Journal of Economic Growth, 2: 61–92 (1997)

Matching, Heterogeneity, and the Evolution of Income Distribution

DARON ACEMOGLU

Massachusetts Institute of Technology

This paper analyzes a model in which firms and workers have to engage in costly search to find a production partner, and endogenizes the skill, job, and wage distributions in this context. The presence of search frictions implies that there are two redistributive forces in the labor market. The first is *mismatch* relative to the Walrasian economy; skilled workers tend to work with lower physical to human capital ratios, and this compresses the earnings differentials. The second is *the opportunity cost* effect; because the opportunity cost of accepting an unskilled worker, which is to forgo the opportunity to employ a skilled worker, is high, unskilled wages are pushed down. The interaction between these two forces leads to a non-ergodic equilibrium process for wage and income inequality. Further, the presence of mismatch reduces the rate of return to physical capital and thus depresses growth. A key prediction of the analysis is that increasing wage inequality is more likely to arise in economies with less frictional labor markets, which is in line with the diverse cross-country patterns observed over the past two decades. Finally, the paper predicts that, as is largely the case with U.S. data, between group and within group wage inequality should move in the same direction.

Keywords: search, matching, mismatch, human capital, growth, wage inequality, income inequality

JEL classification: D31, J41

1. Introduction

Income inequality in the United States has risen considerably over the past two and a half decades. The main component of this change is identified as increased earnings and wage inequality (see for instance, Levy and Murnane, 1992, Juhn et al., 1993). Similar trends are observed in other OECD countries but appear much less pronounced. For instance, between 1979 and 1987, Katz et al. (1995) find that the differential for workers on the 90th and 10th percentiles of the wage distribution increased by 12% in the United States but only by 2% in France. The premium that college graduates earn over noncollege graduates (between-group wage inequality) has also increased in the United States and the United Kingdom but has not changed much in other European countries. Although empirical labor research has achieved reasonable success in explaining the various trends in between-group wage inequality in terms of changes in the relative supply and demand of skills (e.g., Katz and Murphy, 1992), we still lack an understanding of why relative demands and supplies have changed differently in these countries, and also why wages are now more unequal within narrowly defined groups.

As an example consider the diverse wage inequality trends in the United States and France. These trends can be explained to some degree by the fact that over this period (1979 to 87), the growth rate of the relative supply of college educated workers was 0.023 in the United States as compared to 0.050 in France (see Katz et al., 1995). But, both the relative demand for skills, and the changes in relative supplies are to some degree *endogenous*, affected among other things by labor market institutions, and they can be a consequence as much as a cause of the wage and income inequality. Furthermore, changes in the supply and demand of college graduates do not explain why inequality within narrowly defined groups has behaved very differently in these two countries. In fact, many pieces of evidence indicate that supply and demand are not the only factors to consider in the evolution of wage and income inequality. Card (1996) and DiNardo et al. (1996) show that changes in unionization have played an important role in the changing structure of wages, and differences between Europe and the United States are often attributed to their different labor market institutions.

This paper's objective is to offer a simple framework to analyze the relation between inequality, the labor market, and economic performance. The analysis first establishes that the presence of labor market frictions, in the form of costly search and matching, introduces two novel redistributive forces that can contribute to an explanation for why inequality is rising in some countries while remaining unchanged in others. *First*, as a result of the frictions, there will be a certain degree of *mismatch*, and workers of high human capital will work at *lower* physical to human capital ratios than the low human capital workers. Mismatch will tend to make the distribution of income more equal over time (in contrast in the Walrasian analogue of this model, physical to human capital ratios are constant for all workers). The intuitive reason is that firms hoping to employ highly skilled workers choose high levels of physical capital investments and the unskilled workers benefit from these investments when they work for these firms. This mismatch effect not only influences inequality but also growth: mismatch reduces the rate of return to physical capital and depresses investment and growth. *Second*, there is another force impacting on the dynamics of inequality, which I call the *opportunity cost effect* (or the outside option effect). The opportunity cost of employing an unskilled worker for the firm is to forgo the chance of meeting a skilled (more productive) worker next period. In contrast, this opportunity cost is not present when bargaining with a skilled worker; if the firm turns down a skilled worker, the next applicant will be on average less skilled. This differential opportunity cost implies that unskilled wages will be pushed down relative to skilled wages. Moreover, the opportunity cost of employing unskilled workers increases with the degree of inequality (as the gap between skilled and unskilled workers gets larger), and therefore, an economy that starts with a low level of inequality may exhibit decreasing inequality over time and achieve balanced growth, while the same economy with higher income inequality would have experienced increasing wage and income inequality and lower growth. In other words, the impact of labor market frictions on wage determination makes the dynamics of inequality *nonergodic*, and leads to multiple limiting distributions of income and different long-run growth rates.

More important, the analysis predicts that the extent of labor market frictions has a crucial impact on the dynamics of inequality. An economy with less frictional labor markets can experience rising inequality while an economy with more frictions would have decreasing levels of inequality. This prediction is in accord with the cross-country patterns of the past two decades—e.g., the United States and the United Kingdom versus France, Germany, the

Netherlands, and Sweden.[1] Furthermore, the model suggests a negative correlation between the growth rate and the level of inequality as found in the data (see Perotti, 1996). The intuition for this negative relation is due to mismatch in the labor market, rather than the political economy reasons which are often emphasized.

Finally, in most periods, when wage differences increase between workers of different skill levels, they increase among equally skilled workers as well. (See for instance Juhn et al., 1993, and Goldin and Margo, 1992, for the wage compression of the 1950s and 1960s. The late 1970s may be an exception. See section 5). Despite the extent of the changes in this aspect of inequality, in Levy and Murnane's words (1992, p. 1372) "the most important unresolved puzzle concerns the reasons for the almost 20 years trend towards increased within group inequality." This paper offers a new explanation: when inequality of skills increases, firms start creating *a more diverse distribution of jobs*. Due to frictions, workers do not always work at the most appropriate firms; instead, workers of the same skill level sometimes work in jobs of different qualities, and so get paid different wages. Therefore, together with between-group inequality, within-group inequality also increases.

This paper builds on the work of Diamond (1982), Jovanovic (1979), Mortensen (1982), and Pissarides (1985) in which workers and firms have to search for the right partner in the labor market. I extend this framework in three respects: first, both workers and firms are potentially heterogeneous; thus, workers of different skill levels and jobs of different qualities may be simultaneously searching. Second, the composition of jobs and technological progress are endogenized; firms, knowing the distribution of skills in the economy, and the wages they will pay in equilibrium, decide what kind of jobs to open and how much capital to install (see also, Acemoglu, 1996a). Third, as it is the case in the data (for instance, Cameron and Heckman, 1992), the distribution of income is allowed to influence the distribution of skills; in other words, rich parents will be able to buy more education (skills) for their offspring.

Other closely related papers include Banerjee and Newman (1993), Galor and Zeira (1993), Benabou (1996), Durlauf (1996), and Fernandez and Rogerson (1996). All these papers emphasize the importance of the distribution of income and human capital on economic performance and growth, and make predictions regarding the evolution of income and skill distributions. The main difference between my paper and these contributions is the source of the costs of inequality. In Galor and Zeira, Fernandez and Rogerson, and Durlauf, inequality leads to low levels of human capital investments by some agents without any compensating increase in the investment of others, and thus reduces growth. In Benabou (1996), inequality reduces the output of the economy because the unskilled reduce the productivity of high human capital workers. This is similar to the conclusion of my paper. However, this effect is not assumed as part of the technology but derived from the presence of *mismatch* between the jobs and the workers. The important point is that because in this paper the allocation of physical capital to workers and matching are endogenized, there is a new force, the *opportunity cost effect*, which can lead to increasing inequality and multiple long-run equilibria. Finally, Galor and Tsiddon (1996) and Davis (1995) are also related. Galor and Tsiddon (1996) analyze the inequality consequences of waves of technological innovations, which increase and then decrease the return to ability, and they can generate different patterns of inequality over time and across countries. Davis analyzes the impact of bargaining arrangements on wage inequality. Again, the two redistributive forces identified here are not present in these studies.

The plan of the paper is as follows. The next section presents the basic model and characterizes the equilibrium of this economy with competitive labor markets benchmark case. Section 3 analyzes the dynamics of inequality with random matching in the context of a two-class economy, and discusses the impact of labor market efficiency on inequality dynamics. Section 4 analyzes the dynamics of inequality starting from general income distributions. Section 5 turns to different matching technologies and discusses the links between within-group and between-group wage inequality. Section 6 concludes.

2. The Basic Environment and the Walrasian Equilibrium

2.1. Preferences, Timing of Events, Technology

Each agent lives for three periods. The first is youth in which (s)he acquires human capital. In the second, middle age, he works and decides how much to consume. In this period, each agent also conceives a unique offspring, and decides how much to spend on his offspring's education and how much to save for the final, retirement, period of his life. I refer to an agent as of generation t, if he is middle-aged at time t. In each generation, there is a continuum of agents of size 1. The utility function of agent j of generation t is

$$(c_{j,t})^{1-\delta-\gamma} \cdot (e_{j,t+1})^{\delta} \cdot (c^{O}_{j,t+1})^{\gamma} \tag{1}$$

where c denotes consumption, the superscript O is used for old age, and e denotes the education expenditure. The utility function exhibits impure altruism as the agent does not obtain utility from the welfare of his child but from the education he gives him.

All that the agent decides to save during this period is invested at the gross market rate of return R_{t+1}. Thus the budget constraint of the agent is

$$c_{j,t} + \frac{c^{O}_{j,t+1}}{R_{t+1}} + e_{j,t+1} = w_{j,t}, \tag{2}$$

where $w_{j,t}$ is the wage income of individual j, which is his only income when middle-aged.

With this specification, the following simple decision rules for agent j are obtained;

$$\begin{aligned} c_{j,t} &= (1-\delta-\gamma)w_{j,t}, \quad c^{O}_{j,t+1} = R_{t+1}\gamma w_{j,t}, \\ e_{j,t+1} &= \delta w_{j,t}, \quad s_{j,t} = \frac{1}{R_{t+1}}(c^{O}_{j,t+1}) = \gamma w_{j,t}, \end{aligned} \tag{3}$$

where $s_{j,t}$ denotes savings. The human capital of the next generation is determined by the education level (expenditure) given to them by their parents. More precisely,

$$h_{j,t+1} = \begin{cases} e_{j,t+1} & if \quad e_{j,t+1} \geq h_{\min} \\ h_{\min} & if \quad e_{j,t+1} < h_{\min}. \end{cases} \tag{4}$$

In other words, there is a certain level of human capital, $h_{\min}$, below which agents do not fall—e.g., what they learn from compulsory schooling or from their friends. This is a very simple specification for the intergenerational transmission of human capital, and captures the fact that human capital investments are to some degree backward looking because of credit market constraints. In particular, with this specification, the human capital of the offspring is linear in the wage of the parent.[2]

I assume that there is a continuum of firms of measure 1, and that each firm employs one worker, thus these firms can be thought of as jobs. The measure of firms is taken to be equal to that of workers so as to avoid issues of unemployment (see below for a discussion). A firm (job) i, if matched with worker j, produces

$$y_{i,j,t} = Ak_{i,t}^{1-\alpha} h_{j,t}^{\alpha}, \tag{5}$$

where $k_{i,t}$ is the capital associated to this job, and $h_{j,t}$ is the human capital that worker j brings to the relation. This production function exhibits constant returns to scale and complementarity between physical capital of the firm and human capital of the worker. As a result, there are diminishing returns to human capital. In this frictionless (competitive) benchmark case, it is important that the production function has nonincreasing returns to scale, because otherwise a competitive equilibrium would not exist, and also, since there is nothing that pins down jobs in this economy, it should not be possible to produce more output by combining two jobs. The capital that firms will use at time t comes from the savings of generation $t-1$ workers. All workers invest their savings in a mutual fund, which is forced to make zero profits by the threat of potential entry (a perfectly contestable market). The savings are then allocated to firms. Intermediation by the mutual fund removes all uncertainty from the rate of return on savings. I also assume that the economy is small and open; thus, it faces a constant world interest rate R at which the mutual fund can borrow or invest; therefore, the cost of capital to a firm and the rate of return on savings are fixed at R.[3] If firms make positive profits, these accrue to the mutual fund (but in equilibrium there will be no aggregate profits).

