# Networks in Labor Markets: Wage and Employment Dynamics and Inequality*

Antoni Calvó-Armengol† and Matthew O. Jackson‡

October 2001
Revision: August 1, 2005
forthcoming in *The Journal of Economic Theory*§

**Abstract**

We present a model of labor markets that accounts for the social network through which agents hear about jobs. We show that both wages and employment are positively associated (a strong form of correlation) across time and agents. We also analyze the decisions of agents regarding staying in the labor market or dropping out. If there are costs to staying in the labor market, then networks of agents that start with a worse wage status will have higher drop-out rates and there will be a persistent differences in wages between groups according to the starting states of their networks.

Keywords: Networks, Labor Markets, Employment, Unemployment, Wages, Wage Inequality, Drop-Out Rates.

JEL Classification Numbers: D85, A14, J64, J31, J70, L14.

---

*We thank Valentina Bali, Kim Border, Antonio Cabrales, Janet Currie, Isa Hafalir, Eddie Lazear, Massimo Morelli, and David Pérez-Castrillo, for helpful conversations and discussions. We also thank an associate editor and referees for their comments and suggestions. We gratefully acknowledge the financial support of Spain's Ministry of Education under grant BEC2002-02130, Fundación Ramón Areces, the Barcelona Economics Program CREA, and the Lee Center for Advanced Networking at Caltech.

†Universitat Autònoma de Barcelona, Department of Economics, Edifici B, 08193 Bellaterra (Barcelona), Spain. Email: antoni.calvo@uab.es, http://selene.uab.es/acalvo

‡Division of Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena, California 91125, USA. Email: jacksonm@hss.caltech.edu, http://www.hss.caltech.edu/~jacksonm/Jackson.html

§This paper was formerly part of Calvó-Armengol and Jackson (2004). That paper was split into two parts, with the part under the former title looking at a special case of the networks outlined here and focusing on employment dynamics, and this part looking at a more general set of networks and exploring both wage and employment dynamics.

# 1 Introduction

One of the most extensively studied issues in labor economics is the persistent inequality in wages between whites and blacks.[1] Even if one believes any inequality in wages between social groups to be entirely explainable by differences in factors such as education, skills, and drop-out rates; one is still left to explain why those should differ.[2]

The purpose of this paper is to develop a model of how social networks operate in the transmission of job information, and to show that such a model can account for the observed patterns of wages and employment as well as differences in drop-out patterns and their roles in sustaining inequality. An analysis of social networks provides a basis for observing both higher drop-out rates in one race versus another and sustained inequality in wages and employment rates even among those remaining in the labor force.[3]

Our model builds upon a well-established stylized fact: a significant fraction of all jobs are found through contacts. While estimates of the percentage of jobs found through social contacts vary across location and profession, they consistently range between 25 and 80% of jobs in a given profession.[4] We model the transmission of job information among individuals by a function that keeps track of who first heard about a job and who (if anyone) eventually ended up getting an offer for that job. The key condition that we impose on this function is that the expected number of offers that a given agent ends up with is nondecreasing in the wage status of other agents. Allowing agents also to randomly lose jobs, wages can be shown to follow a Markov process, with state transitions depending on the information transmission network. We prove that the resulting stationary distribution is strongly associated; that is, the wages of any path-connected agents are positively correlated under the steady-state distribution. The proof is not as easy as one might expect, as there is a countervailing effect that path-connected agents are sometimes in competition for information about certain jobs. This entails some within period negative correlation among the status of certain agents. So we have to prove that the long run benefits of improved status of friends-of-friends outweighs the short run competition that they might represent. Next, to establish persistent inequality between wages of different types of agents, we analyze drop-out decisions where agents decide whether to enter the labor market or to drop out. We model the drop-out decision as a simultaneous-move game, where agents compare the discounted expected flow of future wages stemming from entering the labor force with the corresponding discounted costs (such as education costs, opportunity costs, skills maintenance, etc.). Because individual wages are positively associated across agents, entry decisions in this entry/drop-out game turn

[1] For instance, see Smith and Welch (1989), Card and Krueger (1992), Chandra (2000), Heckman, Lyons, and Todd (2000). See Farley (1990) for data on other racial- ethnic groups in the U.S.

[2] The extent to which inequality is explainable by such factors is still a point of some debate. See for instance, Darity and Mason (1998) and Heckman (1998).

[3] A social network model of inequality complementary to other theories. For discussion of some other theories and the relation of social networks approach, see Calvó-Armengol and Jackson (2004).

[4] See Ioannides and Loury (2004) for an excellent and extensive survey on the role of social networks in labor markets.

out to be strategic complements. Applying the theory of supermodular games, we deduce that two different social groups with identical job information networks but differing in their starting wage and employment profile will have different drop-outs rates that can be strictly ranked. These differences in drop-out patterns in turn breed persistent differences in wages between the two groups. This theory thus highlights the role of collective employment history in persistent wage inequality across social groups.[5]

This paper has a companion paper: Calvó-Armengol and Jackson (2004), which examines a specific case of the model considered here. The main contributions here are twofold. First, we consider a much more general model both in the passing of job information and in the structure of wages and their relation to job offers. Most applications would fit into the broader class examined here rather than the specific case examined in the companion paper. Second, we study wage dynamics rather than just employment dynamics, which is important as much of the empirical evidence for inequality relates to wage differences between races.

We note that the techniques and results that we have developed here also have applications beyond labor markets. In particular, the association and correlation patterns that we find based on the underlying $p$'s generalize to very different interpretations of the $p$'s $-$ such as the chance that some agent influences the behavior of another agent (such as inducing them to take up smoking, join a club, engage in crime, etc.). Thus, it can be that the results and tools developed here will be of use in modeling a wide variety of social interactions where networked relations are important.

## 2 A Model of Networks in Labor Markets

We begin with a formal description of our model.

### 2.1 Wage and Employment Status

The random variable $W_{it}$ keeps track of the wage of agent $i$ at time $t$. $W_{it}$ takes on values in $I\!R_+$; the unemployment wage is 0. The vector $w_t = (w_{1t}, \ldots, w_{nt})$ is a realization of wage levels at $t$.[6]

We represent random variables by capitol letters and realizations by small letters. The sequence of random variables $\{W_0, W_1, W_2, \ldots\}$ comprises the stochastic process of wage status.

The random variable $S_{it}$ is the employment status of agent $i$ at time $t$. Employment status is derived from wage status. We set $s_{it} = 1$ when $i$ is employed, and $s_{it} = 0$ otherwise. So, the vector $s_t \in \{0, 1\}^n$ is a realization of the employment status at $t$.

[5]This complements a large body of theoretical work built on models of discrimination (e.g., Becker 1957, Arrow 1972), imperfect capital markets (Loury 1981), and local public goods (Benabou 1993, 1996 and Durlauf 1996), among others. See Calvó-Armengol and Jackson (2004) for additional discussion.

[6]While we use the term wages, this random variable might be thought of as representing the expected discounted value of wages in a position. This distinction can be important in situations where some jobs have lower starting salaries but higher overall discounted sum of future wages.

## 2.2 Labor Market Turnover

Labor market turnover proceeds repeatedly through two phases.

- In one phase, agents hear about new jobs. If an agent directly hears about a job vacancy, then she either keeps that information or passes the job on to one of her contacts in the network.
- In the other phase, each currently employed agent $i$ is fired with probability $b_i \in (0,1)$, termed the *breakup rate*.

These phases occur repeatedly over time. The way we index periods is thus un-important. It is convenient to consider the hiring phase first and then the breakup phase.

## 2.3 Specifics of Information Transmission

The job transmission and offer generation is described by a function $p_{ij} : \mathbb{R}^n_+ \to [0,1]$. Here $p_{ij}(W_{t-1})$ is the probability that $i$ originally hears about a job and then it is eventually $j$ that ends up with an offer for that job.

