# THEORIES OF THE DISTRIBUTION OF LABOR EARNINGS

Derek Neal
Sherwin Rosen

Working Paper **6378**

NBER WORKING PAPER SERIES

# THEORIES OF THE DISTRIBUTION OF LABOR EARNINGS

Derek Neal
Sherwin Rosen

Working Paper 6378
http://www.nber.org/papers/w6378

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1998

This paper is a preliminary version of a survey prepared for the Handbook of Income Distribution. We wish to thank participants from the Handbook of Income Distribution Conference in Florence, Italy, March, 1997. We owe special thanks to François Bourguignon, Jim Davies, Robert Gibbons, Lars Hansen, Lawrence Katz, and Casey Mulligan. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Theories of the Distribution of Labor Earnings
Derek Neal and Sherwin Rosen
NBER Working Paper No. 6378
January 1998
JEL Nos. J2, J3

**ABSTRACT**

Several empirical regularities motivate most theories of the distribution of labor earnings. Earnings distributions tend to be skewed to the right and display a long right tail. Further, they are leptokurtic (positive fourth cumulant) and have a fat tail. Mean earnings always exceed median earnings and the top percentiles of earners account for quite a disproportionate share of total earnings. Mean earnings also differ greatly across groups defined by occupation, education, experience, and other observed traits. With respect to the evolution of the distribution of earnings for a given cohort, initial earnings dispersion is smaller than the dispersion observed in prime working years.

We explore several classes of models that address these stylized facts. Stochastic theories begin with distributional assumptions about worker endowments and then examine the stochastic structures that might generate observed features of the aggregate distribution of earnings. Selection models describe how workers allocate their skills to tasks. Because workers choose their best option from a menu of careers, these allocation decisions generate earnings distributions which tend to be skewed. Sorting models provide dynamic versions of selection models and illustrate how gradual learning about endowments leads to sorting patterns that amplify dispersion and generate a skewed distribution of earnings within a cohort of experienced workers.

Human capital theory demonstrates that earnings dispersion is a prerequisite for significant skill investments. Without earnings dispersion, workers would not willingly make the investments necessary for high-skill jobs. Human capital models illustrate how endowments of wealth and talent influence the investment decisions that generate observed distributions of earnings.

Finally, agency models illustrate how wage structures may determine rather than reflect productivity. Tournament theory addresses the long right tails of wage distributions within firms. Efficiency wage models address differences in mean wages across employments that involve different monitoring technologies.

Derek Neal
Department of Economics
University of Chicago
1126 East 59th St.
Chicago, IL 60637
and NBER
derek-neal@uchicago.edu

Sherwin Rosen
Department of Economics
University of Chicago
1126 East 59th St.
Chicago, IL 60637
and NBER
rosen@cicero.spc.uchicago.edu

## Introduction

Theories of earnings distributions have always had strong empirical motivations. The earliest empirical studies of wealth and income discovered a remarkable regularity that is found in all observed earnings distributions in large populations. Earnings distributions (and income distributions more generally) are always skewed to the right. Their density functions are asymmetric and display a long right tail and positive skewness measure (third moment about the mean). They are also leptokurtic (positive fourth moment about the mean) and have a "fat tail." Put differently, mean earnings always exceed median earnings, and the top percentiles of earners account for a disproportionate share of total earnings.

Pareto (1897) and Bowley (1915) pioneered the assembly and empirical analysis of data on personal incomes. Extensive development of microeconomic data sources since WW II led to a more comprehensive development of the subject. Much work attempted to formulate the economic and statistical basis for fitting specific functional forms to empirical distributions. Were earnings capacities log-normally distributed, Paretian, or something else (Staele, 1943, Miller, 1955, Lebergott, 1959)?

Various theories of the distribution of earnings address not only the fat tail of earnings distributions but also several other empirical observations. Earnings differ greatly across groups of workers defined by occupation, education, experience, and other observed traits. Earnings also vary within groups that are observationally similar. Further, earnings dispersion for a particular cohort of workers is greater among experienced workers than workers who are beginning their careers. In a given cohort, much of the eventual inequality in lifetime earnings is not apparent until the workers are well into their careers.

Below, we describe four types of models that offer explanations for some or all of the observations. Stochastic theories begin with distributional assumptions about worker endowments and examine what kinds of stochastic structures might be consistent with the observed aggregate distributions of earnings. These are among the first applications of the theory of stochastic processes in economics. Selection models also begin with assumptions concerning the distribution of endowments, but the decisions of workers concerning the allocation of skills to tasks generates a distribution of earnings. Sorting models describe how workers change these allocations as they learn more their endowments. This sorting process contributes to the evolution of the distribution of earnings for a particular cohort. Human capital theory describes how workers acquire skills. It demonstrates that earnings inequality is a necessity in an economy where some activities require more skill investment than others. These models also illustrate how family resources and natural talents affect the skill investments that generate observed earnings inequality. Finally, agency theory approaches earnings inequality from an entirely opposite direction. Instead of describing how worker decisions and endowments generate a distribution of individual productivities that gives rise to a distribution of earnings, agency models describe how firms choose a distribution of earnings in order to elicit desired levels of individual productivity. Highly skewed wage distributions within firms may be an incentive device.

