$ for all $s$, which implies that PAM maximizes continuation values, since $V$ increases in FSD shifts by Lemma 1. Thus, $\mathcal{T}$ SBM is the natural analogue of the static complementarity assumption, absent further restrictions on $Q$.

A deterministic transition function satisfying this assumption is $f(x, y) = \min\{x, y\}$. Since the Planner’s shadow value function $v$ is increasing by Lemma 1, the sum of shadow continuation values is $2v(\min\{x, y\})$ SPM since:

$$v(\min\{x^{H}, y^{H}\}) + v(\min\{x^{L}, y^{L}\}) \geq v(\min\{x^{H}, y^{L}\}) + v(\min\{x^{L}, y^{H}\}),$$

for all $x^{L} \leq x^{H}, y^{L} \leq y^{H}$. Thus, PAM is statically and dynamically optimal.\(^7\)

While complementarity in production seems reasonable, SBM of $\mathcal{T}$ is actually quite demanding. For example, restrict attention to deterministic transition functions, $f$, and thus $\mathcal{T}(s|x, y) = \mathbb{1}_{s \geq f(x, y)}$. In this case, the following two natural conditions jointly rule out $\mathcal{T}$ SBM: (a) $f(y, x) \geq f(x, y)$ for all $y \geq x$ (i.e. peer effects on human capital accumulation no stronger than own effects) and (2) $f_2 > 0$ (i.e. strictly monotone peer effects). For then $y > x$ implies $f(y, x) > f(y, x) \geq f(x, y) > f(x, x)$; and thus, $1 = \mathbb{1}_{s \geq f(y, y)} + \mathbb{1}_{s \geq f(x, x)} > \mathbb{1}_{s \geq f(x, y)} + \mathbb{1}_{s \geq f(y, x)} = 0$ for $s \in (f(x, x), f(x, y))$.

Having argued that $\mathcal{T}$ SBM is a strong assumption, the next theorem strengthens the assumptions on static output, which allows me to weaken the assumptions on the transition function $\mathcal{T}$.

**Theorem 2** PAM is optimal provided either:

(a) $\int_{z}^{1} \mathcal{T}(s|x)ds$ is SBM, $\int_{z}^{1} \mathcal{T}(s|x, x)ds$ is concave in $x$, and $Q(x, x)$ is convex; or,

(b) $\int_{z}^{1} \mathcal{T}(s|x)ds$ is SBM and either (i) $Q(x, x)$ is concave and $\int_{0}^{x} \mathcal{T}(s|x, x)ds$ is convex in $x$, or (ii) $Q(x, y)$ is bi-concave and $\int_{0}^{x} \mathcal{T}(s|x, y)ds$ is bi-convex in $x$ and $y$.

For an intuition for the dynamic SBM conditions, assume that $V$ rises in ICV shifts in $H$. Then when $\int_{z}^{1} \mathcal{T}(s|x)ds$ is SBM, PAM minimizes $\int_{0}^{z} \mathcal{H}(\mu)(s)ds$ for all $z$, and thus maximizes continuation values $V(\mathcal{H}(\mu))$. As $Q$ is SPM, PAM separately maximizes

\(^7\)Lucas and Moll (2011) consider a growth model with heterogeneous agents and peer effects. Specifically, agents divide their time between producing and searching for other agents. Search is random, and all meetings result in the lower type taking on the human capital of the higher type, i.e. they assume max transitions. Although they do not allow for partnership choice, in the frictionless version of their model sorting on types minimizes continuation values.
static payoffs and continuation values. Similarly, when $V$ rises in the ICX order, PAM maximizes continuation values provided $\int_z^1 \mathcal{T}(s|\cdot)ds$ is SBM for all $z$.

The two parts of the proof (in the Appendix) differ in their approach. Part (b.ii) makes use of Lemma 1, and thus adds a complementarity assumption on transitions to the curvature assumptions used in Lemma 1. Parts (a) and (b.i) replace global curvature assumptions with local curvature assumptions. For example, $Q(x,x)$ convex does not require global bi-convexity of $Q$. The proof using local curvature assumptions is constructive. For concreteness, consider part (a). Assume $V$ is increasing in the ICX order. Then, given $\int_z^1 \mathcal{T}(s|\cdot)ds$ SBM, PAM minimizes $\int_z^1 \mathcal{H}(\mu)(s)ds$ and thus maximizes continuation values $V(\mathcal{H}(\mu))$. Thus, PAM is efficient. Finally, imposing PAM implies a particular Planner value function, which we can show increases in the ICX order given the posited curvature assumptions on $Q(x,x)$ and $\int_z^1 \mathcal{T}(s|x,x)ds$.

**Example: Deterministic Transitions.** Considering deterministic transitions provides intuition for Theorem 2. Recall the shadow Bellman equation defined via the Planner’s complementary slackness conditions (8). Intuitively, PAM will be optimal, provided the associated objective function is SPM. As static output is SPM, complementarity of the shadow continuation value function $\Psi(x,y) \equiv \psi(x|y) + \psi(y|x)$ is sufficient for a SPM objective function. Assuming $v$ twice differentiable, we have

$$
\psi_{12}(x|y) = v'(f(x,y))f_{12} + v''(f(x,y))f_1(x,y)f_2(x,y).
$$

We have assumed $f_1f_2 \geq 0$ and know $v' > 0$ by Lemma 1, so transition complementarity is reinforced by $f$ SPM and $v''$ convex. Recalling (9), $f$ SPM ensures $\int_z^1 \mathcal{T}(s|\cdot)ds$ SBM, since supermodularity is preserved by the max operator. We can ensure convexity of $v$ by assuming $Q$ and $f$ bi-convex (as $f$ bi-convex implies $\int_z^1 \mathcal{T}(s|\cdot)ds$ bi-concave), or alternatively we may assume $Q(x,x)$ and $f(x,x)$ convex.8