The function $p_{ij}$ is a reduced form that can accommodate a variety of situations, including selective passing of information, passing multiple times and/or to multiple agents, and competition among agents for the same job. All that is important for our analysis is to keep track of who first heard about a job and who (if anyone) eventually ended up getting an offer for the job.

Let $p_i(w) = \sum_j p_{ji}(w)$. This is the expected number of offers that $i$ will get when the wage state in the last period is $w$. We take the realizations under $p_{ji}(w)$ and $p_{ki}(w)$ independent.[7]

Let $p$ denote the vector of functions across $i$ and $j$. Let $\overline{w}$ denote the maximum value in the range of wages. The functions $p_{ij}$ are assumed to satisfy the following conditions on their support:

(1) $p_{ji}(w)$ is nondecreasing in $w_{-i}$ and nonincreasing in $w_i$ for every $ji$,

(2) $p_{ji}(w) > 0$ implies $p_{ji}(w_{-i}, w'_i) > 0$ when $w_i \leq w'_i < \overline{w}_i$ for each $ji$; and $p_i(w) > 0$ for any $w$ and $i$ such that $w_i < \overline{w}_i$

(3) if $p_i(w) > p_i(w_{-j}, \widetilde{w}_j)$ for $j \neq i$ and $\widetilde{w}_j$, then $p_i$ is increasing in $w_j$ whenever $w_i < \overline{w}_i$.

(1) encompasses the idea that other agents are (weakly) more likely to directly or indirectly pass information on that will reach $i$ if they are more satisfied with their own position, and also that they might have better access to such information as their situation improves. It also encompasses the idea that other agents are (weakly) less likely to compete with $i$ for an offer if they are more satisfied with their own position. The second requirement is similar but keeps track of $i$'s wage.

[7]Note that this is very different from the realizations under $p_{ij}$ and $p_{ik}$, which will generally be negatively correlated. So we are just assuming that $j$ and $k$ do not coordinate on whether they pass $i$ a job. If indirect passing is present, then this embodies an assumption that the correlation in indirect passing is negligible. This is only a simplifying assumption, as when periods become small the probability of more than one job being in the system at a time becomes negligible.

Note that this allows for $i$ to be more likely to directly hear about a job as $i$'s situation worsens (allowing for a greater search intensity).

(2) requires that if an agent is not at their highest wage level, then there is some probability that they will obtain an offer both directly and from other agents who pass them information. This is important in making sure that $i$ does not have incentives to turn down a job in order to influence the probability of getting multiple offers in some future period.

(3) is a simplifying assumption. This guarantees that if $i$'s probability of hearing about a job is sometimes sensitive to $j$'s status, then it is sensitive to $j$'s status whenever $i$ is not at the highest wage level. This simply allows us to make statements about strictly positive correlations that do not need to be conditioned on particular circumstances.

Let us briefly describe a couple of examples that fit into our model.

First, consider a network of connections among agents described by a weighted and directed graph $g$, an $n \times n$ matrix with $g_{ij} \in \mathbb{R}_+$. Jobs are all identical (e.g., unskilled labor) and wages depend only on whether a worker is employed or not. At the beginning of each period, agent $i$ hears of some available job with probability $a_i \in (0,1)$. When $i$ hears of a job and is unemployed, then $i$ takes the job. If $i$ is employed, then he randomly picks an unemployed acquaintance with weights proportional to $g_{ij}$ and passes the information along. If all direct contacts are employed, then the information is lost. The special case where the $a_i$'s are the same across $i$ and where $g_{ij}$ takes on values either 0 and 1 and $g_{ij} = g_{ji}$ is analyzed in Calvó-Armengol and Jackson (2004). We can also allow agents to relay information in the case that all of their acquaintances are employed.

Next, let jobs be heterogeneous and wages take on different values. Let $\overline{w}_i$ be the highest wage attainable by agent $i$. At the beginning of each period, agent $i$ hears about a job opening offering a wage $w_i$ with probability $a_i^{w_i} \in (0,1)$. If $i$ directly hears about a new job that pays a higher wage than $i$'s current job position, then agent $i$ keeps that information.[8] If the new job does not offer any improvement, then agent $i$ randomly passes the information on to one of his direct contacts with a current wage lower than that of the new job, with weights related to link intensities.

## 2.4 The Determination of Offers, Wages, and Employment

**Determination of Offers**

Let $O_{it}$ be the random variable denoting the number of new opportunities that $i$ has in hand at the end of the hiring process in period $t$. Given $W_{t-1} = w$, the distribution of $O_t$ is governed by the realizations of the $p_{ij}(w)$'s.

**Determination of Employment**

The employment status evolves as follows. If agent $i$ was employed at the end of time $t-1$ and/or receives offers in period $t$, then the agent is employed ($S_{it} = 1$) with probability $(1 - b_i)$ and is unemployed ($S_{it} = 0$) with probability $b_i$. If agent $i$ was unemployed at the end of time $t-1$ and receives no new offers, the agent stays unemployed ($S_{it} = 0$).

[8] A given agent may end up with offers for several jobs, as we discuss next. So holding on to information does not necessarily imply that an agent takes that job.

**Determination of Wages**

The evolution of wages is as follows. The function $w_i : I\!R_+ \times \{0, 1, 2, \ldots\} \rightarrow I\!R_+$ describes the wage that $i$ obtains as a function of $i$'s previous wage and the number of new job opportunities that $i$ ends up with at the end of the hiring phase. It is increasing in past wages, with $w_i(W_{i,t-1}, O_{it}) \geq W_{i,t-1}$.[9] We also assume that $w_i(W_{i,t-1}, O_{it})$ is nondecreasing in the number of new offers received, $O_{it}$, and that $w_i(0, 1) > 0$ so that a new job brings a positive wage. The wage might be increasing in the number of offers an agent has because competition between employers bids the wage up (e.g., see Arrow and Borzekowski (2001)), or simply due to a better match.

We assume that $w_i$ takes on a finite set of values that fall in simple steps so that if $w' > w$ are adjacent in the range of $w_i$, then $w'_i = w_i(w, 1)$. Wages are thus delineated so that an agent may reach the next higher wage level with one offer. We assume that the highest wage an agent may obtain is above 0, that is $\overline{w}_i > 0$, and that $w_i(w', o) \geq w_i(w, o + 1)$ for any $o$ and $w'$ and $w$ such that $w'_i > w_i$. Having a higher wage is thus at least as good as having one additional offer starting from a lower wage (at least in expectations).

The wage of agent $i$ then evolves as $W_{it} = w_i(W_{i,t-1}, O_{it})S_{it}$, where $S_{it}$ keeps track of $i$'s employment status after the breakup phase.

**Networks**

We say that $i$ is *connected* with $j$ if $p_i(w) \neq p_i(w_{-j}, \widetilde{w}_j)$ for some $w$ and $\widetilde{w}_j$.

The term "connected" does not necessarily mean that $i$ and $j$ pass information to each other; it is just that their statuses directly or indirectly affect each other's probability of hearing about a job.[10] Let

$$N_i(p) = \{j \mid i \text{ is connected with } j\}$$

We assume that connections are at least minimally reciprocal, so that $i \in N_j(p)$ if and only if $j \in N_i(p)$. In the absence of such an assumption, all of the nonnegative correlation results that we establish still hold; however, for strictly positive correlations to ensue, it must be that information can have implications that travel sufficiently through the network to have one agent's status affect another.

By keeping track of further levels of "connection" ($i$ is connected to $j$ is connected to $k$...), we partition the set of agents so that all the agents in any element of the partition are path connected to each other. We denote this partition by $\Pi(p)$.

We assume that any element of the partition, $\pi \in \Pi(p)$, contains at least two agents, as completely isolated agents have dynamics of wages and employment that are trivial.

