Assortative Matching, Reputation, and the Beatles Breakup[*]

Axel Anderson[†] Lones Smith[‡]

University of Michigan
Economics Department
Ann Arbor, MI 48109-1220

(first version: June 2000)

this version: November 27, 2001

Abstract

Consider Becker's (1973) classic static matching model, with output a stochastic function of unobserved types. Assume symmetric incomplete information about types, and thus commonly observed Bayesian posteriors. Matching is then assortative in these 'reputations' if expected output is supermodular in types.

We instead consider a standard *dynamic* version of this world, and discover a robust failure of Becker's global result. We show that as the production outcomes grow, assortative matching is neither efficient nor an equilibrium for high enough discount factors. Specifically, assortative matching fails around the highest reputation agents for 'low-skill concealing' technologies. Our theory implies the dynamic result that high-skill matches (like the Beatles) eventually break up.

Our results owe especially to two findings: (a) value convexity due to learning undermines match supermodularity; and (b) for a fixed policy in optimal learning, the second derivative of the value function explodes geometrically at extremes.

[*]Axel wishes to thank the University of Michigan for ongoing financial support, while Lones very much appreciates continual funding from the NSF. We wish to thank Ennio Stacchetti specifically for substantial help with the existence proof, and Canice Prendergast for leading us to think clearly about our anonymity assumption. We have benefited from feedback at the Society for Economic Dynamics and Control (Costa Rica), Michigan, the Royal Dutch Conference on Search & Assignment, Copenhagen, Pennsylvania, and Stanford's 2001 SITE Conference. The most recent version of this paper is available from `www.umich.edu/∼lones`.

[†]Email: `azanders@umich.edu`

[‡]Email: `econ-theorist@earthling.net`; web: `http://www.umich.edu/~lones`

1 Introduction

This paper is co-authored by a tenured faculty member and a graduate student. It is reasonable to assume that the reputation of the faculty member is substantially more established than the graduate student's. The Beatles broke up after it was clear that their members were highly talented, and each member went on to form a new group with individuals having unestablished reputations. While in the first example individuals with established reputations matched with those not established, in the second example partnerships of likes dissolved once reputations were established. These examples motivate two questions: What static matchings can we expect to see when reputations matter? Which partnerships will endure and which will fail?

Becker's (1973) matching model established that in a static Walrasian complete information model, if output depends solely on underlying types, then supermodular match output ensures *positive assortative matching* (PAM). If we extended his work to incomplete information, but retained the static framework, then supermodular expected output would deliver the same result: like reputation agents match together.

But restricting the model to a static setting dodges an important issue, that matches yield not only output but also information about types. One's reputation is often at least as important as the static output produced by the current partnership: The graduate student co-author cares much more about the paper's effect on the market's perception of his ability than about the paper's actual quality. Thus, to investigate the impact of reputational concerns on partner choice, we must embellish Becker's original model with both incomplete information and dynamics. Our goal is to shed light on who matches with whom in such a dynamic, incomplete information setting.

We assume that output is a stochastic function of types — either 'good' or 'bad.' Everyone is summarized by the public posterior chance that he is 'good', which we call his *reputation*. Our main finding is an impossibility result, that PAM fails in the limit for sufficiently patient individuals and rich technologies. We also find that failures of assortative matching occur for those agents with focused reputations. Finally, we flesh out the implications of this for matching dynamics: Partnerships of agents with similar reputations, like the Beatles, will dissolve as they establish themselves as either highly skilled — or not so. Conversely, matches of unlike agents, like tenured faculty with graduate student coauthors, will form in their place.

For an overview, we first find that expected *continuation* values are convex in the reputation of one's partner, information being valuable. Thus, one's own continuation value can be raised by matching with high or low reputation agents. Since those agents with very high or low reputations benefit the least from Bayes-updating, PAM *never* maximizes expected continuation values. More precisely, we show that value convexity due to learning is at cross purposes with the supermodularity that delivers PAM. So even with supermodular expected output, there is a tradeoff between maximizing current payoffs (Becker's effect) and future payoffs (the reputational or informational concern). With enough patience, the latter dominates. Using Becker's PAM result in the final period, we validate this intuition in a stylized *two period model* (Proposition 1).

Proposition 2 provides a simple labor theoretic insight as to *where* the PAM failure occurs. We show, for instance, that PAM fails for high reputations when production is sufficiently 'low-skill concealing' — i.e., it cannot statistically identify both partners as 'bad'. Since matching can only reveal whether both agents are 'good', high types are informationally valuable to middling agents, and should be nonassortatively matched. This static insight has dynamic implications: Partnerships of identical 'good' agents (the Beatles) will eventually break down, as reputations perforce converge to the truth.

In an infinite horizon model, the situation is substantially more complex, and these findings *a priori* hang in the balance. Intuitively, the value function 'flattens out' with rising patience. So as the discount factor tends to 1, match payoffs and information acquired in a match both become vanishingly important. In this way, the static payoff losses may forever outweigh the informational (reputational) gains for any level of patience. We find that under quite general assumptions, given sufficient patience, PAM fails for some agents with well-focused reputations. Our approach to the PAM impossibility result is by contradiction — namely, we study the assortative matching policy, and argue that its associated value function cannot possibly satisfy Bellman's equation. This turns on a new finding relevant to the optimal learning literature: For a fixed policy, the convexity of the value function entirely accumulates at the extremes; specifically, the second derivative explodes at a *geometric rate* near 0 and 1 — namely, where our noted PAM failure occurs (Proposition 3). We characterize this rate, thereby rescuing a close analogue of our earlier skill-concealing conditions in Proposition 2 for PAM to fail at high/low reputation agents (Proposition 4). Finally, these conditions hold for generic production technologies with enough outputs (Proposition 5).

There are a number of extensions of Becker's original work which assume search frictions (see Burdett and Coles (1997), Smith (1997), and Shimer and Smith (2000) for a sampling), but the essential thrust of the results is to develop additional conditions under which PAM still obtains. We are aware of no extension of Becker's original work to include incomplete information about types. And quite unlike the search papers, we show that Becker's finding robustly unravels given the reputational concern.

Our paper also enters a very small literature on equilibrium models of matching with incomplete information. Most relevant and earliest is the work on strategy-proofness with uncertain types in models with small numbers of agents (see Roth and Sotomayor (1990)). We know of only one truly general equilibrium matching model with (albeit asymmetric) incomplete information: Wolinsky (1990) explores information revelation, but in a model of exchange. Finally, Kremer and Maskin (1996) study a failure of PAM in a static complete information general equilibrium model without supermodularity.

The model and Becker's result is found in Section 2. In Section 3, we define a Pareto optimum and competitive equilibrium, establish the Welfare Theorems, and deduce existence. Our theory will thereby apply both to the efficient and equilibrium analysis. In Section 4, we build intuition in a stylized two period setting, in which we deduce our key insight about convexity versus supermodularity. We explore the infinite horizon model in Section 5, where we also develop our value function characterization. Wherever possible, we relegate unilluminating technicalities to the appendix.

2 The Model

A. The Static Economy. We consider the simplest world with uncertain types, where each agent can either be 'good' (G) or 'bad' (B). Only nature knows the types. Agents match in pairs to produce output, and enjoy a symmetric production role. Everyone is risk neutral. There are $N > 1$ distinct output levels q_i possible. For each pair of matched types, there is an implied distribution over output levels. Table 1 summarizes the chances h_i, m_i, and ℓ_i of output q_i (i.e., $\sum_i h_i = \sum_i m_i = \sum_i \ell_i = 1$):

	B	G
B	ℓ_i	m_i
G	m_i	h_i

Table 1: **Stochastic Output.** Probability that the given types produce output q_i

Each of the continuum of individuals has a publicly observed probability $x \in [0, 1]$ of being type G. Call x the *reputation* of the agent. So a match between agents with reputations x and y yields output $q_i \geq 0$ ($i = 1, \ldots, N$) with probability

$$p_i(x, y) = xyh_i + [x(1-y) + y(1-x)]m_i + (1-x)(1-y)\ell_i.$$

The expected output of this match is $f(x, y) = \sum_i q_i p_i(x, y)$. Since $q_i \geq 0$, with $q_i > 0$ for some i, we have $f(x, y) > 0$ when $0 < x, y < 1$, and matching is always optimal.

Recall that a twice differentiable function f is strictly *supermodular* (denoted SPM) *iff* $f_{12}(x, y) > 0$ for all (x, y). Reversing the inequality yields *submodularity* (SBM). A bilinear function like $f(x, y)$ is obviously globally SPM or SBM or amodular (i.e., both). Our results simply require that f not be amodular, which is a knife-edge case.

Assumption 1 (Supermodularity) $\sigma \equiv f_{12}(x, y) = \sum_i q_i(h_i + \ell_i - 2m_i) > 0$. *Hence, output is not stochastically invariant to types: We do not have* $(h_i) = (m_i) = (\ell_i)$.

Assumption 2 *If* $(m_i - \ell_i)/(h_i + \ell_i - 2m_i) = c$ *for a constant* c, *then* $c \notin [0, 1]$.

This 'no confounding' assumption guarantees that for any pair (x, y), an output q_i exists for which the reputations of both agents will change upon its realization.

We assume a (non-probability) density g_0 over reputations $x \in [0, 1]$. Matching is described by a symmetric measure μ on $[0, 1]^2$, where $\mu(X, Y)$ is the measure of matches of (x, y), where $x \in X$ and $y \in Y$, for measurable sets $X, Y \subset [0, 1]$. Also, the marginal $\mu([0, 1], \cdot)$ of μ obeys $\mu([0, 1], x) \leq g_0(x)$ at all x. The static efficient matching $\mu = \mu^*$ yields the constrained one-shot maximum output $\mathcal{V}(g_0) \equiv \int_{[0,1]^2} f(x, y)\mu(dx, dy)$.

Say that x and y are matched if (x, y) lie in the *matching set*, namely the support of μ. Following Shimer and Smith (2000), *positive assortative matching* (PAM) obtains when for any $x \leq x'$ and $y \leq y'$, if matches (x, y') and (x', y) obtain, then so do (x, y) and (x', y'). So if matching sets are singletons, then PAM plus production symmetry reduce to everyone being *self-matched* — that is, each x matched with another x. In this paper, we shall simply mean self-matching when we write PAM.

Theorem 0 (Becker (1973)) *Given Assumption* 1*, PAM is efficient (maximizes* $\mathcal{V}(g_0)$ *for any* g_0*).*[1] *Given* $f_{12}<0$ *(SBM), the efficient matching is negatively assortative.*

Assumptions 1 and 2 shall henceforth remain in force. We thereby maintain the most favorable case for PAM, as we ultimately wish to show that assortative matching fails.

