# Distribution of Ability and Earnings in a Hierarchical Job Assignment Model

Robert M. Costrell

*Massachusetts Executive Office for Administration and Finance and University of Massachusetts at Amherst*

Glenn C. Loury

*Boston University*

We examine the mapping of the distribution of ability onto earnings in a hierarchical job assignment model. Workers are assigned to a continuum of jobs in fixed proportions, ordered by sensitivity to ability. The model implies a novel marginal productivity interpretation of wages. We derive comparative statics for changes in technology and in the distribution of ability. We find conditions under which a more unequal distribution of ability maps onto a more/less unequal distribution of earnings. We also analyze an assignment model with variable proportions and find that in the Cobb-Douglas case, a rise in the inequality of ability always narrows the range of earnings.

## I. Introduction

This is a paper about the role of job assignment in the distribution of earnings. The importance of job assignment can be understood from a simple nondistributional question: Does better information about individual ability raise aggregate output? In a one-job model, the answer is no: better information affects the distribution of income, but not the

We appreciate the helpful comments of seminar participants at the University of Chicago, Harvard University, Northwestern University, the University of Pennsylvania, Washington University, Boston University, the University of Massachusetts, Suffolk University, and the National Bureau of Economic Research Income Distribution Group.

[*Journal of Political Economy,* 2004, vol. 112, no. 6]

total. If better information does raise output, the reason is that there are productivity gains to be had from improving the match between workers and jobs, a phenomenon that can be analyzed only in an assignment model.[1] Specifically, in models such as the one we present, the sensitivity of output to ability varies across jobs, so improved information raises output by better sorting high-ability individuals into more ability-sensitive jobs. The light that job assignment can shed on distributional questions is less well appreciated. Changes in other people's ability may affect our earnings by changing what job we are assigned to, as well as the rewards attached to each job. Understanding the assignment mechanism can help explain such questions as whether and why better information about individual ability—or a rise in inequality of ability itself—increases or decreases the inequality of earnings.

To flesh this out, suppose that production takes place in teams of workers, assigned to varying tasks (or jobs) on the basis of their ability, and that output depends on the quantity or quality of the various tasks performed. What gives an assignment model its analytical force is the fact that the assignment of workers to tasks is endogenous. In models of the type considered in this paper, workers cannot simply be added to any given task without some reshuffling of personnel among other jobs. In the simplest example, there may be some fixed relationship between the number of production and nonproduction workers: one cannot increase without the other. More generally, if there is a hierarchy of jobs, the addition of low-skill workers toward the bottom of the hierarchy will allow other workers to move up the job ladder. Such an assignment process may be thought of as lying in the background of a conventional production function in discrete grades of labor, $F(L_1, \ldots, L_n)$. Spelling out the assignment process imposes structure on the production function. More to the point, it imposes structure on the wage or marginal productivity functions $w_i = F_i(L_1, \ldots, L_n)$, which must be interpreted to include the productivity effect of the endogenous reassignments, upon adding an increment of type $i$ labor.

The mapping of the distribution of ability to that of earnings depends on the following features of the underlying technology we have just described.

First, it depends on the gradient across tasks of how sensitive performance is to ability. Whether that gradient—and, hence, the sensitivity of output to proper job assignment—is concentrated in the high-skill or low-skill jobs will be important. For example, we shall show that if this gradient in skill sensitivity is concentrated among high-skill jobs,

[1] See Hartigan and Wigdor (1989, p. 241ff.) for a clear statement of this point in the employment testing literature. There had been some confusion between the partial and general equilibrium gains from information, a confusion that seems to persist in some popular treatments, such as Herrnstein and Murray (1994, pp. 64, 85–86).

then a rise in inequality of ability will tend to *reduce* inequality of earnings. To sketch the argument, note that the value of an additional low-ability worker is enhanced by the productivity gains of those workers who can now be promoted to more ability-sensitive jobs. Conversely, the value of an additional high-ability worker is reduced by the productivity losses of those workers who must be demoted to make room for him at the top. If these gains and losses are concentrated in high-skill jobs, occupied by workers in the right tail of the distribution, then a rise in the inequality of ability (which raises those workers' ability) will magnify the effect of reassignment. Hence, wages will rise at the bottom and fall at the top, reducing the inequality of earnings. This result involves in an essential way the reassignment of workers and therefore could not be understood outside of an assignment model.

A second feature of the technology is the degree of complementarity of workers in the team across tasks. If the degree of complementarity is high, then two heads are no better than one on any given task without the addition of a full complement of workers on other tasks. The nature of the mapping from ability to earnings depends on this complementarity, on how much better two heads (or pairs of hands) are than one. To analyze this, we advance the assignment literature a major step beyond the traditional assumption of fixed proportions between workers and tasks. Under variable proportions, firms face a nontrivial assignment problem of choosing the optimal density of workers across tasks. We find a simple but powerful relationship between the wage profile and the first-order (Euler) condition for the optimal assignment. This allows us to explore further the conditions under which a rise in the inequality of ability reduces the inequality of earnings. We find that if the degree of complementarity of workers across tasks is reduced from that of fixed proportions to that of Cobb-Douglas, then a rise in the inequality of ability will *always* narrow the wage gap between the top and bottom.[2]

The assignment literature, surveyed by Sattinger (1993), distinguishes between models in which assignment is based on workers' preferences over job characteristics (e.g., Tinbergen 1951), comparative advantage (e.g., Roy 1951; Sattinger 1975; Heckman and Sedlacek 1985; Teulings 1995), and production complementarities across jobs (e.g., Sattinger 1979, 1980). This third type of assignment model, less familiar than the others, is the basis for our paper. As outlined above, the model captures

[2] A third feature of the technology that matters to this mapping is the manner in which performed tasks are combined to generate output. It matters whether performed tasks are highly substitutable (e.g., output is linear in performed tasks) or whether technology is more like O-rings (Kremer [1993], where output is Cobb-Douglas in performed tasks). As Kremer and others have shown, if there is complementarity across tasks, which implies positive cross partials in ability, at least some assortative matching by ability will obtain across firms. In this paper, we confine our attention to perfect substitutability across tasks, with no assortative matching, as discussed further in n. 5.

key features of a hierarchical production technology under which jobs can be ordered by sensitivity to ability, but all must be filled in more or less fixed proportions. Workers of the highest general ability are assigned to the most ability-sensitive jobs and those of least ability are assigned to the least sensitive jobs. This is very different from a comparative advantage model, since the assignment does not rest on heterogeneous skills: ability is one-dimensional in the model's purest form. Conversely, in a comparative advantage model, there is typically no technological constraint on the proportions of workers across jobs. Comparative advantage models have proved quite fruitful for several decades, but there is some recent literature that suggests an erosion of the role of comparative advantage in explaining the distribution of income. This is the finding of Gould (2002), who explains that the rising importance of general or cognitive skills is raising the correlation of skill valuation across broad occupational groups. If so, a model of the type developed here might be increasingly relevant.

Sattinger's (1979, 1980) model is the most direct precedent of ours. He considers a continuum of jobs in which labor of varying quality works with complementary machines of varying size. Workers of higher ability are optimally assigned to jobs with more capital. Sattinger solves for the prices of labor and machine rental of varying quality. The slopes of the wage and rental functions are derived from no-arbitrage conditions, and the levels are set by outside options for labor and capital. He uses this setup to explore the effect of capital heterogeneity on the distribution of labor earnings.

Our model in Sections III and IV has some formal similarities with the Sattinger model because it also assumes fixed proportions between the continua of workers and tasks, that is, a fixed "production hierarchy." This implies that we have the same no-arbitrage condition for the slope of the wage function. But the level of the wage function is determined differently—by the zero-profit condition—because our model focuses on complementarities among types of labor in the same enterprise rather than capital-labor complementarities on jobs that are independent of one another.

Our analysis of this fixed-proportions model offers several major advances. First, we show that wages are in fact marginal productivities of a reduced-form production function (in continuous form), where an increment of labor of a given quality leads to reassignment of all other workers.[3] The effect of that reassignment on output depends critically

[3] Sattinger claims that marginal productivities are undefined in his model because of fixed proportions between capital and labor on any given job. However, this claim implicitly presupposes a definition of marginal productivities as the addition to output associated with a marginal increase in an input, with not only the quantity of other inputs but also their assignment held constant.

on the gradient of ability sensitivity across jobs, in a way that sheds light on the comparative statics of the ability-wage mapping.

Second, we find the assignment model's analogue (in continuous ability) of the production function's cross partials, $F_{ij}(L_1, \ldots, L_n)$. This allows us to characterize Hicks substitutability or complementarity across ability types and show how that depends on the assignment model's technology. This in turn helps us analyze who wins and who loses from changes in the skill mix, such as immigration drawn from one part of the ability distribution or another.

Third (as mentioned above), we ascertain the conditions under which an increase in inequality of ability leads to an increase or decrease in the spread of wages. We show that this analysis directly applies to the issue of improved information about employees, since better information increases the inequality of expected ability. Our results may therefore be pertinent to recent discussions about whether or not testing and other forms of information generate an efficiency-equity trade-off, by improving the match of workers to jobs, while making it easier for firms to distinguish high- from low-ability workers. By providing a general equilibrium analysis of this question, we find conditions under which this trade-off does or does not exist.[4]

Another change in the distribution of abilities arises from integrating workers from two distinct ability distributions. Two different groups, regions, or countries may differ either in their underlying distributions of ability or in the precision of information available about the abilities of their members. We analyze what we believe to be a novel question: If one can estimate more accurately the ability of workers in country (group or region) A than in B, who gains, and how much, when the two countries (groups or regions) integrate into a single market?

In sum, we offer an extensive exploration of the comparative statics of the wage distribution with respect to the distribution of ability (and also technology), thereby extending the assignment literature that is based on fixed proportions between workers and jobs. We then go on, in Section V, to relax the assumption of fixed proportions between workers and tasks, as described above. Finally, in Section VI, we briefly

[4] Rothschild and Stiglitz (1982) provide a model with a continuum of jobs, which posits a quadratic loss in productivity for any mismatch between the individual's ability and the skill requirement of the job. With normally distributed ability, they find that finer information generates higher mean earnings and higher variance, i.e., an efficiency-equity trade-off. By contrast, Heckman and Honore (1990) show that in the Roy model, perfect assignment not only raises mean earnings but reduces the inequality of earnings, compared to random sectoral assignment in the same proportions. This might be interpreted as comparing a perfect test and no test. MacDonald (1982) also analyzes the effect of imperfect substitutes in the production of output (as discussed in n. 2 above). Although MacDonald does show that information raises output, he finds ambiguous effects on wage levels and does not examine the effect on wage inequality.

discuss some limitations of our assumptions and conclude with some reflections on what light our analysis might shed on recent decades' trends in the distribution of income.

## II. A Two-Job Model

In this work we investigate the relationship between the distribution of workers' abilities and the resulting distribution of wages that obtains in competitive equilibrium. A worker's contribution to total firm output depends on his own ability and also on the job he occupies. We begin by finding the equilibrium distribution of wages in a simple two-job model.

The key feature of the model is that jobs must be filled in fixed proportions to one another, $\theta$ and $1-\theta$. Consider a world, for example, of production and supervisory jobs, with a fixed span of control. Alternatively, suppose that production takes place in teams of fixed composition, for example, two-person teams of an electrician and an electrician's helper. One may think of software teams with coders and debuggers in proportions that are hard to vary, or research teams with lead researchers and research assistants, where, again, the quantity of workers cannot be substituted for quality in one or the other type of job.

In the simplest case, output depends only on the ability of those in production jobs; support jobs generate no output, but must be filled to keep the operation going. More generally, suppose that one job is more ability-sensitive than the other, such that output per unit of ability on the two jobs is $\beta_1 > \beta_0 \geq 0$. Thus $\theta$, $\beta_0$, and $\beta_1$ completely characterize the technology.

Let $\mu \in [0, 1]$ denote a worker's ability (or expected ability, given available information). The distribution of ability is exogenous and is characterized by the cumulative distribution function (CDF) $F(\mu)$, so $p = F(\mu)$ denotes the ability quantile of a worker with ability $\mu$. Equivalently, let $\mu(p) \equiv F^{-1}(p)$ denote the ability at rank $p$. Workers will be assigned to job 1 or 0 as $\mu \gtrless \hat{\mu} \equiv \mu(\theta)$. Output is

$$Q = \beta_0 \int_0^\theta \mu(p)dp + \beta_1 \int_\theta^1 \mu(p)dp.$$

It is important to be clear on how the complementarities under consideration in this model (and in the continuous job version below) pertain to the *number* of workers across jobs versus their *abilities*. Under the strict complementarity we have posited (which will be relaxed in Sec. V), the effect on output of adding another worker of any given ability to a given job depends very much on the *number* of workers in the other jobs: they must remain in strict proportion. By contrast, the effect on output of raising the *ability* of a worker in one job is assumed independent of the *abilities* of workers in the other jobs.[5]

The wage schedule is derived from two conditions. The no-arbitrage condition on job assignment gives the slope of the wage function: $W(\mu) - W(\mu') = \beta_i(\mu - \mu')$ for any $\mu$, $\mu'$ on job $i$. Otherwise profits could be raised by substituting one type of worker for another. The no-profit condition, $Q = \int_0^1 W(\mu)dF(\mu)$, sets the level of the wage function. Specifically, these conditions imply

$$W(\mu) = \beta_0\mu + (1 - \theta)(\beta_1 - \beta_0)\hat{\mu}, \quad \mu \leq \hat{\mu}, \tag{1}$$

and

$$W(\mu) = \beta_1\mu - \theta(\beta_1 - \beta_0)\hat{\mu}, \quad \mu \geq \hat{\mu}, \tag{2}$$

as depicted in figure 1.