**Example: Symmetric Beta Transitions.** Let $\eta \in [0,1]$ be distributed according to a beta distribution with matching shape and scale parameters (i.e. a symmetric density around $\eta = 1/2$). WLOG assume $x \leq y$ and let $b(\cdot|x,y)$ be the density of the

8Anderson and Smith (2010) consider sorting for the extreme case of $f(x,y) = \varphi x + (1 - \varphi)y$ for $\varphi \in [0,1]$. As we see in (10), sorting is dynamically efficient iff the shadow value function $v$ is convex.
linearly scaled variable \( x + (y - x)\eta \in [x, y] \), and set transition density \( T_s(s|x, y) = 2b(s|x, y) = [(s-x)(y-s)]^{\alpha-1}(y-x)^{-2\alpha} \), where \( \alpha \geq 0 \) defines the particular symmetric beta distribution. In this class of example, lower values of \( \alpha \) imply types are more persistent, and in the limit \( (\alpha \to 0) \) we recover static types (Becker). We cannot have \( T \) SBM for all \( s \), since \( T_{xy}(s|x, y) \geq 0 \) as \( s \leq (x+y)/2 \).

However, since the expectation of \( x + y \) is linear in \( x \) and \( y \) by construction, so is \( \int_0^1 T(s|\cdot)ds \). Altogether, \( \int_z^1 T(s|\cdot)ds \) is SBM and bi-concave, and so satisfies the premise of Theorem 2 part (a).

6 Dynamics

We have imposed steady state, ruling out aggregate dynamics by assumption. Nonetheless, the model allows for rich micro (individual) dynamics and cohort level dynamics. Since human capital is typically not (fully) observable, I focus on wage dynamics. Toward that end, I first define a market for partners, the associated wages, and establish a decentralization result. Anticipating this decentralization result, overuse notation and define the present discounted value of market wages for a type \( x \) as \( v(x) \). These values must satisfy an accounting identity: The sum of values for any matched pair must equal the \((1-\beta, \beta)\) weighted average of static output and the sum of expected continuation values for that pair. That is, \((7)\) must hold in the market for partners.

Now impose the standard stability condition in bilateral matching markets. A matching \( \mu \) with values \( v \) is immune to bilateral deviations \( \text{iff} \) inequality \((6)\) holds, i.e. no pair can achieve a higher joint value by severing their current assignment and matching with each other instead. And since precisely two agents are required for production, immunity to bilateral deviations is all that is required for stability. Altogether a Steady State Market Equilibrium is a triple \((v, \mu, H)\), where: (i) \((v, \mu)\) is stable \((6)\); (ii) \((v, \mu)\) obey value accounting \((7)\); (iii) \(\mu\) is feasible for \( H: \mu \in \Phi(H) \); and (iv) Steady State obtains: \( H = (1-\sigma)H^0 + \sigma\mathcal{H}(\mu) \). Trivially, the Steady State Pareto Optimum satisfies \((i) \rightarrow (iv)\). Thus, for any steady state equilibrium \((v, \mu, H)\), we can

---

9Direct computation yields \( T_{xy}(s|x, y) \) proportional to:

\[
\left(\frac{y+x-2s}{(y-x)^3}\right)\left(\frac{(s-x)(y-s)}{(x-y)^2}\right)^{\alpha-1}.
\]
define \( w(x|y) \), the equilibrium wage that \( x \) receives when matched to \( y \) as:

\[
(x, y) \in \text{supp}(\mu) \Rightarrow v(x) \equiv (1 - \beta)w(x|y) + \beta \psi(x|y).
\]

Trivially, when PAM obtains, all types receive equilibrium wage \( Q(x, x)/2 \).

Having defined a market wage, consider individual level dynamics. The focus will be on dynamics conditional on PAM. Thus, let \( [X|x] \) be the random variable associated with the PAM transition cdf, \( \mathcal{T}(\cdot|x, x) \), i.e. \( X \) is the random human capital next period for a type \( x \) matched to a type \( x \) this period. Similarly, let \( [W|w] = [W(X)|x = Q^{-1}(w, w)/2] \) be the associated random wage. An immediate implication of PAM is a strong wage persistence result. In particular, \( \mathcal{T}(s|x, x) \) is decreasing in \( x \); and thus, \( [X|x^H] \preceq_{FSD} [X|x^L] \) for all \( x^H \geq x^L \), which along with \( w(x) = Q(x, x)/2 \) increasing in \( x \) yields \( [W|w^H] \preceq_{FSD} [W|w^L] \) for all \( w^H \geq w^L \) and thus \( E[W|w] \) increasing in \( w \). Huggett, Ventura, and Yaron (2006) confirm this monotonicity in the data. In addition to this wage persistence result, the set of assumptions used to establish PAM via Theorem 2 has second order implications for micro wage dynamics.

**Corollary 1 (Micro Wage Dynamics)** We have:

(a) If the premise of Theorem 2 part (a) holds, then \( \lambda[W|w'] + (1 - \lambda)[W|w''] \succeq_{ICX} [W|w] \) for all \( w = \lambda w' + (1 - \lambda)w'' \) and \( E[W|w] \) is increasing and convex in \( w \).