B. The Dynamic Economy. There is a constant *inflow density* $\bar{g}(x)$ of reputation x agents who live forever.[2] Reputation is assumed a sufficient statistic for the entire past history of output production, and only this attribute of an agent is observed. This anonymity assumption is in the general equilibrium spirit of Becker's work. So matching decisions only depend on reputations, and not on identities. Output is produced each period $t = 0, 1, 2, \ldots$. Let $z_i(x, y) \equiv z(q_i, x, y) \equiv p_i(1, y)x/p_i(x, y)$ be agent x's posterior reputation given that a match with y yielded output q_i. The dynamic economy is not a trivial repetition of the static one, by Assumption 1. Let $\langle f, \vec{\mu}\rangle \equiv (1-\beta)\sum_{t=0}^{\infty}\beta^t \int_{[0,1]^2} f(x, y)d\mu_t(x, y)$ for measurable functions f, where $\vec{\mu} = (\mu_t)_{t=0}^{\infty}$.

3 Existence, Welfare Theorems, and Values

A. Pareto Optimum. Given an initial density g_0 over reputations, the planner chooses matching measures $\vec{\mu}$ to maximize the present value of output. He can't match more of any type than is available, and faces a law of motion of the reputation density:

$$\mathcal{V}(g_0) = \sup_{\vec{\mu}}\langle f, \vec{\mu}\rangle \tag{1}$$

$$\text{s.t.} \quad \text{Feasibility:} \quad g_t(y) \geq \int \mu_t(dx, y) \ \forall y \ \forall t \tag{2}$$

$$\text{Bayes Updating:} \ g_{t+1} = \bar{g} + B(\mu_t) \ \forall t \tag{3}$$

where $B(\mu_t)$ is the Bayes-updated density given the matching measure μ_t.[3] In (2), $(\mu_t(.|y))$ are conditional measures[4] for fixed y, i.e. satisfying $d\mu_t(x, y) = d\mu_t(x|y)dy$.

In solving this problem, the planner is forced to trade off higher expected output today and a more informative density over reputations tomorrow. This trade-off is at the heart of this paper. As usual, we can recursively solve (1)–(3), as we next assert.

Theorem 1 *A solution to the planner's problem exists, and* $\mathcal{V}$ *solves the Bellman equation:* $\mathcal{V}(g_t) = \max_{\mu}(1-\beta)\int f d\mu + \beta\mathcal{V}(g_{t+1})$*, for* $t = 1, 2, \ldots$ *subject to* (2) *and* (3)*.*

[1] Becker proved this for the discrete case. For our purposes, Lorentz (1953) is more appropriate as he proved the formal result in the continuum case (albeit unaware of any economic context).

[2] This is nonstandard, as it precludes the possibility of a steady-state, since the mass of agents grows unboundedly in the infinite horizon model. At some notational cost, we could easily assume agents die with some chance after producing output. But our theory is true in or out of steady-state, and so we simply save ourselves the bother of proving one exists. Still, the planner's value will be well-defined, given a constant agent inflow and payoff discounting.

[3] For the precise and natural formula of $B(\mu)$, which is inessential here, consult (16).

[4] We understand that conditional distributions only admit μ-almost sure uniqueness. But we insist for simplicity in (2) on an everywhere identity (i.e., $\forall y$), as we soon shift to the agents' perspective.

The proof by the Theorem of the Maximum and Blackwell's Theorem is appendicized.

Starting to suppress time subscripts, the FOC for this problem are:

$$\begin{array}{rcl} (x,y) \in \mathit{supp}(\mu) \Rightarrow [v(x)+v(y)-(1-\beta)f(x,y)-\beta\Psi^v(x,y)] & = & 0 \\ v(x)+v(y)-(1-\beta)f(x,y)-\beta\Psi^v(x,y) & \geq & 0 \end{array} \tag{4}$$

Here, $v(x)$ is the multiplier on the constraint (2), the shadow value of an agent x. Also, the *expected continuation value of the match* (x, y) plays a central role in this paper:

$$\Psi^v(x,y) \equiv \psi^v(x|y)+\psi^v(y|x), \quad \text{where} \quad \psi^v(x|y) \equiv E[v(z(\tilde{q},x,y))]$$

That is, $\psi^v(x|y)$ is the expected continuation value of agent x when matched with y. Given g_0, a *Pareto optimum* (PO) is abbreviated as a pair $(\vec{\mu}, \vec{v})$ solving the planner's problem, for the infinite value vector $\vec{v} = (v_t(x) : 0 \leq x \leq 1, t = 0, 1, 2, \ldots)$.

The FOCs (4) are then very intuitive. For any matched pair, the sum of the shadow values of the two agents this period equals the total value to the planner of matching them. Further, for *any* pair, the sum of the shadow values today weakly exceeds the value the planner could achieve by matching them.

Observe that since the planner can always self-match any x for whom $g(x) > 0$, and since $f(x, x) \geq 0$, it cannot be optimal to leave any x unmatched; thus, for all x there must exist $y \in \text{supp}\{\mu_t(\cdot|x)\}$. Along with the FOC this implies that:

$$v(x) = \max_y (1-\beta)f(x,y)+\beta\Psi^v(x,y)-v(y) \tag{5}$$

Let $Y(x)$ denote the set of maximizers of this expression given some fixed reputation x. It is well-known (see Theorem 2.8.1 in Topkis (1998)) that if $(1-\beta)f(x,y)+\beta\Psi^v(x,y)$ is SPM, then $Y(x)$ is an increasing set in x. Similarly SBM yields this set decreasing in x. One is thus tempted to seek conditions that guarantee SPM of this expression. Clearly, it suffices that Ψ^v be SPM to establish this. We instead show in §5 that Ψ^v can never be globally SPM, so that a SPM-SBM trade-off emerges for sufficient patience.

B. Competitive Equilibrium. There are different ways to conceptualize a competitive equilibrium (CE). We assume that workers are willing either (*i*) to hire a match partner, offer him a sure wage, and take the output residual as profit; (*ii*) to hire themselves out other to another worker; or (*iii*) to stay unemployed, earning zero.

Let $w_t(x|y)$ be the wage that agent x receives in period t if matched with agent y. Define the infinite wage vector $\vec{w} \equiv \{w_t(x|y) : 0 \leq x, y \leq 1, t = 0, 1, 2, \ldots\}$. We overuse notation, anticipating a welfare theorem to come: Let $v_t(x)$ now denote the optimal discounted sum of wages that a reputation x worker earns from the start of next period onwards. A CE is a triple $(\vec{\mu}, \vec{v}, \vec{w})$ such that $\vec{\mu}$ obeys (2) and (3), while $(\vec{\mu}, \vec{v}, \vec{w})$ satisfy

- Free Entry/Exit: $$\begin{array}{c} (x,y) \in \mathit{supp}(\mu) \Rightarrow w(x|y)+w(y|x)-f(x,y)=0 \\ w(x|y)+w(y|x) \geq f(x,y) \end{array} \tag{6}$$
- Worker Maximization: $$v(x) = \max_y (1-\beta)w(x|y)+\beta\psi^v(x|y) \tag{7}$$

C. Welfare Theorems. The First and Second Welfare Theorems obtain.[5] Observe that while agents' true types are unknown, all information is public. Moreover, the global production technology is linear in measures μ of matched agents, and thus meets the standard convex technology requirement.[6] We appendicize the proofs of Theorems 2–3. One nonstandard feature of these welfare theorems is that we have embellished the allocation $\vec{\mu}$ with the values $\vec{v}$, and so are claiming coincidence of the planner's shadow values and the agents' 'private values'. Hereafter, we consistently refer to both simply as *values*, and the map $x \mapsto v(x)$ as the *value function.*

Theorem 2 *If $(\vec{\mu}, \vec{v}, \vec{w})$ is a CE then $(\vec{\mu}, \vec{v})$ is a PO.*

Theorem 3 *If $(\vec{\mu}, \vec{v})$ is a PO, then $(\vec{\mu}, \vec{v}, \vec{w})$ is a CE, where $\vec{w} = (w_t(x|y))$ solves*

$$w_t(x|y) = f(x, y) + \frac{\beta}{1-\beta}\psi_t^v(y|x) - \frac{v_t(y)}{1-\beta} \tag{8}$$

Reputation x receives a higher fraction of output if *ceteris paribus* (i.e., static output considerations aside) he provides a higher quality signal about his partner's reputation.

Corollary 1 *There exists a CE.*

Having established the equivalence between the set of CE and PO, we can focus on the more tractable social planner's problem.

D. Value Functions. We now prove that map $x \mapsto v(x)$ is convex. This is fundamentally different from the usual convexity of the *planner's* maximum value function.

Lemma 1 *The value function is strictly convex.*

Proof of convexity: Suppose that a binary signal is publicly revealed about the true types in an ε-measure ball around x, so that these agents' reputations increase to near $x' > x$ with chance λ and decrease to near $x'' < x$ with chance $1 - \lambda$. Because beliefs are a martingale, $x = \lambda x' + (1 - \lambda)x''$. The first order change in the planner's value equals the sum of the changes in shadow values, i.e., $\varepsilon[\lambda v(x') + (1 - \lambda)v(x'') - v(x)]$. Since the planner may always ignore this signal, this expression cannot be negative, thus proving weak convexity. Strict convexity obtains because the social planner is not indifferent across all matches, as shown in the Appendix.

We next show that continuation values are convex in one's *partner's* reputation.

[5] Mortensen (1982) is the first paper to show that Welfare Theorems are not to be taken for granted in a matching setting. But his impediment was search frictions, while ours is incomplete information.

[6] On a mathematical level this is not surprising, given the similarity between our problem and an assignment problem. (See Roth and Sotomayor (1990) for a survey of the topic.) Gretsky, Ostroy, and Zame (1992) establish that in a continuum assignment problem, the set of PO, CE, and the Core coincide. They establish their results in a general static setting. Although we relied on their work for insight into our welfare and existence theorems, the dynamic nature of our problem prevented us from applying their results directly.

Lemma 2 *If v is strictly convex, then the expected continuation value function $\psi^v(x|y)$ is strictly convex in x and y.*

Proof: Since v is convex, it is twice differentiable a.e., so we may a.e. twice differentiate $\psi^v(x|y) = E[v(z(\tilde{q}, x, y))]$ in x or y. Let $u = x$ or y. Now $p_i(x, y)$ and $p_i(x, y)z_i(x, y)$ are both bilinear in (x, y). So $p_{iuu}(x, y) = 0$ and $2p_{iu}(x, y)z_{iu}(x, y) + p_i(x, y)z_{iuu}(x, y) = 0$. This yields $\psi^v_{uu}(x, y) = \sum_i p_i(x, y)z_{iuu}(x, y)v''(z_i(x, y)) > 0$, whenever v'' exists. $\square$

4 The Metaphorical Two Period World

To build intuition for our infinite horizon results, we consider a stylized two period model, with payoffs in periods $t = 0$ and $t = 1$ weighted by $1 - \beta \in [0, 1)$ and β. Obviously, $\beta < 1/2$ for a true two period model with period zero more important than period-one payoffs. But investigating what happens as $\beta \to 1$ will capture the basic insights in our infinite horizon model as $\beta \to 1$; however, the continuation value function is *endogenous* in that setting. By contrast, in the two period model, Theorem 0 yields a strictly convex period one continuation function $v_1(x) = f(x, x)/2$, *fixed for any β*. This dodges a hard complication, allowing us to prove an impossibility result.