These expressions have a straightforward marginal productivity interpretation. An additional worker of ability $\mu < \hat{\mu}$ adds $\beta_0\mu$ to output directly on job 0, but also allows the promotion of $1 - \theta$ workers to job 1. These workers will be promoted from the margin between the jobs, so they will have ability $\hat{\mu} \equiv \mu(\theta)$. The output from each of these $1 - \theta$ workers will rise by $(\beta_1 - \beta_0)\hat{\mu}$, so this indirect contribution to output is captured in the second term of (1). Thus low-ability workers earn more than their direct contribution to output. Similarly, an additional worker of ability $\mu > \hat{\mu}$, placed on job 1, earns less than his or her direct contribution to output, $\beta_1\mu$. The reason is that placement in job 1 requires the demotion of $\theta$ workers of ability $\hat{\mu}$ to job 0, with the attendant loss of output, as reflected in the second term of (2).[6]

[5] The model implicitly assumes that performed tasks are perfect substitutes in generating output, as mentioned in n. 2 above: output is simply the sum of performed tasks. For industries adequately approximated by the two-job case with $\beta_0 = 0$, this assumption entails no further loss of generality: a different degree of task substitutability would have no effect. The assumption may also capture essential features in other cases, e.g. (to take perhaps only a slight caricature), academic departments in which output is the (quality-adjusted) sum of articles produced; some departmental jobs that leave less time for research (i.e., low-$\beta$ jobs) must nonetheless be filled. For other industries, however, the degree of cross-task complementarity may be important, as in Kremer's O-ring analysis. This leads to assortative matching, which, the literature has shown, leads to greater inequality of earnings from any given distribution of ability. Our own analysis (available on request) finds that even if we abstract from assortative matching, the relaxation of perfect substitutability across tasks makes it more likely that a mean-preserving spread of ability widens the wage span.

[6] These expressions also show the sense in which the quantity of workers cannot substitute for their quality. If we consider a team of workers with ability below or above $\hat{\mu}$, then integrating over (1) or (2), we see that the value of this team has two terms. The first is the value of the tasks they perform, based on their ability, and the second is based only on their *number*. Thus, in a market for services provided by a labor contractor, the contract does not simply pertain to a certain level of services, but also depends on the number of workers supplying those services. By contrast, this paper's model does not apply to industries such as contracted after-hours office cleaning, where the size of the cleaning crew is irrelevant.

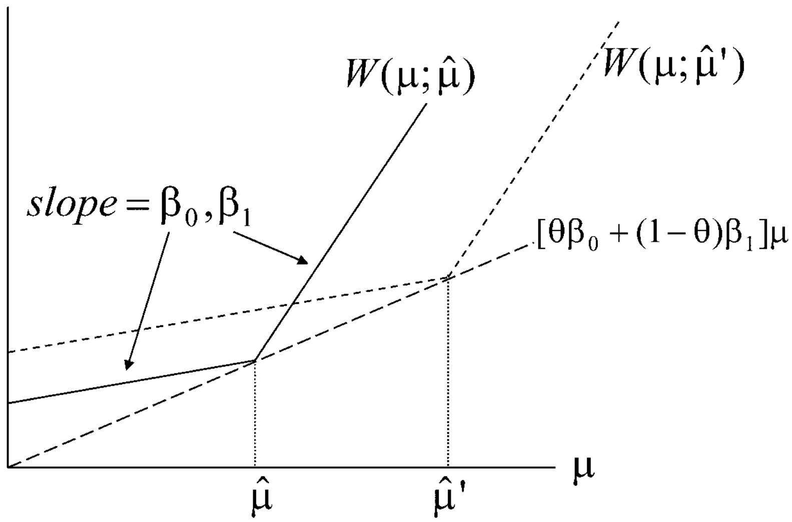


FIG. 1

Note that the technology is one of constant returns to scale: doubling inputs of all ability levels leaves $\hat{\mu}$ and the wage schedule unchanged and doubles output. Thus, as implied by Euler's theorem in the standard production model, marginal productivity payments exhaust output and the zero-profit condition obtains, as stated. There are no externalities at work here, despite the fact that workers receive more or less than the direct value of the tasks they perform. There is, in principle, nothing different going on here than usual, since marginal productivities always depend on the interaction between marginal and incumbent workers of different abilities. This model merely specifies that interaction in a particular way that distinguishes between the direct value of tasks performed by the marginal worker and the effect on the value of tasks performed by incumbent workers, by virtue of the optimal reassignment.

We now consider comparative statics of the wage function. (The results presented here generalize beyond the two-job model to one with a continuum of jobs, presented below.) Technological progress is represented here by a rise in $\beta_0$ or $\beta_1$ or both. This results in an unambiguous widening of the income distribution: for $\mu > \mu'$, $W(\mu) - W(\mu')$ is nondecreasing in $\beta_0$ and $\beta_1$.[7] This result becomes intuitive once one is clear that $\beta_i$ is the *incremental* price of ability for a worker on job $i$; so a rise in $\beta_i$ widens the gap between two workers of different ability.

[7] For individuals on the same job $i$, $W(\mu) - W(\mu') = \beta_i(\mu - \mu')$ rises with $\beta_i$. For individuals in different jobs, $\mu > \hat{\mu} > \mu'$, $W(\mu) - W(\mu') = \beta_1(\mu - \hat{\mu}) + \beta_0(\hat{\mu} - \mu')$ rises with $\beta_0$ and $\beta_1$.

In gauging the effect of technology on the bottom of the distribution, note that $W(0) = (1-\theta)(\beta_1-\beta_0)\hat{\mu}$ depends *entirely* on $\beta$'s gradient $(\beta_1-\beta_0)$. If productivity rises on the low-skill jobs, $W(0)$ falls, perhaps counterintuitively to a simple notion of biased technical progress. That is, one might think that a rise in $\beta_0$ would raise the wages of those with lesser skill, because it does raise their direct contribution to output, $\beta_0\mu$. However, for those with the very least skill, this is outweighed by the fact that a rise in $\beta_0$ reduces the indirect contribution to output, $(1-\theta)(\beta_1-\beta_0)\hat{\mu}$, the productivity gain from promoting workers up the hierarchy. Indeed, in the limit, as $\beta_0 \to \beta_1$, we approach the one-job model, where wages equal ability and $W(0) \to 0$.

We now turn to comparative statics with respect to the ability distribution. Note that in this simple two-job model, the role of the ability distribution in the wage function is *entirely* captured by $\hat{\mu} \equiv F^{-1}(\theta)$, the ability of the worker on the margin between the two jobs. Thus, as figure 1 illustrates, any shift in the ability distribution on fixed support [0, 1] that raises $\hat{\mu}$ to, say, $\hat{\mu}'$ will raise the earnings of low-ability workers and reduce the earnings of high-ability workers, and conversely for any shift that reduces $\hat{\mu}$.

For example, a first-order stochastic improvement in the ability distribution raises (or does not reduce) ability at all quantiles, including ability at the critical $\theta$ quantile, $\hat{\mu}$. Thus a general improvement in ability not only raises output but also improves the distribution of income. The reward to low-ability types rises because they now support the promotions of workers with higher ability than before, and, for the same reason, the reward to high-ability types falls.

Consider instead a more unequal ability distribution, a mean-preserving spread on fixed support. This improves ability in the right tail and reduces ability in the left tail (with possible multiple crossings in between). The effect depends on whether the worker on the margin between jobs, at quantile $\theta$, lies in the upper or lower tail of the ability distribution. More generally, it is useful to think of the technology as one in which the sensitivity to job assignment (gradient of $\beta$) is concentrated in jobs filled by low-skill or high-skill workers. If proper job assignment is most critical in the top jobs (i.e., if most jobs are not skill-sensitive—$\theta$ is high), then the worker on the margin is in the right tail; so a more unequal ability distribution raises $\hat{\mu}$ and *narrows* the wage distribution. Conversely, if assignment matters most toward the bottom of the spectrum ($\theta$ is low), then a more unequal distribution reduces $\hat{\mu}$ and *widens* the wage distribution (see fig. 2).

Finally, let us examine the effect on the income distribution of an increment to the workforce at any arbitrary point of the ability distribution. It is easy to show that this reduces wages of all workers with abilities on the same side of $\hat{\mu}$, so they are (Hicks) substitutes, and those

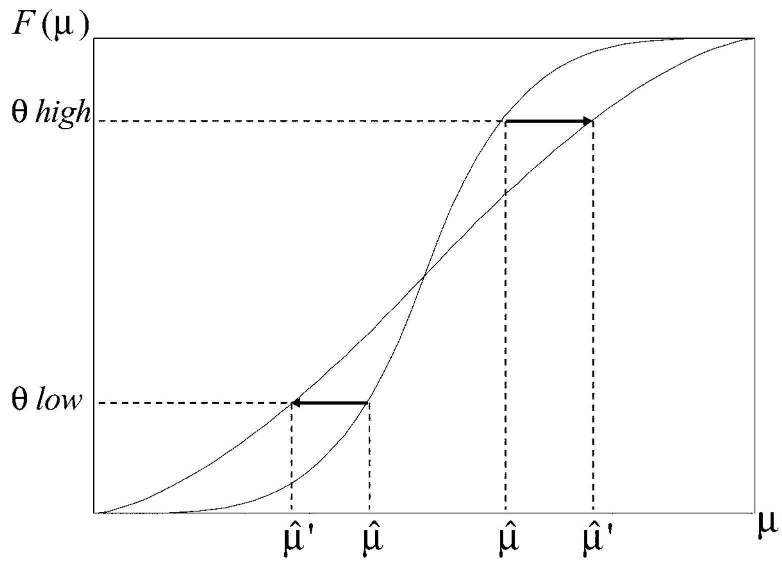


FIG. 2

on the opposite sides of $\hat{\mu}$ are complements. To see this, note first that the addition of a worker with ability $\mu'$ will shift down (up) the quantiles of all workers with ability below (above) $\mu'$: the CDF, $F(\mu)$, will pivot at $\mu'$ as depicted in figure 3. Thus the addition of a worker with any ability level $\mu' > \hat{\mu}$ will raise $\hat{\mu}$. As we have seen, this raises earnings of all workers with ability $\mu < \hat{\mu}$ and reduces earnings of all workers with ability $\mu > \hat{\mu}$. The converse holds for the addition of a worker with ability $\mu' < \hat{\mu}$. As a result, workers of any abilities $\mu$ and $\mu'$ that bracket $\hat{\mu}$ are complements, whereas those on the same side of $\hat{\mu}$ are substitutes.

## III. A Continuum of Jobs in a Fixed Production Hierarchy

### A. *Wage Profile*

In this section and the next, we generalize the model to a continuum of jobs and extend our comparative static results. Let us order jobs, or tasks, within a firm by the sensitivity of output on a task to the ability of workers filling it. Denote the rank order of a task by $t \in [0, 1]$. Define $\beta(t)$ as the direct contribution to output per unit of ability from a worker placed in task $t$, where $\beta(t)$ is nondecreasing.[8] In a fixed production

[8] The two-job model is a special case in which $\beta(\cdot)$ is a step function with a single point of discontinuity.

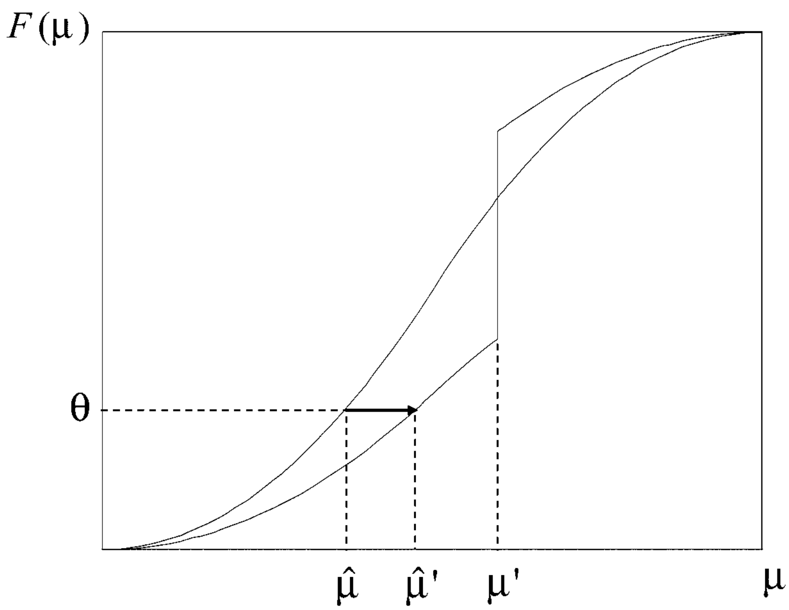


Fig. 3

hierarchy, workers within a firm must be assigned to tasks in a fixed distribution (which we normalize to be uniform, without further loss of generality). With this assumption, $\beta(\cdot)$ completely characterizes the economy's technology.

Higher-ability workers are placed in more ability-sensitive positions;[9] under the assumption of fixed proportions, mapping people to tasks one to one, we have $p = t$. That is, output is maximized by assigning a worker of ability $\mu(p)$ to the task with $\beta(t = p)$. Thus in an efficient assignment, a worker's direct contribution to output is[10]

$$q(p) \equiv \beta(p)\mu(p). \tag{3}$$

[9] If workers of abilities $\mu$ and $\mu'$ are assigned to tasks $t$ and $t'$, respectively, with $\mu' > \mu$, their joint output is $\mu\beta(t) + \mu'\beta(t')$. If they switch positions, their joint output would be $\mu'\beta(t) + \mu\beta(t')$. For there to be no gain from switching, we require $(\mu' - \mu) \cdot [\beta(t) - \beta(t')] \leq 0$, which in turn requires $\beta(t') \geq \beta(t)$.