(b) If the premise of Theorem 2 part (b.i) holds, then \( [W|w] \succeq_{ICV} \lambda[W|w'] + (1 - \lambda)[W|w''] \) for all \( w = \lambda w' + (1 - \lambda)w'' \) and \( E[W|w] \) is increasing and concave in \( w \).

**Proof:** Consider part (a). PAM obtains by Theorem 2. Trivially, \( \int_0^1 (1 - \mathcal{T}(s|x, x))ds \) convex for all \( z \) implies \( \lambda[X|x'] + (1 - \lambda)[X|x''] \succeq_{ICX} [X|x] \) for all \( x = \lambda x' + (1 - \lambda)x'' \). Combining this stochastic ordering on conditional human capital with \( Q(x, x)/2 \) increasing and convex yields \( \lambda[W|w'] + (1 - \lambda)[W|w''] \succeq_{ICX} [W|w] \) by Theorem 4.A.8 part (a) in Shaked and Shanthikumar (2007). As a result \( E[W|w] \) is convex in \( w \). Similar steps establish part (b).

Next consider the dynamics across wage distributions by (age) cohort given PAM. Let \( H^N \) denote the distribution over human capital for all survivors \( N \) periods since birth given PAM, i.e. \( H^N(s) = \int \mathcal{T}(s|x, x)dH^{N-1}(x) \) for all \( N \) with average human capital \( \bar{x}^N \). Then, since \( Q(x, x) \) is strictly increasing, the distribution over wages for
cohort $N$ is uniquely defined by $G^N(w) = H^N(Q^{-1}(2w, 2w))$. The following result characterizes the cohort level wage dynamics accompanying Theorem 2.10

**Theorem 3 (Cohort Wage Dynamics)** We have:

(a) Assume the premise of Theorem 2 part (a) is met and $T(x^0, x^0) \geq_{ICX} H^0$, then $G^{N+1} \succeq_{ICX} G^N$ for all $N$.

(b) Assume the premise of Theorem 2 part (b.i) is met and $H^1 \succeq_{ICV} H^0$, then $G^{N+1} \succeq_{ICV} G^N$ for all $N$.

**Example: Independent Noise.** In empirical applications, independent noise is frequently imposed. In the current context, assume $x_{t+1} = f(x_t, y_t, \xi)$ where $\xi$ is a random variable with cdf $F$, independent of $(x, y)$. Toward a strong dynamic inequality result for this independent noise case, consider the Lorentz order on distributions. Specifically, consider two random variables $X$ and $\hat{X}$ with distributions $H$ and $\hat{H}$, and let $H_0$ and $\hat{H}_0$ be the associated distributions for the normalized random variables $X/E[X]$ and $\hat{X}/E[\hat{X}]$, then we have $\hat{H} \succeq_{LZ} H$ iff $\hat{H}_0 \succeq_{CX} H_0$. The following result illustrates how the theory can be applied to the independent noise case.

**Corollary 2 (Independent Noise)** Assume $Q(x, x)$ convex, $f(x, y, \xi)$ SPM in $(x, y)$, and $f(x, x, \xi)$ convex in $x$, then PAM obtains and $E(W|w)$ is increasing and convex in $w$. If, in addition, $f(x, x, \xi)/x$ and $Q(x, x)/x$ are non-decreasing in $x$, then inequality rises by cohort age, i.e. $G^{N+1} \succeq_{LZ} G^N$ for all $N$.

**Proof:** Given this transition assumption we have:

$$\int_1^{T(s|x, y)ds} = 1 - \int \max\{z, f(x, y, \xi)\}dF(\xi)$$

which is concave along the diagonal $y = x$ as long as $f(x, x)$ is convex and SBM in $(x, y)$, provided $f$ is SPM in $(x, y)$ by Corollary 2.6.2 in Topkis (1998). Thus, by Theorem 2 part (a) PAM obtains, and $E(W|w)$ is increasing and convex in $w$ by Corollary 1.

---

10The stronger monotonicity result $G^{N+1} \succeq_{FSD} G^N$ for all $N$ holds provided $H^1 \succeq_{FSD} H^0$ (established via induction and Theorems 1.A.6 and 1.A.13 part (a) in Shaked and Shanthikumar (2007)).

One could simply assume $H^1 \succeq_{FSD} H^0$ (as in Hopenhayn (1992) Propositions 3 and 4), but this seems an unreasonable assumption in applications. For example, $H^1 \succeq H^0$ for all $H^0$ if $T(s|x, x) = 0$ for all $s < x$. Thus, we focus on monotonicity in the weaker ICV and ICX orders.

11Given normalization $Q(0, 0) = 0$, $Q(x, x)$ increasing and convex implies $Q(x, x)/x$ non-decreasing.
Given PAM and \( f(x, x, \varepsilon)/x \) non-decreasing in \( x \) for all \( \varepsilon \), Theorem 3.A.27 in Shaked and Shanthikumar (2007) yields \( H^{N+1} \succeq_{LZ} H^N \) for all \( N \). Given this order on distributions \( H^N \) and \( Q(x, x)/x \) non-decreasing, we also have \( G^{N+1} \succeq_{LZ} G^N \) for all \( N \) by Theorem 3.A.25 in Shaked and Shanthikumar (2007) as asserted. \( \Box \)

7 Environments

The model speaks to any situation with team production and endogenous group formation, in which group members meaningfully influence one another’s future characteristics. These criteria are salient in the following environments.