**An Economy**

Given a specification of $N$, $p_i$'s, and $b_i$'s, and an initial distribution over states $\mu_0$, the stochastic process of employment $\{S_1, S_2, \ldots\}$ and wages $\{W_1, W_2, \ldots\}$ is completely specified. We refer to the specification of $(N, p, b)$ satisfying the properties that we have outlined as an *economy*.

[9]Below, we discuss a direct extension of the model to include multiple job types, so that wages depend on the number of offers for different job types.

[10]Note that it is also possible that $p_{ij} > 0$, but $i$ and $j$ not be "connected" in cases where $p_i$ does not depend on $w_j$.

# 3 The Dynamics and Patterns of Wage and Employment

It is easy to show that improving the state of an agent's neighbors' wages and/or employment will improve the agent's future wages in the sense of first order stochastic dominance; and similarly, if we add new neighbors to an agents' neighborhood. Deriving correlation patterns among agents' wages is more difficult, and what we turn to now.

## 3.1 Wage Patterns and Dynamics

Suppose that agents are more likely to pass job information to direct connections with lower wages than to those with higher wages. Take an agent who has a low wage, but whose wage is still higher than some other agents who compete with her for information about a job. Then, this agents' next period expected wage may be lower than what she would expect by quitting her job. Indeed, if she were to quit her job, more of her connections may pass information to her, yielding to a positive probability of getting several offers at once. This case is not precluded under the assumptions on $p$. It is due to the fact that the model does not fully separate the arrival of offers over time. This difficulty is overcome when we look at fine enough subdivisions of a period. Then, the probability of obtaining more than one offer becomes negligible compared to the probability of just one offer.

**$T$-period Subvidisions**

Starting from an economy $(N, p, b)$, the *T-period subdivision*, denoted $(N, p^T, b^T)$, is such that $b_i^T = \frac{b_i}{T}$ and $p_{ij}^T = \frac{p_{ij}}{T}$ for each $i$ and $j$.

**Association**

While first order stochastic dominance is well suited for capturing distributions over a single agent's status, we need a richer tool for discussing interrelationships between a number of agents at once. There is a generalization of first order stochastic dominance to random vectors, *association*, introduced into the statistics literature by Esary, Proschan, and Walkup (1967).

A probability measure $\mu$ describing a random vector (e.g., $W$ defined on $I\!R^n$) is *associated* if

$$\mathrm{Cov}_\mu (f, g) \geq 0$$

for all pairs of non-decreasing functions $f : I\!R^n \rightarrow I\!R$ and $g : I\!R^n \rightarrow I\!R$, where $\mathrm{Cov}(f, g)$ is the covariance $E_\mu[f(W)g(W)] - E_\mu[f(W)]E_\mu[g(W)]$.

Association tells us good news in the sense of higher values of $W_i, i \in \{i_1, \ldots, i_\ell\}$ about any subset of agents (here, $\{i_1, \ldots, i_\ell\}$) is good (not bad) news for any other set of agents.

We say that $W_1, \ldots, W_n$ are associated if these are random variables described by an associated measure $\mu$. Independent random variables are associated. Also, if $W$ is a random vector described by $\mu$, then association of $\mu$ implies that $W_i$ and $W_j$ are non-negatively correlated for any $i$ and $j$.

**Strong Association**

We also define a strong version of association, useful to establish strictly positive relationships.

A probability measure $\mu$ describing a random vector on $I\!R^n$ is *strongly associated* relative to the partition $\Pi$ if it is associated, and for any $\pi \in \Pi$ and nondecreasing functions $f$ and $g$

$$\mathrm{Cov}_\mu (f, g) > 0$$

whenever there exist $i$ and $j$ such that $f$ is increasing in $w_i$ for all $w_{-i}$, $g$ is increasing in $w_j$ for all $w_{-j}$, and $i$ and $j$ are path connected under $\Pi$.

Strong association captures the idea that better information about any of the dimensions in $\pi$ leads to strictly higher expectations regarding every other dimension in $\pi$. One implication of this is that $W_i$ and $W_j$ are positively correlated for any $i$ and $j$ in $\pi$.

THEOREM **1** *Consider an economy $(N, p, b)$. For all $T$, let $\mu^T$ be the (unique) steady state distribution of $(N, p^T, b^T)$.*

- *The limit, $\mu$, of the steady state distributions $\mu^T$ is strongly associated relative to $\Pi(p)$. Thus, the wages of any path connected agents are positively correlated under $\mu$ and $\mu^T$ for large enough $T$.*

- *Starting from the steady state distribution, there is a strictly positive correlation between the wage statuses of any path connected agents and at any times. That is, for any times $t$ and $t'$ and large enough $T$,*

$$Cov^T \left[W_{it} W_{jt'}\right] > 0,$$

*where $i$ and $j$ are path connected and $Cov^T$ is the covariance associated with the $T$-period subdivision starting at time 0 under the steady state distribution $\mu^T$.*

The theorem states that any path connected agents have positively correlated wage levels in steady state and across time, and in fact exhibit strong association. The limit of the steady state distributions as $T$ becomes large is a very natural thing to consider, as it is a Poisson birth/death process which would naturally describe the job search. The reason we work with a discrete time approximation is purely for tractability in separating out the hiring and break-up phases.

The proof of Theorem 1 is long and appears in the appendix. The proof can be broken down into several steps. The first step shows that for large enough $T$ the steady state distribution is approximately the same as one for a process where the realizations of $p_{ij}(w)$ across different $j$'s is independent. The idea is that for large enough $T$, the probability that just one job is heard about overwhelms the probability that more than one job is heard about. This is also true under independence. The proof then uses a characterization of steady state distributions of Markov processes by Freidlin and Wentzel (1984) (as adapted to finite processes by Young (1993)). We use the characterization to verify that one can simply keep track of the probabilities of just a single job event to get the approximate steady state distribution for large enough $T$. Next, note that under independence of job hearing, there are no short-run negative conditional correlations. So we can establish that the conclusions of the theorem are true under the independent process. Finally, we come back to show that the same still holds under the true (dependent) process, for large enough $T$.

### 3.2 Employment Patterns and Dynamics

One might conjecture (as we initially did) that it would be a simple corollary to Theorem 1 for employment to exhibit the same positive correlation structure as wages. While the strong association of wages ensures that employment is weakly associated across agents, it is still possible for two agents to have positively correlated wages and yet have their employment status be independent. This is illustrated in the following example.

EXAMPLE **1** *Positive Correlation of Wages but Independence of Employment.*

Let agent $i$'s wages take on three values $\{0, 1, 2\}$ and agent $j$'s wages take on two values $\{0, 1\}$. Let $i$ and $j$ be path connected (but say not connected).[11] Consider a limiting steady state distribution which has the following marginal distribution on $W_i$ and $W_j$:

| | $w_j = 0$ | $w_j = 1$ |
|---|---|---|
| $w_i = 2$ | $\frac{1}{12}$ | $\frac{1}{4}$ |
| $w_i = 1$ | $\frac{1}{4}$ | $\frac{1}{12}$ |
| $w_i = 0$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

Under this distribution, $W_i$ and $W_j$ are positively correlated. Yet, $S_i$ and $S_j$ are independent:

| | $s_j = 0$ | $s_j = 1$ |
|---|---|---|
| $s_i = 1$ | $\frac{1}{3}$ | $\frac{1}{3}$ |
| $s_i = 0$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

This points out that much richer information can be obtained by tracking wages as opposed to employment, which is simply 0-1. For instance, if agents have reasonably high employment rates, then network effects will mainly be observed through their wage dynamics and correlations, as the quality of their jobs may vary dramatically even though their employment status may not.

This type of distribution cannot arise if $p$ is a function of $S$ rather than of $W$. With that added condition we can establish variations of Theorem 1 for employment by similar methods.[12]

## 4 Dropping Out and Long-Run Inequality

Consider the following game endogenizing the network structure. Let $d_i \in \{0, 1\}$ denote $i$'s decision of whether to stay in the labor market. Each agent discounts future wages at a rate $0 < \delta_i < 1$ and pays an expected discounted cost $c_i \geq 0$ to stay in. Agents dropping out get a payoff of zero.