A. Convexity versus Supermodularity. We first deduce an unqualified failure of PAM that is unique to the two period model, but which cleanly demonstrates the opposition between value convexity and match supermodularity. To better flesh out the contrast, observe that if $\psi^v(x|\cdot)$ were type x's match payoff function, then PAM would require that it be SPM on the matching set. We see below that the strict convexity of $\psi^v(x|y)$ in y and the Bayesian sinks at $x = 0, 1$ preclude global supermodularity.

Proposition 1 (*a*) *Let v be strictly convex, and $x \in (0, 1)$. Given the matches $(0, 0)$, (x, x), and $(1, 1)$, the expected continuation value is strictly raised by rematching x with 0 or 1. Specifically,* (*b*) *in the two period model, PAM fails for large enough $\beta < 1$.*

Proof of (*a*)*:* Rematching $x \in (0, 1)$ with 0 changes expected continuation values by

$$A_0 = \psi^v(x|0) - \psi^v(x|x) + \psi^v(0|x) - \psi^v(0|0) = \psi^v(x|0) - \psi^v(x|x)$$

since agent 0 has posterior reputation 0 regardless of partner. Likewise, rematching x with 1 shifts expected values by $A_1 = \psi^v(x|1) - \psi^v(x|x)$. A convenient weighted average of these changes $(1 - x)A_0 + xA_1 = (1 - x)\psi^v(x|0) + x\psi^v(x|1) - \psi^v(x|x)$ is positive by strict convexity of $\psi^v(x|y)$ in y. Hence, either $A_0 > 0$ or $A_1 > 0$. (Note that SPM, however, would instead require $A_0 \le 0$ and $A_1 \le 0$.) So any $x \in (0, 1)$ achieves a higher expected continuation value by matching with 0 or 1 rather than x.

Proof of (*b*)*:* Given the period one value function $v_1(x) = f(x, x)/2$, Assumption 1 yields $v_1''(x) = 2\sum_i(h_i + \ell_i - 2m_i)q_i > 0$; by Lemma 2, ψ^{v_1} is strictly convex in y. The period zero losses from the above rematching are finite, being swamped by the first period gains for β high enough, so PAM is inefficient. $\square$

Examining the proof of Proposition 1 yields a rather striking corollary: In the two period model, if the planner is patient enough and $(0,0)$ and $(1,1)$ are matched in equilibrium, then no other x can be self-matched.

B. The Failure of PAM via Skill-Concealing Conditions. Proposition 1 is a quick illustration of how Bayesian learning of reputations undermines supermodularity; however, it does not suggest whether we should expect the failure for high or low types. Also, it will not prove a robust finding once we abandon the metaphorical two period setting, with its possibly unrealistic negative interest rates.

We next consider a result that does extend to the infinite horizon setting, and that offers labor theory insights — relating the PAM failure to the stochastic productive interaction. Let $h=(h_i), m=(m_i), \ell=(\ell_i)$. Notice that if $m=\ell$ (but of course, not $m=h$, by Assumption 1), then (G,G) matches can be statistically distinguished from (B,G) and (B,B), but the latter two cannot be nuanced: This is an extreme case of a *low-skill concealing* technology. Think of (h,m,ℓ) as sufficiently *low-skill* (resp. *high-skill*) *concealing* if it is close enough to the $m=\ell$ (resp. $m=h$) hyperplane.

Proposition 2 *The expected continuation value* Ψ^{v_1} *is SBM around* $(0,0)$ *or* $(1,1)$ iff*:*

$$\textstyle\sum_i \frac{m_i}{\ell_i^2}\left(m_i^2-2h_i\ell_i+m_i\ell_i\right)>0 \quad \text{or (resp.)} \quad \sum_i \frac{m_i}{h_i^2}\left(m_i^2-2h_i\ell_i+h_im_i\right)>0 \tag{9}$$

Specifically, PAM fails around $(1,1)$ (*resp.* $(0,0)$) *if* β *is high enough and output is sufficiently low-skill (resp. high-skill) concealing.*

To understand the skill concealing conditions (9), assume $h_i=m_i$ for all i. The RHS of inequality (9) is then $S(h,m,\ell)\equiv\sum_i m_i^2(m_i-\ell_i)/\ell_i^2$. As the $m_i-\ell_i$ factor in $S(\cdot)$ is weighted by $m_i^2/\ell_i^2 \gtrless 1$ for $m_i \gtrless \ell_i$, we have $S(\cdot)>\sum_i(m_i-\ell_i)=1-1=0$. Since $S(\cdot)$ is continuous in (h,m,ℓ), the inequality holds near this extreme.

Proof: Since $v_1(x)=f(x,x)/2$, we have $\psi^{v_1}(x,x)$ clearly C^∞ on $(0,1)$, being smoothly defined. Suppressing x and y arguments,

$$\psi^{v_1}_{xy}=\textstyle\sum_i\left[p_{ixy}v_1(z_i)+(p_{ix}z_{iy}+p_{iy}z_{ix}+p_iz_{ixy})v_1(z_i)'+2p_iz_{ix}z_{iy}v_1(z_i)''\right]$$

Define $\kappa(x)\equiv\psi^{v_1}_{xy}(x,x)$. Simple algebraic manipulation establishes that terms in v_1, v_1' and v_1'' each vanish at $(1,1)$ and $(0,0)$.[7] Thus, $\kappa(0)=\kappa(1)=0$. Now $\kappa'(0)<0$ *iff*:

$$v_1''(0)\textstyle\sum_i\frac{m_i}{\ell_i^2}\left(m_i^2-2h_i\ell_i+m_i\ell_i\right)>0$$

proving (9). Similarly, $\kappa'(1)>0$ *iff* $v_1''(1)\sum_i(m_i/h_i^2)\left(m_i^2-2h_i\ell_i+h_im_i\right)>0$.

The proof of ψ^{v_1} SBM around $(1,1)$ for low-skill concealing output is analogous. □

For an intuition, recall that match types are either (G,G), (G,B), or (B,B). If output is purely low-skill concealing, then (G,B) matches cannot be distinguished from (B,B) matches. Clearly, this renders $(x,0)$ matches informationally worthless. More strongly, if $x\in(0,1)$ is self-matched then we will only ever learn $\{(B,B),(G,B)\}$ or (G,G), whereas if x is matched with 1, then we can learn his true type.

[7] We assume $h\gg 0$ and $\ell\gg 0$ in this proof. If $\ell_i=0$ (resp. $h_i=0$) for some i then the cross partial $\kappa(0)=-\infty$ (resp. $\kappa(1)=-\infty$).

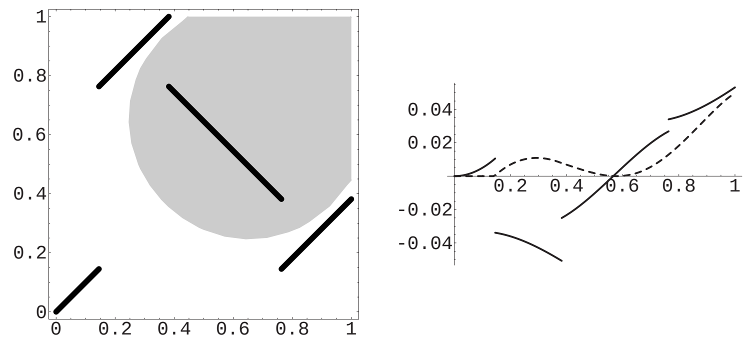

Figure 1: **Two Period Example.** On the left, we depict the shaded SBM total value region (where $H_{xy} < 0$), and the resulting discontinuous optimal matching graph $\mathscr{G} = \{(x, y(x)), 0 \leq x \leq 1\}$ (solid line). On the right, we plot the equilibrium wages $w(x) \equiv w(x|y(x))$ (solid line) — which is discontinuous due to the information rents. We superimpose the surplus in optimal values over assortative matching values $v_0(x) - 0.1x^2 - 0.99\psi^v(x|x)$ (dashed line).

C. An Illustrative Example. We set $\beta = 0.99$ and assume a uniform density over reputations. Let $(q_1, q_2) = (0, 2)$, $h = (1/2, 1/2)$, $m = (1, 0)$, and $\ell = (1, 0)$. Since $m = \ell$, this is a perfectly low-skill concealing case, and thus PAM will fail around $(1, 1)$. Clearly PAM is optimal in the final period, and so $v_1(x) = x^2$. Further, $p_1(x, y) = 1 - xy/2$, $p_2(x, y) = xy/2$, $z_1 = (1 - y/2)x/(1 - xy/2)$, and $z_2 = 1$. So the total match value is $H(x, y) \equiv 0.01xy + 0.99\Psi^{v_1}(x, y)$, where

$$\Psi^{v_1}(x, y) = p_1(x, y)v_1(z_1(x, y)) + p_2(x, y)v_1(z_2(x, y)) = \frac{x^2(1 - y) + y^2(1 - x) + xy}{1 - xy/2}$$

Figure 1 illustrates the solution, showing where $H(x, y)$ is SPM or SBM.

Here's an intuition for the graph. First, by local optimality considerations, the matching set $\mathscr{G}$ is increasing (decreasing), though possibly with jumps, whenever the match value $H(x, y)$ is SPM (SBM). Second, it must later jump downwards since it cannot exit the SPM region on a downward slope. Finally, by the assumed uniform density over reputations, $\mathscr{G}$ has slope ± 1 whenever it is continuous. As we simply wish to illustrate the theory,[8] we do not prove what a computer simulation of a Walrasian tâtonnement process quickly reveals — that there is a single down segment. This avoids a rather lengthy diversion.

At the right in Figure 1, we see that agents with high reputations capture large information rents in less productive matches with lower reputations. This provides their partners a higher quality signal about their own type and thus a higher expected value in period one than they would have otherwise received from matching assortatively.

[8]See Kremer and Maskin (1996) for a one-shot matching model where PAM fails (since SPM fails). Since their paper is half devoted to characterizing such solutions, we do not offer a formal argument.

5 The Infinite Horizon Model

5.1 Extremal Convexity of the Value Function

We now wish to understand the failure of PAM in an infinite horizon setting. In the two period model, the strict convexity of the *given* continuation value v_1 was critical. In the infinite horizon model, the continuation value[9] v_β is a function of β, and flattens out as β approaches 1. Indeed, reputations converge to 0 or 1 (see Lemma 3), and since beliefs are a martingale, the chance that an agent with initial reputation x converges to 1 is x. Thus, as $\beta \to 1$, the value function $v_\beta(x) \to xv(1)+(1-x)v(0)$. So when agents rematch, both current losses (scaled by $1-\beta$) and the gains in expected continuation values vanish as $\beta \to 1$. Thus, Proposition 1 fails; instead, we build on Proposition 2.