I assume that firms hire their capital stock at the beginning of period t and this decision is completely *irreversible*. Also all physical capital depreciates at the end of the period. Note that in this economy, k_t corresponds not only to the capital stock but also to the quality of the equipment, the type of job, the location, etc., and so it is natural that it has a high degree of irreversibility. At the point of investment, while they know the distribution of human capital in the labor market, firms do not necessarily know which worker they will employ. The result of the firms' choices of capital stock will in general give a distribution of jobs and I denote this distribution by $P_t(k)$.

2.2. The Walrasian Equilibrium

In this section, the labor market is frictionless. Thus, all firms compete a la Bertrand for all the workers in the labor market, and the model boils down to an assignment problem (see Sattinger, 1993, for a survey and analysis of this class of models). This implies that workers will be paid their marginal product, and the allocation of the heterogeneous workers to the heterogeneous firms is the same as the one that a Walrasian auctioneer (or a Social Planner) would choose.

Given the distribution of human capital, $F_t(h)$, firms invest in physical capital, which is summarized by $P_t(k)$. Once $P_t(k)$ and $F_t(h)$ are determined, trade takes place in a Walrasian labor market. I use the notation $(h,k) \in \mathcal{P}_t$ to denote a situation in which a firm with capital k hires a worker with human capital h. In an equilibrium, it needs to be the case that

$$\begin{aligned} \pi_t(k) &= Ak^{1-\alpha}h^{\alpha} - Rk - w(h)h = \pi^* \qquad \forall (h,k) \in \mathcal{P}_t \\ &= Ak^{1-\alpha}h^{\alpha} - Rk - w(h)h \leq \pi^* \qquad \forall (h,k) \end{aligned}$$

for some equilibrium level of profits π^* where $w(h)$ is the wage rate per unit of human capital for a worker with human capital h. It is clear that the complementarity of human and physical capital implies that the highest human capital worker is allocated to the highest capital firm, otherwise the above equation cannot be satisfied. This therefore implies that

$$\begin{aligned} \frac{k_t}{h_t} &= \left(\frac{(1-\alpha)A}{R}\right)^{\frac{1}{\alpha}} \\ w_t(h_t) &= \alpha A \left(\frac{(1-\alpha)A}{R}\right)^{\frac{1-\alpha}{\alpha}} h_t. \end{aligned} \tag{6}$$

Also (6) implies that $\pi^* = 0$, therefore all firms are making zero-profits in equilibrium.

I next introduce the variable θ such that $h_{j,t} = \tilde{h}_t \theta_{j,t}$ where $\tilde{h}_t$ is the median level of human capital at time t, and therefore, the distribution of θ measures inequality of skills relative to the median. This distribution can simply be obtained from the distribution of h, and I denote it by $G(\theta)$. Let me also denote the starting distribution of human capital by $F_0(.)$ and the corresponding distribution of inequality by $G_0(.)$. Further, I assume in the rest of the paper that even the poorest agent j^* starts with $h_{j^*,0} \gg h_{\min}$. Then[4];

Proposition 1: *Let $G_0(\theta)$ be the distribution of initial inequality of human capital as defined above. Then, with frictionless labor markets*

(i) *The physical to human capital ratio for all jobs is always given by $\frac{k}{h} = (\frac{(1-\alpha)A}{R})^{\frac{1}{\alpha}}$.*

(ii) *$\forall G_0$, $G_t = G_0$ for all t, thus inequality in this economy self-replicates.*

(iii) *The growth rate of the economy is always $g^C = \delta\alpha A\{(1-\alpha)AR^{-1}\}^{\frac{1-\alpha}{\alpha}} - 1$.*

Proof. See Appendix.

The economy is linear and this is the feature that leads to steady growth as in the model of Rebelo (1991). More importantly for the focus of this paper, given constant returns to scale, the rate of return on human capital is constant; that is, a worker with twice as much human capital as another works with twice as much physical capital and thus earns twice as much. With the preferences as in equation 1, the accumulation rules are also linear. Therefore, the offspring of worker j, who has twice as much capital as j', will also have twice as much capital as the offspring of j'. Naturally, the case of interest for this paper is the one where the economy grows; thus, I would like to ensure that g^C is positive. For this, I assume that $\beta\delta A\{(1-\alpha)(1-\beta)AR^{-1}\}^{\frac{1-\alpha}{\alpha}} > 1$, where β is a constant between 0 and 1, which will be defined in the next section, and this condition will ensure that both the growth rates here and that in the next section are positive.

Note finally that in this economy there are many missing markets: young generations cannot pay their parents to get more human capital in their youth (a form of intergenerational credit constraints). But despite these missing markets, inequality is not harmful to aggregate performance. The economy has the same growth rate irrespective of the level of inequality.

3. Inequality With Random Matching

3.1. Preliminaries

I now analyze the economy outlined in the last section but with a frictional labor market. As before, in every period the distribution of human capital $F_t(h)$ is determined by the bequests of the previous generation, and then knowing this distribution, firms decide how much physical capital to hire. The key difference from the previous case is that it is costly for agents to engage in search activities. This is modeled as follows: after physical capital investments, firms and workers are matched one-to-one; thus, each job will have a worker, and each worker will have a job. At this point either party can terminate the relation and look for a new partner, but this is costly. In particular, workers' and firms' working lives are divided into two segments of length $1-\eta$ and η. If the worker (or the firm) decides to look for another match, they do not produce in the first segment and only meet another partner from the pool of unmatched agents for the second segment of their working lives (middle-age); thereafter, they cannot break the match again.[5] Thus η captures the degree of frictions in the labor market. When $\eta = 0$, mobility is very costly because after a separation no output is produced. As η increases the prospect of separation becomes more attractive, so the frictions which restrict trade in the labor market become less important.

In this section I start with the *extreme assumption* that matching is random. This implies that any two workers, even if they have unequal human capital levels, face exactly the same probability of meeting a given firm in the economy and conversely, the same also applies to firms. This assumption will be relaxed in section 5.

There are a number of other issues to deal with. First, since there are costs to changing partners, each match has some surplus to be shared; thus, a bargaining rule needs to be assumed. I will follow the usual practice of using Nash bargaining (e.g., Pissarides, 1990, or Osborne and Rubinstein, 1990). The second issue is that when η is near 1, some firms and workers will want to separate from their first match and look for another partner. I wish to avoid this issue as it will lead to unemployment and make the analysis much harder. Therefore, throughout the paper, $\eta < \eta^*$ will be assumed so that all firms and workers produce with the first partner they meet.[6] As a result of this assumption that $\eta < \eta^*$, in the economy discussed here there will be no separations (see below), and outside options will influence inequality only through their impact on wages due to the threat of separations. This introduces a minor problem; since there are no separations in equilibrium, an agent who separates will not be able to find a new partner. To avoid this problem, I assume that each worker and firm face a probability ν of not getting matched in the first segment of their life, and this event is independent of their human or physical capital level. I will analyze this economy as $\nu \to 0$; thus, the set of agents who are unmatched will be of measure zero. Also since the event of being unmatched is independent of characteristics, the distribution of the unmatched agents is exactly the same as their initial distributions, i.e., F_t for the workers and P_t for the firms.

With this formulation the wage level of a worker with human capital $h_{j,t}$, who is matched with a firm of physical capital $k_{i,t}$, when the distribution of human and physical capital are respectively $F_t(h)$ and $P_t(k)$, is given by[7]

$$\begin{aligned} w[h_{j,t}, k_{i,t}, F_t(h), P_t(k)] \; = \; & \beta A h_{j,t}^{a} k_{i,t}^{1-\alpha} - \beta(1-\beta)\eta A k_t^{1-\alpha} \int h^{\alpha} dF_t(h) \\ & + (1-\beta)\beta\eta \int A k^{1-\alpha} h_{j,t}^{\alpha} dP_t(k). \end{aligned} \tag{7}$$

The first term is the share of total output that the worker gets, and $0 < \beta < 1$ denotes the bargaining power of the worker. The second is (β times) the threat point of the firm. Because there are no separations in equilibrium, if a firm deviates and separates, it will meet one of the unmatched workers who have a distribution of human capital given by $F_t(h)$. Since this is the last round of matching, it will produce with the worker it meets and obtain a proportion $1-\beta$ of the output. This term is subtracted from the total surplus that the parties share, and since the worker obtains a proportion β of the total surplus, it is also multiplied by β. The third term, which is ($1-\beta$ times) the worker's outside option, is explained similarly as the return to the worker for making one more round of search among the pool of unmatched firms, $P_t(k)$. This amount is subtracted from the total surplus and then also added to the total wage of the worker as his threat point. The sum of these three terms give (7). Then, the firm's profit function can be written as

$$\pi[k_t, F_t(h), P_t(k)] = \int \{A k_t^{1-\alpha} h^{\alpha} - w[h, k_t, F_t(h), P_t(k)]\} dF_t(h) - R k_t, \tag{8}$$

which the firm will maximize by choosing k_t. It is easily seen that (8) is strictly concave in k_t and its derivative with respect to k_t does *not* depend on $P_t(k)$. Thus, when $\eta < \eta^*$ so that there are no separations, all firms choose the same level of physical capital investment equal to

$$k_t = \left[\frac{(1-\beta)(1+\beta\eta)(1-\alpha) \int h^{\alpha} dF_t(h)}{R} \right]^{\frac{1}{\alpha}}. \tag{9}$$

At this level of capital, it is straightforward to check that expected profits are equal to zero. Although some firms will make positive profits, some others will incur losses, and the mutual fund that intermediates all the savings will make zero profits.

There are a number of important features to note about equation 9. First because the firm is unable to capture the full marginal product of its investment, it will always underinvest, and the growth rate of this economy will always differ from the growth rate of the competitive economy, g^C. More interestingly, firm level investments also depend on the distribution of human capital, $F_t(.)$ (see Acemoglu, 1996a). It is immediately clear from (9) that a mean-preserving spread of F_t will reduce investment. The intuition is that a mean-preserving spread, which maintains the total amount of human capital but makes it more unequally distributed, increases *mismatch* and, thus, reduces profits. Mismatch in this economy is reflected by the distribution of human to physical capital ratios. In the Walrasian case, these ratios are always constant. Instead, in this frictional economy, since all firms choose to have the same level of physical capital, G_t also gives the distribution of the human to physical capital ratios, and an increase in inequality in the distribution of human capital increases mismatch. A somehow different way of seeing the intuition is to note that because there are decreasing returns to human capital, a firm loses more from a low human capital worker than it gains from a high human capital worker.

3.2. Analysis of Inequality and Growth Dynamics with Two Classes

To simplify the analysis, I begin with the case in which the economy starts with two groups of workers: rich and poor. I denote the human capital of the rich by h_{2t} and that of the poor by h_{1t}; I will then study the dynamics of $\phi_t = h_{1t}/h_{2t}$, the ratio of human capital and income between the two groups. I denote the proportion of poor agents by λ. This proportion will never change, but poor dynasties may become gradually less poor relative to the rich.