[11] That is, $i$ and $j$ wage statuses *do not* influence each other, but $i$ and $j$ are connected through a chain of agents whose wages statuses *do* influence each other.

[12] Having fixed an initial state $W_0$, an economy induces a Markov chain on the state $W_t$. Note that this does not correspond to a Markov chain on the state $S_t$, as the probability of transitions from $S_t$ to $S_{t+1}$ can still depend on $W_t$ (rather than just $S_t$) and hence on $t$ for a given starting distribution. Nevertheless, as the wage states do form a Markov chain, there is a steady state distribution induced on the wage state $W$. As $S$ is a coarsening of $W$, there is a corresponding steady state distribution on $S$.

An augmented economy is $(N, p, b, c, \delta)$, where $c$ and $\delta$ are vectors of costs and discount rates.

When an agent $i$ exits the labor force, we reset the $p$'s so that $p_{ij}(w) = p_{ji}(w) = 0$ for all $j$ and $w$, but do not alter the other $p_{kj}$'s. The agent who drops out has his or her wage set to zero.[13] Therefore, when an agent drops out, it is as if the agent disappeared from the economy.

Fix an augmented economy $(N, p, b, c, \delta)$ and a starting state $W_0 = w$. A vector of decisions $d$ is an *equilibrium* if for each $i \in \{1, \ldots, n\}$, $d_i = 1$ implies

$$E\left[\sum_t \delta_i^t W_{it} \,|W_0 = w, d_{-i}\right] \geq c_i,$$

and $d_i = 0$ implies the reverse inequality.

The "drop-out" game is supermodular (see Topkis (1979)) which leads to the following lemma.

LEMMA **1** *Consider any augmented economy $(N, p, b, c, \delta)$ and state $W_0 = w$. There exists $T'$ such that for any $T$-period subdivision of the economy ($T \geq T'$), there is a unique equilibrium $d^*(w)$ such that $d^*(w) \geq d$ for any other equilibrium $d$.*

We refer to the equilibrium $d^*(w)$ in Lemma 1 as the *maximal* equilibrium.

THEOREM **2** *Consider any augmented economy $(N, p, b, c, \delta)$. Consider two starting wages states, $w' \geq w$ with $w \neq w'$. There exists $T'$ such that the set of drop-outs under the maximal equilibrium following $w'$ is a subset of that under $w$ for any $T$-period subdivision ($T \geq T'$); for some specifications of the costs and discount rates the inclusion is strict. Moreover, if $d^*(w)_i = d^*(w')_i = 1$, then the distributions of $i$'s wages and employment $W_{it}$ and $S_{it}$ for any $t$ under the maximal equilibrium following $w'$ first order stochastic dominate those under the maximal equilibrium following $w$, with strict dominance for large enough $t$ if $d^*(w)_j \neq d^*(w')_j$ for any $j$ who is path connected to $i$. In fact, for any increasing $f : \mathbb{R}_+^n \to \mathbb{R}$ and any $t$*

$$E^T\left[f(W_t)\,|W_0 = w', d^*(w')\right] \geq E^T\left[f(W_t)\,|W_0 = w, d^*(w)\right],$$

*with strict inequality for some specifications of $c$ and $\delta$.*

Theorem 2 shows how persistent inequality can arise between two otherwise similar groups with different initial employment conditions.

## 4.1 Multiple Job Types

The model we have presented does not differentiate between different types of jobs: wages depend only on current wage levels and the number of received offers. In some situations, a worker might be qualified for different types of jobs and might even have different networks for different types of jobs. The model is easily extended to accommodate such situations by simply keeping track

[13] This choice is not innocuous, as we must make some choice as to how to reset the function $p_{kj}$ when $i$ drops out, as this is a function of $w_i$. How we set this has implications for agent $j$ if agent $j$ remains in the economy.

of offers for different types of jobs via different $p$'s. Wages are then a function of the best offer received in a given period. The monotonicity conditions readily extend as do all of our results. As an illustration, consider a world with two types of jobs −low-skilled and high-skilled, where high-skilled jobs pay a higher wage rate than low-skilled jobs. If an agent is unemployed and receives offers only for low-skilled jobs, then he or she will take a low skilled job. If an agent hears about a high-skilled job, then he or she will take a high-skilled job. If an agent is unemployed, then he or she will only pass on information about low-skilled jobs if he or she also happens to get information about a high-skilled job. If an agent is already employed, then the probability of passing on information about both low-skilled and/or a high-skilled job is higher. The monotonicity conditions still naturally hold.[14]

## References


[1] Aliprantis C.D. and K.C. Border (2000) *Infinite Dimensional Analysis*, Springer-Verlag.

[2] Arrow, K. and R. Borzekowski (2001) "Limited Network Connections and the Distribution of Wages," mimeo: Stanford University.

[3] Calvó-Armengol, A. and M.O. Jackson (2004) "The Effects of Social Networks on Employment and Inequality," *American Economic Review*, vol. 94, no. 3, June, pp 426-454.

[4] Card, D. and Krueger (1992) "School Quality and Black-White Relative Earnings: A Direct Assessment," *Quarterly Journal of Economics*, 151-200.

[5] Chandra, A. (2000) "Labor-Market Dropouts and the Racial Wage Gap: 1940-1990," *American Economic Review*, Papers and Proceedings, 333-338.

[6] Darity, W.A. and P.L. Mason (1998) "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender," *Journal of Economic Perspectives*, 12, 63-90.

[7] Esary, J.D., F. Proschan, and D.W. Walkup (1967) "Association of Random Variables, with Applications," *Annals of Mathematical Statistics*, 38, 1466-1474.

[8] Farley, R. (1990) "Blacks, Hispanics, and White Ethnic Groups: Are Blacks Uniquely Disadvantaged?," *American Economic Review*, Papers and Proceedings, 80, 237-241.

[9] Freidlin, M. and A. Wentzel (1984) *Random Perturbations of Dynamical Systems*, Springer-Verlag, New-York.

[10] Heckman, J.J. (1998) "Detecting Discrimination," *Journal of Economic Perspectives*, 12, 101-116.


[14] One can also integrate people only qualified for low-skilled jobs into the high-skilled information passing without any difficulties.

[11] Heckman, J., T.M. Lyons and P.E. Todd (2000) "Understanding Black-White Wage Differentials, 1960-1990," *American Economic Review*, Papers and Proceedings, 344-349.

[12] Ioannides, Y. and L.D. Loury (2004) "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, 42, Dec., 1056-1093.

[13] Smith and Welch (1989) "Black Economic Progress after Myrdal," *Journal of Economic Literature*, 27, 519-564.

[14] Szekli, R. (1995) *Stochastic Ordering and Dependence in Applied Probability*, Springer-Verlag (Lecture Notes in Statistics # 97), New York.

[15] Topkis, D. (1979) "Equilibrium points in nonzero-sum $n-$person submodular games," *SIAM Journal of Control and Optimization* 17, 773-787.

[16] Young, H.P. (1993) "The Evolution of Conventions," *Econometrica* 61, 57-84.

## Appendix

The following definitions and lemmas are useful in the proof of Theorem 1.

$\mu$ and $\nu$ are two probability measures on a state space that is a subset of $I\!R^n$. $\mu$ *dominates* $\nu$ if

$$E_{\mu}\left[f\right] \geq E_{\nu}\left[f\right]$$

for every non-decreasing function $f : I\!R^n \to I\!R$.[15] The domination is *strict* if strict inequality holds for some non-decreasing $f$. When $n = 1$, domination reduces to first order stochastic dominance.