We proceed entirely by contradiction, assuming PAM and thereby working with a *time-independent* value function v_β — yet in our possibly nonstationary model! Indeed, agent x always receives the wage $f(x,x)/2$ with PAM. Using the resulting value function v_β, we find hard-to-satisfy necessary conditions for the efficiency of PAM. Towards our goal, we now argue that v''_β geometrically explodes near 0 and 1 under PAM. This will compensate for the flattening out of the value function in β. Understanding learning with very patient individuals is important (see Easley and Kiefer (1988)), and yet we do not believe that the extremal accumulation of convexity has been noticed, yet alone characterized as we do now. Hence, this result has stand-alone merit in its own right.

Proposition 3 *Assume PAM and $\chi_\beta \equiv \beta \sum_i m_i^2/\ell_i > 1$. Implicitly define $\alpha_\beta \in (0,1)$ by $1 \equiv \beta \sum_i \ell_i (m_i/\ell_i)^{2-\alpha_\beta}$. Then $\alpha_\beta \to 1$ as $\beta \to 1$ and $v''_\beta(x) \sim c_\beta x^{-\alpha_\beta}$ as $x \to 0$, for*[10] *some $c_\beta > 0$. (An analogous result is valid as $x \to 1$, defining $\chi'_\beta \equiv \beta \sum_i m_i^2/h_i > 1$.)*

Proof: Our proof uses some nonstandard tools, and so is mostly included in the text.

Step 1 *$\beta \sum_i \ell_i (m_i/\ell_i)^{2-\alpha_\beta} \equiv 1$ defines an increasing function $\beta \mapsto \alpha_\beta$, with $\alpha_1 = 1$.*

Proof: If $\alpha = 1$, then $\sum_i \ell_i (m_i/\ell_i)^{2-\alpha} = 1$, and if $\alpha = 0$ then this sum is $\sum_i m_i^2/\ell_i > 1$. It suffices then to show that $\sum_i \ell_i (m_i/\ell_i)^{2-\alpha}$ falls in α. By the Power Mean Inequality, or Theorem 96 in Hardy, Littlewood, and Polya (1952), $\mathcal{M}_\theta(x) = (\sum_i \omega_i x_i^\theta)^{1/\theta}$ is rising in θ if $\sum_i \omega_i = 1$. Setting $\theta = 2-\alpha$, $\omega_i = \ell_i$ and $x_i = m_i/\ell_i$ yields the desired result.□

• **Proof that v'' Explodes Near 0.** Simplify notation: $\xi(x) \equiv f(x,x)$, $\zeta_i(x) \equiv z_i(x,x)$, and $\pi_i(x) \equiv p_i(x,x)$. Given PAM, v_β satisfies the following 'policy equation':

$$v_\beta(x) = (1-\beta)\xi(x) + \beta \sum_i \pi_i(x) v_\beta(\zeta_i(x)) \tag{10}$$

We now affinely transform the value function, defining $\phi(x) \equiv v_\beta(x) - v_\beta(0) - v'_\beta(0)x$. Thus, $\phi(x)$ is the deviation of v_β from its best linear approximation at $x=0$. Observe

[9] We now remove time subscripts, and instead add a β subscript to highlight the dependence of v_β on discounting. We also switch to the more compact notation ψ^β and Ψ^β rather than ψ^{v_β} and Ψ^{v_β}.

[10] Notational reminder: $\phi(x) \sim g(x)$ near 0 iff $\lim_{x\to 0} \phi(x)/g(x) = 1$ iff $\phi(x) = g(x)(1+o(1))$.

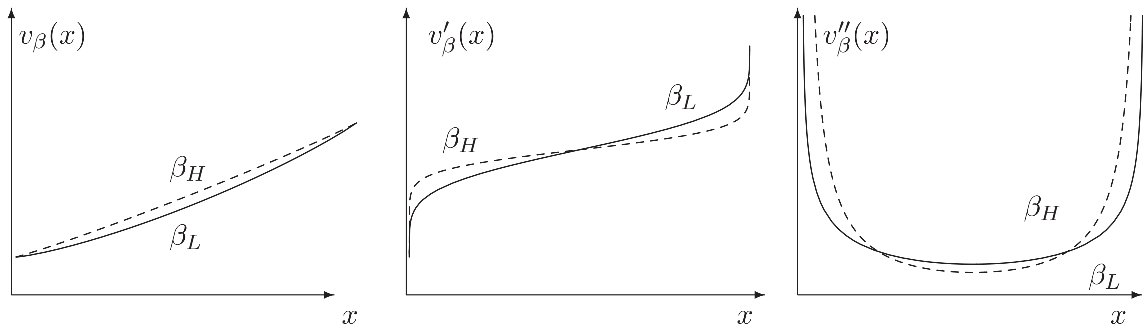

Figure 2: v_β**,** v'_β**, and** v''_β**.** This graph illustrates $\lim_{x\to 0} v''_\beta(x) = \infty$. (Note $\beta_H > \beta_L$.)

how this preserves $\phi''(x) = v''_\beta(x)$. Now, evaluating (10) and its derivative at $x = 0$ yields $v_\beta(0) = \xi(0)$ and $v'_\beta(0) = \xi'(0)$ for all $\beta < 1$. Since $\xi(x) \equiv \sigma x^2 + \xi'(0)x + \xi(0)$,

$$\phi(x) = (1-\beta)\sigma x^2 + \beta \textstyle\sum_i \pi_i(x)\phi(\zeta_i(x)) \tag{11}$$

Twice differentiate (11) — valid a.e. by convexity of v_β (Lemma 1) — to get:

$$\varphi''(x) = T(\varphi'')(x) \equiv \gamma(\phi(x), \phi'(x), x) + \beta \textstyle\sum_i \pi_i(x)\zeta'_i(x)^2 \varphi''(\zeta_i(x)) \tag{12}$$

an operator equation with *fixed point* $\varphi'' = \phi''$. Here, $\gamma(x) \equiv \gamma(\phi, \phi', x)$ is defined by:

$$\gamma(\phi(x), \phi'(x), x) \equiv 2(1-\beta)\sigma + \beta \textstyle\sum_i [p''_i(x)\phi(\zeta_i(x)) + (2p'_i(x)\zeta'_i(x) + \pi_i(x)\zeta''_i(x))\phi'(\zeta_i(x))]$$

(We thus fix the limits ϕ and ϕ' in the operator equation $\varphi'' = T(\varphi'')$ in (12) for ϕ''.) Naively evaluating (12) at $x = 0$, we have $v''_\beta(0+) = 2(1-\beta)\sigma + v''_\beta(0+)\sum_i \beta\ell_i(m_i/\ell_i)^2$. As the sum $\chi_\beta > 1$ by premise, and $v''_\beta \geq 0$ (Lemma 1), the solution $v''_\beta(0+)$ explodes.

- **Intuition for the Rate of Explosion of v''.** Multiply (12) by x^α to get:

$$\varphi''_\alpha(x) = \gamma(\phi, \phi', x)x^\alpha + \textstyle\sum_i w_{\alpha i}(x)\varphi''_\alpha(\zeta_i(x)) \tag{13}$$

where $\varphi''_\alpha(x) \equiv x^\alpha \varphi''(x)$ and $w_{\alpha i}(x) \equiv \beta\pi_i(x)\zeta'_i(x)^2(x/\zeta_i(x))^\alpha$. But since $w_{\alpha i}(0) = \beta\ell_i(m_i/\ell_i)^{2-\alpha}$, we have $\sum_i w_{\alpha_\beta i}(0) = 1$. In fact, $\sum_i w_{\alpha i}(0) \gtrless 1$ respectively as $\alpha \gtrless \alpha_\beta$, just as in the proof of Step 1. Finally, $\sum_i w_{\alpha i}(x)$ is continuous in x, so that

$$\textstyle\sum_i w_{\alpha i}(x) \gtrless 1 \text{ as } \alpha \gtrless \alpha_\beta \text{ so in a neighborhood } [0, \epsilon] \ (\epsilon = \epsilon(\alpha) > 0) \tag{14}$$

So $\lim_{x\to 0} x^\alpha v''_\beta(x) = 0$ (resp. ∞) for $\alpha > \alpha_\beta$ (resp. $\alpha < \alpha_\beta$). This is suggestive, but in no way proves the Proposition, since other functional forms, like $x^{\alpha_\beta}\log x$, also satisfy the twin limits. Towards a formal proof, we first establish the asymptotic form of ϕ.

Step 2 *For some* $0 < c_\beta < \infty$, *we have* $\phi(x) \sim \frac{c_\beta}{(2-\alpha_\beta)(1-\alpha_\beta)} x^{2-\alpha_\beta}$.

Proof: Write (11) as $\phi(x) = T\phi(x)$, where $T : \mathscr{C}[0,1] \mapsto \mathscr{C}[0,1]$, the space of continuous functions on $[0,1]$. Since $\beta < 1$, T is a contraction in the sup-norm. Let $\varphi = \phi$ be the unique fixed point of $\varphi(x) = T\varphi(x)$. Also, note that T is a monotonic operator.

Step 2-a *There exists* $\underline{a} > 0$ *such that* $\underline{a}x^{2-\alpha_\beta} \leq \phi(x)$.

Step 2-b *There exists* $\bar{a} > 0$ *and* $b > 0$ *such that* $\phi(x) \leq \bar{a}x^{2-\alpha_\beta} - bx^2$.

Steps 2-a and 2-b are proven in the Appendix. Now, $\phi(x)x^{\alpha_\beta-2}$ is trivially nonnegative, and bounded above as $x \downarrow 0$, by Step 2-b. Being continuous, it has some limit at 0, say $c_\beta/[(2-\alpha_\beta)(1-\alpha_\beta)]$. Moreover, Step 2-a forces $c_\beta > 0$. Finally, since ϕ is also continuous, the map $x \mapsto \phi(x)x^{\alpha_\beta-2}$ is continuous at $x = 0$. □

We have found the tail asymptotic order on the value function, namely $\phi(x) \propto x^{2-\alpha_\beta}$, and thereby $v_\beta(x) - \xi(0) - \xi'(0)x \propto x^{2-\alpha_\beta}$. This seems tantalizingly close to our goal of $v''_\beta(x) \propto x^{-\alpha_\beta}$. But whereas integrating asymptotic relations is a fully valid exercise, differentiation requires regularity conditions on the derivative (results known as 'Tauberian Theorems'). Fortunately, by §7.3 in De Bruijn (1958), monotonicity is one such condition for our context, with an implicitly defined function.

Step 3 *Near* $x = 0$*, we have* $\phi'(x) \sim c_\beta x^{1-\alpha_\beta}/(1-\alpha_\beta)$.

Proof: Where it exists, the first derivative ϕ' is increasing, by convexity of the value function, Lemma 1. Hence, De Bruijn's condition is met, provided ϕ' exists near 0, as next established in Step 3-a (whose proof by contraction methods is appendicized.) □

Step 3-a *There exists* $\epsilon' > 0$ *such that* ϕ' *is continuous on* $[0, \epsilon']$.

Now, being convex on $[0, 1]$, the value function is a.e. twice differentiable; however, the easy deduction of Step 3 is simply not an option for the second derivative, since ϕ'' is not monotonic (see Figure 2). We proceed down a different route. Step 3-a rules out any 'kinks' near 0, where the first derivative jumps up. The next result asserts that near 0, the second derivative is not merely nonnegative (when it exists), but is in fact locally continuous. Trivially, this implies that it everywhere exists near 0.