From equation 7, wages of each group, $j = 1, 2$, can be written as

$$\begin{aligned} w_{jt} &= \beta A k_t^{1-\alpha} h_{jt}^{\alpha} - \beta(1-\beta)\lambda\eta A k_t^{1-\alpha} h_{1t}^{\alpha} - \beta(1-\beta)(1-\lambda)\eta A k_t^{1-\alpha} h_{2t}^{\alpha} \\ &\quad + \beta(1-\beta)\eta A k_t^{1-\alpha} h_{jt}^{\alpha}. \end{aligned} \tag{10}$$

Now, as long as h_{1t} is away from $h_{\min}$, the law of motion of $\phi_t = h_{1t}/h_{2t}$ is;

$$\phi_{t+1} = \frac{h_{1,t+1}}{h_{2,t+1}} = \frac{\delta w_{1t}}{\delta w_{2t}} = \frac{[1+(1-\beta)(1-\lambda)\eta]\phi_t^{\alpha} - (1-\beta)(1-\lambda)\eta}{1+(1-\beta)\lambda\eta - (1-\beta)\lambda\eta\phi_t^{\alpha}}. \tag{11}$$

Therefore, a simple nonlinear first-order difference equation describes the evolution of inequality. In other words, equation 11 gives the law of motion of the human capital ratio of the poor to that of the rich, and from (9) and (10) wages in any given period can be determined. Equation 11 has a stationary point at full equality, $\phi = 1$. The questions to ask are Are there others? Is this stationary point stable?

Proposition 2: (i) *If* $\eta = 0$, $\forall \phi_0$, ϕ_t *monotonically converges to* $\phi_\infty = 1$ *and the growth rate* g_t *monotonically converges (from below) to* $g_\infty^* = \beta\delta A\{(1-\alpha)(1-\beta)AR^{-1}\}^{\frac{1-\alpha}{\alpha}} - 1$.

(ii) *If* $\eta > 0$ *and* $\alpha[1+\eta(1-\beta)] < 1$, *then* $\exists\, \phi(\eta) \in (0,1)$: $\forall \phi_0 \in (\phi(\eta), 1]$, ϕ_t *monotonically converges to* $\phi_\infty = 1$ *and the growth rate* g_t *converges (from below) to* $g_\infty^{**} = \beta\delta A\{(1-\alpha)(1-\beta)(1+\beta\eta)AR^{-1}\}^{\frac{1-\alpha}{\alpha}} - 1$. $\forall \phi_0 < \phi(\eta)$, ϕ_t *monotonically converges to* $\phi_\infty = \phi_{\min} > 0$, *and* g_t *converges (from above) to* $g_\infty = 0$.

(iii) *If* $\eta > 0$ *and* $\alpha[1+\eta(1-\beta)] > 1$, *then* $\exists\, \phi^* \in (0,1]$ *such that* $\forall \phi_0 < \phi^*$, $\phi_t \to \phi_\infty = \phi_{\min}$ *and* $g_\infty = 0$. *If* $\phi^* < 1$, *then* $\forall \phi_0 \in (\phi^*, 1)$, *we have* $0 < \phi_\infty < 1$ *and* $0 < g_\infty < g_\infty^*$.

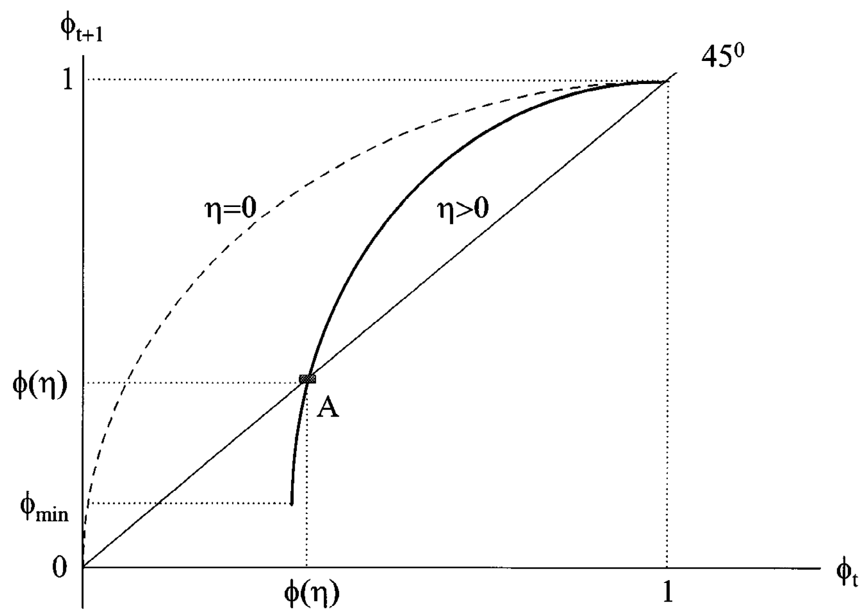

Figure 1.

Proof. See Appendix.

Therefore, as long as $\eta > 0$ so that opportunity cost effect is present, there always exists a range where wage and income inequality will be rising over time. The smaller ϕ is, the larger the difference between rich and poor, and the more likely the economy is to be in this range. Furthermore, as long as η and α are not too large, there will exist a range in which inequality will be decreasing over time. Finally, the growth rate of the economy is a negative function of the level of inequality because of mismatch: the presence of mismatch reduces the return to capital; hence, with higher inequality, investment and growth are lower. Therefore, in this economy, inequality is costly due to mismatch, but also mismatch can make inequality eventually disappear.[8]

The technical intuition of the result can be obtained from Figure 1; the broken line represents $\eta = 0$, and all possible inequality levels (all ϕ's) are in the basin of attraction of full equality. However, as soon as $\eta > 0$, (11) is negative at $\phi = 0$; thus, it starts below the 45° line. If (11) approaches $\phi_t = 1$ from below the 45° line, full equality will be globally unstable. On the other hand, to be able to approach it from above, this curve needs to cut the 45° line at some $\phi(\eta)$ (point A) as in Figure 1, and now only points to the right of point A are in the basin of attraction of full equality. To the left of A, wage inequality increases over time. Whether this curve approaches $\phi_t = 1$ from above or below 45° can be determined by looking at

$$\frac{\partial \phi_{t+1}(\phi_t = 1)}{\partial \phi_t} = \alpha + \alpha(1-\beta)\eta. \tag{12}$$

For (12) to be less than 1 (i.e., local stability of $\phi = 1$), the condition in case (ii) of Proposition 2 needs to hold.[9]

Intuitively, (i) is the case where the frictions are so high that there is no *opportunity cost or outside option* effect on wages. The wage rate is simply βy for each worker where y is the output produced by the worker and the firm. Since all firms choose the same level of capital, high-human-capital workers (h_2) produce more than low-human-capital workers (h_1) and receive higher wages. However, given decreasing returns to human capital (constant returns to scale), a worker who has twice as much human capital does not produce twice as much output. Since the accumulation rules (h_{t+1} as a function w_t) are linear, this means that his offspring is not twice as rich, and therefore the gap is getting narrower. This process will gradually take the economy to full equality. Expressed differently, when $\eta = 0$, the only redistributive force caused by labor market frictions is *mismatch*. Compared to the frictionless economy where the human to physical capital ratio was constant in all firms, in this economy high human capital agents are working with lower physical to human capital ratios than low human capital agents. Also note that this redistributive force is not directly related to the fact that workers are not receiving their marginal product in the employment relation. Even with wages equal to marginal product, if high human capital agents worked at lower physical to human capital ratios, inequality would get compressed over time. Thus, the term mismatch.

Next consider $\eta > 0$. There is now an additional redistributive force, *the opportunity cost effect*, which increases the relative wages of the skilled workers. Intuitively, a firm, when bargaining with a skilled worker, has a low outside option or opportunity cost because if it leaves the worker, on average the next worker will be worse. In contrast, it has a strong bargaining position against unskilled workers because the next worker will be on average better.[10] This opportunity cost effect, which redistributes from the poor to the rich, is stronger when the gap between the skilled and the unskilled is larger. Therefore, the mismatch effect dominates at low levels of inequality while the outside option effect dominates at high levels.

A possible conjecture after realizing the presence of two counteracting effects could have been that the mismatch effect would always dominate the opportunity cost effect. After all outside options are strongest when $\eta = 1$ and there is infinite sampling, which is the case of the frictionless economy (see Gale, 1987), and there inequality never changes. But this conjecture is wrong. Why? The intuition lies in realizing that the distribution of jobs is *endogenous* in this economy, and outside options are determined by the *composition* of jobs. In the frictionless economy, there are firms who prefer to employ the unskilled workers (low physical capital firms targeting to employ low skill workers—see also Section 5.1). In contrast, in the economy of this section all firms *dislike* employing unskilled workers (but given that $\eta < \eta^*$, they are happy *not* to segregate into two groups, one targeting the unskilled). This depresses the relative wages of the unskilled below their level in the frictionless economy.

Note also that as the economy converges to full equality, the growth rate of the economy is increasing until it finally reaches g_{∞}^{**}. Conversely, as the economy converges to maximal inequality the growth rate is decreasing. The intuition of this result can be obtained from expression (9). In this economy, accumulation is the engine of growth, and the wider the inequality of skills is the less investment there is, because firms anticipate the mismatch will reduce their profitability. This is the feature that leads to a negative link between inequality and growth (see Perotti, 1996, for the cross-country evidence on this respect).

3.3. Remarks and Discussion

First, note that the results in Proposition 2 are intimately linked to mismatch: the fact that workers of different levels of skills are employed at different human to physical capital ratios. There is evidence that mismatch between the skills of workers and the requirements of firms is a ubiquitous phenomenon in the real world (e.g., the literature on undereducation and overeducation, see Rumberger, 1981). An interesting empirical piece for us is a careful paper using the PSID by Sicherman (1991). He finds that workers who have more education than required for the job (overeducated workers) earn more than other workers doing the same job but less than a typical worker with the same education level and characteristics. Conversely, undereducated workers earn less than others doing the same job but more than a typical worker with the same characteristics. Although the evidence could also be interpreted as overeducated workers having less "unobserved capital," this is not consistent with the rest of Sicherman's results; overeducated workers have a significantly higher probability of moving to a better job, which means that on average these workers are truly overqualified for the job they are performing and also, this reallocation is quite slow implying that mismatch is not a very transitory phenomenon.

Second, note that increasing the number of workers per firm would not change the results as long as there is diminishing marginal product of human capital, i.e., the return to the human capital of one worker is decreasing once all other workers and irreversible attributes of the job are in place. The interesting question is whether a large firm can avoid the mismatch problems. For instance, a firm planning to hire two workers can open one skilled and one unskilled job. But this will clearly not solve all problems. When two skilled or two unskilled workers arrive, there will again be mismatch. It can also be thought that in the limit, if a firm can employ a continuum of workers, there will be no uncertainty regarding the qualifications of new employees. However, in practice there are limits as to how large firms can become, especially when they have diverse jobs and work forces. And even large firms do not have a large number of job openings at the same time; for instance, a large firm in need of an engineer in a given period would face the same problems as here.

Third, as noted in the introduction, diminishing marginal product of human capital play an important role in this result (see for instance, Benabou, 1996). This assumption is completely standard in this context: in the absence of decreasing returns to human capital, the Walrasian anlogue of this economy analyzed in Section 2 would not be well-defined. This comparison with the Walrasian economy also demonstrates that the key results are not driven by technological externalities but are due to the frictions in the labor market.

Fourth, note that in this economy human capital investments are completely backward looking; parents, when deciding how much education to give to their offspring, do not consider the rate of return on human capital. When human capital investments are forward-looking, with high inequality, the incentives to invest in human capital will also be high, and this will act as a countervailing force (see Acemoglu, 1996a, for forward-looking human capital investments in a model of frictional labor market).

Finally, it is important to recall that the economy is analyzed in the range $\eta < \eta^*$ where firms accept the first applicant who comes along, and there is no unemployment. Some of the results are sensitive to this assumption. The case in which firms turn down workers is analyzed in a simpler setting in Acemoglu (1996b). There, I find that a mean preserving spread of the distribution of skills again leads to more inequality of earnings, but not necessarily to more mismatch, nor to a lower rate of return on physical capital. The qualification that the results here apply when the economy is away from the frictionless case is an important caveat to bear in mind when interpreting the results.