Lemma **2** *Consider two measures $\mu$ and $\nu$ on $I\!R^n$ which have supports that are a subset of a finite set $W \subset I\!R^n$. $\mu$ dominates $\nu$ if and only if there exists a Markov transition function $\phi : W \to \mathcal{P}(W)$ (where $\mathcal{P}(W)$ is the set of probability measures on $W$) such that*

$$\mu(w') = \sum_{w} \phi_{ww'}\nu(w),$$

*where $\phi$ is a dilation (that is $\phi_{ww'} > 0$ implies that $w' \geq w$). Strict domination holds if $\phi_{ww'} > 0$ for some $w' \neq w$.*

Thus, $\mu$ derives from $\nu$ by an "upwards" shift of mass under the partial order $\geq$ over $w \in W$.

**Proof of Lemma 2:** This follows from Theorem 18.40 in Aliprantis and Border (2000). ▮

The set of subsets of states such that if one state is in the event then all states with at least as high wages (person by person) are also in is:

$$\mathcal{E} = \{E \subset W \mid w \in E, w' \geq w \Rightarrow w' \in E\}.$$

Variations of the following useful lemma appear in the statistics literature (e.g., see Section 3.3 in Esary, Proschan and Walkup (1967)). We include a proof of this version for completeness.

[15] We can take the probability measures to be Borel measures and $E_{\mu}[f]$ simply represents the usual $\int_{I\!R^n} f(x)d\mu(x)$.

LEMMA **3** *Consider two measures $\mu$ and $\nu$ on $W$. Then, $\mu(E) \geq \nu(E)$ for every $E \in \mathcal{E}$, if and only if $\mu$ dominates $\nu$. Strict domination holds if and only if the first inequality is strict for at least one $E \in \mathcal{E}$. The measure $\mu$ is associated if and only if $\mu(EE') \geq \mu(E)\mu(E')$ for every $E$ and $E' \in \mathcal{E}$. The association is strong (relative to $\Pi$) if the inequality is strict whenever $E$ and $E'$ are both sensitive to some $\pi \in \Pi$.*[16]

**Proof of Lemma 3:** First, suppose that for every $E \in \mathcal{E}$:

$$\mu(E) \geq \nu(E). \tag{1}$$

Take any $f$ non-decreasing. Enumerate $r_1, \ldots, r_K$ the elements on its range, $r_K > r_{k-1} \ldots > r_1$. Let $E_K = f^{-1}(r_K)$. $f$ non-decreasing implies that $E_K \in \mathcal{E}$. Inductively, define $E_k = E_{k+1} \cup f^{-1}(r_{k-1})$. Clearly, $E_k \in \mathcal{E}$. Note that[17]

$$f(w) = \sum_k (r_k - r_{k-1}) I_{E_k}(w).$$

Thus,

$$E_\mu(f(W_t)) = \sum_k (r_k - r_{k-1})\mu(E_k) \qquad \text{and} \qquad E_\nu(f(W_t)) = \sum_k (r_k - r_{k-1})\nu(E_k).$$

By (1), $E_\mu(f(W_t)) \geq E_\nu(f(W_t))$ for every non-decreasing $f$. This implies dominance. If $\mu(E) > \nu(E)$ for some $E$, then $E_\mu(I_E(W_t)) > E_\nu(I_E(W_t))$, and strict dominance follows.

Next let us show the converse. Suppose that $\mu$ dominates $\nu$. For any $E \in \mathcal{E}$, let $f(w) = I_E(w)$. This is a non-decreasing function. Thus, $E_\mu(I_E(W_t)) \geq E_\nu(I_E(W_t))$, and so $\mu(E) \geq \nu(E)$.

To see that strict dominance implies that $\mu(E) > \nu(E)$ for some $E$, note that under strict dominance we have some $f$ for which

$$E_\mu(f(W_t)) = \sum_k (r_k - r_{k-1})\mu(E_k) > E_\nu(f(W_t)) = \sum_k (r_k - r_{k-1})\nu(E_k).$$

Since $\mu(E_k) \geq \nu(E_k)$ for each $E_k$, this implies that we have strict inequality for some $E_k$.

The proof for association (and strong association) is a straightforward extension of the above proof that we leave to the reader (or see Esary, Proschan and Walkup (1967)). ∎

LEMMA **4** *Let $\mu$ be associated and have full (finite) support on values of $W$. If $f$ is nondecreasing and is increasing in $W_i$ for some $i$, and $g$ is a nondecreasing function which is increasing in $W_j$ for some $j$, and $Cov_\mu(W_i, W_j) > 0$, then $Cov_\mu(f, g) > 0$.*

**Proof of Lemma 4:** We first prove the following Claim.

CLAIM **1** *Let $\mu$ be associated and have finite support. If $f$ is an increasing function of $W_i$ which depends only on $W_i$, and $g$ is an increasing function of $W_j$ which depends only on $W_j$, and $Cov_\mu(W_i, W_j) > 0$, then $Cov_\mu(f(W), g(W)) > 0$.*

[16] $E$ is sensitive to $\pi$ if its indicator function is. A nondecreasing function $f : \mathbb{R}^n \to \mathbb{R}$ is *sensitive to* $\pi \in \Pi$ (relative to $\mu$) if there exist $x$ and $\widetilde{x}_\pi$ such that $f(x) \neq f(x_{-\pi}, \widetilde{x}_\pi)$ and $x$ and $x_{-\pi}, \widetilde{x}_\pi$ are in the support of $\mu$.

[17] $I_E$ is the indicator function of $E$.

**Proof of Claim 1**: We write

$$\mathrm{Cov}_\mu (W_i, W_j) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{Cov}_\mu \left( I_{W_i}(s), I_{W_j}(t) \right) dsdt,^{18}$$

where $I_{W_i}(s) = 1$ if $W_i > s$, and $I_{W_i}(s) = 0$, otherwise. By assumption, $\mathrm{Cov}_\mu (W_i, W_j) > 0$. Therefore, $\mathrm{Cov}_\mu \left( I_{W_i}(\bar{s}), I_{W_j}(\bar{t}) \right) > 0$ for a set of $\bar{s}, \bar{t}$'s. Also,

$$\mathrm{Cov}_\mu (f(W_i), g(W_J)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{Cov}_\mu (I_f(s), I_g(t)) \, dsdt, \tag{2}$$

where $I_f(s) = 1$ if $f(W_i) > s$, and $I_f(s) = 0$, otherwise. For each $\bar{s}$ as described above, there exists some $s'$ such that $I_{W_i}(\bar{s}) = 1$ if and only if $I_f(f(s')) = 1$, and similarly for $\bar{t}$, $g$, and $t'$. Therefore, $\mathrm{Cov}_\mu (I_f(f(s')), I_g(g(t'))) > 0$. Given the finite support of $W$, the sets of such $\bar{s}, \bar{t}$'s and corresponding $s', t'$'s are unions of closed intervals with nonempty interiors. By association also we know that $\mathrm{Cov}_\mu (I_f(f(\bar{s})), I_g(g(\bar{t}))) \geq 0$ for any $s, t$. Since this expression is positive on a set with positive measure, and everywhere nonnegative, it follows from (2) that $\mathrm{Cov}_\mu (f, g) > 0$. ∎

Next consider $f$ that is increasing in $W_i$, but might also depend on $W_{-i}$. Label the possible wage levels of $i$ by $w_i^k$ where $w_i^1 = 0$ and $w_i^K = \overline{w}_i$. Let $\gamma = \min_{K \geq k > 1, w_{-i}} f(w_i^k, w_{-i}) - f(w_i^{k-1}, w_{-i})$. By the increasing property of $f$ it follows that $\gamma > 0$. Define $f'(w_i^k) = f(0, \ldots, 0) + k\frac{\gamma}{2}$. Let $f''(w) = f(w) - f'(w_i)$. It is easily checked that $f''$ is non-decreasing. Similarly define $g'$ and $g''$ for $g$ relative to $W_j$. Then

$$Cov(f, g) = Cov(f'', g'') + Cov(f'', g') + Cov(f', g'') + Cov(f', g').$$

By association, each expression is nonnegative. By Claim 1 the last expression is positive. ∎

Fix the economy $(N, p, b)$. Let $P^T$ denote the matrix of transitions between different $w$'s under the $T$-period subdivision. So $P^T_{ww'}$ is the probability that $W_t = w'$ conditional on $W_{t-1} = w$.