Step 4 *There exists* $\epsilon \in (0, \epsilon')$ *such that the solution* φ''_{α_β} *to* (13) *is continuous on* $(0, \epsilon]$.

Proof: Consider instead (13) for fixed $\alpha > \alpha_\beta$. Recalling (14), $\sum_i w_{\alpha i}(x) < 1$ on some interval $[0, \epsilon]$, with $\epsilon \leq \epsilon'$ WLOG. Also, γ is continuous on $[0, \epsilon']$ by Step 3-a, and $\zeta_i(0) = 0$. So by Lemma 4 in Choczewski (1961),[11] φ''_α is continuous on $[0, \epsilon]$. Since $\varphi''_{\alpha_\beta} = x^{\alpha_\beta - \alpha}\varphi''_\alpha(x)$, we have φ''_{α_β} continuous on the left open interval $(0, \epsilon]$. □

We can now deduce our intuited asymptotic expression for the second derivative ϕ''.

Step 5 *We have* $\phi''(x) \sim c_\beta x^{-\alpha_\beta}$*, and therefore* $v''_\beta(x) \sim c_\beta x^{-\alpha_\beta}$.

Proof: By Step 4, $\varphi''_{\alpha_\beta}(x) \equiv x^{\alpha_\beta}\phi''(x)$ is continuous near 0. If $\varphi''_{\alpha_\beta}(x)$ converges to a finite constant c as $x \downarrow 0$, we are done — for then $\varphi''_{\alpha_\beta}(x) \sim c$; this means that $\phi''(x) \sim cx^{-\alpha_\beta}$, and $c = c_\beta$. Otherwise, $\varphi''_{\alpha_\beta}(x)$ explodes near 0, so that for all $M > 0$, there exists $\eta_M > 0$ with $x^{\alpha_\beta}\phi''(x) > M$ on $[0, \eta_M]$. Since $\phi'(0) = 0$,

$$\phi'(\eta_M) = \int_0^{\eta_M} \phi''(t)dt \geq \int_0^{\eta_M} Mt^{-\alpha_\beta}dt = M\eta_M^{1-\alpha_\beta}/(1-\alpha_\beta)$$

Since M is arbitrarily large for small η_M, this violates Step 3. Thus, $c = c_\beta < \infty$. □

[11] Lemma 4 also assumes that $\zeta_i(x) > x$ (false here) to strengthen his conclusion beyond continuity.

5.2 Non-Assortative Matching and Match Dynamics

A. The Failure of PAM. The major result of the paper is a sequel to Proposition 2. By simple substitution, the new high-skill concealing conditions below[12] (15a) and (15b) are met for open neighborhoods around the same hyperplanes $h = m$ and $\ell = m$ as (9). This result therefore admits the same intuition as Proposition 2.

Proposition 4 *In the infinite horizon model, there exists $\beta^* \in (0,1)$ such that for any $\beta \in (\beta^*, 1)$, PAM fails in a neighborhood of $(0,0)$ or $(1,1)$ iff:*

$$\sum_i \left[(m_i - h_i) \log \left(\frac{m_i}{\ell_i} \right) + \frac{m_i^2 - h_i \ell_i}{\ell_i} \right] > 0 \tag{15a}$$

or (respectively)

$$\sum_i \left[(m_i - \ell_i) \log \left(\frac{m_i}{h_i} \right) + \frac{m_i^2 - h_i \ell_i}{h_i} \right] > 0 \tag{15b}$$

Specifically, PAM fails around $(0,0)$ (resp. $(1,1)$) if β is high enough and output is sufficiently high-skill (resp. low-skill) concealing.

The proof is appendicized. For an overview, in the proof of Proposition 2 for the two period model, we established submodularity by explicitly evaluating the cross partial κ of Ψ^{v_1} at the extremes $x = y = 0$ and $x = y = 1$. In the infinite horizon model, the cross partial Ψ^β explodes as $x \to 0$ and $x \to 1$ (whenever $\chi_\beta > 1$, as it turns out). Hence, we must proceed asymptotically. How *fast* does Ψ^β explode? In the appendix, we use our approximation for v''_β to evaluate this cross partial near $x = 0$.

A natural question is: How commonly are inequalities (15a) and (15b) satisfied? They are satisfied as $N \to \infty$, as seen in the following proposition (proof appendicized).

Proposition 5 *With an atomless probability measure over the parameter space, the chance that PAM fails at one extreme converges to one as $N \to \infty$.*[13]

For insight, we argue here that the LHS of (15a) is a positive definite sum, plus one with a positive expectation. Indeed, the second sum $\sum_i (m_i^2 - h_i \ell_i)/\ell_i = \sum_i (m_i^2/\ell_i) - 1 = \sum_i (m_i/\ell_i)(m_i - \ell_i) > 0$, as the summands command weight $\gtreqless 1$ whenever $\gtreqless 0$. Assume that h, m, ℓ are each independently generated by a uniform measure on the simplex, denoted Δ_N. Consider the first sum. The map $(m,h) \mapsto (m-h)\log(m/\ell)$ is strictly convex in (m,h). By Jensen's inequality and $Em_i = Eh_i$, if $u_i \equiv \log(m_i/\ell_i)(m_i - h_i)$ then $E(u_i) > \log[(Em_i)/\ell_j](Em_i - Eh_i) = 0$, for all fixed ℓ_j. But we cannot apply a Law of Large Numbers as we have dependence across i among the summands ($\sum_i h_i = 1$, etc.), and the domain $(\Delta_N)^3$ changes with N. The appendicized proof uses Chebyshev's inequality to show that as $N \to \infty$: (a) $(1/n)\sum_i u_i \Rightarrow 0$ in probability, and (b) the second sum on the LHS of (15a) converges in probability to $c > 1$.

[12] While highly correlated, neither pair of conditions is implied by the other.

[13] Simulations suggest *extremely* rapid convergence. With uniformly generated parameters, (15a) and (15b) are simultaneously violated 43, 18, 5, and 1 of 1 billion times for the $N = 3, 4, 5, 6$ cases.

To underscore the probabilistic nature of this result, we observe that there are rare cases in which PAM obtains at 0 *and* 1. For a parameterized family, let $h = (3\varepsilon^2, 1/2 - 3\varepsilon^2, 1/2)$, $m = (\varepsilon, 1-2\varepsilon, \varepsilon)$, and $\ell = (1/2, 1/2 - 3\varepsilon^2, 3\varepsilon^2)$. Then Ψ^β is SPM at both extremes. Indeed, each LHS of (15a) and (15b) has the form (*constant*) + $\log(\varepsilon)/2 + O(\varepsilon)$, which is clearly negative for small ε. For intuition, note that this reduces to $h = (0, 1/2, 1/2)$, $m = (0, 1, 0)$, and $\ell = (1/2, 1/2, 0)$ for $\varepsilon = 0$. So self-matching all agents with reputations x near zero or one reveals much information about the actual types of those matched. Indeed, there is a roughly equal chance that both partners are immediately updated to surely G or surely B for these matches.

B. Long Run Match Dynamics. Our focus until now has been on a failure of PAM *in the large*. We now focus on the small. Indeed, in our title we promise to opine on the Beatles breakup. Not wishing to disappoint avid Beatles fans, we come to our discussion of that famous split. To this end, we first ask whether types are fully revealed in the limit. We apply a law of large numbers result for controlled processes (due to Easley and Kiefer (1988)) to establish in the appendix:

Lemma 3 *Fix an agent. Let x_t be his reputation at time t. Then $\langle x_t \rangle$ is a martingale, and $x_t \to 0$ or 1 with ex ante chances $1 - x_0$ and x_0 (i.e., when his type is B or G).*

Intuitively, Assumption 2 implies that in each period information is revealed about any agent's type (G or B). Thus, his type is eventually revealed. As limit beliefs must respect the martingale property, the limit distribution has the form given in Lemma 3.

Proposition 6 *Assume $\beta \in (0, 1)$ is high enough and output is sufficiently low-skill (resp. high-skill) concealing. Assume an individual's true type is in fact G (resp. B). Then, almost surely, he will eventually no longer assortatively match.*

Proposition 6 only speaks about one individual, since our model cannot distinguish between his matches with like-reputation agents. The problem with applying our insights to an observably long-lived partnership such as the Beatles is that matching here is assumed anonymous. That is, the relevant state space is not $\{G, B\}$, as we have, but rather $\{(G, G), (B, B), (G, B), (B, G)\}$. Of course, with a continuum of agents, completely dispensing with anonymity is infeasible. But we can relax anonymity somewhat.

Assumption 3 *The output history of currently matched individuals is observable.*

This captures the informational essence of long-lived matches, while adhering as much as possible to the general equilibrium spirit. As long as two agents have been matched, their outputs belong to recorded history; but if they break up, only their reputations remain. This generalization allows us to speak of partnerships in a meaningful sense. Yet an agent's reputation is still a sufficient statistic for the information revealed from all his previous matches. Given the production symmetry, any previous insights will carry over to this richer information structure. Lemma 3 and Proposition 6 imply:

Corollary 2 *Assume β high enough and production sufficiently low-skill (resp. high-skill) concealing. If both agents' types are G (resp. B), they will eventually break up.*

6 Conclusion

This paper merges insights from two fields: matching and learning. We have started with Becker's well-known paper on *static* matching theory, and added an economically important *dynamic* concern — reputation. We have found that this overturns Becker's finding of global PAM under supermodularity assumptions. With sufficient patience, PAM cannot arise in a stylized two period model. In the infinite horizon model, an unqualified failure no longer holds. Still, we show that PAM fails for high reputation agents if production is sufficiently low-skill concealing (the Beatles case). These conditions admit simple intuitions, and are so weak that one almost must hold — that is, for randomly chosen production technologies, with chance tending to one as the number of outputs N grows — and in practice, almost always at very low N, such as 3 or 4. Individuals are more likely to assortatively match when their reputations are low, while stars are more likely to match with those having unestablished reputations. Finally, this yields the dynamic result that matched stars eventually must split.

En route, our analysis has yielded two key theoretical insights that should prove valuable in future dynamic matching work. First, supermodular continuation values is at odds with convexity. Second, the convexity of the value function in our Bayesian learning model (and thus perhaps others) explodes near 0 and 1 at a geometric rate.