3.4. The Importance of Institutions and Implications for Cross-country Trends

Labor Market Efficiency. Labor market institutions differ across countries. Large differences between U.S. and European labor markets are often emphasized and the less efficient labor markets of European countries are suggested as the reason for the limited increase in inequality (e.g., Katz et al., 1995, and Bertola and Ichino, 1995). In the model, the parameter η is a measure of the degree of frictions in the labor market. The greater is η, the less are mobility costs and the closer are wages to marginal product. Thus a labor market with easier turnover and less unionization can be thought of as one with higher η. From equations 11 and 12, the higher is η, the more likely is inequality to increase as long as we are in the range $\eta < \eta^*$ (thus, the economy is sufficiently away from the frictionless case).

In particular, the system is more likely to have a stable region for lower η. A reduction in η shifts the solid curve in Figure 1 up and the point A to the left, and therefore, enlarges the basin of attraction of full equality (the region of decreasing wage inequality). Thus consider two economies a and b, identical except that $\eta_b < \eta_a < \eta^*$ (i.e., a has lower search costs), then from Figure 1, a will always have more inequality—a lower ϕ_t—than economy b. Moreover, there exists a set of initial inequality levels $F_0 = F_0^a = F_0^b$ such that starting from these inequality levels, b will converge to full equality and steady growth, while economy a converges to maximum inequality and no growth.

In other words, since the main role of higher η in this economy is to redistribute income from the poor to the rich (or slow down the reverse redistribution), a higher level of η is harmful to equalization of incomes, thus it creates more inequality and more mismatch.

Implication 3.1: *Inequality is more likely to increase with more efficient labor markets.*

Implication 3.2: *More efficient labor markets can lead to poor long-run performance.*

These implications are in line with the view that the rise in wage and income inequality in the United States is due to more efficient labor market institutions (easy turnover, low firing costs and weak unions); however, they question the conventional wisdom that this is necessarily a desirable outcome.

Also, this analysis suggests that a change in η can be a potent driving force for the increasing wage inequality. Both in the United States and the United Kingdom, the power of unions was severely reduced at the beginning of 1980s, and during the same period inequality increased sharply (see DiNardo, et al., 1996). As the reduction in the power of unions could be captured as an increase in η, this model predicts that it can take the economy from the region of falling inequality to one of widening inequality.

Redistribution. Redistribution is generally thought to reduce growth due to its adverse incentive effects. However, Perotti (1996) finds that across countries, more redistribution is positively correlated with growth during the post-war period. This is the prediction of models like Benabou (1996) and others discussed previously. It is also a prediction of my model, but the mechanism again works through mismatch, and is thus worth outlining.

Consider a tax rate τ on wage earnings, which is then redistributed lump-sum among all agents. This can be interpreted as redistributive labor income taxation, or public schooling where rich households subsidize the education of poorer households.

Now the dynamics of inequality can be characterized as:

$$\begin{aligned}\phi_{t+1} &= \frac{h_{1,t+1}}{h_{2,t+1}}\frac{\delta w^p_{1t}}{\delta w^p_{2t}} = \frac{(1-\tau)w_{1t}+\lambda\tau w_{1t}+(1-\lambda)\tau w_{2t}}{(1-\tau)w_{2t}+\lambda\tau w_{1t}+(1-\lambda)\tau w_{2t}} \\ &= \frac{w_{1t}-\tau(1-\lambda)(w_{2t}-w_{1t})}{w_{2t}-\tau\lambda(w_{2t}-w_{1t})}, \end{aligned} \tag{13}$$

where w^p denotes the post tax income and as before, w denotes the actual wages and the subscripts are again used to distinguish the two groups. Now substituting from (10)

$$\phi_{t+1} = \frac{[1+(1-\beta)(1-\lambda)\eta]\phi_t^\alpha-(1-\beta)(1-\lambda)\eta+(1-\lambda)\tau[1+(1-\beta)\eta](1-\phi_t^\alpha)}{+(1-\beta)\lambda\eta-(1-\beta)\lambda\eta\phi_t^\alpha-\lambda\tau[1+(1-\beta)\eta](1-\phi_t^\alpha)} \tag{14}$$

This difference equation, like (11), has a steady state at $\phi = 1$. Stability now depends on;

$$\frac{\partial\phi_{t+1}(\phi=1)}{\partial\phi_t} = \alpha[1+(1-\beta)\eta-\tau(1+(1-\beta)\eta)] \tag{15}$$

Therefore, the economy with redistributive taxation will tend to be more stable because the slope of the curve $\phi_{t+1}(\phi_t)$ near the steady state $\phi = 1$ is always decreasing in τ. Also,

$$\phi_{t+1}(\phi_t=0) = \frac{-(1-\beta)(1-\lambda)\eta+\tau(1-\lambda)[1+(1-\beta)\eta]}{1+(1-\beta)\lambda\eta-\tau\lambda[1+(1-\beta)\eta]} \tag{16}$$

is increasing in τ, and $\phi_{t+1}(\phi_t = 0)$ is no longer always negative, and for high enough values of τ, the system may become globally stable. Thus;

Implication 3.3: *Redistribution reduces the forces which lead to inequality in the pretax wage distribution, thus stabilizes the system. It also tends to increase growth in the long-run.*

It is straightforward to see that the same analysis can be repeated for a temporary redistribution, and the conclusion would be that a temporary redistribution can change the long-run dynamics of inequality and growth in this economy. This is naturally a consequence of the multiple limiting income distributions with different long-run growth rates.

Overall, these comparative static results suggest that societies with less redistributive tax systems and more emphasis on private funding of education, again the United States and the United Kingdom should have had a more pronounced increase in their *pretax* wage and income inequality, but perhaps also lower long-run growth rates.

4. Inequality Dynamics With General Income Distributions

In this section I will demonstrate that the results obtained so far generalize to economies with more general starting distributions of inequality than two classes. This analysis will also yield some new predictions.

Wages are still given by (7), which is equivalent to assuming that η is smaller than an appropriately defined η^{**}, so that there are no separations along the equilibrium path. The economy will now start with an arbitrary distribution $F_0(h)$ and a corresponding distribution of relative wealth $G_0(\theta)$. As long as the poorest agent is away from $h_{\min}$, the dynamics of inequality can be determined from the following equation;

$$\theta_{j,t+1} = \frac{h_{j,t+1}}{\tilde{h}_{t+1}} = \frac{\beta[1+(1-\beta)\eta]h_{j,t}^{\alpha} - \beta(1-\beta)\eta\int h^{\alpha}dF_t(h)}{\beta[1+(1-\beta)\eta]\tilde{h}_t^{\alpha} - \beta(1-\beta)\eta\int h^{\alpha}dF_t(h)}$$
$$= \frac{\beta[1+(1-\beta)\eta]\theta_{j,t}^{\alpha} - \beta(1-\beta)\eta\int \theta^{\alpha}dG_t(\theta)}{\beta[1+(1-\beta)\eta] - \beta(1-\beta)\eta\int \theta^{\alpha}dG_t(\theta)}, \quad (17)$$

where recall that $\tilde{h}_t$ is the median of the human capital distribution at time t. In contrast, when the poorest agent hits $h_{\min}$, then for that agent we have $\theta_{j,t} = \theta_{\min} = \frac{h_{\min}}{\tilde{h}_t}$. Therefore, the evolution of inequality is determined by a dynamic functional equation. Although it is not possible to solve this type of equation, a number of useful features of the dynamics of inequality can be characterized. The main results are summarized in Proposition 3 (proof in Appendix).

Proposition 3: A) *Suppose* $\eta = 0$, *then* $\forall G_0$, G_t *converges monotonically to full equality and the growth rate of the economy converges monotonically to* g^*_∞.

B) *Suppose* $\eta > 0$. *Then,*

(i) *Full equality is always a stationary distribution.*

(ii) *There always exists another stationary distribution with two or three groups. One group is with positive measure at* $h_{\min}$*; one group is with positive measure at some* h *above the median; and one more group may also have positive measure at the median.*

(iii) *No stationary distributions other than the ones in (i) and (ii) exist.*

(iv) *If* $\alpha[1+(1-\beta)\eta] < 1$, *then full equality is locally stable and otherwise, it is not.*

(v) *Maximal inequality with two groups is always locally stable.*

With $\eta = 0$ only the mismatch effect is present as before, and this leads to global stability of full equality. With $\eta > 0$, full equality is always a stationary distribution as inspection of (17) shows immediately. Also, the same condition as in the two-class economy ensures the local stability of this stationary distribution. To see what other stationary distributions are possible we can make use of Figure 2. This figure plots (17) with $\theta_{j,t+1}$ on the vertical and $\theta_{j,t}$ on the horizontal axis for a given distribution of inequality at time t (i.e., for a given value of the integral $\int \theta^{\alpha} dG_t(\theta)$). The curve is upward sloping and strictly concave. It always intersects the 45° line at $\theta = 1$ and never for any $\theta < 1$, but it may intersect 45° one more time at some $\theta > 1$. A stationary distribution must have the property that $\theta_{j,t+1} = \theta_{j,t}$, for all j, thus there cannot be a stationary distribution with positive weight at a level below the median unless (17) does not apply, i.e. a group at $h_{\min}$. Also, by the same argument, the stationary distribution can have positive weight at most on one group above the median.

4.2. Which Distributions Are Likely to Lead to Increasing Wage Inequality

To answer this question, we need to investigate the role that the integral term in equation (17) plays in determining convergence. I will therefore determine how the conditions for a "poor" agent to get richer are affected by the integral term. First

Lemma 1: *Suppose infimum*$\{\theta_t\} < \alpha^{\frac{1}{1-\alpha}}$, *then* $\exists\bar{\theta}_t$, *such that* $\theta_t < \bar{\theta}_t$ *get poorer and* $\theta \in (\bar{\theta}_t, 1)$ *get richer.* $\bar{\theta}_t$ *is increasing in* $\int \theta^{\alpha} dG_t(\theta)$.

Proof. See Appendix.

The economic intuition of this result is straightforward. The integral term is related to the outside option (opportunity costs) of firms—observe that when $\eta = 0$, the integral terms disappear. It the outside option of firms is sufficiently high—a large value of the integral—then there exists a section at the bottom of the distribution who have to accept a very low wage to get employed, and therefore, this group will get poorer while the rest of the population converges to a higher level of skills and income.

When is the integral term high? Intuitively, the outside option of firms will be high if they expect to get a highly skilled worker with a high likelihood; therefore, when the distribution of skills is fairly equal. Since it is the differential opportunity costs (outside options) of firms which introduces the forces towards increasing wage inequality, convergence to full equality is most difficult when the distribution is skewed to the right—that is, when there are a large number of rich agents with a smaller group of agents who are sufficiently poor relative to them. When this is the case, the integral term is large, and there is not much