Let $P^T_{wE} = \sum_{w' \in E} P^T_{ww'}$.

Lemma 5 *Consider an economy $(N, p, b)$. Consider $w' \in W$ and $w \in W$ such that $w' \geq w$, and any $t \geq 1$. Then there exists $T'$ such that for all $T \geq T'$ and $E \in \mathcal{E}$*

$$P^T_{w'E} \geq P^T_{wE}.$$

*Moreover, if $w' \neq w$, then the inequality is strict for at least one $E$.*

**Proof of Lemma 5:** Let us say that two states $w'$ and $w$ are adjacent if there exists $\ell$ such that $w'_{-\ell} = w_{-\ell}$ and $w'_\ell > w_\ell$ take on adjacent values in the range of $\ell$'s wage function.

We show that

$$P^T_{w'E} \geq P^T_{wE}.$$

[18] See, for instance, Corollary B in Section 3.1 of Szekli (1995). As $\mu$ has finite support, these integrals trivially exist.

for large enough $T$ and adjacent $w$ and $w'$, as the statement then follows from a chain of comparisons across such $w'$ and $w$. Let $\ell$ be such that $w'_\ell > w_\ell$. By definition of two adjacent wage vectors, $w'_i = w_i$, for all $i \neq \ell$. We write

$$P^T_{wE} = \sum_o Prob^T_w(W_t \in E | O_t = o) Prob^T_w(O_t = o),$$

where $Prob^T_w$ is the probability conditional on $W_{t-1} = w$. Note that by property (1) of $p$, $p_{\ell j}(w') \geq p_{\ell j}(w)$ for all $j \neq \ell$. Also since $w'_k = w_k$ for all $k \neq \ell$ property (1) also implies that $p_{ij}(w') \geq p_{ij}(w)$ for all $j \neq \ell$ and for all $i$. These inequalities imply that $Prob^T_{w'}(O_{-\ell,t})$ dominates $Prob^T_w(O_{-\ell,t})$. It is only $\ell$, whose job prospects may have worsened.

Since $w'_\ell > w_\ell$, given our assumption on wages (that $w_i(w', o) \geq w_i(w, o+1)$ for any $o$ and $w'$ and $w$ such that $w'_i > w_i$), it is enough to show that for any $a$, $Prob^T_{w'}(O_{\ell,t} \geq a) \geq Prob^T_w(O_{\ell,t} \geq a+1)$. This holds for large enough $T$, given the independence of different realizations of $p_{j\ell}$ and $p_{i\ell}$ for $i \neq j$ and property (2) of $p$, as then the probability of any given number of offers is of a higher order than that of a greater number of offers (regardless of the starting state).[19]

To see the strict domination, consider $E = \{w | w_\ell \geq w'_\ell\}$. Since (for large enough $T$) there is a positive probability that $\ell$ hears 0 offers under $w$, the inequality is strict. ∎

Given a measure $\xi$ on $W$, let $\xi P^T$ denote the measure induced by multiplying the $(1 \times n)$ vector $\xi$ by the $(n \times n)$ transition matrix $P^T$. This is the distribution over states induced by a starting distribution $\xi$ multiplied by the transition probabilities $P^T$.

Lemma **6** *Consider an economy $(N, p, b)$ and two measures $\mu$ and $\nu$ on $W$. There exists $T'$ such that for all $T \geq T'$, if $\mu$ dominates $\nu$, then $\mu P^T$ dominates $\nu P^T$. Moreover, if $\mu$ strictly dominates $\nu$, then $\mu P^T$ strictly dominates $\nu P^T$.*

**Proof of Lemma 6:**

$$[\mu P^T](E) - [\nu P^T](E) = \sum_w P^T_{wE} \left(\mu_w - \nu_w\right).$$

By Lemma 2 we rewrite this as

$$[\mu P^T](E) - [\nu P^T](E) = \sum_w \sum_{w'} \nu_{w'} \phi_{w'w} P^T_{wE} - \sum_w \nu_w P^T_{wE}.$$

As the second term depends only on $w$, we rewrite that sum on $w'$. Then, since $\phi$ is a dilation (and $\phi_{w'w} > 0$ only if $w \geq w'$) we can sum over $w \geq w'$:

$$[\mu P^T](E) - [\nu P^T](E) = \sum_{w'} \nu_{w'} \left( \sum_{w \geq w'} \phi_{w'w} P^T_{wE} - P^T_{w'E} \right).$$

Lemma 5 implies that for large enough $T$, $P^T_{wE} \geq P^T_{w'E}$ whenever $w \geq w'$. Thus since $\phi_{w'w} \geq 0$ and $\sum_{w \geq w'} \phi_{w'w} = 1$, the result follows.

[19] This holds provided $w'_\ell < \overline{w}_\ell$, but in the other case, the agent is already at the highest wage state and so the claim is verified.

Suppose that $\mu$ strictly dominates $\nu$. It follows from Lemma 2 that there exists some $w \neq w'$ such that $\phi_{w'w} > 0$. By Lemma 5, there exists some $E \in \mathcal{E}$ such that $P^T_{wE} > P^T_{w'E}$. Then $[\mu P^T](E) > [\nu P^T](E)$ for such $E$, implying (by Lemma 3) that $\mu P^T$ strictly dominates $\nu P^T$. ∎

**Proof of Theorem 1:** Recall that $P^T$ denotes the matrix of transitions between different $w$'s. Since $P^T$ is an irreducible and aperiodic Markov chain, it has a unique steady state distribution that we denote by $\mu^T$. The steady state distributions $\mu^T$ converge to a unique limit distribution (see Young (1993)), which we denote $\mu^*$.

Let $\overline{P}^T$ be the transition matrix where the process is modified as follows. Starting in state $w$, in the hiring phase each agent $i$ hears about a new job (and at most one) with probability $\frac{p_i(w)}{T}$ and this is *independent* of what happens to other agents, while the breakup phase is as before with independent probabilities $\frac{b_i}{T}$ of losing jobs. Let $\overline{\mu}^T$ be the associated (again unique) steady state distribution, and $\overline{\mu}^* = \lim_T \overline{\mu}^T$ (which is well-defined as shown in the proof of Claim 2 below).

The following claims establish the theorem.

Claim **2** $\overline{\mu}^* = \mu^*$.

Claim **3** $\overline{\mu}^*$ *is strongly associated.*

The following lemma is useful in the proof of Claim 2.

Let $P$ be a transition matrix for an aperiodic irreducible Markov chain on a finite state space $Z$.

For any $z \in Z$, let a $z$-tree be a directed graph on the set of vertices $Z$, with a unique directed path leading from each state $z' \neq z$ to $z$. Denote the set of all $z$-trees by $\mathcal{T}_z$. Let

$$p_z = \sum_{\tau \in \mathcal{T}_z} \left[ \times_{z',z'' \in \tau} P_{z'z''} \right]. \tag{3}$$

Lemma **7** Freidlin and Wentzel (1984):[20] *If $P$ is a transition matrix for an aperiodic, irreducible Markov chain on a finite state space $Z$, then its unique steady state distribution $\mu$ is described by*

$$\mu(z) = \frac{p_z}{\sum_{z' \in Z} p_{z'}},$$

*where $p_z$ is as in (3) above.*

**Proof of Claim 2:** Given $w \in W$, we consider a special subset of the set of $\mathcal{T}_w$, which we denote $\mathcal{T}^*_w$. This is the set of $w$-trees such that if $w'$ is directed to $w''$ under the tree $\tau$, then $w'$ and $w''$ are adjacent. As $P^T_{w',w''}$ goes to 0 at the rate $1/T$ when $w'$ and $w''$ are adjacent,[21] and other transition probabilities go to 0 at a rate of at least $1/T^2$, it follows from Lemma 7 that $\mu^T(w)$ may be approximated for large enough $T$ by

$$\frac{\sum_{\tau \in \mathcal{T}^*_w} \left[ \times_{w',w'' \in \tau} P^T_{w'w''} \right]}{\sum_{\widehat{w}} \sum_{\tau \in \mathcal{T}^*_{\widehat{w}}} \left[ \times_{w',w'' \in \tau} P^T_{w'w''} \right]}.$$

[20] See Chapter 6, Lemma 3.1; and also see Young (1993) for the adaptation to discrete processes.