[10] Nothing of substance would be altered by generalizing (3) to $q(p) \equiv \alpha(p) + \beta(p)\mu(p)$. The output that is independent of ability, $\alpha(\cdot)$, affects only the zero-profit condition, so the integral over $\alpha(\cdot)$ gets folded into $w(0)$, as derived below. That is, $\alpha(\cdot)$ affects only the level of $w(\cdot)$, not its slope. No restrictions need be placed on $\alpha(\cdot)$, so there need be no presumption that the most ability-sensitive job is also the job that generates the most output. But even if not, the worker filling that job, with the highest ability, will still earn the highest wage.

Total output in the hierarchy, $Q$, is

$$Q(F) = \int_0^1 q(p)dp = \int_0^1 \mu\beta(F(\mu))dF(\mu). \quad (4)$$

What is the distribution of wages for workers in such hierarchies? We use two conditions to derive the wage profile, $w(p)$, the wage as a function of the worker's ability quantile. (We distinguish $w(p)$ from $W(\mu)$, which gives the wage as a function of ability itself.) First, from the no-arbitrage condition we find the slope of the profile,

$$w'(p) = \beta(p)\mu'(p), \quad (5)$$

which equals the derivative of $\beta(t)\mu(p)$ with respect to $p$, evaluated at $t = p$. This follows from observing that no profitable arbitrage opportunity exists if and only if, for all $p$, $p'$,

$$\beta(p)\mu(p) - w(p) \geq \beta(p)\mu(p') - w(p')$$

or, equivalently,

$$\beta(p)[\mu(p) - \mu(p')] \geq w(p) - w(p') \geq \beta(p')[\mu(p) - \mu(p')].$$

We assume (though nothing important hinges on this) that the ability distribution is atomless with full support on [0, 1], so that $F(\cdot)$ is continuous and strictly increasing on [0, 1]; thus $\mu(\cdot) \equiv F^{-1}(\cdot)$ is differentiable almost everywhere on [0, 1]. Divide by $p - p'$ and let $p \rightarrow p'$ to get (5). Note that $\beta(p)$ is the marginal return to ability for someone of ability level $\mu(p)$, $dw(p)/d\mu(p)$.[11]

We can see from (5) and (3) that the wage profile is flatter than the productivity profile of workers' direct contribution to output:

$$w'(p) = \beta(p)\mu'(p) \leq q'(p) = \beta(p)\mu'(p) + \beta'(p)\mu(p).$$

The higher wage earned by someone slightly up the hierarchy reflects his higher ability, not the higher job placement per se; the gain in output from promoting a worker of ability $\mu(p)$ to a higher job, $\beta'(p)\mu(p)$, does not accrue to that worker, but rather (as we shall see) to those who make that promotion possible.

The second condition is the zero-profit condition, $\int_0^1 q(p)dp = \int_0^1 w(p)dp$. This immediately implies that the wage profile crosses the productivity profile from above: those in low job placements earn more

[11] It immediately follows that wages are a convex function of ability, since the marginal return to ability, $dW/d\mu = \beta(F(\mu))$, is nondecreasing. The wage is linear in $\mu$ *on any task*, but a higher $\mu$ moves one up to a higher task. This convexity of $W(\mu)$ skews the distribution of wages to the right, relative to the distribution of ability, consistent with the empirical phenomenon that has motivated a very long previous literature (e.g., Mayer [1960], Sattinger [1975], and Rosen [1982], to name just a few).

than their direct contribution to output, and those in high job placements earn less.

These two conditions (assured, e.g., by Bertrand-competitive wage-setting firms) suffice to establish the wage profile, by the following result.

Proposition 1. Let the ability distribution be atomless with full support on [0, 1]. Then the fixed hierarchy model of production yields the competitive wage profile:

$$w(p) = \beta(p)\mu(p) + \int_0^1 \int_{z=p}^{y} \mu(z)d\beta(z)dy. \tag{6}$$

*Proof.* Since $w'(p) = \beta(p)\mu'(p)$,

$$w(p) = w(0) + \int_{z=0}^{p} \beta(z)d\mu(z). \tag{7}$$

Integrating by parts, we get

$$w(p) = w(0) + \beta(p)\mu(p) - \int_{z=0}^{p} \mu(z)d\beta(z). \tag{8}$$

Using the zero-profit condition, we find

$$w(0) = \int_0^1 \int_{z=0}^{y} \mu(z)d\beta(z)dy. \tag{9}$$

Substituting in (8) gives

$$w(p) = \beta(p)\mu(p) + \int_0^1 \left[\int_{z=0}^{y} \mu(z)d\beta(z) - \int_{z=0}^{p} \mu(z)d\beta(z)\right]dy$$

and the result stated in (6). Q.E.D.

*Discussion.*—The wage profile, given in (6), clearly shows the relationship to the productivity profile, $q(p)$, since that is given by the first term in (6). The second term, $w(p) - q(p)$, decreases in $p$: low-wage workers earn more than their direct contribution to output, and high-wage workers earn less, as stated earlier. Further insight into this relationship can be gleaned from the following marginal productivity interpretation of (6).

As in the two-job model, (1) and (2), the second term of (6), $\int_0^1 \int_{z=p}^{y} \mu(z)d\beta(z)dy$, is the indirect contribution to (or detraction from) output. It is the effect of reallocating workers up and down the hierarchy in order to restore the one-to-one assignment of workers to tasks, following the introduction of an additional worker of ability $\mu(p)$. The $d\beta(z)$ term represents the productivity gains/losses from the reassignments, as did the $\beta_1 - \beta_0$ terms in (1) and (2). The only new bit of analysis here is to understand how the rate of reassignment varies up and down the hierarchy and how this is represented by the double integral.

Consider $w(0)$, the wage of an individual with ability $\mu = 0$. His indirect (indeed, only) contribution to output is the productivity gain from pushing higher-ability workers up the hierarchy, given in (9), $\int_0^1 \int_{z=0}^{y} \mu(z) d\beta(z) dy$. The addition of a mass of workers at $\mu = 0$ leads to the most intensive rate of reassignment at or near rank zero. Further upward reassignments diminish in intensity the further up the hierarchy we go, until workers are once again uniformly distributed across tasks.[12] This diminution of the rate of reassignment as we move away from the point of insertion is captured by the double integral, since the greatest weight is accorded to $\mu(z) d\beta(z)$ for the lowest values of $z$. In economic terms, this means that wages at the bottom of the distribution depend *most* heavily on the sensitivity of output to job assignment in the *lower-skilled* jobs, and less so in the higher-skilled jobs.

Conversely, at the opposite end of the wage profile, we have

$$w(1) = \beta(1) - \int_0^1 \int_{z=y}^{1} \mu(z) d\beta(z) dy. \tag{10}$$

Wages at the top of the profile are less than the direct contribution to output because workers must be demoted to lower positions, with reduced output, to accommodate an additional worker at the top. Again, the most intensive reallocation occurs at jobs that are closest to that of the new worker—the top jobs, in this case. More generally, for a worker of rank $p \in (0, 1)$, (6) shows that the indirect contribution to output consists of the productivity gains and losses from reallocating workers up and down the hierarchy from task $t = p$, with the intensity of the reallocation diminishing as one moves farther from job $p$, occupied by the worker in question, toward jobs 0 and 1.[13]

[12] Consider a simple case of three jobs (with parameters $\beta_1$, $\beta_2$, and $\beta_3$) filled by three workers, with abilities $\mu_1 = 0$, $\mu_2$, and $\mu_3$. (This can be represented in the present model by step functions for $\beta(\cdot)$ and $\mu(\cdot)$, which have two [common] points of discontinuity.) The addition of a zero-ability worker at the bottom of the hierarchy allows the scale of operations to expand to one and one-third workers on each job. The $\mu_2$ worker on job 2 now reallocates one-third of his time to the top job, for a productivity gain of $\frac{1}{3}\mu_2(\beta_3 - \beta_2) = \frac{1}{3}\mu_2\Delta\beta_2$. The $\mu_1$ worker on job 1 reallocates two-thirds of his time to job 2, for a gain of $\frac{2}{3}\mu_1\Delta\beta_1$. Thus the rate of reassignment is greatest (two-thirds vs. one-third) closest to the rank of new workers.

[13] An alternative way of seeing the reallocation process, directly related to (6), is to consider a wave of reallocations from job $p$ to the adjacent job, and so on, to job $y$. The output gain (or loss) from this wave is $\int_{z=p}^{y} \mu(z) d\beta(z)$. If we insert at job $p$ a mass $dp$ of workers with ability $\mu(p)$, then one restores the uniform assignment of workers to jobs with a series of waves, just like the one described, with the endpoints of such waves, $y$, distributed uniformly on [0, 1]. That is precisely what the double integral in (6) describes.

Note that although workers of high ability earn less than their direct contribution to output, firms cannot gain ex post by bidding these workers away from one another. Any such move would require shifting downward the assignments of less able incumbents. The attendant reduction in output would absorb all of the apparent gain from the acquisition. Similarly, although low-ability workers earn more than their direct contribution to output, firms cannot gain by firing such workers, since this would necessitate (after rescaling the level of the firm's operation) the demotion of others to take up their tasks, with the attendant reduction in their output.

We have thus provided an explanation of why the wages of members of an academic department (say) may be less variable than their productivities in any particular task. If agents work in teams with varied tasks that require the application of person-hours in fixed relative proportions, the total returns to ability cannot be proportional to agents' direct contributions to output. This is so even when, as in our model, more able agents are proportionately better at every conceivable task, because adding such an agent to the team incurs the cost of reassigning all the other team members.

### B. *Substitutes and Complements*

We now turn to the question of Hicks substitutability or complementarity, the effect of an increase in one type of labor on the wage of another. The wage effects of any change in the distribution of ability—the effect of immigration, educational tracking, and so forth—rest on these relationships. Under a production function with discrete ability types, we would examine the cross partials, $F_{ij}(L_1, \ldots, L_n)$. Here, we shall perform the analogous exercise pertinent to a continuous ability distribution and consider how these relationships depend on the technology, $\beta(\cdot)$. Specifically, we consider the addition to the labor force $L$ of a mass of labor $\Delta_p$, with ability $\mu_p$, and find the effect on the wage of a worker with ability $\mu_r$. (The notation is chosen to indicate that $\mu_p$ and $\mu_r$ are constants that correspond to the ability at quantiles $p$ and $r$, prior to the addition of $\Delta_p$.)

The key to the analysis is the effect that an increment of labor $\Delta_p$ has on the assignment of ability types to quantiles above and below it. As figure 3 depicts, the insertion of labor raises the ability levels assigned to lower quantiles and reduces those assigned to higher ones. Specifically, the assignment of ability $\mu$ to quantile $z$, $\mu(z; \Delta_p)$, is implicitly given by

$$z = F(\mu)\left(\frac{L}{L + \Delta_p}\right) \quad \text{for } \mu < \mu_p$$

and

$$1 - z = [1 - F(\mu)]\left(\frac{L}{L + \Delta_p}\right) \quad \text{for } \mu > \mu_p.$$

The assignment of ability to any quantile $z$ responds to an infinitesimal $\Delta_p$ as follows:

$$\frac{\partial \mu(z;\ \Delta_p)}{\partial \Delta_p} = \begin{cases} \mu'(z)\dfrac{z}{L} & z < p \\ -\mu'(z)\dfrac{1-z}{L} & z > p. \end{cases} \tag{11}$$

We can now derive the marginal product for a worker of ability $\mu_r$. Since output is given by

$$Q(\Delta_r) = (L + \Delta_r)\int_0^1 \beta(z)\mu(z;\ \Delta_r)dz,$$

we have

$$\frac{\partial Q}{\partial \Delta_r} = \int_0^1 \beta(z)\mu(z)dz + \int_0^r \beta(z)\mu'(z)zdz - \int_r^1 \beta(z)\mu'(z)(1-z)dz, \tag{12}$$

using (11). The marginal productivity interpretation of (6) can then readily be verified (see the Appendix):

$$\frac{\partial Q}{\partial \Delta_r} = \beta(r)\mu_r + \int_0^1\int_r^y \mu(z)\beta'(z)dzdy = W(\mu_r) = w(r).$$

We now turn to the result we are looking for (see the Appendix for the proof).

PROPOSITION 2. For quantiles $p < r$,

$$\frac{\partial}{\partial \Delta_p}\left(\frac{\partial Q}{\partial \Delta_r}\right) = \frac{\partial W(\mu_r)}{\partial \Delta_p} = \int_0^1\int_r^y \frac{\partial \mu(z;\ \Delta_p)}{\partial \Delta_p}d\beta(z)dy$$

$$= \left(\frac{1}{L}\right)\left[-\int_0^p z^2\mu'(z)d\beta(z) + \int_p^r z(1-z)\mu'(z)d\beta(z) - \int_r^1 (1-z)^2\mu'(z)d\beta(z)\right]. \tag{13}$$

By Shepherd's lemma, this expression also gives us $\partial W(\mu_p)/\partial \Delta_r$.

*Discussion.*—The net effect on $W(\mu_r)$ of adding workers of ability $\mu_p$ works entirely through the $\mu_r$ worker's indirect marginal productivity.