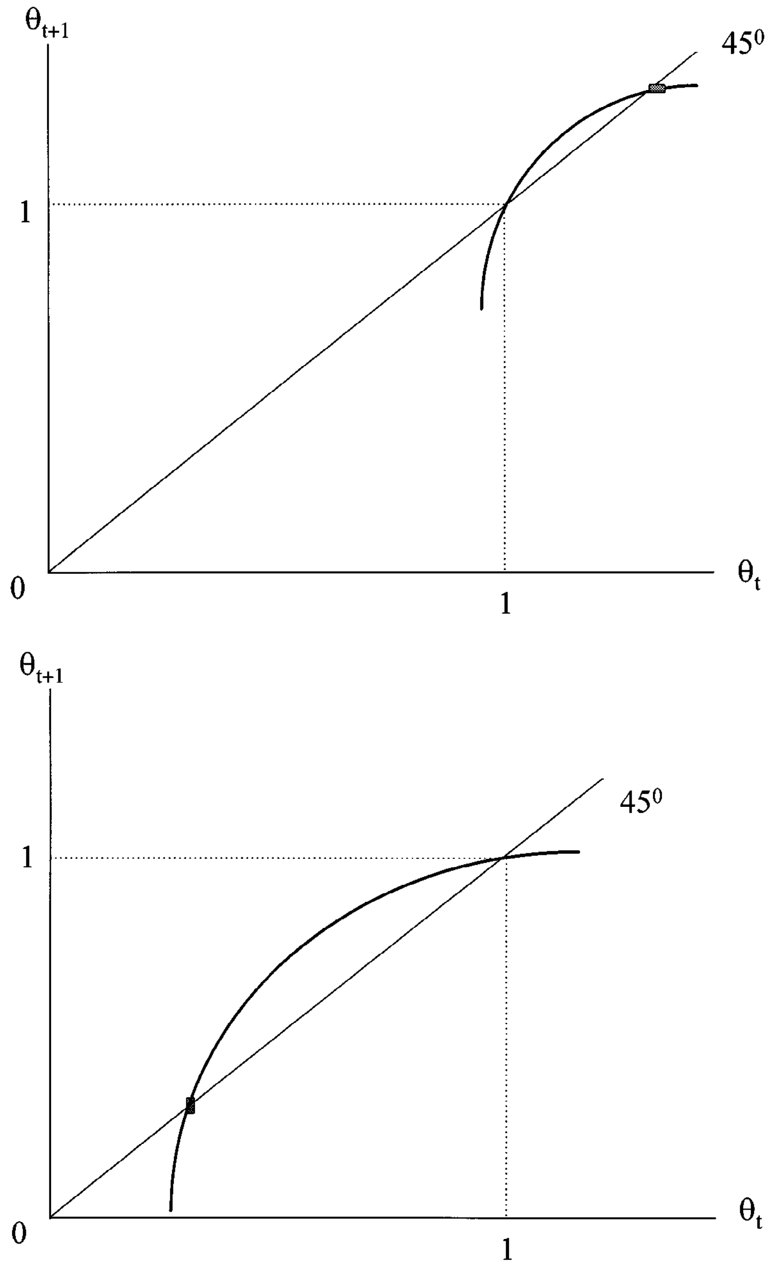

Figure 2.

uncertainty regarding the human capital of the workers that the firm can get at the second round of matching. Consequently, the poorest agents will have to accept low wages and thus get even poorer over time.

Implication 4.1: *Wage inequality is more likely to increase when the distribution of income and skills are skewed to the right.*

4.3. Inequality Cycles

Proposition 3 only dealt with stationary distributions. However, a dynamic system may also settle into a cycle. Whether this economy can generate endogenous cycles is of some interest as it illustrates the interaction of the counteracting forces in the model. The discussion around Lemma 1 suggests that inequality cycles may be possible in this model. Intuitively, poor agents are more likely to get poorer when income inequality is limited, but when they get poorer income inequality increases, and because the integral term in (17) falls, they may again get richer. However, with only two-groups, Proposition 2 gave the complete characterization, and there were no cycles. In fact, we will find that two groups is a special case and with more general starting distributions of income cycles are a generic possibility.

I prove the possibility of two-period-cycles by constructing an example. Consider the following economy consisting of three groups: $\theta^H > \theta^M = 1 > \theta^L$ with proportions λ^H, λ^M, and λ^L such that $\lambda^L + \lambda^M + \lambda^H = 1$, and to ensure that agents in the middle group are the median, suppose that $\lambda^L < 1/2$ and $\lambda^H < 1/2$. Figure 3 draws the possibility of a two-cycle. In odd numbered periods inequality increases and the share of high group goes up from θ_2^H to θ_1^H. Similarly, the share of the Low group goes from θ_2^L to θ_1^L. In contrast, in even numbered periods the share of the Low group goes up to $\theta_2^L > \theta_1^L$ and that of the High falls to $\theta_2^H < \theta_1^H$. Since inequality is higher in odd numbered periods, the curve in Figure 3 that describes transitions is the broken one. In contrast, inequality decreases in even numbered periods, thus the solid curve in Figure 3 applies, and this curve is a tilted image of the broken one around $\theta = 1$. The intuition for this figure follows from Lemma 1. When the integral increases, the poor get poorer and the rich get richer, and hence the curve tilts around $\theta = 1$ from the broken to the solid curve.

It can also be verified that when there are more than three groups, three-period cycles are possible and thus from Sarkovski's Theorem (see Grandmont, 1985, Li and York, 1975) cycles of any periodicity and chaotic behavior can arise in this economy.

Proposition 4: (i) *In a three group economy, there always exists an open set of vectors* $(\lambda^L, \lambda^H, \theta_1^L, \theta_1^H, \theta_2^L, \theta_2^H)$, *thus a corresponding set of starting values, which lead to a two-period cycle.*

(ii) *In a four group economy, there exists an open set of starting distributions which lead to cycles of any periodicity.*

Proof. See Appendix.

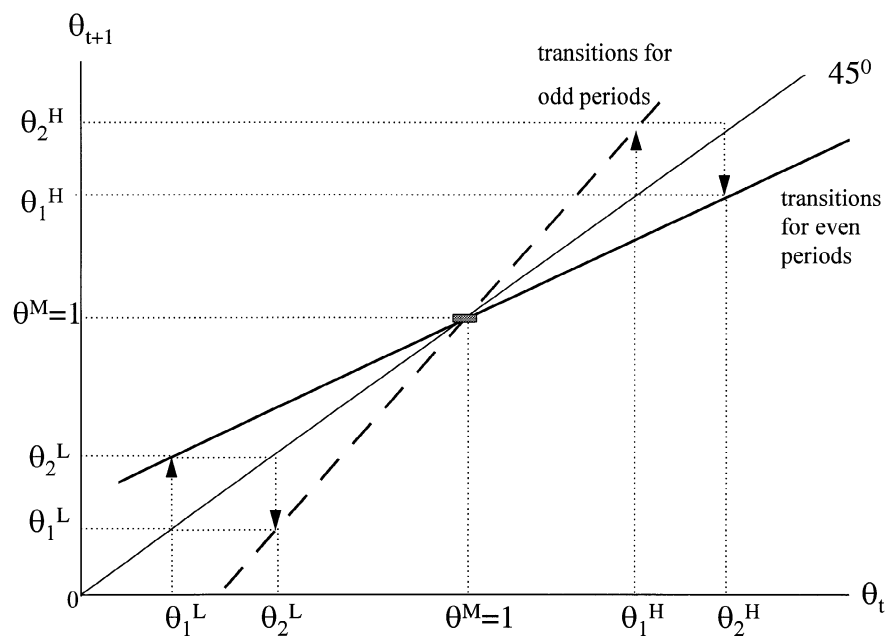

Figure 3.

5. More General Matching Technologies, Mismatch and Implications for Within Versus Between Group Inequality

The premise that an increase in heterogeneity will lead to more mismatch has played an important role in the analysis so far and was derived for the special case of random matching. However, it is quite clear that for many situations random matching is not an appropriate assumption. For instance, skilled workers often do not look for unskilled jobs nor do firms post vacancies that are open to all skill categories. Of equal importance is that so far a change in the skill composition affected the investment of firms but not the diversity of jobs that were offered, and this is again unrealistic. This section will analyze these issues. An interesting new prediction will also follow from the analysis: as higher inequality leads to more types of jobs becoming available, the between and within group measures of inequality will covary as in the data.

I start by defining the polar extreme to the random matching technology: efficient matching. If the matching technology of the economy is efficient, then the highest skilled worker is allocated to the firm with the highest amount of physical capital, the second most skilled is allocated to the second highest, and so on. If there is a separation, then the same rule applies within the set of separated workers. This is clearly the same allocation that the Walrasian auctioneer chose in the equilibrium of section II. More formally,

Definition: *Let us define for every worker* j, $\Omega_W(j) = \int_{s \in S_W : h_s > h_j} ds$ *and similarly for each firm* i, $\Omega_F(i) = \int_{s \in S_F : k_s > k_i} ds$ *where* S_F *and* S_W *are the sets of firms and workers who are looking for a match. Then* $\Omega_F(i)$ *and* $\Omega_W(j)$ *are respectively the ranks of firm* i *and worker* j *in the set of firms, and in the set of workers looking for a match. Let us also again use the notation* $(i, j) \in \mathcal{P}$ *if* i *and* j *will be matched together. The matching technology is efficient iff* $(i, j) \in \mathcal{P} \Leftrightarrow$

either (i) $\Omega_W(j) = \Omega_F(i)$

or (ii) if $\Omega_W(j) < \Omega_F(i) \Rightarrow \forall i^*$: $\Omega_F(i^*) < \Omega_F(i)$ *and* $(i^*, j^*) \in \mathcal{P}$, *then* $\Omega_W(j^*) \leq \Omega_W(j)$.

or (iii) if $\Omega_W(j) < \Omega_F(i) \Rightarrow \forall j^*$: $\Omega_W(j^*) < \Omega_W(j)$ *and* $(i^*, j^*) \in \mathcal{P}$, *then* $\Omega_F(i^*) \leq \Omega_F(i)$.

Note that (ii) and (iii) take care of the case in which there are atoms in the distribution of human or physical capital.

The matching technology I will use is a hybrid between the efficient and random technologies. In particular, I assume that a *random* proportion $(1 - q)$ of firms and workers are chosen to match randomly among themselves, and the remaining q of the firms and workers are matched efficiently. As a motivation consider the case where a worker looks at all firms before deciding which one to apply to. However, there is only a probability q that he will correctly assess the firm's type, and otherwise, with probability $(1 - q)$ he will be in effect applying to a random firm.

The exact timing of events is as follows: as before, first $F_t(h)$ is determined from the bequest decisions of the previous generation. Then firms choose their physical capital which determines $P_t(k)$. Subsequently, a proportion q of firms and workers are randomly drawn for efficient matching, and match efficiently among themselves, and the remaining firms and workers match randomly. Finally, wages are determined. To simplify the analysis in this section, I assume $\eta = 0$, thus workers receive a proportion β of the output they produce.

For the analysis it is also important that, since there is a continuum of workers, the subsamples selected for random and efficient matching are identical to the overall distributions. Further, I use the notation $(k', h') \in \mathcal{P}_t$, to denote h' and k' matching together at time t, if they were both chosen to match *efficiently*. Then

Lemma 2: *For all* $q > 0$, $\forall (k', h') \in \mathcal{P}_t$, $F_t(h') = P_t(k')$.

Proof. See Appendix.

This lemma states that if a firm and a worker will work together when selected for efficient matching, then they must have *exactly* the same rank in their respective distributions, i.e., in terms of the definition in the above paragraph, the equilibrium will always be in case (i). It is not surprising that given efficient matching and complementarities between human and physical capital, a high-skill worker and low-physical-capital firm will not match together. However, the result of the lemma is considerably stronger than this: it rules out the possibility of *clustering* by firms; for instance, the possibility that two firms choose the same level of physical capital and one matches with a high skill worker, and the other with a less skilled one is ruled out. Why?

Suppose two firms of the same capital level k expect to match with two workers of different levels of human capital in the case when they are both selected for efficient matching. Since $q > 0$, there is a positive probability that both firms will be selected for efficient matching, and therefore, either one can increase its ex ante profits by increasing its investment by a small amount (to $k+\epsilon$), and ensure that it will match with the higher human capital worker in case it is selected for efficient matching. For all $q > 0$, we can find ϵ small enough that this is profitable, hence the result of no clustering. Moreover, the subsamples of firms and workers selected for efficient matching are identical to the initial distributions F_t and P_t, therefore these distributions must have the same form as each other (that is, F_t must be a one-to-one transformation of P_t), thus we know a lot about the equilibrium distributions.

Now the profits of a firm with physical capital k are given by

$$\begin{aligned}\pi_t(k) \;=\; & (1-q)(1-\beta)Ak^{1-\alpha}\int \nu^{\alpha} dF_t(\nu) \\ & + q(1-\beta)Ak^{1-\alpha}h^{\alpha} - Rk \quad s.t. \quad (k,h)\in\mathcal{P}_t \end{aligned} \tag{18}$$

The first term is the expected revenue when the firm is selected for random matching, and the second is the revenue that the firm will obtain when it matches efficiently. It is clear that all firms have to make the same level of profits. Thus, $\pi_t(k) = \pi_t$ for all levels of investment k that are chosen in equilibrium.