[21] Note that under property (3) of $p$, since $w'$ and $w''$ are adjacent, it must be that $P^T_{w',w''} \neq 0$.

Moreover, note that for large $T$ and adjacent $w'$ and $w''$, $P^T_{w'w''}$ is either $\frac{b_i}{T}+o(1/T^2)$ (when $w'_i > w''_i$) or $\frac{p_i(w')}{T}+o(1/T^2)$ (when $w'_i < w''_i$), where $o(1/T^2)$ indicates a term that goes to zero at the rate of $1/T^2$. For adjacent $w'$ and $w''$, let $\widetilde{P}^T_{w'w''} = \frac{b_i}{T}$ when $w'_i > w''_i$, and $\frac{p_i(w')}{T}$ when $w'_i < w''_i$.[22] It then follows that

$$\mu^*(w) = \lim_{T\to\infty} \frac{\sum_{\tau\in\mathcal{T}^*_w}\left[\times_{w',w''\in\tau}\widetilde{P}^T_{w'w''}\right]}{\sum_{\widehat{w}}\sum_{\tau\in\mathcal{T}^*_{\widehat{w}}}\left[\times_{w',w''\in\tau}\widetilde{P}^T_{w'w''}\right]}. \tag{4}$$

By a parallel argument, this is the same as $\overline{\mu}^*(w)$. ∎

**Proof of Claim 3:** Equation 4 and Claim 2 imply that

$$\overline{\mu}^*(w) = \lim_{T\to\infty} \frac{\sum_{\tau\in\mathcal{T}^*_w}\left[\times_{w',w''\in\tau}\widetilde{P}^T_{w'w''}\right]}{\sum_{\widehat{w}}\sum_{\tau\in\mathcal{T}^*_{\widehat{w}}}\left[\times_{w',w''\in\tau}\widetilde{P}^T_{w'w''}\right]}.$$

Multiplying top and bottom of the fraction on the right hand side by $T$, we find that

$$\overline{\mu}^*(w) = \frac{\sum_{\tau\in\mathcal{T}^*_w}\left[\times_{w',w''\in\tau}\widehat{P}_{w'w''}\right]}{\sum_{\widehat{w}}\sum_{\tau\in\mathcal{T}^*_{\widehat{w}}}\left[\times_{w',w''\in\tau}\widehat{P}_{w'w''}\right]}, \tag{5}$$

where $\widehat{P}^T$ is set as follows. For adjacent $w'$ and $w''$ (letting $i$ be the agent for whom $w'_i \neq w''_i$) $\widehat{P}^T_{w'w''} = b_i$ when $w'_i > w''_i$, and $p_i(w')$ when $w'_i < w''_i$,[23] and $\widehat{P}^T_{w'w''} = 0$ for non-adjacent $w'$ and $w''$.

The proof of the claim is then established via the following steps.

**Step 1:** $\overline{\mu}^*$ is associated.

**Step 2:** $\overline{\mu}^*$ is strongly associated.

**Proof of Step 1:** We show that for any $T$ and any associated $\mu$, $\mu\overline{P}^T$ is associated. From this, it follows that if we start from an associated $\mu_0$ at time 0 (say an independent distribution), then $\mu_0(\overline{P}^T)^k$ is associated for any $k$. Since $\overline{\mu}^T = \lim_k \mu_0(\overline{P}^T)^k$ for any $\mu_0$ (as $\overline{\mu}^T$ is the steady-state distribution), and association is preserved under (weak) convergence,[24] this implies that $\overline{\mu}^T$ is associated for all $T$. Then again, since association is preserved under (weak) convergence, this implies that $\lim_T \overline{\mu}^T = \overline{\mu}^*$ is associated.

So, let us now show that for any $T$ and any associated $\mu$, $\nu = \mu\overline{P}^T$ is associated. By Lemma 3, we need to show that

$$\nu(EE') - \nu(E)\nu(E') \geq 0 \tag{6}$$

for any $E$ and $E'$ in $\mathcal{E}$. Write

$$\nu(EE') - \nu(E)\nu(E') = \sum_w \mu(w)\left(\overline{P}^T_{wEE'} - \overline{P}^T_{wE}\nu(E')\right).$$

Since $W_t$ is independent conditional on $W_{t-1} = w$, it is associated.[25] Hence,

$$\overline{P}^T_{wEE'} \geq \overline{P}^T_{wE}\overline{P}^T_{wE'}.$$

[22] We take $T$ high enough such that all coefficients of the transition matrix $\widetilde{P}$ are between 0 and 1.

[23] If $p_i(w') > 1$ for some $i$ and $w'$, we can divide top and bottom through by some fixed constant to adjust, without changing the steady state distribution.

[24] See, for instance, P5 in Section 3.1 of Szekli (1995).

[25] See, for instance, P2 in Section 3.1 of Szekli (1995).

Substituting into the previous expression we find that

$$\nu(EE') - \nu(E)\nu(E') \geq \sum_w \mu(w)\overline{P}^T_{wE}\left(\overline{P}^T_{wE'} - \nu(E')\right). \tag{7}$$

Under the properties of the $p_{ij}$'s, both $\overline{P}^T_{wE}$ and $\left(\overline{P}^T_{wE'} - \nu(E')\right)$ are non-decreasing functions of $w$. Thus, since $\mu$ is associated, it follows from (7) that

$$\nu(EE') - \nu(E)\nu(E') \geq \left[\sum_w \mu(w)\overline{P}^T_{wE}\right]\left[\sum_w \mu(w)\left(\overline{P}^T_{wE'} - \nu(E')\right)\right].$$

Then since $\sum_w \mu(w)\left(\overline{P}^T_{wE'} - \nu(E')\right) = 0$ (by the definition of $\nu$), the above inequality implies (6). ∎

**Proof of Step 2:** We have already established association. Thus, we need to establish that for any $f$ and $g$ that are increasing in some $w_i$ and $w_j$ respectively, where $i$ and $j$ are path connected,

$$Cov_{\overline{\mu}^*}(f,g) > 0.$$

By Lemma 4 it suffices to verify that

$$Cov_{\overline{\mu}^*}(W_i, W_j) > 0$$

For any transition matrix $P$, let $P_{wij} = \sum_{w'} P_{ww'}w'_iw'_j$, and similarly $P_{wi} = \sum_{w'} P_{ww'}w'_i$. Thus these are the expected values of the product $W_iW_j$ and the wage $W_i$ conditional on starting at $w$ in the previous period, respectively.

Let

$$Cov^T_{ij} = \sum_w \overline{\mu}^T(w)\overline{P}^T_{wij} - \sum_w \overline{\mu}^T(w)\overline{P}^T_{wi}\sum_{w'}\overline{\mu}^T(w')\overline{P}^T_{w'j}.$$

It suffices to show that for each $i, j$ for all large enough $T$, we have $Cov^T_{ij} > 0$.

The matrix $\overline{P}^T$ has diagonal entries $\overline{P}^T_{ww}$ which tend to 1 as $T \to \infty$ while other entries tend to 0. Thus, we use a closely associated matrix, which has the same steady state distribution, but for which some other entries do not tend to 0.