**Marital Sorting.** This first application is a literal dynamic extension of Becker (1973).\(^{12}\) A married couple with productivities \((x_t, y_t)\) produces expected lifetime earnings \( Q(x_t, y_t) \) and a pair of children with productivities \((x_{t+1}, y_{t+1})\). The intergenerational earnings mobility literature finds prima facie evidence for intergenerational transmission of labor market productivity: high (low) earning parents tend to have high (low) earning children.\(^{13}\) This correlation may be the result of genetic inheritance (intellect, health, etc.) and/or parental socialization (work ethic). In either case, the model captures this dependence by assuming the parent-to-child transition cdf \( T(\cdot|x, y) \) is first order increasing in the types of both parents \( x \) and \( y \).

The empirical literature typically assumes a direct relationship between parent and child earnings (often log-linear) without making assumptions about the underlying transmission mechanism. In contrast, Solon (2004) posits a model of human capital transmission in which children’s human capital depends both on inherited traits and parental investment in their childrens' human capital (a choice variable). Given optimal parental investment choices, child human capital \( f(x, y, \varepsilon) \) is linear in parental human capital \((x, y)\) with additive noise. Trivially, \( f(x, y, \varepsilon) \) is SPM in \((x, y)\) and \( f(x, x, \varepsilon) \) is convex in \( x \). If the lifetime earnings of couples \( Q(x, y) \) is SPM and convex.

\(^{12}\)Search and informational frictions are surely of first order importance in marriage markets. As with Becker’s original matching model, this dynamic generalization should be considered a benchmark. \(^{13}\)See Solon (1999) for a survey. Chadwick and Solon (2002), Huang, Li, Liu, and Zhang (2009), and Güell, Mora, and Telmer (2015) find a positive correlation between the extent of sorting and the intergenerational correlation in earnings.
along the diagonal $y = x$, then by Corollary 2 PAM obtains and the expected lifetime earnings of children are increasing and convex in parents' lifetime earnings. And since $f(x, x, \varepsilon)/x$ is a constant, inequality (measured by the Lorentz order) rises across generations provided $Q(x, x)/x$ is non-decreasing in $x$.

**Labor Markets.** The model can be interpreted as matching within a firm (internal labor markets) or between firms (external labor markets). First, consider the internal labor market interpretation. In this case, the distribution over types $H_t$ captures the current skill distribution within a firm. The Planner groups workers together in each period to maximize the discounted stream of aggregate output. A worker’s future skill depends on her skill today and the skill of her teammate. The skill transition cdf $\mathcal{T}$ may capture learning how to produce from more effective co-workers, and/or transference of work habits (positive or negative).

Most studies of peer effects within organizations focus on contemporaneous effects (for example, free riding). Chan, Li, and Pierce (2014) is an exception. This paper documents statistically and economically significant dynamic peer effects in the internal labor market for cosmetic counter sales. For example, a new salesperson working alone in the first week gains 2% productivity (sales revenue) in the following week, while working with a co-worker with twice the productivity yields an additional 5% average week-to-week productivity gain. In contrast, working with low ability peers has a negative impact on future productivity. Altogether, the transition cdf $\mathcal{T}$ is again monotone in own type and partner type. I am not aware of any study that speaks to the convexity or complementary properties of $\mathcal{T}$ in the workplace.

For the external labor market interpretation, each matched pair is a metaphorical firm as in the decentralized market considered in Section 6. Intuitively, the basic structure of peer effects between co-workers in external labor markets should be similar to those in internal labor markets, although I would anticipate the magnitude would be smaller across firms (due to firm specific learning, not captured in the current model). For anecdotal evidence, we need look no further than our own labor market. Concerned about their future career trajectory, rookie assistant professors overwhelmingly accept offers from top departments over offers from middle ranked departments, even if the salary and teaching loads are identical.
Schooling. There is a large literature on peer effects in schooling. Typically, individual performance in a given period (ex. school year) is modelled as a function of prior performance and prior peer performance (within the same classroom, dorm room, school, etc). Many of these studies have found that individual student performance is stochastically increasing in the prior performance of peers. A common empirical restriction is independent noise and peer effects “linear-in-means.” In our metaphorical two-student classrooms this translates as $f(x, y, \varepsilon)$ linear in $x$ and $y$, which yields $\int_{x}^{1} T(s|x, y)ds$ SBM and bi-concave as shown in the proof of Corollary 2. Thus, in this case peer effects favor sorting students by prior performance.

Two examples of non-linear models considered in the literature (see Hoxby and Weingarth (2006)) are Bad Apples (a single poor student has a disproportionate negative effect) and Shining Light (a single good student has a disproportionate positive effect). Focusing on the deterministic case for simplicity, an extreme version of Bad Apples is $f(x, y) = \min(x, y)$. More generally, if a bad student harms a good student more than a good student helps a bad student, i.e. $f(x, x) - f(x, y) > f(y, x) - f(y, y)$ for all $y < x$, then $f$ is SPM, which reinforces sorting by (10). Similarly, the Shining Lights model is consistent with $f$ SBM with extreme $f(x, y) = \max(x, y)$. Sacerdote (2011) summarizes “Using nonlinear models, one prevalent finding is larger peer effects in which high ability students benefit from the presence of other high ability students,” which suggests a SPM transition function $f$. However, there is certainly no empirical consensus on the structure of peer effects in schooling.

While the model sheds light on the forces behind sorting in schooling, the interpretation of static output $Q$ is problematic. A Planner would place some weight on the flow utility of students, but much of the gain to education is better modelled as a terminal reward. Thus, a finite horizon version of the model with a terminal reward seems like a better fit for this schooling application. Intuitively, the basic insights from the current infinite horizon model extend to the finite horizon model. Indeed, we have established this already for a two period model with terminal reward (5) in Section 3.