A Omitted Proofs

A.1 Existence, Welfare Theorems, and Values

• **Proof of Theorem 1.** Equip $Z \equiv \mathcal{L}_\infty([0,1])$ and $X \equiv \mathcal{L}_\infty([0,1]^2)$ with the standard norm topology. The dual X^* of X is the space of bounded measures on $[0,1]^2$. We endow X^* with the weak* topology. Let $\Phi : Z \to X^*$ be the correspondence that captures constraint (2):

$$\Phi(g) = \{\mu \in X^* : \lambda_g(A) \geq \mu(A \times [0,1]) \geq 0 \ \forall A \text{ measurable}\}, \ g \in Z$$

where $\lambda_g(A) \equiv \int_A g d\lambda$, and λ is Lebesgue measure. Define the Bayes operator B : $X^* \to Z$ by

$$B(\mu)(z) = \int \rho(z,x,y) d\mu(x,y), \ \mu \in X^*, \ z \in [0,1] \tag{16}$$

where $\rho(z,x,y)$ is the easily computed probability that x updates to z when matched with y plus the probability that y updates to z when matched with x. Let

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

where we have endowed W with the standard norm $\|\mathcal{V}\| = \sup_{\|g\| \leq 1} |\mathcal{V}(g)|$. Let $\Gamma(\mathcal{V},\mu) \equiv (1-\beta)\int f d\mu + \beta\mathcal{V}(\overline{g} + B(\mu))$, where the argument of the latter $\mathcal{V}$ captures constraint (3). Define the Bellman operator T, as follows: $T\mathcal{V}(g) = \max_{\mu \in \Phi(g)} \Gamma(\mathcal{V},\mu)$, where $g \in Z$. We prove below that T maps W into itself and is a contraction. Thus, it has a unique fixed point $\mathcal{V}$, by the Banach Fixed Point Theorem, solving (1), (2), (3).

Claim 1 Φ *is a continuous and compact valued correspondence.*

Proof: Alaoglu's Theorem states that if $\Phi(g) \subset X^*$ is weak* closed, bounded, and convex then $\Phi(g)$ is weak* compact. Convexity and boundedness are immediate. Let $\mathbb{I}_Y$ be the characteristic function of the set Y. Let $\mathbb{B}^A_{[a,b]} = \{\mu : a \leq \int \mathbb{I}_{A\times[0,1]} d\mu \leq b\}$, and likewise define notation for open intervals and half-open and half-closed intervals. Note that $\Phi(g) = \bigcap_A \mathbb{B}^A_{[0,\lambda_g(A)]}$ and that $\mathbb{I}_{A\times[0,1]} \in X$. By definition, $\mathbb{B}^A_{[0,\lambda_g(A)]}$ is weak* closed, and therefore $\Phi(g)$ is weak* closed.

We now show that this correspondence is upper and lower hemi-continuous (u.h.c. and l.h.c.). Now Φ is point closed, and we can assume WLOG that it maps into a compact subset[14] of X^*, say with upper bound $M < \infty$. Thus, we only need show that Φ has the closed graph property to prove u.h.c. Now, Φ has the closed graph property if for any $g \in Z$: $\mu \notin \Phi(g)$ implies that there exists an open set $\mathcal{O}$ that contains μ such that $\mathcal{O} \cap \Phi(g) = \emptyset$. But $\mu \notin \Phi(g)$ if $\mu > \lambda_g(A)$ for some A. The result follows from continuity of $\lambda_g(A) = \int \mathbb{I}_{A\times[0,1]} d\mu$ in μ.

For l.h.c., WLOG we only consider (basis) open sets of the form $\mathcal{O} = \bigcap_{k=1}^m \mathbb{B}^{A_k}_{(a_k,b_k)}$. Pick $\mu \in \mathcal{O} \cap \Phi(g)$, and let $\mu_\epsilon(A \times [0,1]) \equiv \mu(A \times [0,1]) - \epsilon\lambda(A)$ for all A. We claim that there exists $\delta, \epsilon > 0$ such that $\mu_\epsilon \in \mathcal{O} \cap \Phi(\hat{g})$ for all $\hat{g} \in Z$ with $\|\hat{g} - g\|_\infty < \delta$. Pick any such $\hat{g}$. For ϵ small enough, $\mu_\epsilon \in \mathcal{O}$. To show $\mu_\epsilon \in \Phi(\hat{g})$, i.e. $\lambda_{\hat{g}}(A) \geq \mu_\epsilon(A \times [0,1])$ for all A, first note that $|\lambda_g(A) - \lambda_{\hat{g}}(A)| < \delta\lambda(A)$ for all A, as shown below:

$$\begin{aligned}|\lambda_g(A) - \lambda_{\hat{g}}(A)| &= |\textstyle\int_A g d\lambda - \int_A \hat{g} d\lambda| \leq \int_A |g - \hat{g}| d\lambda \leq \int_A \sup_{x\in A} |g(x) - \hat{g}(x)| d\lambda \\ &= \sup_{x\in A} |g(x) - \hat{g}(x)| \lambda(A) \leq \|g - \hat{g}\|_\infty \lambda(A) < \delta\lambda(A)\end{aligned}$$

Thus, $\lambda_{\hat{g}}(A) > \lambda_g(A) - \delta\lambda(A)$, so that $\lambda_g(A) - \delta\lambda(A) \geq \mu(A \times [0,1]) - \epsilon\lambda(A)$ suffices. Since $\lambda_g(A) \geq \mu(A \times [0,1])$, it is enough that $\delta\lambda(A) \leq \epsilon\lambda(A)$, or $\delta \leq \epsilon$. □

Claim 2 $T : W \to W$.

Proof: The mapping clearly preserves boundedness and homogeneity. We now show that T preserves continuity. First, $\Gamma(\mathcal{V}, \mu)$ is weak* continuous in $(\mathcal{V}, \mu) \in W \times X^*$. Indeed, since $f \in X$, $\mu \mapsto \int f d\mu$ is weak* continuous on X^*. Similarly $\rho \in X$ yields $\mu \mapsto B(\mu)$ weak* continuous on X^*. For each $\mathcal{V} \in W$, the composition $\mathcal{V}(B(\mu))$ is continuous in μ. Thus, $\Gamma(\mathcal{V}, \mu)$ is continuous. Also, the constraint correspondence is continuous and compact valued by Claim 1. Then by Robinson and Day (1974) — a generalization of Berge's Theorem of the Maximum — $T\mathcal{V}$ is continuous. □

Claim 3 T *is a contraction.*

Proof: Indeed, T is monotonic and $T(\mathcal{V} + c) = T\mathcal{V} + \beta c$, where $0 < \beta < 1$ and c is real. Thus, T is a contraction by Blackwell's Theorem.[15] □

[14] Of course, the population size is unbounded, absent deaths. But $\Phi(g)$ and thereby $T\mathcal{V}(g)$ are linear functions of g. So we can work on a compact set of g's, and extend Φ and $T\mathcal{V}$ linearly.

[15] We are indebted to Ennio Stacchetti for providing the key insights for this proof. Any errors and preference for hemi- over semi-continuity are, of course, our responsibility.

- **Proof of Theorem 2:** Assume that $(\vec{\mu}, \vec{v}, \vec{w})$ is a CE, but $\vec{\mu}$ is not a PO. Thus, there exists feasible $\vec{\nu}$ with $\langle f, \vec{\nu}\rangle > \langle f, \vec{\mu}\rangle$. Define $w^y(x,y) = w(y|x)$. By definition of a CE and (6), we have $\langle \vec{w} + \vec{w}^y, \vec{\mu}\rangle = \langle f, \vec{\mu}\rangle$ and $\langle \vec{w} + \vec{w}^y, \vec{\nu}\rangle \geq \langle f, \vec{\nu}\rangle$.[16] Hence, $\langle \vec{w} + \vec{w}^y, \vec{\nu}\rangle > \langle \vec{w} + \vec{w}^y, \vec{\mu}\rangle$. By symmetry, $\langle \vec{w}, \vec{\nu}\rangle = \langle \vec{w}^y, \vec{\nu}\rangle$, and so $\langle \vec{w}, \vec{\nu}\rangle > \langle \vec{w}, \vec{\mu}\rangle$.

Let $\hat{g}_t$ be the density associated with matching $\vec{\nu}$, and define $\rho_t(z, x, \vec{\mu})$ as the revised probability that x at time 0 updates to z at time t. By worker maximization (7),

$$\sum_t \beta^t \int \int_{z \in supp\, g_t} w_t(z|y)\rho_t(z,x,\vec{\mu}) \frac{d\mu_t(z|y)}{g_t(z)} dy \geq \sum_t \beta^t \int \int_{z \in supp\, g_t} w_t(z|y)\rho_t(z,x,\vec{\nu}) \frac{d\nu_t(z|y)}{\hat{g}_t(z)} dy$$

Multiply both sides by $g_0(x)$, integrate over x, and note $g_t(z) = \int \rho_t(z, x, \vec{\mu}) g_0(x) dx$. This yields $\langle \vec{w}, \vec{\mu}\rangle \geq \langle \vec{w}, \vec{\nu}\rangle$, which contradicts $\langle \vec{w}, \vec{\nu}\rangle > \langle \vec{w}, \vec{\mu}\rangle$. Thus μ is a PO.

To establish that $\vec{v}$ is a multiplier in the planner's problem for the given (efficient) $\vec{\mu}$, we show that $(\vec{\mu}, \vec{v})$ satisfies the planner's FOC. Take any matched pair (x, y). If we sum the worker maximization conditions (7) for x and y we obtain:

$$v(x) + v(y) = (1-\beta)(w(x|y) + w(y|x)) + \beta \Psi^v(x,y)$$

Since $w(x|y) + w(y|x) = f(x,y)$, the planner's FOC (4) is satisfied for this matched pair. Now take any (x, y) (not necessarily matched). Worker maximization (7) implies:

$$v(x) \geq (1-\beta)w(x|y) + \beta\psi^v(x|y) \qquad \text{and} \qquad v(y) \geq (1-\beta)w(y|x) + \beta\psi^v(y|x)$$

Summing these two inequalities and applying (6) yields:

$$v(x) + v(y) \geq (1-\beta)(w(x|y) + w(y|x)) + \beta\Psi^v(x,y) \geq (1-\beta)f(x,y) + \beta\Psi^v(x,y) \ \square$$

- **Proof of Theorem 3:** Let $(\vec{\mu}, \vec{v})$ be a PO. Note that (2) and (3) are satisfied by assumption. Plugging the given wages (8) into (7), we see the given $\vec{\mu}$ satisfies (7) by the alternative representation of the planner's FOC (5). Summing the specified wages for any matched pair and multiplying both by $1-\beta$:

$$(1-\beta)(w(x|y) + w(y|x)) = (1-\beta)f(x,y) + (1-\beta)f(x,y) + \beta\Psi^v(x,y) - v(x) - v(y)$$

Along with the FOC of the social planner's problem (4), this implies that (6) holds. $\square$

- **Proof of Lemma 1:** We show that v cannot be linear, or piecewise linear.