Let

$$\underline{P}^T_{ww'} = \begin{cases} T\overline{P}^T_{ww'} & \text{if } w \neq w' \\ 1 - \sum_{w'' \neq w} T\overline{P}^T_{ww''} & \text{if } w' = w. \end{cases}$$

One can directly check that the unique steady state distribution of $\underline{P}^T$ is the same as that of $\overline{P}^T$, and thus also that

$$Cov^T_{ij} = \sum_w \overline{\mu}^T(w)\underline{P}^T_{wij} - \sum_w \overline{\mu}^T(w)\underline{P}^T_{wi}\sum_{w'}\overline{\mu}^T(w')\underline{P}^T_{w'j}.$$

Note also that transitions are still independent under $\underline{P}^T$. This implies that starting from any $w$, the distribution $\underline{P}^T_w$ is associated and so

$$\underline{P}^T_{wij} \geq \underline{P}^T_{wi}\underline{P}^T_{wj}.$$

Therefore,

$$Cov_{ij}^T \geq \sum_w \overline{\mu}^T(w)\underline{P}_{wi}^T\underline{P}_{wj}^T - \sum_w \overline{\mu}^T(w)\underline{P}_{wi}^T \sum_{w'} \overline{\mu}^T(w')\underline{P}_{w'j}^T.$$

Note that $\underline{P}_{wi}^T$ converges to $\widetilde{P}_{wi}$, where $\widetilde{P}_{wi}$ is the rescaled version of $\widehat{P}$ (defined in the proof of Claim 2),

$$\widetilde{P}_{ww'} = \begin{cases} T\widehat{P}_{ww'} & \text{if } w \neq w' \\ 1 - \sum_{w'' \neq w} T\widehat{P}_{ww''} & \text{if } w' = w. \end{cases}$$

It follows that

$$\lim_{T\to\infty} Cov_{ij}^T \geq \sum_w \overline{\mu}^*(w)\widetilde{P}_{wi}\widetilde{P}_{wj} - \sum_w \overline{\mu}^*(w)\widetilde{P}_{wi} \sum_{w'} \overline{\mu}^*(w')\widetilde{P}_{w'j}.$$

Thus, to complete the proof, it suffices to show that

$$\sum_w \overline{\mu}^*(w)\widetilde{P}_{wi}\widetilde{P}_{wj} > \sum_w \overline{\mu}^*(w)\widetilde{P}_{wi} \sum_{w'} \overline{\mu}^*(w')\widetilde{P}_{w'j}. \tag{8}$$

Viewing $\widetilde{P}_{wi}$ as a function of $w$, this is equivalent to showing that $\mathrm{Cov}(\widetilde{P}_{wi}, \widetilde{P}_{wj}) > 0$. From Step 1 we know that $\overline{\mu}^*$ is associated. We also know that $\widetilde{P}_{wi}$ and $\widetilde{P}_{wj}$ are both non-decreasing functions of $w$.

First let us consider the case where $j \in N_i(p)$.[26] We know that $\widetilde{P}_{wi}$ is increasing in $w_i$, and also, given the assumptions on $p$, that $\widetilde{P}_{wi}$ is increasing ni $w_j$ for $j \in N_i(p)$. Similarly, $\widetilde{P}_{wj}$ is increasing in $w_j$. (8) then follows from Lemma 4 (where we apply it to the case where $W_i = W_j$), as both $\widetilde{P}_{wi}$ and $\widetilde{P}_{wj}$ are increasing in $w_j$.

Next, consider any $k \in N_j(p)$. Repeating the argument above, since $\widetilde{P}_{wj}$ is increasing $w_j$ we apply Lemma 4 again to find that $W_i$ and $W_k$ are positively correlated. Repeating this argument inductively leads to the conclusion that $W_i$ and $W_k$ are positively correlated for any $i$ and $k$ that are path connected. ∎

The first part of Theorem 1 now follows from Claim 3 since $\mu^T \to \overline{\mu}^*$.We now show the second part. We know from Claim 3 that $\overline{\mu}^*$ is strongly associated. The result then follows by induction using Lemma 6,[27] and then taking a large enough $T$ so that $\mu^T$ is close enough to $\overline{\mu}^*$ for the desired strict inequalities to hold. ∎

**Proof of Lemma 1:** Consider what happens when an agent $i$ drops out. The resulting $w'$ is dominated by the $w$ if that agent does not drop out. Furthermore, from Lemma 6 for large enough $T$, the next period wage distribution over other agents when the agent drops out is dominated by that when the agent stays in, if one were to assume that the agent were still able to pass job information on. This domination then easily extends to the case where the agent does not pass any job information on. Iteratively applying this, the future stream of wages of other agents is dominated when the agent drops out relative to that where the agent stays in. This directly

[26] If $i$ is such that $N_i(p) = \emptyset$, then strong association is trivial. So we treat the case where at least two agents are path connected.

[27] While Lemma 6 does not state that the strict inequalities are preserved on given elements of the partition $\Pi(p)$, it is easy extension of the proof to see that this is true.

implies that the drop-out game is supermodular. The lemma then follows from the theorem by Topkis (1979). ∎

**Proof of Theorem 2:** Let $w \geq w'$ and $d \in \{0,1\}^n$. We first show that for large enough $T$

$$E^T\left[f(W_t)\,|W_0 = w', d\right] \geq E^T\left[f(W_t)\,|W_0 = w, d\right].$$

Lemma 5 implies that for a fine enough $T$-period subdivision and for every non-decreasing $f$,

$$E^T\left[f(W_1)\,|W_0 = w', d\right] \geq E^T\left[f(W_1)\,|W_0 = w, d\right].$$

Lemma 6 and a simple induction argument then establish the inequality for all $t \geq 1$. The inequality is strict whenever $f$ is increasing and $w' > w$.

Next, let $d \geq d'$. For a fine enough $T$-period subdivision and for every non-decreasing $f$, given that drop-outs have wages set to the lowest level it follows that

$$E^T\left[f(W_1)\,|W_0 = w, d'\right] \geq E^T\left[f(W_1)\,|W_0 = w, d\right]$$

As before, the inequality extends to all $t \geq 1$ by induction. Again, $f$ increasing and $d' > d$ imply a strict inequality. Combining these observations, we find that for large enough $T$ when $w' \geq w$ and $d' \geq d$

$$E^T\left[f(W_t)\,|W_0 = w', d'\right] \geq E^T\left[f(W_t)\,|W_0 = w, d\right] \tag{9}$$

Consider the maximal equilibrium $d^*(w)$. By (9), for large enough $T$ and all $t$

$$E^T\left[W_{it}\,|W_0 = w', d^*(w)\right] \geq E^T\left[W_{it}\,|W_0 = w, d^*(w)\right]$$

Thus,

$$\sum_t \delta_i^t E^T\left[W_{it}\,|W_0 = w', d^*(w)\right] \geq \sum_t \delta_i^t E^T\left[W_{it}\,|W_0 = w, d^*(w)\right]$$

If $d^*(w)_i = 1$, then

$$\sum_t \delta_i^t E^T\left[W_{it}\,|W_0 = w', d^*(w)\right] \geq \sum_t \delta_i^t E^T\left[W_{it}\,|W_0 = w, d^*(w)\right] \geq c_i$$

and so also for all $d' \geq d^*(w)$, if $i$ is such that $d^*(w)_i = 1$, then

$$\sum_t \delta_i^t E^T\left[W_{it}\,|W_0 = w', d'\right] \geq c_i. \tag{10}$$

Set $d'_i = d^*(w)_i$ for any $i$ such that $d^*(w)_i = 1$. Fixing $d'$ for such $i$'s, find a maximal equilibrium at $w'$ for the remaining $i$'s, and set $d'$ accordingly. By (10), it follows that $d'$ is an equilibrium when considering all agents. It follows that $d' \geq d^*(w)$. Given the definition of maximal equilibrium, it then follows that $d^*(w') \geq d' \geq d^*(w)$. ∎