First, the addition of $\mu_p$ workers raises the ability of individuals assigned to lower quantiles, so the demotions over $(0, p)$ required to accommodate a worker of $\mu_r$ become more costly. This effect tends to *reduce* $W(\mu_r)$, as the first term within the brackets of (13) indicates. The second term concerns the quantile range $(p, r)$. Since the addition of $\mu_p$ workers decreases ability over that interval, the demotions in that range to accommodate a worker of $\mu_r$ become less costly, thereby *raising* $W(\mu_r)$. Finally, the addition of $\mu_p$ workers decreases the abilities assigned to higher quantiles, too $(r, 1)$. This reduces the productivity gains from the promotions supported by $\mu_r$ workers, *reducing* $W(\mu_r)$, as the third term indicates. Thus the first and third terms contribute to Hicks substitutability and the second term to complementarity.

Naturally, the more similar two workers are (i.e., the closer $p$ is to $r$), the more likely they are to be substitutes. In the limiting case, as $p \rightarrow r$, the second term of (13) vanishes, and we simply have diminishing marginal productivity. At the opposite extreme, as $p \rightarrow 0$ and $r \rightarrow 1$, the first and third terms vanish, and we have complementarity: top-ability workers are helped by an increased supply of the least skilled, and conversely.

Consider more closely the addition of low-ability workers, near $\mu = 0$, for example, from the low-skilled range of immigrants. What determines how many low-skill workers are hurt versus high-skill workers who are helped? In this model, the nature of the technology determines the answer. Examining (13), one can show that if $\beta(\cdot)$ is convex, then more low-skill workers are likely to be hurt than if $\beta(\cdot)$ were concave. That is, more people of low ability would be hurt by additional workers of lesser ability if the gradient of $\beta(\cdot)$—the sensitivity of output to proper job assignment—were concentrated in the high-skill jobs. The reason is that under such a technology, the main effect of new low-skill workers, which reduces ability levels assigned to all quantiles, is the adverse effect on the value of promotions supported by most workers, up through the high-skill jobs. Conversely, if the technology is such that the sensitivity of output to proper job assignment is concentrated in low-skill jobs, then the infusion of very-low-skill workers helps most workers by reducing the cost of demotions required to accommodate them.

Similarly, we may consider the addition of high-ability types, near $\mu = 1$, for example, from the top end of the immigrant spectrum. Again, the concavity or convexity of $\beta(\cdot)$ affects the gains and losses. More people of low ability would be *helped* by additional high-ability workers, to the extent that the sensitivity of output to job assignment is concentrated in high-skill jobs. Here, the main effect of the rise in ability is to raise the value of promotions through the high-skilled jobs, which are supported by most workers. Again, the converse holds for the opposite type of technology.

In Section IV, we analyze broader changes in the ability distribution, using different methods. However, it is worth bearing in mind that the effects of *any* change in the ability distribution ultimately rest on a series of point insertions such as those we have analyzed here.

## IV. Comparative Statics of Technology and Ability in the Fixed Hierarchy Model

### A. *The Distributional Effect of Technical Progress*

Technical progress, a rise in output for given inputs, is represented by a rise in $\beta(\cdot)$. This widens the income distribution, a result that generalizes from the two-job model.

Proposition 3. Technical progress widens the wage gap between the top and the bottom, $w(1) - w(0)$, and, indeed, between individuals on either side of any affected tasks: for any $p > p'$ such that $\beta(z)$ rises for some $z \in (p', p)$ and does not fall, $w(p) - w(p')$ rises.

The result follows directly from (5):

$$w(p) - w(p') = \int_{z=p'}^{p} \beta(z) d\mu(z). \tag{14}$$

Since $\beta(\cdot)$ is the marginal return to ability, a rise in $\beta(\cdot)$ widens the gap between wages of workers of different ability, as in the two-job model.

The effect of technical progress on wages of the least skilled, $w(0)$, is also similar to that found in the two-job model: it depends on what happens to the gradient of $\beta(\cdot)$. This can be immediately seen by writing (9) as $w(0) = \int_0^1 \int_{z=0}^{y} \mu(z)\beta'(z)dzdy$. A parallel shift in $\beta(\cdot)$ has no effect on $w(0)$. If technical progress is concentrated in the higher-skilled jobs, then $\beta(\cdot)$ gets steeper and $w(0)$ rises. Conversely, if technical progress is concentrated in the low-skilled jobs, this flattens $\beta(\cdot)$ and reduces $w(0)$. As in the two-job case, this possibly surprising result is easily explained. The rise in $\beta(\cdot)$ for jobs filled by the least skilled raises their direct contribution to output, but vanishingly so as $\mu(p) \to 0$. This is outweighed by the fact that a rise in $\beta(\cdot)$ for the low $t$'s toward that of the high $t$'s reduces the productivity gain from shifting workers up the hierarchy, which is the zero-ability worker's sole contribution to output. As $\beta(\cdot)$ flattens out entirely, we approach the one-job model and $w(0)$ vanishes.

### B. *The Distributional Effect of Improved Ability*

Consider an improvement in the endowment of ability. That is, consider a CDF on the distribution of $\mu$, $G(\cdot)$, that exhibits first-order stochastic dominance over $F(\cdot)$: $G(\mu) \le F(\mu)$ for all $\mu \in [0, 1]$. Then we can immediately establish the following result.

Proposition 4. A first-order stochastic improvement in the distribution of ability raises $w(0)$ and reduces $w(1)$.

*Proof.* Since $F(\cdot)$ and $G(\cdot)$ are nondecreasing,

$$G(\mu) \le F(\mu) \rightarrow \mu(z;\ G) \equiv G^{-1}(z) \ge F^{-1}(z) \equiv \mu(z;\ F)$$

for all $z \in [0, 1]$. The result then follows by inspection of (9) and (10). Q.E.D.

The logic is the same as we found in the two-job model. A rise in the population's ability makes the promotions supported by a zero-ability worker more valuable, and it makes more costly the demotions that are required to accommodate a top-ability worker.

Thus an improvement in the population's ability distribution has the opposite effect on the wage span from an improvement in technology. As we saw, technical progress raises the price of incremental ability, $\beta(\cdot)$, widening the wage gap $w(1) - w(0) = \int_{z=0}^{1} \beta(z)\mu'(z)dz$. By contrast, a general rise in ability between $\mu = 0$ and 1 reduces $\mu'(\cdot)$ at high $z$'s and raises it at low $z$'s, thereby redistributing the gradient of ability toward the lower-skilled jobs, where it fetches a lower price. Thus a general rise in ability reduces the average price over the full span of ability, narrowing the wage span.

### C. *The Output and Distributional Effects of Ability Dispersion*

We now turn to the paper's main question: Under what conditions does a more unequal ability distribution map onto a wider or narrower wage distribution? The question can be motivated by a comparison of economies with identical means but different dispersions of ability, as might be occasioned by more or less stratified processes of human capital development (e.g., educational tracking, according to its critics).

The question also arises when considering the effect of improved information. Suppose that employers do not observe a worker's ability, but receive an imperfectly informative signal (e.g., a test or school attendance records).[14] Each worker's expected ability is assessed conditional on the information available about him. Since output is linear in worker abilities, expected output depends only on the distribution of expected abilities. That is, it makes no difference whether the workers

[14] We keep matters simple by assuming that workers and employers have the same information, so workers too are uncertain about how well they will perform. We continue to assume that worker abilities are exogenous. Thus we are neglecting here the important issue of how changing the quality of information available to employers about worker abilities affects the incentives that workers have to acquire skills. We further discuss this issue briefly in Sec. VI.

assigned to a given job all have the same known ability or have an array of unknown abilities with the same average. Thus, when the distributions are taken as a whole, there is no difference for risk-neutral firms between facing a population with a given distribution of known abilities and facing a different population with the same distribution of conditional expected abilities (but of course a different, more disperse, distribution of underlying unknown abilities). Both cases result in the same expected output and the same wage distribution.[15]

Now, better information will lead to a different, more disperse, distribution of expected ability in the worker population. We can formally establish a rather general and quite useful result (see the Appendix for the proof) on the precise sense in which better information leads to a more unequal distribution of expected ability.[16]

THEOREM 1. Consider two tests of worker abilities, and suppose that test 1 is more informative than test 2 in the sense of Blackwell (1953). Then the population distribution of estimated mean productivities induced by test 1 (viewed as a random variable ex ante) is riskier (in the sense of second-order stochastic dominance) than that induced by test 2.

The significance of the result is this: to do comparative statics of the effect of “improved information” in a model with uncertain worker abilities, it is sufficient to study the effect of a mean-preserving spread of abilities in a corresponding model in which abilities are known. That is, our analysis of the effects of a more unequal (i.e., riskier) ability distribution can be interpreted as applying either to the underlying distribution itself or to an improvement in information about each worker’s place in it.[17]

Either way, however, we confine ourselves to the case of fixed support $\mu \in [0, 1]$. Under the informational interpretation, this means that we shall compare situations in both of which there is sufficient information to assign individuals to the bottom and top of the distribution. This rules out the case of zero information, where the distribution of expected ability is concentrated at the mean, a restriction that should be borne in mind below.

[15] Rank workers of unknown ability $\alpha$ by the value of their signal $s$, scaled in percentiles. Then expected output is $E(Q) = \int_0^1 \beta(s)E[\alpha|s]\,ds$. This is isomorphic to the certainty case, $Q = \int_0^1 \beta(p)\mu(p)\,dp$, for a hypothetical population in which the distribution of known ability, $\mu(p)$, is identical to the first population’s distribution of expected ability, $E[\alpha|s]$.

[16] This theorem generalizes the well-known case of the normal testing model, where the posterior distribution of expected ability is widened by a reduction in the noise of the test (less regression to the mean). Cornell and Welch (1996) present a related result, but our theorem is somewhat stronger.

[17] Some readers of earlier versions of this paper have commented that improved information affects earnings only early in one’s career, before one’s true abilities are established on the job. The point is well taken, so the noninformational interpretation may be more compelling for some readers.

We may now establish an important result: A mean-preserving rise in the riskiness of the ability distribution raises total output and, therefore, the mean wage. We can see this intuitively in the simple case of a single crossing of CDFs, where ability rises in the right tail of the distribution and drops in the left tail. The rise in ability occurs for those filling the nearly top jobs, which are the most ability-sensitive. Thus the output effect here outweighs that from the drop in ability among those filling the ability-insensitive jobs. This result holds more generally for a mean-preserving spread with possibly multiple crossings (see the Appendix for the proof).

Proposition 5. Suppose that the ability distribution $G(\cdot)$ is "riskier" than $F(\cdot)$, but the two distributions have the same mean. Then $Q(G) \geq Q(F)$.

We turn now to our principal question: How does a more unequal distribution of ability, or better information about abilities, affect the equilibrium degree of wage inequality? With regard to better information, it is often assumed that there is a sacrifice in equity that is necessarily incurred for the efficiency gains (proposition 5) from improving the quality of matches between workers and jobs. The following comparative static analysis should shed some light on the conditions under which this equity-efficiency trade-off holds or whether better information advances both efficiency and equity goals. From a noninformational viewpoint, our result will address the question of whether a more unequal ability distribution necessarily implies more wage inequality.

Let $w_F(p)$ denote the wage profile under ability distribution $F$. From (14), we have

$$\Delta w_F \equiv w_F(1) - w_F(0) = \int_0^1 \beta(y) dF^{-1}(y), \tag{15}$$

where $\Delta w_F$ denotes the wage span. We can also write[18]

$$w_F(0) = \int_0^1 \int_y^1 [\beta(z) - \beta(y)] dz dF^{-1}(y) \tag{16a}$$

[18] To see this, integrate (9) by parts to get

$$w_F(0) = \int_0^1 F^{-1}(z)\beta(z)\, dz - \int_0^1 \int_{y=0}^{z} \beta(y)\, dF^{-1}(y)\, dz$$

$$= \int_0^1 \int_{y=0}^{z} [\beta(z) - \beta(y)]\, dF^{-1}(y)\, dz.$$

Changing the order of integration gives the result in (16a). Substituting into (15) gives the expression for $w_F(1)$ in (16b).

and

$$w_F(1) = \int_0^1 \left[ y\beta(y) + \int_y^1 \beta(z)dz \right] dF^{-1}(y). \tag{16b}$$

Suppose now that $F(\cdot)$ and $G(\cdot)$ are two distinct distributions of estimated abilities in the worker population. Invoking second-order dominance, we shall write "$G$ is more unequal (or riskier) than $F$" (hereafter $G$ MUT $F$) if $\int_0^1 [G(y) - F(y)]dy = 0$ and $\int_0^\mu [G(y) - F(y)]dy \geq 0$ for all $\mu \in [0, 1]$. We may now state the main result for the fixed hierarchy model.

PROPOSITION 6. Let $F$ and $G$ be atomless with full support on [0, 1]. Let $G$ MUT $F$: (i) if $\beta(\cdot)$ is concave, then $w_G(0) \leq w_F(0)$ and $\Delta w_G \geq \Delta w_F$; (ii) if $\beta(\cdot)$ is convex, then $w_G(1) \leq w_F(1)$ and $\Delta w_G \leq \Delta w_F$.

*Proof.* The results follow immediately from the following lemma.