Next the first-order condition for a typical firm can be written as

$$\begin{aligned} & (1-\alpha)(1-q)(1-\beta)Ak^{-\alpha}\int \nu^{\alpha} dF_t(\nu) + (1-\alpha)q(1-\beta)Ak^{-\alpha}h^{\alpha} \\ & \quad - R + \alpha q(1-\beta)Ak^{1-\alpha}h^{\alpha-1}\frac{dh}{dk}|_{\mathcal{P}_t} = 0, \end{aligned} \tag{19}$$

where the last term captures the fact that when the firm increases its investment, with probability q, it is also improving the human capital of the worker it will match with. Using equation 19, it can be established that

Lemma 3: *For $q = 1$, the physical to human capital ratio, $\mu(k)$, for workers matching efficiently is constant. For all $q < 1$, this ratio, $\mu(k)$, is decreasing in k.*

Proof. See Appendix.

This is a very strong result. It demonstrates that for $q < 1$, as well as for the workers matching randomly, there will be wage compression across skill groups among those allocated to efficient matching. Let us first understand the result of the constant human to physical capital ratio with full efficient matching ($q = 1$). In this case the economy is very close to the Walrasian allocation with the only difference that, due to ex post bargaining, firms underinvest. In fact, a firm knows that it can match with the highest skill worker if it chooses a high enough level of physical capital. It will only be happy not to do so when it is making the same level of profits from another worker, and this requires that all firms have the same human to physical capital ratio. Now consider $q < 1$, and suppose $\mu(k)$ is constant so that all firms matching efficiently have the same human to physical capital ratio. Recall that in this case, each firm can also be selected for random matching, and a firm with a high level of physical capital will lose more from being randomly allocated—recall the strict concavity of the profit function in the random matching case as in equation 8. Thus to be compensated, such a firm, when matching efficiently, should make higher profits, that is, it should work with a higher human to physical capital ratio.

Next, given mismatch established in Lemma 3, the next two results can be proved:

Lemma 4: *For all $q < 1$, a mean-preserving spread of $F_t(.)$ reduces investment and output.*

Proposition 5: *Suppose $\eta = 0$. Consider an initial human capital distribution F_0 and a corresponding distribution of relative wealth G_0.*

(i) *Suppose $q = 1$, then $\forall G_0$, $G_t = G_0$ for all t, thus inequality self-replicates and the growth rate $g_t = g^*_\infty$ for all t.*

(ii) $\forall\, 0 < q < 1$ *and* $\forall\, G_0$, $G_t \to G_\infty$ *where G_∞ exhibits full equality. The growth rate of the economy, g_t, is always less than g^*_∞ and monotonically converges to it.*

Proof. See Appendix.

In this economy with hybrid matching technology the key results of the Section 3 hold; in particular, as long as $q < 1$, the high skilled workers *always* produce with lower physical to human capital ratio than low skilled workers.[11] Furthermore, the degree of efficiency of the matching technology is another measure of labor market efficiency, and a higher q (more efficient matching technology) makes inequality more long-lasting. Therefore this result complements the one obtained in Section 3.4 that more efficient labor markets are more likely to lead to higher wage inequality.

5.1. Between Versus Within Group Inequality

Finally, some new implications regarding the interaction of between group and within group inequality can also be drawn from the analysis of this section. Let me first define the expected wage of a worker with human capital h_j at time t by $W_t(h_j)$. This wage depends on the equilibrium choices of firms. In particular, denote the equilibrium distribution of firms by $P_t(k)$ again, and consider k_j be such that $(h^j, k^j) \in \mathcal{P}_t$. Then

$$W_t(h_t) = (1-q)\beta A(h_j)^\alpha \int k^{1-\alpha} dP_t(k) + q\beta A(h_j)^\alpha (k_j)^{1-\alpha}. \qquad (20)$$

And clearly for $h_1 < h_2$, $W_t(h_2) - W_t(h_1) > 0$. If these two skill groups are observable, this difference is what will be measured in data as the "skill" premium or as the between group wage inequality.

Within group wage inequality is also present in this economy; the wage distribution for workers who have human capital h_1 has a mass of q at $\beta A(h_1)^\alpha (k_1)^{1-\alpha}$ and the rest of its mass is distributed as $\beta A(h_1)^\alpha k^{1-\alpha}$ where k has distribution $P_t(k)$. Now, take a mean-preserving spread of F_t. This will mean that there is more inequality and in fact the gap between the less and the more skilled is larger, hence more *between group* inequality. But when F_t undergoes a mean-preserving spread, then from Lemma 2, P_t also becomes more spread (though its mean falls too), and the *within* group wage inequality increases. Intuitively, higher between group inequality leads to a more diverse distribution of jobs. Yet, sometimes (with probability $1 - q$) two identical workers end up randomly allocated to these jobs, and hence, more diversity in these jobs implies a larger wage gap between these two identical workers.

Finally, although it is often the case, between and within group inequality do not always move together. For instance, in the late 1970s, the United States experienced an increase in within group inequality but not in between group wage inequality (see Juhn et al., 1993). Such a pattern can be explained if q is changing, which will be a consequence of the changing organization of firms and labor markets. From (20), an increase in q will increase the wages of the h_2 (high skill) group more than that for the h_1 (low skill) group, leading to an increase in between group inequality. It will also increase the proportion of firms and workers matching efficiently; thus, it will generally lead to a compression in within group inequality.

6. Concluding Comments

This paper has presented a general equilibrium model with endogenous job and skill distribution. The model provides a simple framework with which to analyze the interaction between labor markets and the evolution of income and wage inequality. This investigation has revealed that labor market frictions introduce two new redistributive forces. The first one, which I dubbed mismatch, redistributes from rich to poor workers while the second, the opportunity cost effect, redistributes from the poor to the rich. Moreover, the second effect becomes stronger when there is more inequality, and this led to the conclusion that the dynamics of income and wage inequality are nonergodic: inequality is decreasing at low levels but can increase starting from high initial levels of inequality.

This framework has also enabled some simple comparative statics. Most important, wage inequality is more likely to increase in economies with more efficient labor markets, and within and between group inequality should move together. Both of these predictions are consistent with the broad patterns of inequality dynamics in the data.

This paper is a first attempt at a framework for the theoretical analysis of wage inequality and labor market organization. As such, it leaves many issues unresolved. First, I have restricted attention to parameter values for which there are neither separations nor disagreements along the equilibrium path. Therefore, this model is not well-suited to analyze issues of unemployment and inequality, which are interesting and important. Introducing separations makes the model technically much more complicated (see for instance the models of search with ex ante heterogeneity of Sattinger, 1995, or Shimer and Smith, 1996), but this is an important extension to consider (see Acemoglu, 1996b). Second, it is natural to question whether segmentation in the labor market is likely to arise as a way of limiting mismatch (and whether this type of model will predict endogenous segmentation in the labor market). Finally, this model poses a number of new empirical questions; can we find a good measure of mismatch? Is mismatch higher in labor markets with more inequality? Is performance in labor markets with higher inequality worse?

Appendix: Proofs of Lemmas and Propositions

Proof of Proposition 1

(i) Take the equilibrium with the physical to human capital ratio as in Proposition 1 and suppose all workers are paid their marginal product as in (6). Then a firm with capital level k makes profits equal to

$$\begin{aligned} \pi(k) &= Ak^{1-\alpha}h^{\alpha} - Rk - w(h)h \\ &= \frac{1}{1-\alpha}R - Rk - \frac{\alpha}{1-\alpha}Rk, \\ &= 0 \end{aligned} \tag{A1}$$

irrespective of the value of k. Thus at the allocation where all firms are allocated according to the Walrasian rule (see the definition in section V.1) and k/h is constant, all firms make zero profit. Given the value of h, each firm would also exactly choose this level of capital ratio since

$$\pi'(k) = 0 \text{ at } k = \left(\frac{(1-\alpha)A}{R}\right)^{\frac{1}{\alpha}} h. \tag{A2}$$

Thus no firm wants to deviate and change its capital stock, and the allocation is an equilibrium. At no other capital stock , a firm would be maximizing profit, thus there is no equilibrium in which firms are at a different capital ratio.

(ii) Given a constant physical to human capital ratio, we can write the wage of worker with human capital $h_{j,t}$ as

$$w(h_{j,t}) = \alpha A \left(\frac{(1-\alpha)A}{R}\right)^{\frac{1-\alpha}{\alpha}} h_{j,t}. \tag{A3}$$

Thus from (3) in the text,

$$h_{j,t+1} = \delta\alpha A \left(\frac{(1-\alpha)A}{R}\right)^{\frac{1-\alpha}{\alpha}} h_{j,t}, \tag{A4}$$

for all j. Next, $\theta_{j,t} = \frac{h_{j,t}}{\tilde{h}_t}$ and since all transitions are linear, $\tilde{h}_{t+1} = \delta\alpha A(\frac{(1-\alpha)A}{R})^{\frac{1-\alpha}{\alpha}}\tilde{h}_t$ and hence, $\theta_{j,t+1} = \theta_{j,t}$.

(iii) From the above equations, the human capital of all agents grows at the rate

$$g^C = \delta\alpha \left(\frac{(1-\alpha)A}{R} \right)^{\frac{1-\alpha}{\alpha}} - 1, \tag{A5}$$

and given that the capital ratio is constant all the time, this is the rate at which firm level capital stocks grow too, hence also the output growth rate. □

Proof of Proposition 2

(i) Consider Figure 1 in the text; stationary points (distributions) correspond to the intersection of $\phi_{t+1}(\phi_t)$ with the 45° line and $\phi_{t+1}(\phi_t = 1) = 1$, thus full equality is always a stationary point.

Now from (11) in the text,

$$\phi_{t+1}(\phi_t = 0) = \frac{-(1-\beta)(1-\lambda)\eta}{1 + \eta(1-\beta)\lambda}. \tag{A6}$$

This is negative for all $\eta > 0$ and equal to zero for $\eta = 0$.

$$\begin{aligned} \frac{\partial\phi_{t+1}}{\partial\phi_t} = {} & \frac{[\alpha\phi^{\alpha-1} + (1-\beta)(1-\lambda)\eta\alpha\phi_t^{\alpha-1}] \times [1-(1-\beta)\lambda\eta\phi_t^{\alpha} + (1-\beta)\lambda\eta]}{[1-(1-\beta)\lambda\eta\phi_t^{\alpha} + (1-\beta)\lambda\eta]^2} \\ & + \frac{[(1-\beta)\lambda\eta\alpha\phi_t^{\alpha-1}] \times [\phi_t^{\alpha} + (1-\beta)(1-\lambda)\eta\phi_t^{\alpha} - (1-\beta)(1-\lambda)\eta]}{[1-(1-\beta)\lambda\eta\phi_t^{\alpha} + (1-\beta)\lambda\eta]^2}. \end{aligned} \tag{A7}$$

This derivative in (A7) always exists and is positive, and when $\eta = 0$, (11) starts as negative in Figure 1 and never falls below the 45° line. Thus the system is globally stable. The growth rate at the limit is obtained straightforwardly from (9) and (4) for $\eta = 0$.

(ii) Next for all $\eta > 0$, the curve in Figure 1 starts negative at $\phi = 0$ and thus if it is going to approach $\phi = 1$ from above, it must cut the 45° line at least once before $\phi = 1$. Evaluating (A7) at $\phi_t = 1$, we get (12). This implies that for $\alpha[1 + (1-\beta)\eta] < 1$, the function in Figure 1 cuts the 45° line at $\phi_t = 1$ from above (with a slope of less than 1). Differentiation shows that either $\partial\phi_{t+1}^2/\partial\phi_t^2 < 0$ or $\partial\phi_{t+1}^2/\partial\phi_t^2 > 0$ but then $\partial^3\phi_{t+1}/\partial\phi_t^3$ is unambiguously positive [full details available upon request]. Therefore, the function in Figure 1 cannot turn from convex to concave and hence, the shape of the function can only be as in Figure 1.