---

14See Sacerdote (2011) for a recent survey.

15For example, Hoxby and Weingarth (2006) reject the linear-in-means model, as well as versions of the bad apples and shining lights model.
Neighborhood Sorting. Significant peer effects within neighborhoods have also been documented for crime, educational attainment, teen fertility, self esteem, etc.\footnote{Table 2 in Durlauf (2004) summarizes a wide range of empirical analysis of such neighborhood peer effects. Becker was an early pioneer in the general theory of social interactions (Becker 1974).} For some of these applications types could be interpreted as the propensity to engage in certain behaviors (ex. commit a crime), while for other applications types could be a scalar measure of a trait (ex. self esteem). The precise interpretation of static production $Q$ also depends on the application. For traits that have a large impact on flow utility (exs. self-esteem and obesity), $Q$ could be a flow value as in the current model. For other applications, like crime, a simple model posits a low payoff $\ell$ to society for each crime committed in a given period and a high payoff $h > \ell$ otherwise. Then letting $q(x, y)$ be the chance that a type $x$ does not commit a crime when matched to a type $y$, we have $Q(x, y) = \ell + (q(x, y) + q(y, x))(h - \ell)$.

For these applications, the decentralized market modeled in Section 6 is harder to justify. An interesting extension would be to embed the current model of dynamic peer effects in a model of neighborhood choice, in which partial transferability occurs via adjustment in property values. I conjecture that the basic insights (what factors drive sorting), should extend to such a model.\footnote{Indeed, we should expect the conditions derived here in the perfectly transferable case to be necessary, but not sufficient for a robust characterization of sorting in the partially transferable utility case, as Legros and Newman (2007) established in the static case.}

8 Suggestive Discussion of an OLG Model

As mentioned in the Introduction, Jovanovic (2014) characterizes equilibria with positive sorting in an OLG matching model. Since young and old agents play asymmetric roles in an OLG model, the current results are not directly applicable to his framework. Nonetheless, Theorem 2 can make sense of his results.

In Jovanovic’s model, type evolution (only relevant for the young) takes place due to learning (unaffected by partner choice) and deterministic training. As he acknowledges, equilibria with positive sorting need not exist in his model. That is, he does not establish when positive sorting obtains in his model. Instead, he assumes positive sorting and characterizes the implied equilibrium.
While he does not establish when positive sorting obtains in the general version of his model, he does solve for an equilibrium involving positive sorting given Log-Normal signals for young worker types, and Cobb-Douglas static production function $x^{1-\rho}\tau^\rho$ and training function $Ax^{1-\theta}\tau^\theta$, where $x$ is the known type of the old worker and $\tau$ the (imperfectly known) type of the young worker. It turns out that the resulting production and transition functions satisfy the premise of Theorem 2; and so, it is not surprising that a positive sorting equilibrium obtains in his model for this case.

First, consider the simpler perfect information case, where the signal of ability $y$ equals $x$ for all young agents. In this case, Jovanovic’s static production function $Q(x, y) = x^{1-\rho}y^\rho$ is SPM and linear along the diagonal $y = x$. Given perfect information, Jovanovic’s training function is a deterministic version of this paper’s transition function, i.e. $f(x, y) = Ax^{1-\theta}y^\theta$, also SPM and linear along the diagonal. Thus, $\int_z^1 T(s|x, y)ds$ will be SBM and $\int_z^1 T(s|x, x)ds$ will be concave by (9). Altogether, we can easily confirm that the premise of Theorem 2 part (a) is satisfied.

In the imperfect information case, Jovanovic assumes that $\tau$ is Log-Normal and that signals $y$ are log linear: $\log(y) = \log(\tau) + \zeta$, where $\zeta$ is normal with zero mean. Since $\tau$ is unobservable, consider sorting on the signal $y$.\footnote{As the conditional cdf over types $\tau$ given signal $y$ is increasing in $y$ (in a FSD sense), this is a natural analogue of sorting in this incomplete information world.} Static expected production is then: $Q(x, y) = x^{1-\rho}E[\tau^\rho|y]$. Explicitly evaluating with the log-normal distributional assumptions, we find: $Q(x, y) = x^{1-\rho}y^\rho e^{\rho^2/2}$, and so expected output remains SPM and linear along the diagonal.

Turning to the transition function with the given Log-Normal signals and Cobb-Douglas training function, we calculate:

$$T(s|x, y) = \Upsilon(\theta^{-1}(\log(s) - (\theta \log(y) + (1 - \theta) \log(x)) - \log A))$$

where $\Upsilon(z)$ is the CDF of a mean zero normal random variable and $y$ is the commonly observed signal of the young. Explicitly evaluating $T_{12}(s|x, y)$ and $T_{xx}(s|x, x)$ we can see that they both satisfy single crossing in $s$, positive for low $s$ and negative for high $s$. Further, since type transitions are a martingale (conditional on commonly observed signals), $\int_{-\infty}^\infty T_{12}(s|x, y) = \int_{-\infty}^\infty T_{xx}(s|x, x) = 0$. Altogether, $\int_z^\infty T(s|x, y)$ is SBM and $\int_z^\infty T(s|x, x)$ is concave. The premise of Theorem 2 part (a) continues to hold with
Log-Normal signals.

9 Conclusion

This paper characterized value functions in a dynamic partnership model with peer effects on human capital dynamics. The resulting Planner ranking over (continuation) distributions allowed for a characterization of assortative matching. While a theory of sorting that relies only on complementarity conditions is possible in this environment, I argued that the requisite assumptions were unusually strong. A robust sorting theory requires curvature assumptions on static production (also provided). These curvature assumptions alone pinned down wage dynamics for individuals. Dynamics in the age distribution by cohort required some additional structure.