LEMMA 1. $G$ MUT $F$ and $\psi(\cdot)$ convex [concave] imply

$$\int_0^1 \psi(y)dG^{-1}(y) \leq [\geq] \int_0^1 \psi(y)dF^{-1}(y).$$

*Proof of lemma* 1. Since abilities lie in the unit interval and both $F(\cdot)$ and $G(\cdot)$ are atomless with full support, we conclude that $F^{-1}(0) = G^{-1}(0) = 0$, $F^{-1}(1) = G^{-1}(1) = 1$, and both $F^{-1}(\cdot)$ and $G^{-1}(\cdot)$ are continuous and strictly increasing on [0, 1]. Taking these inverse functions as CDFs in their own right, we further conclude that $G$ MUT $F$ implies $F^{-1}$ MUT $G^{-1}$ (see the proof of proposition 5). So the inequality of the lemma is the principal characterization result for second-order stochastic dominance (the expectation of a convex function is no less under a riskier distribution). Q.E.D.

To complete the proof of proposition 6, employ (15) and (16) to see that

$$w_F(0) = \int_0^1 \psi_0(y)dF^{-1}(y), \quad \text{for } \psi_0(y) \equiv \int_y^1 [\beta(z) - \beta(y)]dz,$$

$$w_F(1) = \int_0^1 \psi_1(y)dF^{-1}(y), \quad \text{for } \psi_1(y) \equiv y\beta(y) + \int_y^1 \beta(z)dz,$$

and

$$\Delta w_F = \int_0^1 \psi_2(y)dF^{-1}(y), \quad \text{for } \psi_2(y) \equiv \psi_1(y) - \psi_0(y) = \beta(y).$$

Since $\beta(\cdot)$ is nondecreasing, $\beta(\cdot)$ convex implies that $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are convex, whereas $\beta(\cdot)$ concave implies that $\psi_0(\cdot)$ is convex and $\psi_2(\cdot)$ is concave. This proves proposition 6.

Proposition 6 assesses the impact of a more unequal ability distribution (or a more informative test of worker abilities) on the equilibrium range of wages. If output per unit of ability, $\beta(\cdot)$, is a convex (concave) function of a worker's position in the hierarchy, then a greater dispersion of ability (on fixed support) lowers wages for the most (least) skilled workers and compresses (widens) the wage span.

The way that $\beta(\cdot)$'s curvature conditions the effect of ability dispersion on $w(0)$, $w(1)$, and $\Delta w$ can be understood by noting first that such dispersion raises the ability of those in the high-skilled jobs and reduces it in the low-skilled jobs. Recall that $w(0)$ represents the productivity gains from promoting workers up the hierarchy, especially through the low-skill jobs, since that is where the intensity of the rate of promotions is greatest upon the addition of a $\mu = 0$ worker. Thus, by reducing the ability of workers on these jobs, a rise in ability dispersion tends to reduce the productivity gains from these promotions, provided that the gradient $\beta'(\cdot)$ in these jobs is at least as high as in the more skilled jobs. Thus, if $\beta(\cdot)$ is concave, $w(0)$ falls with a rise in ability dispersion; indeed, $w(0)$ falls relative to $w(1)$, so the wage span widens. Conversely, the earnings distribution is compressed by a more unequal ability distribution when $\beta(\cdot)$ is convex. Here the gradient is steepest among the high-skill jobs, whose occupants are the ones most intensively demoted from the addition of a $\mu = 1$ worker. Since their ability rises with a more unequal distribution, the cost of demoting them rises; so $w(1)$ falls and the wage span narrows.[19]

The same logic holds when we consider technologies in which $\beta(\cdot)$ is S-shaped rather than convex or concave, as in the two-job model. What matters is the location of the near-vertical region of the S, that is, the steepest region of $\beta(\cdot)$. This is the region in which the gains or losses from reassigning workers are concentrated, so that is where the distributional effect of a change in the ability distribution will play out. If this region lies in the low-skilled jobs, where a more unequal ability distribution reduces $\mu$, then $w(0)$ will fall and $w(1)$ will rise. Conversely, a more unequal ability distribution will raise $w(0)$ and reduce $w(1)$ if job assignment is most sensitive in the high-skilled jobs. That is why, in our two-job model, the effect of a more unequal ability distribution depended critically on whether $\theta$, the proportion of low-skill jobs, was low or high.

[19] For linear $\beta(\cdot)$, wages fall at both extremes by the same amount, so $\Delta w = 0$. However, the mean wage rises (by proposition 5), so the middle class gains from ability dispersion.

### D. *The Output and Distributional Effects of Economic Integration*

The final change we analyze in the distribution of abilities arises from the economic integration of distinct populations. Let there be two populations with distributions of estimated ability denoted by $F_a(\mu)$ and $F_b(\mu)$, respectively. Think of them as representing workers in different regions of a country, distinct racial or ethnic groups, or different nations. The underlying ability distributions may differ, or the test of worker abilities may be more informative about one group than the other. Consider what would happen if the two populations were to merge, as a result of the economic integration of their labor markets, assuming that a common technology, represented by $\beta(\cdot)$, prevails both before and after integration. How would output and the distribution of income be affected?

Let $\lambda$ be the fraction of *a*'s in the total population, $0 < \lambda < 1$. Then the merged population's ability distribution is given by $F_\lambda(\mu) \equiv \lambda F_a(\mu) + (1 - \lambda)F_b(\mu)$. So, if group *a*'s ability distribution first-order stochastically dominates that of group *b*, the merged group lies in between: $F_a$ FOSD $F_b \Leftrightarrow F_a$ FOSD $F_\lambda$ FOSD $F_b$. Thus proposition 4 implies the following corollary.

COROLLARY 1. When group *a*'s ability distribution FOSD that of group *b*, economic integration results in $w_b(1) > w_\lambda(1) > w_a(1)$ and $w_a(0) > w_\lambda(0) > w_b(0)$.

This result may help explain why it is the elites within relatively backward populations, along with the lower classes of relatively advanced groups, who sometimes resist labor market integration. Each is the relatively scarce "factor" within its own group; as the Stolper-Samuelson theorem suggests, these are precisely the workers harmed by economic integration. Corollary 1 also shows how economic integration tends to narrow wage dispersion, by raising the overall minimum and lowering the overall maximum wages.

Similarly, if we merge groups with the same mean but unequal distributions, we have $F_a$ MUT $F_b \Leftrightarrow F_a$ MUT $F_\lambda$ MUT $F_b$, so proposition 6 implies the following corollary.

COROLLARY 2. When group *a*'s ability distribution MUT that of group *b*, economic integration results in $w_b(1) > w_\lambda(1) > w_a(1)$ if $\beta(\cdot)$ is convex and $w_b(0) > w_\lambda(0) > w_a(0)$ if $\beta(\cdot)$ is concave.

Thus, whether $\beta(\cdot)$ is convex or concave, integration always has the effect of moving at least one extreme of the overall wage distribution toward the mean. It is interesting to consider this result as applying to two groups with the same underlying ability distribution but different degrees of observability. Those at one or both extremes of the wage distribution in the less accurately observed group, *b*, stand to lose from integration with the more accurately observed group.

Whatever its distributional consequences, integration of two such groups—or *any* two groups—must raise each group's average wage (see the Appendix for the proof).

Proposition 7. Regardless of the technology or the ability distributions in the two populations, both groups benefit, on average, from economic integration:

$$\int_0^1 W_\lambda(\mu)\,dF_i(\mu) \geq \int_0^1 W_i(\mu)\,dF_i(\mu) \quad \text{for each group } i.$$

*Discussion.*—Economic integration of two groups affects the wage of a worker in group $i$ by embedding him in a population with different abilities and by changing his job assignment. *When the assignment is held constant,* the change in his coworkers' abilities has no effect on the worker's direct contribution to output, but it affects his indirect marginal productivity by changing the value of the promotions and demotions associated with inserting this worker into the job hierarchy.[20] For group $i$ as a whole, however, the indirect marginal productivities must still net out to zero, since the effects of promotions enabled by adding a worker at one position are washed out by the demotions required to accommodate the marginal worker at another, regardless of the ability distribution.

Thus it is the change in job assignments that accounts for the effect of economic integration on the income of group $i$ as a whole. Consider a worker at quantile $p$ before the merger who is reassigned to quantile $F_\lambda(\mu_i(p))$ afterward. So, if $p < F_\lambda(\mu_i(p))$, integration causes that worker to be promoted to a higher position in the job hierarchy, raising the direct marginal productivity component of his wage. At the same time, to accommodate that promotion, there is an equal and opposite movement down the hierarchy of all those at positions $z \in [p, F_\lambda(\mu_i(p))]$, intermediate between that worker's old and new assignments. The rise in the worker's direct marginal productivity outweighs the reduction in his indirect marginal productivity from the accommodating demotions. The reason is that the demotions occur for workers of lesser ability, $\mu_\lambda(z) < \mu_i(p)$, so the productivity effect of moving them down through the gradient of $\beta(z)$ is less than that of promoting the worker in question up through the same gradient. Conversely, for a worker who is demoted upon integration, the direct loss in marginal productivity is outweighed by the indirect gains in marginal productivity from the promotion of those of higher ability, $\mu_\lambda(z) > \mu_i(p)$. For each group taken as a whole, therefore, income must rise, as proposition 7 states.

[20] This is the sole effect on $w_i(0)$ and $w_i(1)$ in corollaries 1 and 2 above, since those workers are not in fact reassigned.

Note that the reassignment maximizes output for the merged group, but not necessarily for each group separately. However, if the direct output of (say) group *a* declines because of a net shift to lower positions, group *a* will be more than compensated by an increase in its indirect marginal productivity. This increase in the indirect marginal productivity is, in effect, a transfer from group *b*. That is, group *b*, whose direct output has risen, compensates group *a* for the reassignments that made this possible.

## V. Production in Variable Proportion Hierarchies

### A. *Crowding, Nonuniform Assignment, Arbitrage, and No-Profit Conditions*

We now relax the assumption of strict complementarity in the number (or density) of workers across tasks, that is, of fixed proportions. We may think of fixed proportions as a technology characterized by extremely deleterious effects of crowding workers into tasks, relative to the number of workers in other tasks. That is, any deviation from the fixed proportions that reassigns workers from one small range of tasks to another reduces output in the former but does not raise output at all in the latter. The obvious relaxation, which we adopt, is to suppose that the effect of crowding is not so extreme as to block *any* additional output, but rather that raising the proportion of workers assigned to a small range of tasks raises the output from those tasks less than proportionately.

Formally, we relax the assumption of uniform assignment, $p = t$, by letting a nonuniform CDF $\Phi(t)$ denote the proportion of the workforce assigned to task $t$ or lower: $p = \Phi(t)$. Let $\varphi(t) \equiv \Phi'(t)$ be the assignment density of workers to tasks. Let $\varphi dt$ natural units of labor on tasks $(t, t + dt)$ generate $h(\varphi)dt$ "effective" units of labor, where $h(\varphi)$ is a smooth neoclassical production function,[21] $h(0) = 0$, $h'(\cdot) > 0$, and $h''(\cdot) < 0$. Normalize $h(1) = 1$ such that effective units of labor equal natural units when workers are assigned uniformly across tasks. Note that the elasticity of $h$, $\eta(\varphi) \equiv \varphi h'(\varphi)/h(\varphi) < 1$, so effective labor rises less than proportionately to the density of natural units, the relaxation of fixed proportions that we were seeking. Output is thus given by

$$Q = \int_0^1 \beta(t)\mu[\Phi(t)]h[\varphi(t)]dt. \tag{17}$$

[21] We may think of $h(\varphi)$ as being derived from a more conventional production function, based on absolute employment density (with constant returns), rather than being directly based on the relative employment density $\varphi$. Details are available on request.

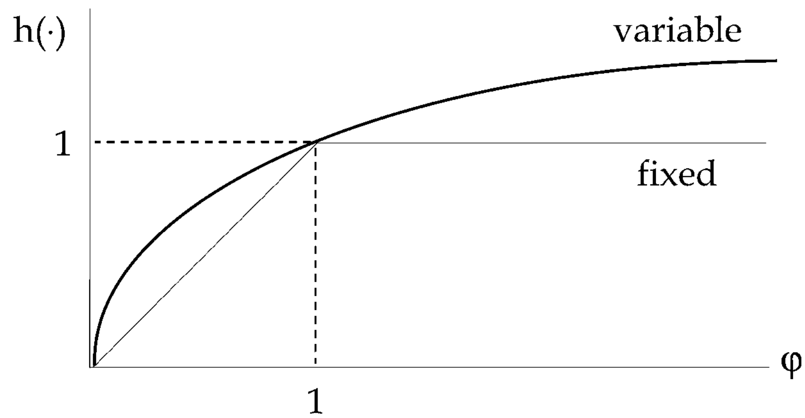


FIG. 4

Note that fixed proportions is a limiting case in which $h(\varphi) \to \min(\varphi, 1)$, the Leontief technology (see fig. 4). In this case, firms will choose the uniform assignment, $\Phi(t) = t$, $\varphi(t) = 1$, for all $t$, since any departure from this means $\varphi(t) > 1$ over some tasks, which wastes labor. We can then trivially change variables from $t$ to $p$, giving us (4).

More generally, for non-Leontief $h(\varphi)$, it will be useful to express the output from tasks $(t, t + dt)$ as the output from corresponding workers $(p, p + dp)$. To implement the change of variables from $t$ to $p$, define the assignment of tasks to workers as the inverse of the assignment of workers to tasks. That is, the proportion of tasks assigned to workers of ability quantile $p$ or lower is $t = T(p) \equiv \Phi^{-1}(p)$. Workers at quantile $p$ are assigned tasks with density $\tau(p) \equiv T'(p) = 1/\varphi[T(p)]$. We thus have

$$Q = \int_0^1 \beta[T(p)]\mu(p)h\left[\frac{1}{\tau(p)}\right]\tau(p)dp. \tag{18}$$

This expression is readily interpreted. Since $h[1/\tau(p)]\tau(p)$ is the effective units of labor per natural unit of labor, the integrand is simply output per natural unit of labor.