As a result, the curve must cut the 45° line once at some point A. let us call the horizontal coordinate of A, $\phi(\lambda)$. Then, $\forall\,\phi_t \in (\phi(\eta), 1)$, $d[\phi_{t+1}, 1] < d[\phi_t, 1]$ where $d[.,.]$ is the Euclidean distance between two points. Thus, all points to the right of A are in the basin of attraction of full equality; or in other words, $\phi_0 \in (\phi(\eta), 1)$, $\phi_t \to \phi_\infty = 1$. And again the growth rate at the limit is given by (9) and (4); and before this limit is reached, inspection of (9) shows that investment is less, thus growth is lower.

Similarly, $\forall\phi_t \in (\phi_{\min}, \phi(\eta))$ where $\phi_{\min}$ will be defined below, $d[\phi_{t+1}, 1] > d[\phi_t, 1]$ and thus for all ϕ_0 to the left of A, we diverge from full equality. The dynamics for both groups are defined by continuous functions, and $\{\phi_t\}$ forms a monotonic sequence and is defined between 0 and 1, thus over a closed and bounded set. This implies that $\{\phi_t\}$ must have a convergent subsequence, therefore, we must converge to a certain point. Since $h_{1,t}$ is bounded below by $h_{\min}$, this sequence $\{\phi_t\}$ can only tend to zero, if $h_{2,t}$ goes to infinity. Thus we need to check whether growth in $h_{2,t}$ can be sustained with a proportion $(1-\lambda)$ of the agents accumulating and the rest at the lower bound $h_{\min}$. In this case the law of motion of h_{2t} would be given by

$$\begin{aligned} h_{2,t+1} &= \delta w_{2t} = \beta h_t^{\alpha} k_t^{1-\alpha} - \beta(1-\beta)\eta k_t^{1-\alpha}[\lambda h_{\min}^{\alpha} + (1-\lambda)h_{2t}^{\alpha}] \\ &\quad + (1-\beta)\beta\eta A k^{1-\alpha} h_{2t}^{\alpha}. \end{aligned} \tag{A8}$$

I will show that $h_{2,t}$ growing without a bound (at a constant rate) is not possible by contradiction: $h_{2,t}$ can grow only if k_t is linear in $h_{2,t}$. So let us suppose that k_t is linear in $h_{2,t}$. This would imply from (A8) that $h_{2,t+1} = Bh_{2,t} + Ch_{2,t}^{\alpha}$ where B and C are suitably defined constants. But this implies that constant growth is not possible for $h_{2,t}$ and there exists a unique level of h_2 such that $h_{2,t+1} = h_{2,t} = h_2$. Similarly, if k_t were an everywhere concave function of $h_{2,t}$, the same conclusion would apply a fortiori. From (9), k_t is immediately seen to be a concave function of $h_{2,t}$ and therefore, there is a unique level of h_2 such that $h_{2t} = h_{2,t+1} = h_2$ and thus the human capital of the rich class cannot grow for ever when the poor hit their lower bound, $h_{\min}$. Then, $\phi_{\min} = h_{\min}/h_2$ is a stationary point. By definition there can be no other stationary point when $\phi_t < \phi(\eta)$. Thus, the economy starting with a level of inequality more than $\phi(\eta)$ exhibits no long-run growth. This proves part (ii).

(iii) Finally, I turn to the alternative configuration with $\alpha[1+(1-\beta)\eta] > 1$. In this case, full equality is not a stable ergodic set. If the function never intersects the 45° line before $\phi = 1$, the unique ergodic set is maximum inequality which corresponds to $\phi^* = 1$ and the economy has a unique stable limiting distribution that is the one described above in (ii). Alternatively, if the curve intersects the 45° line, it must do so twice, thus $\phi^* < 1$ and then there are two locally stable ergodic sets; (a) maximum inequality without any growth and (b) another locally stable ergodic set with a certain degree of inequality but also positive growth but at a rate less than g_{∞}. This proves case (iii). □

Proof of Proposition 3

A) $\eta = 0$, then

$$\theta_{j,t+1} = \beta\theta_{j,t}^{\alpha} \text{ for } h_{j,t} \gg h_{\min}. \tag{A9}$$

Thus $\theta_{jt} > 1$ implies that $\theta_{j,t+1} < \theta_{j,t}$ and for $\theta_{jt} < 1$, $\theta_{j,t+1} > \theta_{j,t}$, hence convergence to full equality.

B) (i) $\theta_{j,t} = 1, \forall j$ implies that $\theta_{j,t+1} = 1\,\forall j$. Therefore, full equality is a stationary point.

(ii) $\theta_{j,t} = 1$, implies that $\theta_{j,t+1} = 1$ irrespective of the distribution. Thus, the group at the median stays there. Take a proportion of the population λ_L to be at $h_{\min}$ and suppose a proportion λ_H is above the median at h_2. This will be a stationary distribution iff

$$\begin{aligned} h_2 &= \beta[1+(1-\beta)\eta]h_2^{\alpha} - \beta(1-\beta)\eta[\lambda_H h_2^{\alpha} + (1-\lambda_H-\lambda_L) + \lambda_L h_{\min}^{\alpha}] \\ h_{\min} &\geq \beta[1+(1-\beta)\eta]h_{\min}^{\alpha} - \beta(1-\beta)\eta[\lambda_H h_2^{\alpha} + (1-\lambda_H-\lambda_L) + \lambda_L h_{\min}^{\alpha}]. \end{aligned} \tag{A10}$$

Thus any vector $(\lambda_L, \lambda_H, h_2)$ that satisfies (A10) will constitute a stationary distribution and given these two equations such vectors always exist.

(iii) This follows from the inspection of equation (17) and Figure 2. (17) can have at most one intersection with the 45° above $\theta = 1$, hence at most one group above the median. Also it cannot have any intersections below $\theta = 1$, thus no group below the median other than one at $h_{\min}$. Finally, we can have a third group at the median.

(iv) Take $\theta_t < 1$. For convergence we need $\theta_{t+1} > \theta_t$. This implies

$$\frac{\beta[1+(1-\beta)\eta]\theta_t^\alpha - \beta(1-\beta)\eta\int\theta^\alpha dG_t(\theta)}{\beta[1+(1-\beta)\eta] - \beta(1-\beta)\eta\int\theta^\alpha dG_t(\theta)} > \theta_t, \tag{A11}$$

or, equivalently,

$$\Theta^- = \beta[1+(1-\beta)\eta](\theta_t^\alpha - \theta_t) - \beta(1-\beta)\eta(1-\theta_t)\int\theta^\alpha dG_t(\theta) > 0. \tag{A12}$$

Similarly, for $\theta_t > 1$, convergence requires $\theta_{t+1} < \theta_t$, thus

$$\Theta^+ = \beta[1+(1-\beta)\eta](\theta - \theta^\alpha) + \beta(1-\beta)\eta(\theta - 1)\int\theta^\alpha dG_t(\theta) > 0. \tag{A13}$$

For local stability, the relevant conditions can be written as

$$\begin{aligned}&\Theta^-(\theta_t^i = 1\,\forall i \neq j, \theta_t^j \to 1^-) > 0\\ &\Theta^+(\theta_t^i = 1\,\forall i \neq j, \theta_t^j \to 1^+) < 0.\end{aligned} \tag{A14}$$

These in turn are equivalent to $\frac{\partial\Theta^-(\theta_t^i=1)}{\partial\theta_t} < 0$ and $\frac{\partial\Theta^+(\theta_t^i=1)}{\partial\theta_t} > 0$. Thus, local stability requires

$$\frac{\partial\Theta^-(\theta_t^i = 1)}{\partial\theta_t} = \beta(\alpha-1) + \beta(1-\beta)\eta(\alpha-1) + \beta(1-\beta)\eta < 0, \tag{A15}$$

and

$$\frac{\partial\Theta^+(\theta_t^i = 1)}{\partial\theta_t} = \beta(1-\alpha) + \beta(1-\beta)\eta(1-\alpha) + \beta(1-\beta)\eta > 0. \tag{A16}$$

(A15) and (A16) are equivalent to each other and in turn to our condition for the stability of full equality in the text $\alpha + \alpha(1-\beta)\eta < 1$.

(v) The diagrammatic representation of the solution to (A10) corresponds to Figure 2b (an intersection above the median is necessary), hence the local stability. □

Proof of Lemma 1

From (A12), we obtain

$$\frac{\partial\Theta^-}{\partial\theta_t} = \beta[1+(1-\beta)\eta](\alpha\theta_t^{\alpha-1} - 1) + \beta(1-\beta)\eta\int\theta^\alpha dG_t(\theta). \tag{A17}$$

This expression is unambiguously negative if $\theta_t < \alpha^{\frac{\alpha}{1-\alpha}}$. That is if relative inequality is sufficiently large, it is most difficult to close the gap between the poorest agents and the mean level of income (and skills). This expression is also increasing in the integral term, thus its zero, $\bar{\theta}_t$, is decreasing in the amount of inequality. □

Proof of Proposition 4

(i) From the text, we can see the conditions for a cycle to be

$$\theta_1^H = \frac{\beta[1+(1-\beta)\eta](\theta_2^H)-\beta(1-\beta)\eta[\lambda^L(\theta_2^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_2^H)^\alpha]}{\beta[1+(1-\beta)\eta]-\beta(1-\beta)\eta[\lambda^L(\theta_2^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_2^H)^\alpha]} \tag{A18}$$

$$\theta_1^L = \frac{\beta[1+(1-\beta)\eta](\theta_2^L)^\alpha-\beta(1-\beta)\eta[\lambda^L(\theta_2^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_2^H)^\alpha]}{\beta[1+(1-\beta)\eta]-\beta(1-\beta)\eta[\lambda^L(\theta_2^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_2^H)^\alpha]} \tag{A19}$$

$$\theta_2^H = \frac{\beta[1+(1-\beta)\eta](\theta_1^H)-\beta(1-\beta)\eta[\lambda^L(\theta_1^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_1^H)^\alpha]}{\beta[1+(1-\beta)\eta]-\beta(1-\beta)\eta[\lambda^L(\theta_1^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_1^H)^\alpha]} \tag{A20}$$

$$\theta_2^L = \frac{\beta[1+(1-\beta)\eta](\theta_1^L)^\alpha-\beta(1-\beta)\eta[\lambda^L(\theta_1^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_1^H)^\alpha]}{\beta[1+(1-\beta)\eta]-\beta(1-\beta)\eta[\lambda^L(\theta_1^L)^\alpha+(1-\lambda^L-\lambda^H)+\lambda^H(\theta_1^H)^\alpha]}. \tag{A21}$$

Thus the problem of finding a cycle is to find a vector $(\lambda^L, \lambda^H, \theta_1^L, \theta_1^H, \theta_2^L, \theta_2^H)$ to satisfy these four equations, (A18)–(A21) while also $\lambda^L < 1/2$ and $\lambda^H < 1/2$. Pick a vector (λ^L, λ^H), then for all values of α, β, η, (A18)–(A21) are continuous and map from the bounded, closed, convex set $[0, 1]^3$ into itself. Thus by Brouwer's fixed-point theorem such a vector will always exist. Thus a two-cycle always exists. Since we can choose different values of (λ^L, λ^H) in an open set, the claim is established.

(ii) This part of the proposition follows part (i) with six equations and three inequalities in nine unknowns. The details are omitted. □

Proof of Lemma 2

First suppose $F(.)$ has no atoms. I will first show that in this case $P(k)$ cannot have any atoms either. Then, by definition of the efficient matching technology, the highest human capital will match with the highest physical capital and so on and thus, by definition, $F(h) = P(k)\ \forall (h, k) \in \mathcal{P}$.

I will now show that $P(k)$ has no atoms by contradiction. Suppose $P(k)$ has an atom at k', then $\exists$ a set of firms $\sigma(k')$ such that all $i \in \sigma(k')$ have the same capital level. Since $F(.)$ has no atoms, with positive probability q, firms in $\sigma(k')$ will match with workers of human capital levels h_1 and h_2 where $h_2 > h_1$. But, this cannot be an equilibrium because one of the firms in $\sigma(k')$ could increase its capital by ϵ and make sure that it meets h_2 whenever it is selected for efficient matching. For all $q > 0$, we can find a small enough ϵ such that this strategy is profitable.