A Critical Contraction Mapping

Let $Z$ be the space of cdfs on $[0,1]$ endowed with the discrepancy metric defined on the associated measures, $||H - \tilde{H}|| = \sup_A |\lambda_H(A) - \lambda_{\tilde{H}}(A)|$. Using the norm $||V|| = \sup_{H\in Z} |V(H)|$ on value functions, define:

$$V = \{ V : Z \rightarrow \mathbb{R} : V \text{ is homogeneous of degree 1, continuous, and } ||V|| < \infty \}.$$  

Then for all $H \in Z$ and $U \in V$, define the Bellman operator $B$:

$$BU(H) = \max_{\mu \in \Phi(H)} \left[ (1 - \beta) \int_{[0,1]^2} Q(x,y)d\mu(x,y) + \delta U \left( (1 - \sigma)H^0 + \sigma \mathcal{H}(\mu) \right) \right]. \quad (11)$$

Anderson and Smith (2010) establish that the operator $B$ is a contraction. Thus, I characterize the unique fixed point $V = BV$ using standard contraction techniques.

B Value Characterization: Proof of Lemma 1

We establish the FSD and ICV ranking. The ICX proof follows symmetric steps.  

Step 0: Dilations. Let $P$ be a Markov transition kernel, i.e. $P(x|x')$ gives the
probability that \( x' \) transitions to \( s \leq x \). Blackwell (1953) established the following familiar result: We have \( H \succ_C \hat{H} \) if and only if there exists a Markov transition kernel \( P \) satisfying:

\[
H(x) = \int P(x|x')d\hat{H}(x') \quad \text{and} \quad \int xP(dx|x') = x' \quad \forall \ x'.
\]

In fact, this is a special case of a more general result.\(^{19}\) Let \( \succeq_D \) be any integral stochastic order. The Markov transition function \( P \) is a \( D \)-dilation, if and only if \( P(\cdot|x) \succeq_D \Delta_x \) for all \( x \), where \( \Delta_x \) is the degenerate distribution with unit mass at \( x \). Then \( H \succeq_D \hat{H} \) if and only if there exists a \( D \)-dilation \( P \) such that:

\[
H(x) = \int P(x|x')d\hat{H}(x').
\]

**Step 1:** \( V(\hat{H}) \geq V(H) \) for all \( \hat{H} \succ_{FSD} H \). I show that if \( U \) is non-decreasing in FSD shifts, then \( BU \) is increasing in FSD shift; and thus, since \( B \) is a contraction, the unique fixed point \( V = BV \) also increases in FSD shifts.

Assume arbitrary \( U \in \mathcal{V} \) such that \( U(H_1) \geq U(H_2) \) for all \( H_1 \succeq_{FSD} H_2 \). Now, choose arbitrary \( \hat{H} \succ_{FSD} H \). Then by Step 0, there exists a Markov kernel \( P \), such that:

\[
\hat{H}(x) = \int P(x|x')dH(x') \quad \text{and} \quad P(x|x') = 0 \quad \forall \ x < x'.
\]

Now, assume \( \mu \in \Phi(H) \) solves the Bellman Equation (11) given \( U \) and \( H \), and form \( \hat{\mu} \in \Phi(\hat{H}) \):\(^{20}\)

\[
\hat{\mu}(x, y) = \int P(x|x')P(y|y')d\mu(x', y').
\]

Since \( Q \) is increasing in each argument and \( P \) is a FSD shift, static output is higher under \( \hat{\mu} \) than \( \mu \): \( \int Qd\hat{\mu} \geq \int Qd\mu \). Likewise, since \( T(s|x, y) \) is non-increasing in \( x \) and in \( y \) for all \( s \), we have:

\[
\mathcal{H}(\hat{\mu})(s) = \int T(s|x, y)d\hat{\mu}(x, y) \leq \int T(s|x, y)d\mu(x, dy) = \mathcal{H}(\mu)(s).
\]

Thus, \( \mathcal{H}(\hat{\mu}) \succeq_{FSD} \mathcal{H}(\mu) \); and so, \( \sigma \mathcal{H}(\hat{\mu}) + (1 - \sigma)H^0 \succeq_{FSD} \sigma \mathcal{H}(\mu) + (1 - \sigma)H^0 \). And thus:

\[
U(\sigma \mathcal{H}(\hat{\mu}) + (1 - \sigma)H^0) \geq U(\sigma \mathcal{H}(\mu) + (1 - \sigma)H^0).
\]

\(^{19}\)One reference is Meyer (1966) pages 239-246.

\(^{20}\)Technically, \( \hat{\mu} \) maintains the same Copula as \( \mu \), and simply adjusts for the shift in the human capital CDF given by \( P \). Thus, \( \mu \in \Phi(H) \), ensures \( \hat{\mu} \in \Phi(\hat{H}) \).
Since \( \hat{\mu} \) need not solve the Bellman Equation (11) for \( \hat{H} \) and \( U \), we have shown:

\[
BU(\hat{H}) = \max_{\mu \in \Phi(\hat{H})} \left( (1 - \beta) \int Qd\mu + \delta U((1 - \sigma)H^0 + \sigma\mathcal{H}(\mu)) \right)
\]

\[
\geq (1 - \beta) \int Qd\hat{\mu} + \delta U((1 - \sigma)H^0 + \sigma\mathcal{H}(\hat{\mu}))
\]

\[
\geq (1 - \beta) \int Qd\mu + \delta U((1 - \sigma)H^0 + \sigma\mathcal{H}(\mu)) = BU(H)
\]

The unique fixed point \( V = BV \) must also obey \( V(\hat{H}) \geq V(H) \) for all \( \hat{H} \succ_{FSD} H \).