We can now immediately generalize proposition 1. Define

$$b(p) \equiv \beta[T(p)]h\left[\frac{1}{\tau(p)}\right]\tau(p), \tag{19}$$

the output per unit of ability for a worker of ability $\mu(p)$, *for any given* *assignment function.* Then the same arbitrage and zero-profit conditions used in (6) give us

$$w(p) = b(p)\mu(p) + \int_0^1 \int_{z=p}^{y} \mu(z)db(z)dy. \tag{20}$$

That is, when we substitute $b(\cdot)$ for $\beta(\cdot)$, the wage profile (6) generalizes from the fixed uniform assignment to any given assignment function, whether or not it is optimal.

### *B. Optimal Assignment*

For any given wage schedule, profit-maximizing firms will distribute workers across jobs to maximize output. Formally, the choice of $\Phi(\cdot)$ can be posed in the standard calculus of variations format, to maximize

$$Q = \int_0^1 \hat{q}(t, \Phi, \varphi)dt, \tag{21}$$

where

$$\hat{q}(t, \Phi, \varphi) \equiv \beta(t)\mu(\Phi)h(\varphi),$$

with boundary conditions $\Phi(0) = 0$ and $\Phi(1) = 1$. The Euler equation, $\partial\hat{q}/\partial\Phi = d(\partial\hat{q}/\partial\varphi)/dt$, is

$$\beta(t)\mu'[\Phi(t)]h[\varphi(t)] = \frac{d}{dt}\{\beta(t)\mu[\Phi(t)]h'[\varphi(t)]\} \quad \forall t \tag{22}$$

or, equivalently,

$$\int_{t_0}^{t_1} \beta(t)\mu'[\Phi(t)]h[\varphi(t)]dt = \beta(t_1)\mu[\Phi(t_1)]h'[\varphi(t_1)] - \beta(t_0)\mu[\Phi(t_0)]h'[\varphi(t_0)]. \tag{23}$$

The right-hand side is the marginal benefit of reducing the density of workers at task $t_0$ and raising the density at task $t_1$, a task with higher marginal productivity; the left-hand side is the marginal cost of this reallocation of workers from lower to higher tasks, which is to replace the workers in intervening tasks with workers of lower ability, drawn from lower tasks.

Further intuition can be gleaned by carrying out the differentiation and rearranging:

$$\varphi' = \left(\frac{h'}{-h''}\right)\left\{\frac{\beta'}{\beta} - \frac{\mu'}{\mu}\left[\frac{(1-\eta)\varphi}{\eta}\right]\right\}, \tag{24}$$

where $\eta \equiv \varphi h'/h < 1$. The left-hand side, the slope of the optimal assignment density, is the rate at which crowding should rise at higher tasks. The right-hand side points to three factors. First, the curvature of the effective labor production function, $-h''$, measures the cost of uneven crowding due to diminishing returns, so it is inversely related to $|\varphi'|$.[22] Second, the benefit of pushing more workers up the hierarchy is the higher return to any given ability, $\beta'/\beta$. Third, the cost of pushing more workers up the hierarchy is that it crowds workers of higher ability: the steeper the ability gradient $\mu'/\mu$ is, the greater the cost of crowding more workers onto higher tasks. The slope of the optimal assignment density reflects these costs and benefits.

### C. *Wage Profile at Optimal Assignment*

The no-arbitrage and zero-profit conditions imply wage profile (20) for any given assignment, optimal or not. At the optimal assignment, we offer the following nice result.

PROPOSITION 8. The variable proportions model yields the competitive wage profile:

$$w(p) = \mu(p)b(p)\eta\left[\frac{1}{\tau(p)}\right] + \int_0^1 \mu(y)b(y)\left\{1 - \eta\left[\frac{1}{\tau(y)}\right]\right\}dy. \tag{25}$$

*Proof.* The Euler condition (23) can be rewritten with a change of variables:

$$\beta[T(y)]\mu(y)h'\left[\frac{1}{\tau(y)}\right] - \beta[T(p)]\mu(p)h'\left[\frac{1}{\tau(p)}\right] = \int_{z=p}^{y} b(z)d\mu(z). \tag{26}$$

[22] The extreme case is Leontief ($h(\varphi) \to \min(\varphi, 1)$), where $h'' \to -\infty$ around $\varphi = 1$; so the optimal assignment is uniform ($\varphi' \equiv 0$), as stated earlier. The opposite limiting case is perfect substitutability, $h(\varphi) \to \varphi$, for all $\varphi$. Here, $h'(\varphi) \equiv \eta(\varphi) \equiv 1$ and $h''(\varphi) \equiv 0$, so

$$-h''\varphi' = 0 < h'\left\{\frac{\beta'}{\beta} - \frac{\mu'}{\mu}\left[\frac{(1-\eta)\varphi}{\eta}\right]\right\} = \frac{\beta'}{\beta},$$

and (24) fails for any nondegenerate assignment. The reason is that under perfect substitutability, crowding has no effect, so all workers are piled onto task $t = 1$, where productivity is highest.

Integrating $\int_{z=p}^{y} b(z)d\mu(z)$ by parts and rearranging, we get

$$\int_{z=p}^{y} \mu(z)db(z) = b(y)\mu(y)\left\{1 - \eta\left[\frac{1}{\tau(y)}\right]\right\} - b(p)\mu(p)\left\{1 - \eta\left[\frac{1}{\tau(p)}\right]\right\}. \quad (27)$$

Substituting into (20) gives the desired result. Q.E.D.

The interpretation of (25) is straightforward. The output per worker of ability $\mu(p)$, on the task to which he is assigned, is $b(p)\mu(p)$. Of that, the share attributable to that worker is simply the elasticity of effective labor $\eta[1/\tau(p)]$. That is, his marginal contribution to output on that task is

$$\beta[T(p)]\mu(p)h'\left[\frac{1}{\tau(p)}\right] = b(p)\mu(p)\eta\left[\frac{1}{\tau(p)}\right],$$

the first term in (25). The remaining portion of $b(p)\mu(p)$, $1 - \eta[1/\tau(p)]$, is attributable to workers on other tasks, since they reduce the relative crowding on the task in question. Conversely, the second term of (25) is the contribution of worker $p$ to the output on tasks assigned to all other workers because of complementarity throughout the job structure.

Of course, the marginal productivity interpretation of (6), as generalized in (20), still holds as well. This interpretation, it will be recalled, is the addition to output from adding a worker of $\mu(p)$ to the appropriate task *while reassigning other workers to maintain the given assignment proportions.* The marginal productivity interpretation of (25), just discussed, considers the addition to output *without reassigning other workers.* These two assignments yield equal results at the optimum, by the envelope theorem.

### D. *Comparative Statics in the Cobb-Douglas Case*

Consider the case $h(\varphi) = \varphi^{\eta}$, where $\eta$, the elasticity of effective labor with respect to natural units, is constant. This can be considered a Cobb-Douglas production function, with unitary elasticity of substitution between labor on tasks and a task-specific factor.

Explicit solutions can be found in this case (see the Appendix). From these solutions immediately follow the comparative statics with respect to technology.

PROPOSITION 9. In the Cobb-Douglas case, any change in the productivity function $\beta(\cdot)$ that raises (reduces) output (i) raises (reduces) the wage paid at any quantile and (ii) widens (narrows) the wage gap between any two quantiles.

Let us compare these results with the fixed proportions case. As discussed in Section IV*A*, if technical progress is concentrated in the low-skilled jobs, flattening $\beta(\cdot)$, $w(0)$ is reduced under fixed proportions. But under constant elasticity, part i of proposition 9 states that technical progress must raise all wages, including $w(0)$. Here there is no such thing as technical progress that is so biased against any skill group that it can reduce its wages. This is similar to the classic result for the Cobb-Douglas production function in nonassignment models.

To understand the difference between the Cobb-Douglas and fixed-proportions cases, note from (20) that under variable proportions, $w(0)$ depends on the gradient of $b(\cdot)$, not just $\beta(\cdot)$. In the Cobb-Douglas case, there is enough substitutability that any technical improvement that flattens $\beta(\cdot)$ is offset by optimizing reassignment that steepens $b(p) = \beta[T(p)]\tau(p)^{1-\eta}$ over at least some quantiles, thereby raising $w(0)$.

Part ii of proposition 9 states that the interquantile wage differential, $w(p) - w(p')$, $p > p'$, must widen with any output-raising change in $\beta(\cdot)$. This is a bit stronger than the corresponding proposition 3. Under fixed proportions, the wage gap is widened between any two quantiles on either side of any improved tasks, but here the wage gap is widened between *all* quantiles, even if the improved tasks are located elsewhere in the hierarchy. Again, note that the incremental price of ability is $b(p)$, not $\beta(p)$. Part ii of proposition 9 implies that in the Cobb-Douglas case, if $\beta(p)$ rises over *any* interval, $b(p)$ rises over *all* intervals, by virtue of the optimizing reassignment. This widens $w(p) - w(p')$.

Comparative statics for the ability distribution (see the Appendix) are given by the following proposition.

Proposition 10. In the Cobb-Douglas case, any change in the distribution of (expected) ability on [0, 1] that raises (reduces) output (i) raises (reduces) $w(0)$, (ii) reduces (raises) $w(1)$, and (iii) narrows (widens) the wage span between any two workers of given abilities, $W(\mu) - W(\mu')$. A mean-preserving spread of the ability distribution raises output, so it raises $w(0)$, reduces $w(1)$, and narrows all $W(\mu) - W(\mu')$.

Proposition 10 shows that the fixed-proportions results on first-order stochastic improvements in the ability distribution (proposition 4) hold for the Cobb-Douglas case as well. Also, the output-enhancing effect of a mean-preserving spread in the ability distribution is not surprising, since proposition 5 obtained this result for fixed assignment; so a reoptimizing assignment only reinforces this.

What is different is that proposition 10 finds that a mean-preserving spread raises $w(0)$ and reduces $w(1)$, narrowing $w(1) - w(0)$ (and all other fixed-ability wage spans), *independent of the shape of* $\beta(\cdot)$, quite unlike the fixed-proportions case (proposition 6), where the shape of $\beta(\cdot)$ is critical. Specifically, under concave $\beta(\cdot)$, with fixed assignment, a mean-preserving spread of abilities reduces $w(0)$ and widens $w(1) - w(0)$, so the reoptimizing assignment reverses these results. The effect of information—and ability inequality more generally—is thus more assuredly egalitarian under Cobb-Douglas crowding technology than under fixed proportions.

### *E. CES Crowding Technology*

We have found opposite effects on the wage span from a mean-preserving spread in abilities under Cobb-Douglas and Leontief crowding technologies ($h = \varphi^{\eta}$ and $h \to \min[\varphi, 1]$) for concave $\beta(\cdot)$. This raises the question of what happens for other degrees of curvature of $h(\varphi)$. The question can be sharply posed by noting, from (25), that

$$w(1) - w(0) = \beta(1)h'[\varphi(1)]. \tag{28}$$

The behavior of the wage span turns entirely on what happens to $\varphi(1)$. In the Cobb-Douglas case, a more unequal ability distribution always raises $\varphi(1)$ and reduces $w(1) - w(0)$. What happens, and why, for less flexible crowding technologies?

A mean-preserving spread raises the ability of those who are near—but not at—the top of the distribution. This raises the cost of crowding these individuals on the tasks to which they are assigned, relative to the costs of crowding individuals in other tasks, both above and below them. How should this crowding be alleviated? Some shift in density down the hierarchy is obviously worth doing, since ability has declined in some of the lower quantiles, thereby diminishing the costs of crowding there. The question at hand is when there should also be some reallocation up the hierarchy, raising $\varphi(1)$.

Consider the constant elasticity of substitution (CES) crowding function, $h(\varphi) = [(1 - a) + a\varphi^{-r}]^{-1/r}$, $r \geq 1$.[23] As $r \to 0$ and $r \to \infty$, we have the Cobb-Douglas and Leontief cases already analyzed. In general, an explicit solution for $\varphi(1)$ and the wage span is not available, but some insights can be obtained, beginning with the special case of constant $\beta$ (see the Appendix).

PROPOSITION 11. *Constant $\beta$ and CES h.* (i) For any given distribution of $\mu$, there exists $r^* \in (0, 1)$ such that, for $r \in (-1, r^*)$, the wage span $w(1) - w(0)$ falls with any mean-preserving spread of $\mu$; (ii) as $r \to \infty$, the effect of any mean-preserving spread of $\mu$ on $w(1) - w(0)$ approaches zero from above.

Part ii of proposition 11 indicates that if the crowding effects are strong (substitution is limited), then the crowding of those near the

[23] The constants $a$ and $1 - a$ preserve the normalization $h(1) = 1$. Also, $\eta(1) = a$. The elasticity of substitution between labor on a task and a task-specific fixed factor is $1/(1 + r)$.

top of the hierarchy, whose ability has risen from a mean-preserving spread, is alleviated entirely by a shift down the hierarchy. This reduces $\varphi(1)$ and widens $w(1) - w(0)$. But if the crowding effects are weak (more substitution), then part i of proposition 11 indicates that the optimal response will also include reallocating workers from near the top to the very highest tasks, marginally raising the crowding of those top workers whose ability has not risen by much. This raises $\varphi(1)$ and reduces $w(1) - w(0)$.