Claim 4 *The value function v cannot be linear.*

Proof: If v is linear, then Ψ^v is amodular, and match values are strictly SPM (as f is strictly SPM), so that PAM obtains. But then $v(x) = (1-\beta)f(x,x)/2 + \beta\Psi^v(x,x)$. As f is strictly SPM, $f(x,x)$ is strictly convex, which contradicts v globally linear. $\square$

Claim 5 *If v is linear over some interval, then v is linear.*

Proof: Being convex (Lemma 1), v is continuous. Any maximal interval of linearity in $[0,1]$ is closed, say $[\underline{x}, \bar{x}]$. By continuity of $z_i(x,y)$ and Assumption 2, $\exists \varepsilon \in (0, \bar{x} - \underline{x})$, such that $\forall\ y\ \exists\, i$ s.t. $z_i(\bar{x} - \varepsilon, y) > \bar{x}$. The logic of Lemma 2 yields Ψ^v strictly convex at $\bar{x} - \varepsilon$ for all y. But $v(\bar{x} - \varepsilon) = \max_y (1-\beta)f(\bar{x} - \varepsilon, y) + \beta\Psi^v(\bar{x} - \varepsilon, y) - v(y)$ by (5), where the maximand is strictly convex at $\bar{x} - \varepsilon$, contradicting v linear on $[\underline{x}, \bar{x}]$. $\square$

[16] We abuse notation by using a vector of measurable functions as a *first* argument in $\langle \cdot, \cdot \rangle$.

A.2 Extremal Convexity of the Value Function

Define $\Upsilon(x) \equiv \beta \sum_i \pi_i(x)(\zeta_i(x)/x)^{2-\alpha_\beta}$ and $\hat{\Upsilon}(x) \equiv \beta \sum_i \pi_i(x)(\zeta_i(x)/x)^2$.

Claim 6 *$\Upsilon(x) = 1 + O(x)$, and $\hat{\Upsilon}(x) = \chi_\beta + O(x)$, recalling $\chi_\beta = \beta \sum_i m_i^2/\ell_i > 1$.*

Proof: This follows directly from taking a Taylor Series about $x = 0$. $\square$

Claim 7 *$0 < \underline{c} \equiv \min_x \Upsilon(x) < 1$, $0 < \underline{\hat{c}} \equiv \min_x \hat{\Upsilon}(x)$, and $\bar{c} \equiv \max_x \Upsilon(x) < \infty$.*

Proof: Clearly, Υ and $\hat{\Upsilon}$ are continuous on $[0, 1]$. Further, $\pi_i(x)$ and $\zeta_i(x)/x$ are positive on $[0, 1]$. So $\underline{c} > 0$ and $\underline{\hat{c}} > 0$. Finally, $\Upsilon(0) = 1$, and $\Upsilon'(0) < 0$ forces $\underline{c} < 1$. $\square$

Claim 8 *If $\Upsilon(x) \geq \beta$, then $\hat{\Upsilon}(x) > \Upsilon(x)$, and $\underline{d} \equiv \min\{\hat{\Upsilon}(x) - \Upsilon(x) : \Upsilon(x) \geq \beta\} > 0$.*

Proof: Since not all $u_i \equiv \zeta_i(x)/x$ are equal, $(\sum_i p_i u_i^2)^{1/2} > (\sum_i p_i u_i^{2-\alpha_\beta})^{1/(2-\alpha_\beta)}$, by Theorem 96 in Hardy, Littlewood, and Polya (1952). Simple algebra yields $\hat{\Upsilon}(x)/\beta > (\Upsilon(x)/\beta)^{2/(2-\alpha_\beta)} \geq \Upsilon(x)/\beta$, the second inequality owing to $\alpha_\beta \in (0, 1)$ and $\Upsilon(x)/\beta \geq 1$. Finally, $\hat{\Upsilon}(x) - \Upsilon(x)$ attains its minimum $\underline{d} > 0$, being continuous on $[0, 1]$. $\square$

- **Proof of Step 2-a.** Let $\varphi_0(x) = \underline{a}x^{2-\alpha_\beta}$, where $\underline{a}$ is determined below. By the monotonicity of T, it suffices that $T\varphi_0(x) \geq \varphi_0(x)$, or:

$$(1 - \beta)\sigma x^2 + \underline{a}x^{2-\alpha_\beta}\Upsilon(x) \geq \underline{a}x^{2-\alpha_\beta} \tag{17}$$

Claim 9 *There exists $\varepsilon > 0$ such that (17) holds for all $x \leq \varepsilon$ and $x \geq 1 - \varepsilon$.*

Proof: We now prove this in the neighborhood of $x = 0$. (The $x = 1$ case is similar.) By Claim 6, inequality (17) obtains *iff*

$$(1 - \beta)\sigma + \underline{a}x^{-\alpha_\beta}(1 + O(x)) \geq \underline{a}x^{-\alpha_\beta}$$

Rearrangement yields $(1 - \beta)\sigma + \underline{a}x^{-\alpha_\beta}O(x) \geq 0$, valid since $O(x^{1-\alpha_\beta}) = o(1)$. $\square$

Claim 10 *Inequality (17) holds for all $x \in [\varepsilon, 1 - \varepsilon]$.*

Proof: Replacing $\Upsilon(x)$ by its lower bound $\underline{c}$ from Claim 7, we need $(1-\beta)\sigma + \underline{a}x^{-\alpha_\beta}\underline{c} \geq \underline{a}x^{-\alpha_\beta}$ to prove (17). Equivalently, $x^{\alpha_\beta}(1 - \beta)\sigma \geq (1 - \underline{c})\underline{a}$, whose LHS is minimized on $[\varepsilon, 1 - \varepsilon]$ at $x = \varepsilon$. It suffices that $\underline{a} \leq \varepsilon^{\alpha_\beta}(1 - \beta)\sigma/(1 - \underline{c})$, which is positive. $\square$

- **Proof of Step 2-b.** We need $T\varphi_0(x) \leq \varphi_0(x)$, where $\varphi_0(x) = \bar{a}x^{2-\alpha_\beta} - bx^2$, or

$$(1 - \beta)\sigma x^2 + \beta \sum_i \pi_i(x)[\bar{a}\zeta_i(x)^{2-\alpha_\beta} - b\zeta_i(x)^2] \leq \bar{a}x^{2-\alpha_\beta} - bx^2$$

This expression is seen to be equivalent to $(\bar{a}, b) \in S(x)$ for all x, where

$$S(x) \equiv \{(\bar{a}, b) \in \mathbb{R}^2 | (1 - \beta)\sigma \leq \bar{a}x^{-\alpha_\beta}(1 - \Upsilon(x)) - b(1 - \hat{\Upsilon}(x))\}$$

Claim 11 *If* $S_1 \equiv \{(\bar{a}, b) | (1-\beta)\sigma < b(\chi_\beta - 1)\}$ *then* $S_1 \subset S(x)$ *for all* $x \in [0, \delta]$ $(\delta > 0)$.

Proof: Replacing $\Upsilon(x)$ and $\hat{\Upsilon}(x)$ in $S(x)$ by their asymptotic forms (i.e. in some $[0, \delta]$) yields: $(1-\beta)\sigma \le \bar{a}x^{-\alpha_\beta}O(x) - b(1-\chi_\beta + O(x))$. Since $x^{-\alpha_\beta}O(x) = o(1)$, if $(1-\beta)\sigma < b(\chi_\beta - 1)$, or equivalently $(\bar{a}, b) \in S_1$, then $(\bar{a}, b) \in S(x)$. □

Claim 12 *Define* $S_2 \equiv \{b > \bar{a}\delta^{-\alpha_\beta}\}$ *and* $S_3 \equiv \{\sigma \le \bar{a} + b(\underline{d} - 1 + \beta)/(1-\beta)\}$. *Then* $S_2 \cap S_3 \cap \{\Upsilon(x) > \beta\} \subset S(x)$ *for all* $x \ge \delta$.

Proof: Since the RHS of the inequality in $S(x)$ is rising in $\hat{\Upsilon}(x)$, we can replace $\hat{\Upsilon}(x)$ by its lower bound $\Upsilon(x) + \underline{d}$ (Claim 8) to get the following sufficient condition for $S(x)$:

$$(1-\beta)\sigma \le (\bar{a}x^{-\alpha_\beta} - b)(1 - \Upsilon(x)) + b\underline{d} \tag{18}$$

If $(\bar{a}, b) \in S_2$, then $\bar{a}x^{-\alpha_\beta} - b \le 0$ for all $x \ge \delta$. The following inequality suffices for (18): $(1-\beta)\sigma \le (\bar{a} - b)(1-\beta) + b\underline{d}$. Rearranging we find $(\bar{a}, b) \in S_3$. □

Claim 13 *Define* $S_4 \equiv \{\sigma + b(1 - \underline{\hat{c}})/(1-\beta) \le \bar{a}\}$. *Then* $S_4 \cap \{\Upsilon(x) \le \beta\} \subset S(x)$ *for all* $x \ge \delta$.

Proof: The RHS of the inequality in $S(x)$ is falling in $\Upsilon(x)$ and rising in $\hat{\Upsilon}(x)$, so set $\hat{\Upsilon}(x) = \underline{\hat{c}}$ (the minimum established in Claim 7) and set $\Upsilon(x) = \beta$. Thus, $(1-\beta)\sigma \le \bar{a}(1-\beta) - b(1 - \underline{\hat{c}})$ implies $S(x)$. Rearranging, this becomes $(\bar{a}, b) \in S_4$. □

Finally, we prove $S = S_1 \cap S_2 \cap S_3 \cap S_4 \neq \emptyset$. Indeed, $(\bar{a}, b) \in S_1 \cap S_3$ for all large enough $\bar{a}, b$. Next, $(\bar{a}, b) \in S_2$ iff b exceeds a linear function of a with adjustable slope, while $(\bar{a}, b) \in S_4$ iff b less than a *given* linear function of $\bar{a}$. We can choose the adjustable slope in S_2 low enough (large δ) such that $S_2 \cap S_4 \neq \emptyset$, since it contains the large values of $\bar{a}, b$ lying inside $S_1 \cap S_3$. We thus find $(\bar{a}, b) \in S_1 \cap S_2 \cap S_3 \cap S_4$. □

- **Proof of Step 3-a.** Differentiating equation (11) yields:

$$\phi'(x) = \eta(\phi(x), x) + \beta \textstyle\sum_i \pi_i(x)\zeta_i'(x)\phi'(\zeta_i(x))$$

where $\eta(\phi(x), x) \equiv 2(1-\beta)\sigma x + \beta \sum_i p_i'(x)\phi(\zeta_i(x))$. From this we form the functional equation:

$$\varphi'(x) = \eta(\phi(x), x) + \beta \textstyle\sum_i \pi_i(x)\zeta_i'(x)\varphi'(\zeta_i(x)) \tag{19}$$

Notice that we have used ϕ as an argument in η, while allowing φ' to vary. Clearly, $\varphi' = \phi'$ solves this functional equation. Multiply both sides of (19) by $x^{\alpha-1}$ for some $\alpha > \alpha_\beta$ to get:

$$\varphi_\alpha'(x) = \eta(\phi(x), x)x^{\alpha-1} + \textstyle\sum_i \hat{w}_{\alpha i}(x)\varphi_\alpha'(\zeta_i(x)) \tag{20}$$

where $\varphi_\alpha'(x) \equiv \varphi'(x)x^{\alpha-1}$ and $\hat{w}_{\alpha i}(x) \equiv \beta \sum_i \pi_i(x)\zeta_i'(x)(x/\zeta_i(x))^{\alpha-1}$. Since $\sum_i \hat{w}_{\alpha i}(0) = \sum_i w_{\alpha i}(0) = \beta \sum_i \ell_i(m_i/\ell_i)^{2-\alpha}$, and $\hat{w}_{\alpha i}$ is continuous, we have $\sum_i \hat{w}_{\alpha i}(x) < 1$ on some $[0, \epsilon']$, just as in (14). Finally, η is continuous on $[0, \epsilon']$ and $\zeta_i(0) = 0$. Altogether, Lemma 4 in Choczewski (1961) implies that the solution to (20) uniquely exists on $[0, \epsilon']$, and is continuous. Consequently, $\phi'(x)$ also is continuous on $[0, \epsilon']$. □