**Step 2:** \( v \) is increasing. Relocate a small mass \( \delta \) of the \( H \) distribution around \( x \) to \( x + h \), where \( h > 0 \) is feasible and arbitrary. The slope of \( V(H) = \int v(x)H(dx) \) in \( \delta \) is proportional to \( v(x + h) - v(x) \), at \( \delta = 0 \). Since \( V(H) \) rises in any FSD shift of \( H \), this must be strictly positive. So \( v(x) \) is everywhere increasing.

**Step 3:** If \( Q \) is bi-concave and \( \int_0^2 T(s|\cdot)ds \) is bi-convex then \( V \) increases in ICV shifts. Assume arbitrary \( U \in V \) such that \( U(H_1) \geq U(H_2) \) for all \( H_1 \succeq_{ICV} H_2 \). Now, choose arbitrary \( \hat{H} \succeq_{ICV} H' \). Then by Shaked and Shanthikumar (2007) Theorem 4.A.6, there exists \( H \) such that: \( H \succeq_{FSD} H' \) and \( H \succeq_{CX} \hat{H} \). By Step 1, \( BU(H) \geq BU(H') \). Thus, we seek to show that \( BU(\hat{H}) \geq BU(H) \), so that altogether \( BU(\hat{H}) \geq BU(H') \), and the unique fixed point obeys \( V(\hat{H}) \geq V(H') \).

By Step 0, \( \hat{H} \succeq_{CX} \hat{H} \) implies that there exists Markov kernel \( P \) satisfying:

\[
H(x) = \int P(x|x')d\hat{H}(x') \quad \text{and} \quad \int xP(dx|x') = x' \forall x'.
\]

Now, assume \( \mu \in \Phi(H) \) solves the Bellman Equation (11) given \( U \) and \( H \), and form \( \hat{\mu} \in \Phi(\hat{H}) \):

\[
\mu(x, y) \equiv \int P(x|x')P(y|y')d\hat{\mu}(x', y').
\]

Then since \( P(\cdot|x) \) is a mean-preserving spread at all \( x \) and \( Q \) is bi-concave, Jensen’s
inequality yields static output higher under $\hat{\mu}$ than $\mu$:

\[
\int Qd\mu = \int \int \int Q(x, y)P(dx|x')P(dy|y')d\hat{\mu}(x', y') \\
\leq \int Q\left(\int xP(dx|x')dx, \int yP(dy|y')dy\right)d\hat{\mu}(x', y') \\
= \int Q(x', y')d\hat{\mu}(x', y') = \int Qd\hat{\mu}.
\]

Likewise for continuation distributions we have:

\[
\int_0^z \mathcal{H}(\mu)(s)ds = \int_0^z \mathcal{T}(s|x, y)d\mu(x, y)ds \\
= \int \int \int \int_0^z \mathcal{T}(s|x, y)P(dx|x')P(dy|y')d\hat{\mu}(x', y')ds \\
\geq \int \int \int_0^z \mathcal{T}\left(s|\int xP(dx|x')dx, \int yP(dy|y')dy\right)d\hat{\mu}(x', y')ds \\
= \int \int \int_0^z \mathcal{T}(s|x', y')d\hat{\mu}(x', y')ds = \int_0^z \mathcal{H}(\hat{\mu})(s)ds
\]

where the inequality follows from $P(\cdot|x)$ a mean-preserving spread at all $x$ and $\int_0^z \mathcal{T}(s|x, y)ds$ bi-convex. Thus, $\mathcal{H}(\hat{\mu}) \succ_{ICV} \mathcal{H}(\mu)$. But, since $U$ increases in ICV shifts, then we have:

\[
U \left(\sigma\mathcal{H}(\hat{\mu}) + (1 - \sigma)H^0\right) \geq U \left(\sigma\mathcal{H}(\mu) + (1 - \sigma)H^0\right).
\]

Altogether, both static output and the continuation value is higher under given $\hat{H}$ and matching $\hat{\mu}$ than under $H$ with matching $\mu$. Recalling that $\mu \in \Phi(H)$ solved the Bellman Equation (11) given $U$ and $H$, we have shown $BU(\hat{H}) \geq BU(H)$, as required.

**Step 4:** $V \downarrow CX \Rightarrow v$ is concave. For any $x$ in the support of $H$, equally spread a small fraction $\varepsilon$ of the $H$ distribution near $x$ to $x \pm h$, where $h > 0$ is feasible and arbitrary. The slope of $V(H) = \int v(x)H(dx)$ in $\varepsilon$ is proportional to $[v(x + h) + v(x - h)]/2 - v(x)$, at $\varepsilon = 0$. Since $V(H)$ falls in any mean-preserving of $H$, this must be negative. So the planner’s shadow value $v(x)$ is everywhere concave.
C Assortative Matching Proofs

Proof of Theorem 1: Since $Q$ is SPM, PAM maximizes static output. Since $T(s|\cdot)$ is SBM and $\mathcal{H}(\mu)(s) = \int T(s|\cdot)d\mu$, PAM solves $\min_{\mu \in \Phi(H)} \mathcal{H}(\mu)(s)$ for all $s$. Since $V$ increases in FSD shifts, this in turn implies that PAM also solves $\max_{\mu \in \Phi(H)} V(\mathcal{H}(\mu))$. Altogether, PAM separately maximizes static output and continuation values.