Now suppose that $F(.)$ has an atom at h' and denote the measure of this by $f(h')$. To show that $F(h) = P(k)\ \forall (h, k) \in \mathcal{P}$, it is sufficient to prove that (i) if $(h', k') \in \mathcal{P}$ and

$(h', k'') \in \mathcal{P}$, then $k' = k''$; *and* (ii) if $(h', k') \in \mathcal{P}$ and $(h'', k') \in \mathcal{P}$, then $h' = h''$. [PS: in other words, $P(.)$ must have an atom of exactly the same size at k' which is the level of physical capital that matches with h' when matching efficiently].

(i) Suppose $\exists\, k''$, such that $(h', k'') \in \mathcal{P}$. but given the human capital at h', the profits $\pi(k)$ is concave, thus either k' or k'' is not profit maximizing. Hence a contradiction.

(ii) Suppose $\exists\, i \in \sigma(k')$ such that $(h, k') \in \mathcal{P}$ for h different from h'. If $h > h'$ then some of $i \in \sigma(k')$ would increase its investment by ϵ and match with h with probability q. If $h < h'$, then i would increase its investment by ϵ and match with h'. Hence a contradiction. □

Proof of Lemma 3

(i) Take two workers with different human capital levels $h_1 < h_2$ and denote the capital stocks that these two workers will work respectively with by $k_1 < k_2$ and the corresponding physical to human capital ratios by μ_1 and μ_2. From the fact that both firms will make the same profit level, we can write

$$k_1[(1-\beta)A\mu_1^{-\alpha} - R] = k_2[(1-\beta)A\mu_2^{-\alpha} - R]. \tag{A22}$$

Since $k_2 > k_1$, as long as both firms are making positive profits, $\mu_1 < \mu_2$, thus the physical to human capital ratio is decreasing. And if the profits are equal to zero, $\mu_1 = \mu_2$, i.e. all firms have constant human to physical capital ratios (this can also be checked directly from (19)).

Now re-write (19) by noting that $dk/dh \leq \mu(k)$, i.e. a unit increase in capital will increase human capital by more than the current human to physical capital ratio [this is by definition of the human to physical capital ratio being increasing from (A22)]. Hence,

$$(1-\alpha)(1-\beta)A\mu(k)^{-\alpha} + \alpha(1-\beta)A\mu(k)^{1-\alpha}\mu(k)^{-1} \leq R. \tag{A23}$$

But this implies that the firm is making negative profits. Thus all firms must be making zero-profits and $\mu(k)$ is constant. Since $\mu(k)$ is constant and independent of the distribution of human capital, the rate of return on physical capital and total output are independent of heterogeneity.

(ii) Take a range of values of k and evaluate (19) in the text and let us call this $T(k, h)$. Take k_1 and $k_2 > k_1$ and $h_1 < h_2$ such that $(h_1, k_1) \in \mathcal{P}$ and $(h_2, k_2) \in \mathcal{P}$. Then, by definition $T(h_1, k_1) = T(h_2, k_2) = 0$. $T(.,.)$ has three terms (not counting R which is constant): the first term is decreasing in k and does not depend on h, this therefore implies that the sum of the other terms have to be higher for k_2 than for k_1. This implies that either $\mu(k)$, the physical to human capital ratio, has to be decreasing in k or dh/dk have to be increasing in k. But for k_1 and k_2 sufficiently close to each other $\mu(k)$ is decreasing if and only if dh/dk is increasing (since by Lemma 2 dh/dk is continuous) and therefore $\mu(k)$ has to be decreasing. □

Proof of Lemma 4

This follows from equation 19. □

Proof of Proposition 5

(i) $q = 1$: since human capital ratios are constant from Lemma 2, wages are linear in human capital, thus inequality self-replicates as in Proposition 1.

(ii) From Lemma 3, wages are a concave function of human capital, thus the argument of Proposition 3 for the case of $\eta = 0$ applies. □

Acknowledgments

I have benefited from comments of Abhijit Banerjee, Roland Benabou, Alberto Bisin, Olivier Blanchard, Ricardo Caballero, Peter Diamond, Bengt Holmstrom, Kevin Lang, Thomas Piketty, Chris Pissarides, Steve Pischke, Jim Robinson, Andrew Scott, Robert Shimer, Enis Uysal, Jaume Ventura, four anonymous referees of the *Journal of Economic Growth* and participants at various seminars and conferences. Part of this research has been funded by National Science Foundation Grant No. SBR-9602116.

Notes

1. An alternative approach is to explain the cross-country patterns as the result of different technology shocks for different countries rather than different institutional structures. This approach is however not very satisfactory since OECD countries are likely to be affected by the same shocks. Indeed Card et al. (1995) find that the measures of skilled biased technological change to be identical for the United States, Canada, and France. They also find that despite the much higher relative wages for unskilled workers, the increase in unskilled employment has not been any slower in France (and Canada) than in the United States, which is not easily reconciled with the most popular account of the increase in inequality, the skill-biased technological view.
2. Appendix B in Acemoglu (1995) demonstrated the robustness of the results to a general transmission rule of the form $h_{t+1} = \psi(w_t, h_t)$. Note also that if (1) were changed such that parents cared about their offspring's human capital rather than the education expenditure, (4) would still apply as the law of motion of human capital and the dynamics would not change at all, but the decision rules in (3) would be nonlinear.
3. An alternative is to have a linear household technology with a rate of return R. Then a simple condition, namely $[(1-\alpha)A/R]^{1/\alpha}\alpha\delta \leq \gamma$, ensures that the economy always invests some funds in this linear technology and thus fixes the cost of capital at R. In the absence of a linear technology and other investment opportunities, all the results would hold except those regarding the growth rates. This is of course natural; growth in this economy is driven by the total amount of investment and in the absence of an alternative investment opportunity, all that is saved will be invested, and moreover, with logarithmic preferences, savings are independent of the rate of return—i.e. (4)—hence the economy would achieve the same growth rate irrespective of the rate of return on capital.
4. If the poorest agent had less than $h_{\min}$, then in the first period there would be a contraction in inequality and inequality would self-replicate forever from then on.
5. Appendix C in Acemoglu (1995) shows that all results here are unchanged if parties can change as many partners as they like, i.e., infinite sampling. The analysis here is limited to one round of separation to simplify the expressions and the discussion in the text. The advantage of having infinite sampling is that as $\eta \to 1$, the economy would converge to the perfectly competitive case analyzed in the last section (e.g., Gale, 1987)

and this shows that the only difference between the economy of Section 2 and the one analyzed here is labor market frictions.

6. This is trivially true when $\eta = 0$. The expression for η^* is derived in Appendix C of Acemoglu (1995).
7. See Appendix C of Acemoglu (1995) for a derivation of this bargaining rule from first principles. Note that given the preferences as in equation 1, workers are risk-neutral, which is key to this simple form of rent-sharing.
8. I will not compare the results to a constrained efficient allocation because this allocation would crucially depend on the discount factor used by the social planner. But it can be shown that when matching is random, and $\eta > \eta^*$, the planner would also choose the same level of physical capital for all firms (but this level of capital would be higher), and then she will redistribute output to regulate the evolution of the skill distribution.
9. Note that when the poor hit $h_{\min}$, the curve no longer applies and the system converges to $\phi_{\min}$—see proof of Proposition 2 the Appendix.
10. A different way of expressing the same intuition is that the firm has the same outside option when bargaining against the skilled and unskilled workers but this outside option is larger *relative to the value of the match* in the case of a match with an skilled worker.
11. The limiting case of fully efficient technology, $q = 1$, is obviously unrealistic as it relies on an invisible hand arranging the right matches.

References

Acemoglu, D. (1995). "Matching, Heterogeneity, and the Evolution of Income Distribution," MIT Working Paper No. 95-25.

Acemoglu, D. (1996a). "A Microfoundation for Increasing Returns in Human Capital Accumulation," *Quarterly Journal of Economics* 111, 779–804.

Acemoglu, D. (1996b). "Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence," CEPR Discussion Paper No. 1459.

Banerjee, A., and A. Newman. (1993). "Occupational Choice and the Process of Development," *Journal of Political Economy* 59, 427–453.

Benabou, R. (1996). "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance," *American Economic Review* 86, 584–609.

Bertola, G., and A. Ichino. (1995). "Wage Inequality and Unemployment: US versus Europe," *NBER Macroannual*.

Cameron, S. V., and J. Heckman. (1993). "Determinants of Young Male Schooling and Training Choices," NBER Working Paper, 4327.

Card, D. (1996). "The Effects of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica* 64, 957–981.

Card, D., F. Kramartz, and T. Lemieux. (1995). "Changes in the Relative Structure of Wages and Employment: A Comparison of the United States, Canada, and France," Mimeo.

Davis, S. (1995). "The Quality Distribution of Jobs in Search Equilibrium," University of Chicago, Mimeo.

Diamond, P. (1982). "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies* 49, 217–227.

DiNardo, J., N. Fortin, and T. Lemieux. (1996). "Labor Market Institutions and the Distribution of Wages, 1973–1992," *Econometrica* 64, 1001–1044.

Durlauf, S. (1996). "A Theory of Persistent Income Inequality," *Journal of Economic Growth* 1, 75–93.

Fernandez, R., and R. Rogerson. (1996). "Income Distribution, Communities and the Quality of Public Education," *Quarterly Journal of Economics* 3, 135–164.

Gale, D. (1987). "Limit Theorems for Markets with Sequential Bargaining," *Journal of Economic Theory* 43, 20–54.

Galor, O., and J. Zeira. (1993). "Income Distribution and Macroeconomics," *Review of Economic Studies* 60, 35–52.

Galor, O., and D. Tsiddon. (1996). "Technological Progress, Mobility and Economic Growth," *American Economic Review*, forthcoming.

Goldin, C., and R. A. Margo. (1992). "The Great Compression: The Wage Structure in the United States at Mid-Century," *Quarterly Journal of Economics* 197, 1–34.

Grandmont, J. M. (1985). "Endogenous Competitive Business Cycles," *Econometrica* 54, 995–1045.

Jovanovic, B. (1979). "Job Matching and the Theory of Turnover," *Journal of Political Economy* 87, 972–990.

Juhn, C., K. M. Murphy, and B. Pierce. (1993). "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* 101, 410–442.

Katz, L., G. W. Lovemen, and D. G. Blanchflower. (1995). "A Comparison of Changes in the Structure of Wages in Four Countries." In R. Freeman and L. Katz (eds), *Differences and Changes in Wage Structures*. NBER and University of Chicago Press.

Levy, F., and R. Murnane. (1992). "US Earnings Level and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature* 3, 1333-1381.

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

Li, T., and J. A. Yorke. (1975). "Period Three Implies Chaos," *American Mathematical Monthly* 82, 985–992.

Osborne, M. J., and A. Rubinstein. (1990). *Bargaining and Markets*. San Diego: Academic Press.

Perotti, R. (1996). "Growth, Income Distribution and Democracy: What the Data Say," *Journal of Economic Growth* 1, 149–188.

Pissarides, C. A. (1990). *Equilibrium Unemployment Theory*. Basil Blackwell, Oxford, UK.

Rebelo, S. (1991). "Long Run Policy Analysis and Long Run Growth," *Journal of Political Economy* 99, 500–521.

Rumberger, R. W. (1981). *Overeducation in the US Labor Market*. New York: Praeger.

Sattinger, M. (1993). "Assignment Models of Distribution of Earnings," *Journal of Economic Literature* 3, 831–880.

Sattinger, M. (1995). "Search and the Efficient Assignment of Workers to Jobs," *International Economic Review* 36, 283–302.

Shimer, R., and L. Smith. (1996). "Heterogeneity, Matching and Search," MIT mimeo.

Sicherman, N. (1991). "Overeducation in the Labor Market," *Journal of Labor Economics* 9, 101–122.