This basic result, for constant $\beta$, is modified when $\beta(\cdot)$ slopes up. That is, in addition to the previous analysis of how to minimize the costs of crowding workers of varying abilities, we must now also consider the benefits of putting the higher-ability workers in the more ability-sensitive jobs. Intuitively, this would seem to tilt the optimal reallocation upward after a mean-preserving spread, compared to the constant-$\beta$ case, in order to exploit the rise in ability in the higher fractiles. Thus the net result on $\varphi(1)$ and the wage span depends not only on the severity of crowding but now also on the shape of $\beta(\cdot)$—its concavity or convexity—as we saw under fixed proportions.

To summarize this line of thinking, we should expect the following:[24] (i) If the cost of crowding is low, then a mean-preserving spread of ability leads to a rise in density at the very top jobs, narrowing the wage span, regardless of the shape of $\beta(\cdot)$, as we saw in the Cobb-Douglas case. The reason the shape of $\beta(\cdot)$ does not matter here is that since density shifts upward with the ability spread under constant $\beta$, it does so a fortiori when $\beta(\cdot)$ slopes up (regardless of its curvature) as just discussed. (ii) If the cost of crowding is high, then the shape of $\beta(\cdot)$ matters. For constant $\beta$, a mean-preserving spread of ability leads to a drop in density in the top jobs, widening the wage span; but this need not hold when $\beta(\cdot)$ slopes up. The condition under which more unequal ability will narrow the wage span depends on how steep $\beta(\cdot)$ is in higher-skilled tasks relative to the lower-skilled tasks, that is, how convex $\beta(\cdot)$ is. If $\beta(\cdot)$ is very steep over the high-skilled tasks, then it is very costly to reallocate workers down the hierarchy in order to alleviate the crowding on those fractiles in which ability has risen; it is better to reallocate at least some of the workers upward, raising density on the top jobs and reducing the wage span.

## VI. Conclusion

It is commonplace among labor economists to think of job ladders, where workers move up and down the hierarchy to fill needed slots.[25]

[24] While we offer no formal theorems in support of this intuition, it does accord with numerical explorations using a stepwise $\beta$ function. Details are available on request.

[25] For example, Bishop (1998, p. 6) writes that "most high and intermediate level jobs

Our model draws out the wage implications of this job structure, including the productivity gains/losses from such reassignments that are indirectly attributable to workers elsewhere in the hierarchy.

Our paper has analyzed the implications of such a job assignment model for the comparative static effects of changes in technology and the ability distribution on the distribution of wages. The results have been summarized in the Introduction and need not be repeated here. We would, however, invoke these results with some caution, given the limits of our assumptions, chief of which is that of exogenous ability. Under perfect observability, the model generalizes nicely, and we are able to show that investment in ability is efficient. However, if there is imperfect observability, investment is suboptimal and, more interesting, may in some cases become even more suboptimal when information becomes less imperfect. If so, we would have a situation in which more information has adverse effects on both efficiency and distribution, whereas under exogenous ability, information unambiguously enhances efficiency (proposition 5).

We conclude with some reflections on what insights our analysis might conceivably contribute to the discussion of recent decades' trends in the distribution of income. Some portion of the deteriorating real wage at the bottom of the distribution is ascribed to changes in the ability distribution, including the effects of immigration. Our model is certainly consistent with such findings (Sec. III*B*) and may offer an additional interpretation of the phenomenon: a bulge of unskilled workers depresses wages in that part of the spectrum by depressing the gains from promoting those immediately above oneself. By the same token, the model certainly suggests that improvements in the bottom of the educational distribution would ameliorate the income distribution, again through the logic of the assignment model. On the other hand, the model indicates that if such improvements were to be made at the expense of maintaining achievement at the highest levels of the ability distribution, then the benefits to the income distribution are by no means assured (propositions 6 and 10).

Technical change, of course, is widely understood to have driven much of the last two decades' distributional trends. Our model offers some potential insights into the mechanism by which this might have occurred. If productivity gains are concentrated in the least skilled jobs, flattening $\beta(\cdot)$, this would have an adverse effect on the bottom of the wage distribution. Has computerization, for example, had its greatest impact on clerical output and administrative support functions? Obvi-

are filled by people moving up from below …. This starts a chain of vacancies that may eventually generate an entry level opening for poorly educated workers who lack a history of steady employment."

ously, our model cannot answer these empirical questions, but it can help pose them and interpret the results.

## Appendix

*Proof That $w(r)$ Equals the Marginal Productivity of a Worker with Ability $\mu_r$*

From (12),

$$\frac{\partial Q}{\partial \Delta_r} = \int_0^r \beta(z)\mu'(z)dz - \int_0^1 \beta(z)\mu'(z)(1-z)dz + \int_0^1 \beta(z)\mu(z)dz$$
$$= \int_0^r \beta(z)\mu'(z)dz - \int_0^1 \beta(z)\frac{d}{dz}[\mu(z)(1-z)]dz.$$

Integrating by parts, we have

$$\frac{\partial Q}{\partial \Delta_r} = \beta(r)\mu_r - \int_0^r \mu(z)d\beta(z) + \int_0^1 \mu(z)(1-z)d\beta(z).$$

But

$$\int_0^1 \mu(z)(1-z)d\beta(z) = \int_0^1\left[\int_z^1 \mu(z)dy\right]d\beta(z) = \int_0^1\left[\int_0^y \mu(z)d\beta(z)\right]dy.$$

Thus, substituting, we get

$$\frac{\partial Q}{\partial \Delta_r} = \beta(r)\mu_r + \int_0^1\left[\int_0^y \mu(z)d\beta(z) - \int_0^r \mu(z)d\beta(z)\right]dy$$
$$= \beta(r)\mu_r + \int_0^1\int_r^y \mu(z)d\beta(z)dy = W(\mu_r) = w(r),$$

as was to be shown. Q.E.D.

*Proof of Proposition 2*

Let $W(\mu_r;\ \Delta_p)$ explicitly denote $W(\mu_r)$'s dependence on $\Delta_p$:

$$W(\mu_r;\ \Delta_p) = \beta(z(\mu_r;\ \Delta_p))\mu_r + \int_0^1\int_{z(\mu_r;\Delta_p)}^y \mu(z;\ \Delta_p)d\beta(z)dy.$$

We can immediately establish that

$$\frac{\partial}{\partial \Delta_p}\left(\frac{\partial Q}{\partial \Delta_r}\right) = \frac{\partial W(\mu_r;\ \Delta_p)}{\partial \Delta_p} = \int_0^1\int_r^y \frac{\partial \mu(z;\ \Delta_p)}{\partial \Delta_p}d\beta(z)dy.$$

Reversing the order of integration, we get

$$\frac{\partial}{\partial \Delta_p}\left(\frac{\partial Q}{\partial \Delta_r}\right) = -\int_0^r \int_0^z \frac{\partial \mu(z;\ \Delta_p)}{\partial \Delta_p} dy d\beta(z) + \int_r^1 \int_z^1 \frac{\partial \mu(z;\ \Delta_p)}{\partial \Delta_p} dy d\beta(z)$$

$$= -\int_0^r z \frac{\partial \mu(z;\ \Delta_p)}{\partial \Delta_p} d\beta(z) + \int_r^1 (1-z) \frac{\partial \mu(z;\ \Delta_p)}{\partial \Delta_p} d\beta(z)$$

$$= \left(\frac{1}{L}\right) \cdot \left[-\int_0^p z^2 \mu'(z) d\beta(z) + \int_p^r z(1-z)\mu'(z) d\beta(z) - \int_r^1 (1-z)^2 \mu'(z) d\beta(z)\right],$$

using (11), as was to be shown. Q.E.D.

*Proof of Theorem 1*

Consider the following version of the standard Bayesian model of statistical inference. Firms have common prior beliefs about ability $\alpha \in [0, 1]$ in the population from which they hire, given by probability density $\pi(\alpha)$. Firms observe a *common* signal $s \in [0, 1]$ for each worker. The information this signal provides about unobserved ability depends on the likelihood $L(s|\alpha)$, which gives the density of the signal, conditional on ability, for every ability level. Firms use Bayes' rule to formulate posterior beliefs about each worker's productivity.

In terms of statistical decision theory, the likelihood $L(\cdot|\cdot)$ defines an *experiment.* (Below we use $L$ to denote a likelihood function and the "experiment" or "test" that it characterizes.) A classic paper by Blackwell (1953) defines and usefully characterizes what it means for one experiment to be more informative than another. Intuitively, $L^1$ "is more informative than" $L^2$ (hereafter, $L^1$ MIT $L^2$) if the signals reported from $L^2$ are a "garbled" version of the signals reported from $L^1$—that is, if there exists a family of probability densities $b(s_2|s_1)$ such that, for all $\alpha$, $L^2(s_2|\alpha) = \int_0^1 b(s_2|s_1)L^1(s_1|\alpha)ds_1$. The term $b(s_2|s_1)$ shows how the signal $s_1$ from $L^1$ is randomly garbled into a reported signal $s_2$ from $L^2$. Blackwell proved that $L^1$ MIT $L^2$ in this sense if and only if every expected utility maximizer would prefer to observe the signal $s_1$ from experiment $L^1$ rather than the signal $s_2$ from experiment $L^2$, before making a decision, the payoff from which depends on the unknown state of the world. Theorem 1, based on Blackwell's utilitarian definition of MIT, can be proved using his result on garbling.

Let $s_i$ be an arbitrary outcome from test $L^i$, $i = 1, 2$; let $\alpha$ be unknown worker ability; let $\bar{\mu} = E[\alpha]$, the a priori mean ability; let $\mu_i(s_i) = E[\alpha|s_i]$, the conditional mean ability given score $s_i$ on test $L^i$; and, finally, let $\tilde{\mu}_i$ denote the random variable $\mu_i(s_i)$ (as envisioned before $L^1$ is given), the population distribution of estimated ability.

Obviously $E[\tilde{\mu}_1] = E[\tilde{\mu}_2] = \bar{\mu}$. It thus suffices to show that $E[v(\tilde{\mu}_1)] \geq E[v(\tilde{\mu}_2)]$ for all convex functions, $v(\cdot)$. By Blackwell's theorem, there exists a garbling of the signal from $L^1$ that generates the signal from $L^2$. That is, the distribution of the experimental outcome $s_2$ in $L^2$ can be regarded as having been generated by first running $L^1$ to get some outcome $s_1$ and then randomly reporting a possible outcome $s_2$, according to some probability density $b(s_2|s_1)$, which can be given independently of the true state of nature, $\alpha$.

Now, "invert the garble" as follows: for each $s_2$ under $L^2$, consider $b^{-1}(s_1|s_2)$, the density of $\tilde{s}_1$, the unknown outcome under $L^1$ from which $s_2$ was produced via the garbling.[26] Then, using this conditional probability distribution, we have that

$$\mu_2(s_2) \equiv E[\alpha|s_2] = E[E[\alpha|\tilde{s}_1]|s_2] = E[\mu_1(\tilde{s}_1)|s_2].$$

But now, Jensen's inequality and the law of iterated expectations imply

$$E[v(\tilde{\mu}_2)] = E[v(E[\mu_1(\tilde{s}_1)|s_2])] \leq E[E[v(\mu_1(\tilde{s}_1))|s_2]] = E[v(\mu_1(\tilde{s}_1))] = E[v(\tilde{\mu}_1)],$$

as was to be shown. Q.E.D.

*Proof of Proposition 5*

The hypothesis implies that, for all $\mu \geq 0$,

$$\int_0^\mu [G(z) - F(z)dz] \geq 0$$

and

$$\int_0^1 [G(z) - F(z)]dz = 0.$$

From this we conclude that $\int_\mu^1 [G(z) - F(z)]dz \leq 0$ for all $\mu \geq 0$. Integrate by parts to see that, for all $\mu'$, where $G(\mu') = F(\mu') \equiv \rho$ (i.e., at all points at which the distributions cross), it is the case that

$$\Gamma(\rho) \equiv \int_\rho^1 [G^{-1}(x) - F^{-1}(x)]dx = -\int_{\mu'}^1 [G(z) - F(z)]dz \geq 0.$$

Then, for $\Gamma(\cdot)$ so defined, we have $\Gamma(\rho) \geq 0$ whenever $d\Gamma/d\rho = 0$. Since $\Gamma(0) = \Gamma(1) = 0$, it must be that $\Gamma(\rho) \geq 0$ for all $\rho \in [0, 1]$. Integrate by parts, and recall that $\beta(\cdot)$ is nondecreasing, to get

$$Q(G) - Q(F) = \int_0^1 \int_p^1 [G^{-1}(x) - F^{-1}(x)]dxd\beta(p) = \int_0^1 \Gamma(p)d\beta(p) \geq 0.$$

Q.E.D.