Proof of Theorem 2: Given $Q$ SPM, PAM maximizes static output. Thus, we need only show in each case that PAM maximizes continuation values. We prove part (b). The proof of part (a) follows the same logic as part (b.i).

Proof of part (b.i): Assume arbitrary $U \in \mathcal{V}$ such that $U(H_1) \geq U(H_2)$ for all $H_1 \succeq_{ICV} H_2$. I claim this has two implications given the premise of (b.i):

- Step 1: PAM solves the induced Planner’s Problem (11)
- Step 2: $B$ preserves monotonicity: $BU(H_1) \geq BU(H_2)$ for all $H_1 \succeq_{ICV} H_2$

Thus, the unique fixed point obeys $V(H_2) \geq V(H_1)$ for all $H_1 \succeq_{ICV} H_2$ (by Step 2) and PAM is optimal (by Step 1).

Step 1: Since $\int_0^\infty T(s|\cdot)ds$ is SBM, PAM solves $\min_{\mu \in \Phi(H)} \int_0^\infty \mathcal{H}(\mu)(s)ds$. That is, PAM minimizes the weight in the tails of the continuation distribution $\mathcal{H}(\mu)$, which implies that PAM solves $\max_{\mu \in \Phi(H)} U(\mathcal{H}(\mu))$ since $U$ increases in the ICV order. Altogether PAM separately maximizes static production and the continuation value, and thus, solves the Planner’s Problem (11) for any distribution $H$. Altogether, for all $U \in \mathcal{V}$ increasing in the ICV order we conclude:

$$BU(H) = \frac{(1 - \beta)}{2} \int Q(x, x)dH(x) + \delta U \left( (1 - \sigma)H^0 + \sigma \mathcal{H}^P(H) \right)$$ (12)

where $\mathcal{H}^P(H)(s) \equiv \int T(s|x, x) \ dH(x)$ for all $s$ is the continuation distribution given PAM.

Step 2: Choose arbitrary $H_1 \succeq_{ICV} H_2$. Given $Q(x, x)$ increasing and concave:

$$\int Q(x, x)dH_1(x) \geq \int Q(x, x)dH_2(x)$$ (13)
Similarly, since $\int_{z}^{1} T(s|x,x)ds$ is decreasing and convex:

$$\int_{0}^{z} H^P(H_1)(s)ds = \int_{0}^{z} T(s|x,x)ds \, dH_2(x) \leq \int_{0}^{z} T(s|x,x)ds \, dH_1(x) = \int_{0}^{z} H^P(H_2)(s)ds$$

i.e. $H^P(H_1) \succeq_{ICV} H^P(H_2)$ and thus since $U$ increases in ICV shifts by assumption:

$$U(\sigma H^P(H_1) + (1 - \sigma) H^0) \geq U(\sigma H^P(H_2) + (1 - \sigma) H^0).$$

Altogether, both static output and the continuation value in (12) is higher for $H_1$ than $H_2$ and thus, $BU(H_1) \geq BU(H_2)$.

**Proof of part (b.ii):** Given $Q$ bi-concave and $\int_{0}^{z} T(s|\cdot)ds$ bi-convex, $V$ is increasing in ICV shifts by Lemma 1. Given $\int_{0}^{z} T(s|\cdot)ds$ SBM for all $z$ PAM solves $\min_{\mu \in \Phi(H)} \int_{0}^{z} H(\mu)(s)ds$ for all $z$, and thus since $V$ is increasing in the ICV order, PAM solves $\max_{\mu \in \Phi(H)} V(H(\mu))$. Altogether PAM separately maximizes static output and continuation values.

**D Cohort Wage Dynamics: Proof of Theorem 3**

Prove by induction to establish part (a). PAM obtains by Theorem 2. Then since $\int_{z}^{1} T(s|x,x)ds$ is concave in $x$, we have:

$$\int_{z}^{1} \int_{0}^{1} (1 - T(s|x))dH^0(x)ds = \int_{0}^{1} \int_{z}^{1} (1 - T(s|x))ds \, dH^0(x) \geq \int_{z}^{1} (1 - T(s|x^0,\bar{x}^0))ds$$

Altogether, $H_1 \succeq_{ICX} T(\cdot|x^0,\bar{x}^0) \succeq_{ICX} H^0$.

To establish the inductive step, note that $\int_{z}^{1} (1 - T(s|x,x))ds$ is increasing and convex in $x$ by assumption, thus $H^{N+1} \succeq_{ICX} H^N$ implies:

$$\int_{z}^{1} (1 - H^{N+2}(s))ds = \int_{0}^{1} \left[ \int_{z}^{1} (1 - T(s|x,x))ds \right] \, dH^{N+1}(x) \geq \int_{0}^{1} \left[ \int_{z}^{1} (1 - T(s|x,x))ds \right] \, dH^N(x) = \int_{z}^{1} (1 - H^{N+1}(s))ds$$

26
That is, $H^{N+2} \succeq_{ICX} H^{N+1}$ whenever $H^{N+1} \succeq_{ICX} H^N$. Altogether, $H^{N+1} \succeq_{ICX} H^N$ for all $N$. But then, since $w(x) = Q(x, x)/2$ is increasing and convex, $G^{N+1} \succeq_{ICX} G^N$ for all $N$ by Theorem 4.A.8 part (a) in Shaked and Shanthikumar (2007).

The proof of part (b) follows symmetric steps, except $H^1 \succeq_{ICV} H^0$ is asserted rather than proven from more elementary conditions.

References