A.3 Non-Assortative Matching

• **Proof of Proposition 4.** Computing the exploding cross partial derivative, we have

$$\kappa_\beta(x) \equiv \Psi^\beta_{12}(x,x) = \textstyle\sum_i (\sigma_i v_\beta(z_i) + a_i(x) v'_\beta(z_i) + b_i(x) v''_\beta(z_i)), \tag{21}$$

where $\sigma_i = h_i - 2m_i + \ell_i$, $a_i(x) = h_i - m_i + O(x)$, and $b_i(x) = [(h_i\ell_i - m_i^2)m_i/\ell_i^2]x + O(x^2)$. Proposition 3 proved that $v''_\beta(x) = c_\beta x^{-\alpha_\beta}(1+o(1))$. Substituting the integrated asymptotic expressions for v_β, v'_β into (21), and using $z_i(x) = (m_i/\ell_i)x + O(x^2)$, yields:

$$\begin{aligned}
\sum_i \sigma_i v_\beta(z_i) &= \sum_i \sigma_i \left[\frac{c_\beta}{(1-\alpha_\beta)(2-\alpha_\beta)} z_i^{2-\alpha_\beta} + v'_\beta(0) z_i + v_\beta(0) + o(x^{2-\alpha_\beta})\right] \\
&= \left(\textstyle\sum_i \sigma_i (m_i/\ell_i)\right) v'_\beta(0) x (1+o(1)) \\
\sum_i a_i(x) v'_\beta(z_i) &= \sum_i (h_i - m_i + O(x)) \left[v'_\beta(0) + \frac{c_\beta}{1-\alpha_\beta} z_i^{1-\alpha_\beta}(1+o(1))\right] \\
&= \frac{c_\beta}{1-\alpha_\beta} \sum_i (h_i - m_i) \left(\frac{m_i}{\ell_i}\right)^{1-\alpha_\beta} x^{1-\alpha_\beta}(1+o(1)) \\
\sum_i b_i(x) v''_\beta(z_i) &= c_\beta \sum_i \left[\frac{(h_i\ell_i - m_i^2)m_i}{\ell_i^2} x + O(x^2)\right] \left(\frac{m_i}{\ell_i} x\right)^{-\alpha_\beta} (1+o(1)) \\
&= c_\beta \sum_i \frac{h_i\ell_i - m_i^2}{\ell_i} \left(\frac{m_i}{\ell_i}\right)^{1-\alpha_\beta} x^{1-\alpha_\beta}(1+o(1))
\end{aligned}$$

The lowest order term $x^{1-\alpha_\beta}$ in (21) has coefficient $c_\beta[R_1(\alpha_\beta)+(1-\alpha_\beta)R_2(\alpha_\beta)]/(1-\alpha_\beta)$, where $R_1(\alpha_\beta) \equiv \sum_i (h_i - m_i)(m_i/\ell_i)^{1-\alpha_\beta}$ and $R_2(\alpha_\beta) \equiv \sum_i [(h_i\ell_i - m_i^2)/\ell_i](m_i/\ell_i)^{1-\alpha_\beta}$. We want $\kappa_\beta(x) < 0$ in a neighborhood of 0, for large enough $\beta < 1$. To prove this, it suffices that $R(\alpha) \equiv R_1(\alpha) + (1-\alpha)R_2(\alpha) < 0$ for large enough $\alpha < 1$. Since $R(1) = 0$, we are done if $R'(1) > 0$. For then $\kappa_\beta(x) < 0$ for β near 1 for x small enough. Finally, differentiation reveals that $R'(1)$ is the LHS of (15a), which is strictly positive. □

• **Proof of Proposition 5.** Given the atomless assumption, it is WLOG to assume that $h^n = (h^n_1, h^n_2, \ldots, h^n_n)$ is a r.v. uniformly distributed (i.e. Lebesgue measure λ) on Δ_n. Thus $(n-1)!$ is the joint density, with marginals $d\lambda_i(h^n_i) = (n-1)(1-h^n_i)^{n-2}$ and $d\lambda_{ij}(h^n_i, h^n_j) = (n-1)(n-2)(1-h^n_i-h^n_j)^{n-3}$.

We wish to show that measure of parameters (h, m, ℓ) for which (15) fails (so the opposite weak inequality holds) vanishes as n increases. That is, if we define $s^1_n = (1/n)\sum_i \log(m^n_i/\ell^n_i)(m^n_i - h^n_i)$ and $s^2_n = (1/n)\sum_i (m^n_i)^2/\ell^n_i$, then $Pr[s^1_n + s^2_n > 1] \to 1$.

Claim 14 *The first sum in* (15) *vanishes in probability:* $s^1_n \Rightarrow 0$.

Proof: First $Es^1_n \to 0$. Indeed, straightforward calculation using the densities given above yields that $Es^1_n = ((n-1)/n^2)(\varsigma + \Gamma'(n)/\Gamma(n))$. Here ς is Euler's constant and

$\Gamma(n) \equiv \int_0^\infty s^{n-1}e^{-s}ds$ is the Gamma function, i.e. $\Gamma'(n) = \int_0^\infty \log(s)s^{n-1}e^{-s}ds$. It is known that (the 'ψ function') $\Gamma'(n)/\Gamma(n) \sim \log n$ as $n \to \infty$, and so $Es_n^1 \to 0$.

Then by routine calculations, $var(s_n^1) \to 0$. Thus, $Pr[|s_n^1 - Es_n^1| \geq \varepsilon] \to 0\ \forall\varepsilon$. Finally, Chebyshev's inequality states that $Pr[|s_n^1 - Es_n^1| \geq \varepsilon] \leq var(s_n^1)/\varepsilon^2\ \forall\varepsilon$. $\square$

Claim 15 *The second sum in* (15) *converges in probability to some* $c \geq 2$*:* $s_n^2 \Rightarrow c \geq 2$.

Proof: Define $\tilde{\ell}_i^n = \ell_i^n$ if $\ell_i^n \geq 1/n^2$ and $1/n^2$ otherwise. Let $\tilde{s}_n^2 = s_n^2$ with ℓ_i^n replaced with $\tilde{\ell}_i^n$, and note that $s_n^2 \geq \tilde{s}_n^2$. It suffices to prove $\tilde{s}_n^2 \Rightarrow 2$.

We first claim that $E\tilde{s}_n^2 \to 2$. To this end, note that $E(1/\tilde{\ell}_i^n) = n^2 + o(n^2)$. To see this let $\rho^n = 1 - (1 - 1/n^2)^{n-1}$ be the chance that $\ell_i \leq 1/n^2$. Then $E(1/\tilde{\ell}_i^n) = n^2\rho^n + \int_{1/n^2}^1 (n-1)(1-s)^{n-2}/sds$. The latter term is bounded above by $(1-\rho^n)n^2$, and $\rho^n \to 1$, proving the result. Finally, $E(m_i^n)^2 = 2/n(n+1)$, so that $E\tilde{s}_n^2 \to 2$.

It thus suffices that $Pr[|\tilde{s}_n^2 - E\tilde{s}_n^2| \geq \varepsilon] \to 0\ \forall\varepsilon$. By similar reasoning as above, $E(1/(\tilde{\ell}_i^n)^2) = n^4 + o(n^4)$, and $E((1/\tilde{\ell}_i^n)(1/\tilde{\ell}_j^n)) = n^4 + o(n^4)$. Also, $E((m_i^n)^4) = 24/(n(n+1)(n+2)(n+3))$, and $E((m_i^n)^2(m_j^n)^2) = 4/(n(n+1)(n+2)(n+3))$. Then

$$var(\tilde{s}_n^2) = \frac{1}{n}E\frac{(m_i^n)^4}{(\tilde{\ell}_i^n)^2} + \frac{n-1}{n}E\frac{(m_i^n m_j^n)^2}{\tilde{\ell}_i^n\tilde{\ell}_j^n} - \left(E\frac{(m_i^n)^2}{\tilde{\ell}_i^n}\right)^2 = \frac{4n^2(n-1)}{(n+1)^2(n+2)(n+3)} + o(n)$$

using the independence of m^n from $\tilde{\ell}^n$ — i.e., $var(\tilde{s}_n^2) \to 0$. Apply Chebyshev. $\square$

• **Proof of Lemma 3.** By Easley and Kiefer (1988), controlled stochastic processes may only converge to potentially confounding beliefs, where the belief remains unchanged. Now, if $x \neq 0$ then $z_i(x,y) = x$ *iff* $p_i(1,y) = p_i(x,y)$. Since $\partial p_i(x,y)/\partial x$ is constant in x, this requires $\partial p_i(x,y)/\partial x = (h_i + \ell_i - 2m_i)y + m_i - \ell_i = 0$, or $y = (m_i - \ell_i)/(h_i + \ell_i - 2m_i)\ \forall\ i$, contrary to Assumption 2. But absent interior potentially confounding beliefs, the long run distribution g_∞ has support $\{0,1\}$. As reputation is a martingale, the weights are $(1-x_0, x_0)$. $\square$

References

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

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

Choczewski, B. (1961): "On Continuous Solutions of Some Functional Equations of the N-th Order," *Annales Polonices Mathematica*, 11, 123–32.

De Bruijn, N. G. (1958): *Asymptotic Methods in Analysis*. North-Holland.

Easley, D., and N. Kiefer (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica*, 56(5), 1045–1064.

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

Hardy, G., J. Littlewood, and G. Polya (1952): *Inequalities*. Cambridge, MA: Cambridge University Press, second edn.

Kremer, M., and E. Maskin (1996): "Wage Inequality and Segregation by Skill," *NBER Working Paper*, (5718).

Lorentz, G. (1953): "An Inequality for Rearrangements," *American Mathematical Monthly*, 60, 176–179.

Mortensen, D. (1982): "Property Rights and Efficiency in Mating, Racing, and Related Games," *American Economic Review*, 72(3), 968–979.

Robinson, S., and R. Day (1974): "A Sufficient Condition for Continuity of Optimal Sets in Mathematical Programming," *Journal of Mathematical Analysis and Applications*, 45, 506–511.

Roth, A., and O. Sotomayor (1990): *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*. Cambridge, MA: Cambridge University Press.

Shimer, R., and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

Smith, L. (1997): "The Marriage Model with Search Frictions," MIT working Paper.

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

Wolinsky, A. (1990): "Information Revelation in a Market with Pairwise Meetings," *Econometrica*, 58(1), 1–23.