[26] Specifically,

$$b^{-1}(s_1|s_2) = \frac{\int_0^1 b(s_2|s_1)L^1(s_1|\alpha)\pi(\alpha)d\alpha}{\int_0^1\int_0^1 b(s_2|\sigma)L^1(\sigma|\alpha)\pi(\alpha)d\alpha d\sigma}.$$

*Proof of Proposition 7*

Consider the wages earned by a given worker at some fixed quantile $p$ of the ability distribution for group $i \in \{a, b\}$, with and without integration. Using proposition 1, we know that, without integration, this worker earns

$$w_i(p) = \beta(p)\mu_i(p) + \int_0^1 \int_{z=p}^{y} \mu_i(z)d\beta(z)dy.$$

Under integration this same worker earns

$$w_\lambda(F_\lambda(\mu_i(p))) = \beta(F_\lambda(\mu_i(p)))\mu_i(p) + \int_0^1 \int_{z=F_\lambda(\mu_i(p))}^{y} \mu_\lambda(z)d\beta(z)dy.$$

Hence, the impact of integration on the mean wage earned by workers in group $i$ is given by

$$\begin{aligned}\int_0^1 [w_\lambda(F_\lambda(\mu_i(p))) - w_i(p)]dp &= \int_0^1 \mu_i(p) \cdot [\beta(F_\lambda(\mu_i(p))) - \beta(p)]dp \\ &\quad + \int_0^1 \int_0^1 \left[ \int_{z=F_\lambda(\mu_i(p))}^{y} \mu_\lambda(z)d\beta(z) \right. \\ &\quad \left. - \int_{z=p}^{y} \mu_i(z)d\beta(z) \right] dydp \\ &= \int_0^1 \int_{z=p}^{F_\lambda(\mu_i(p))} [\mu_i(p) - \mu_\lambda(z)]d\beta(z)dp \\ &\quad + \int_0^1 \int_0^1 \left\{ \int_{z=p}^{y} [\mu_\lambda(z) - \mu_i(z)]d\beta(z) \right\} dydp \\ &= \int_0^1 \int_{z=p}^{F_\lambda(\mu_i(p))} [\mu_i(p) - \mu_\lambda(z)]d\beta(z)dp \geq 0.\end{aligned}$$

To establish the inequality above, we have used the following facts: (i)

$$\int_{z=F_\lambda(\mu_i(p))}^{y} \mu_\lambda(z)d\beta(z) = \int_{z=p}^{y} \mu_\lambda(z)d\beta(z) - \int_{z=p}^{F_\lambda(\mu_i(p))} \mu_\lambda(z)d\beta(z);$$

(ii)

$$\mu_i(p) \cdot [\beta(F_\lambda(\mu_i(p))) - \beta(p)] = \int_{z=p}^{F_\lambda(\mu_i(p))} \mu_i(p)d\beta(z);$$

(iii) for any function $G: [0, 1]^2 \rightarrow R$ and satisfying $G(p, y) = -G(y, p)$ for all $(p, y) \in [0, 1]^2$, it must be that $\int_0^1\int_0^1 G(p, y)dydp = 0$ (in this case $G(p, y) \equiv \int_{z=p}^{y} [\mu_\lambda(z) - \mu_i(z)]d\beta(z)$); and (iv) $\mu_\lambda(z) \leq \mu_i(p)$ if and only if $z \leq F_\lambda(\mu_i(p))$.

We have thus demonstrated that the mean wage earned by the workers in group $i \in \{a, b\}$ does not fall after integration, as was to be shown. (Obviously, this result extends to the integration of any number of population groups.) Q.E.D.

*Solution 1 (to the Cobb-Douglas Case $h(\varphi) = \varphi^{\eta}$)*

Write the control problem for $T(p)$: $Q = \int_0^1 q(p, T, \tau)dp$, where $q(p, T, \tau) \equiv \beta(T)\mu(p)h(1/\tau)\tau$. The Euler equation, $\partial q/\partial T = d(\partial q/\partial \tau)/dp$, can be used to show

$$\frac{d}{dp}\ln[\mu(p)] = \left(\frac{\eta}{1-\eta}\right)\frac{d}{dp}\ln[\beta(T(p))] + \eta\frac{d}{dp}\ln[\tau(p)] + \left(\frac{\eta}{1-\eta}\right)\frac{d}{dp}\ln\left[\eta\left(\frac{1}{\tau(p)}\right)\right].$$

In the constant-$\eta$ case (i.e., Cobb-Douglas), the last term drops out, and we integrate to find

$$\frac{1}{\eta}\ln[\mu(p)] = \left(\frac{1}{1-\eta}\right)\ln[\beta(T(p))] + \ln[\tau(p)] + k_0,$$

where $k_0$ and $k_1$ below are constants of integration. Exponentiating gives

$$\mu(p)^{1/\eta} = k_1\tau(p)\beta[T(p)]^{1/(1-\eta)} = k_1\frac{d}{dp}\left[\int_0^{T(p)}\beta(y)^{1/(1-\eta)}dy\right].$$

Integrating above and using the facts that $T(0) = 0$ and $T(1) = 1$, we conclude that

$$\frac{\int_0^{T(p)}\beta(y)^{1/(1-\eta)}dy}{\int_0^1\beta(y)^{1/(1-\eta)}dy} = \frac{\int_0^p\mu(y)^{1/\eta}dy}{\int_0^1\mu(y)^{1/\eta}dy}\quad\forall p.$$

Differentiating both sides with respect to $p$ and noting that $b(p) = \beta[T(p)]\tau(p)^{1-\eta}$, we get

$$\mu(p)b(p) = \mu(p)^{1/\eta}\cdot\left[\frac{\int_0^1\beta(y)^{1/(1-\eta)}dy}{\int_0^1\mu(y)^{1/\eta}dy}\right]^{1-\eta}.$$

Substituting into $Q = \int_0^1\mu(p)b(p)dp$ and into (25) gives ($a$)

$$Q = \left[\int_0^1\beta(y)^{1/(1-\eta)}dy\right]^{1-\eta}\cdot\left[\int_0^1\mu(y)^{1/\eta}dy\right]^{\eta}$$

and ($b$)

$$w(p) = Q\cdot\left\{(1-\eta) + \eta\cdot\left[\frac{\mu(p)^{1/\eta}}{\int_0^1\mu(y)^{1/\eta}dy}\right]\right\}.$$

Proposition 9 follows immediately from inspection of points $a$ and $b$.

*Proof of Proposition 10*

Part i of proposition 10 follows immediately from point $b$ above, since $w(0) = (1-\eta)Q$. To show parts ii and iii of proposition 10, use points $a$ and $b$ above to write

$$w(1) = \left[\int_0^1 \beta(y)^{1/(1-\eta)}dy\right]^{1-\eta} \cdot \left\{(1-\eta)\cdot\left[\int_0^1 \mu(y)^{1/\eta}dy\right]^{\eta} + \eta\cdot\left[\int_0^1 \mu(y)^{1/\eta}dy\right]^{\eta-1}\right\}$$

and

$$W(\mu) - W(\mu') = \eta\cdot\left[\frac{\int_0^1 \beta(y)^{1/(1-\eta)}dy}{\int_0^1 \mu(y)^{1/\eta}dy}\right]^{1-\eta}\cdot(\mu^{1/\eta} - \mu'^{1/\eta}).$$

The effect of the ability distribution on $Q$ is conveyed directly by the term $\int_0^1 \mu(y)^{1/\eta}dy$. Since this term is less than one, $w(1)$ varies inversely with it, as does $W(\mu) - W(\mu')$. Thus parts ii and iii of proposition 10 immediately follow. Finally, since $\int_0^1 \mu(y)^{1/\eta}dy = \int_0^1 \mu^{1/\eta}dF(\mu)$ is the expected value of a convex function of $\mu$, a mean-preserving spread in the distribution of $\mu$ raises output, with the corresponding effects given in parts i–iii of proposition 10. Q.E.D.

*Solution 2 (to the CES Case, $h = [(1-a) + a\varphi^{-r}]^{-1/r}$, Constant $\beta$)*

The Euler equation is

$$\frac{\mu'}{\mu}(p) = \left[\frac{\eta(1)}{1-\eta(1)}\right]\frac{\beta'}{\beta}[T(p)]\tau(p)^{1+r} + (1+r)\left\{\frac{\eta(1)}{[1-\eta(1)]\tau(p)^{-r}+\eta(1)}\right\}\frac{\tau'}{\tau}(p).$$

For $\beta' \equiv 0$, integrate from $p$ to 1 to find

$$\ln[\mu(z)]\big|_p^1 = (1+r)\eta(1)\int_{z=p}^1 \frac{d\tau(z)}{[1-\eta(1)]\tau(z)^{1-r}+\eta(1)\tau(z)}$$

$$= \frac{1+r}{r}\cdot\ln\{\eta(1)\tau(z)^r + [1-\eta(1)]\}\big|_p^1.$$

Exponentiating and rearranging gives ($c$)

$$\tau(p) = \left\{\frac{k\cdot\mu(p)^{r/(1+r)} - [1-\eta(1)]}{\eta(1)}\right\}^{1/r},$$

where $k \equiv \eta(1)\tau(1)^r + [1-\eta(1)]$ is chosen to satisfy $\int_0^1 \tau(p)dp = 1$. For $r \in (-1, 0]$, point $c$ holds for all $p \in [0, 1]$ (the term in braces is nonnegative, since $k \geq 1-\eta(1)$ and $\mu(p) \leq 1$). For $r \in (0, \infty)$, point $c$ holds for all $p \in [p_{\min}, 1]$ and $\tau(p) = 0$ for $p \in [0, p_{\min}]$, where $k\cdot\mu(p_{\min})^{r/(1+r)} = 1-\eta(1)$. Note that $t = T(p) = 0$ for $p \in [0, p_{\min}]$, so $\Phi(0) = p_{\min}$. That is, the cost of crowding is sufficiently pronounced here that a mass of workers is dumped on task 0, rather than crowd higher-ability workers on their tasks.

*Proof of Proposition 11*

From point $c$, consider $\tau$ as a function of $\mu$ and $\tau(1)$. Note that $\tau$ is increasing in $\tau(1)$, for all $r$, and so is $\int_0^1 \tau dp$. Thus, if a mean-preserving spread of $\mu$ raises (lowers) $\int_0^1 \tau dp$, $\tau(1)$ must fall (rise) to restore $\int_0^1 \tau dp = 1$, and $w(1) - w(0)$ must fall (rise). For $r \in (-1, 0]$, it can be readily verified that $\tau$ is convex in $\mu$. Therefore, $\int_0^1 \tau dp$ is the expected value of a convex function of $\mu$, so it rises with a mean-preserving spread in $\mu$, and $w(1) - w(0)$ must fall. For $r > 0$, $\tau$ is convex in $\mu$ on $[0, 1]$ iff it is convex as $\mu \rightarrow 1$. This is true as $r \rightarrow 0$, but not as $r \rightarrow 1$. Hence for any given distribution of $\mu$ (and therefore any given $k$), there exists $r^* \in (0, 1)$ such that $\tau$ is convex in $\mu$ for $r \leq r^*$, and part i of proposition 11 follows. For $r \geq 1$, $\tau$ has a convex kink at $\mu(p_{\min})$ but is concave for $\mu > \mu(p_{\min})$. As $r \rightarrow \infty$, $p_{\min} \rightarrow 0$. That is, $\tau(\mu(p))$ converges on unity from a concave function that emanates from a point approaching the origin, so any mean-preserving spread of $\mu$ reduces $\int_0^1 \tau dp$ (infinitesimally) and widens $w(1) - w(0)$. Q.E.D.

## References


Bishop, John H. 1998. "Is Welfare Reform Succeeding?" Working Paper no. 98-15. Ithaca, N.Y.: Cornell Univ., School Indus. and Labor Relations.

Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *Ann. Math. Statis.* 24 (June): 265–72.

Cornell, Bradford, and Ivo Welch. 1996. "Culture, Information, and Screening Discrimination." *J.P.E.* 104 (June): 542–71.

Gould, Eric D. 2002. "Rising Wage Inequality, Comparative Advantage, and the Growing Importance of General Skills in the United States." *J. Labor Econ.* 20 (January): 105–47.

Hartigan, John A., and Alexandra K. Wigdor, eds. 1989. *Fairness in Employment Testing: Validity, Generalization, Minority Issues, and the General Aptitude Test Battery.* Washington, D.C.: Nat. Acad. Press.

Heckman, James J., and Bo E. Honore. 1990. "The Empirical Content of the Roy Model." *Econometrica* 58 (September): 1121–49.

Heckman, James J., and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *J.P.E.* 93 (December): 1077–1125.

Herrnstein, Richard J., and Charles Murray. 1994. *The Bell Curve: Intelligence and the Class Structure in American Life.* New York: Free Press.

Kremer, Michael. 1993. "The O-Ring Theory of Economic Development." *Q.J.E.* 108 (August): 551–75.

MacDonald, Glenn M. 1982. "Information in Production." *Econometrica* 50 (September): 1143–62.

Mayer, Thomas. 1960. "The Distribution of Ability and Earnings." *Rev. Econ. and Statis.* 42 (May): 189–95.

Rosen, Sherwin. 1982. "Authority, Control, and the Distribution of Earnings." *Bell J. Econ.* 13 (Autumn): 311–23.

Rothschild, Michael, and Joseph E. Stiglitz. 1982. "A Model of Employment Outcomes Illustrating the Effect of the Structure of Information on the Level and Distribution of Income." *Econ. Letters* 10 (3–4): 231–36.

Roy, Andrew D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Econ. Papers* 3 (June): 135–46.

Sattinger, Michael. 1975. "Comparative Advantage and the Distributions of Earnings and Abilities." *Econometrica* 43 (May): 455–68.

———. 1979. "Differential Rents and the Distribution of Earnings." *Oxford Econ. Papers* 31 (March): 60–71.

———. 1980. *Capital and the Distribution of Labor Earnings.* Amsterdam: North-Holland.

———. 1993. "Assignment Models of the Distribution of Earnings." *J. Econ. Literature* 31 (June): 831–80.

Teulings, Coen N. 1995. "The Wage Distribution in a Model of the Assignment of Skills to Jobs." *J.P.E.* 103 (April): 280–315.

Tinbergen, Jan. 1951. "Some Remarks on the Distribution of Labour Incomes." In *International Economic Papers*, no. 1, edited by Alan T. Peacock et al. London: Macmillan (for Internat. Econ. Assoc.).