## 1 The Characteristics Skew of Earnings Distributions

In the physical and biological worlds, distributions tend to be symmetric and bell-shaped. But in the social world, size distributions tend to be asymmetric and right- skewed. Thus, we begin with stochastic process theories and other models that focus specifically on generating earnings distributions with long right tails. We then turn to models that address skewness while generating richer behavioral implications.

From the central limit theorem we know that normal distributions arise from sums of many independent and identically distributed components, irrespective of how the components themselves are distributed. To apply that argument to earnings and other long-tailed distributions in economics, it is necessary to hypothesize that a person's earning capacity is generated as a product of independent increments, as in

$$y = \prod \epsilon_i, \tag{1.1}$$

where the $\epsilon_i$'s reflect the myriad factors that determine personal productivity (Roy, 1950). With enough independent $\epsilon$-components, log(y) tends to the Normal distribution whatever the distribution of the $\epsilon_i$'s (Atchinson and Brown, 1957). The distribution of the natural units (the antilog) has a long right tail. However, observed earnings distributions tend to have tails that are thicker and longer than the log normal. For instance, the Pareto distribution has "infinite" second, third, and higher order moments, and fits the upper tail of earnings distributions quite well.

### 1a Stochastic Process Theories

The stochastic process approach explicitly addressed the "long-tail problem" (see especially the accounts of Steindl, 1968, Lydall, 1968 and Pen, 1971). This literature built up a "generating process" for the overall distribution from more elementary micro components. However, the components themselves have no behavioral content. The basic idea is related to the binomial approximation to the normal distribution: a log normal distribution is generated by a random walk process in which the percentage change in a person's earnings is distributed independently of earnings itself. But since the earnings level is the product of random variables realized up to that time, such a process implies ever growing variance in the overall distribution of log earnings. Some stabilizing force that offsets the tendency for increasing variance is necessary for the process to converge to a stationary steady-state distribution.

Various economists devised different auxiliary stabilizing hypotheses to solve this problem. (Kalecki, 1945; Rutherford, 1955; Simon, 1955, Wold and Whittle, 1957). Champernowne (1953, 1973) established the basic method. He replaced the random walk with a Markov process where workers face identical, fixed transition probabilities of moving between exogenously defined (log) earnings classes. Markov chains also have the variance-increasing property unless the transition probabilities are appropriately restricted. Champernowne's theory controls the variance of the diffusion by restricting the average transition to be downward, toward lower incomes. Intuitively, the idea of intergenerational turnover of workers justified this restriction (such turnover is explicit in Rutherford, 1955): higher earning older workers drop out of the process at retirement and are replaced, in a sense, with lower earning new entrants. It is slightly astonishing that a simple specific form of Champernowne's restriction implies Pareto's law as the stationary distribution.

Singh and Maddala (1976) pursued a related hazard function approach. If F(x) is the cumulative probability distribution of random variable x and f(x) is its density, 1-F(x) is called the survival function, and $h(x) = f(x)/[1 - F(x)]$ is its hazard (failure) rate. A decreasing hazard seems necessary to produce the kind of probability masses observed in the upper tails of earnings distributions. For example, the normal has a hazard that is strictly increasing in x, but the hazard of the Pareto distribution is strictly decreasing (it is $h(x) = \alpha/x$). The hazard of the log normal is not monotone. It first increases and then decreases in x. Singh and Maddala produce a new class of distributions by specifying that the proportional hazard is logistic, or alternatively that the hazard itself follows a $\text{sech}^2$ law. These new distributions closely fit the overall distribution of family incomes.

1b Correlated Increments and Common Factors

Another way to account for the extra concentration of mass in the tail (relative to the log normal) is to allow the $\epsilon_i$ components in equation (1.1) to be positively correlated. The law of large numbers no longer applies, but it is intuitively clear why a multiplicative structure in (1.1) and positive correlation leads to a long right tail in the distribution of y. Think of (1.1) as a "production function" for earnings capacity, with the $\epsilon$'s regarded as "inputs." Then a person who has $100\lambda$ percent more of each component input earns $(1+\lambda)^n$ times more money. The effect of $\lambda$ is magnified many times over when n is large.

For instance, consider a statistical factor-analytic structure with a single factor $\lambda$, interpreted as generalized ability or "IQ." Suppose personal endowments of the specific $\epsilon$-inputs follow $\epsilon_i = k_i(1+\lambda)$, where $k_i$ are independently distributed. The common factor $\lambda$ induces positive correlation between $\epsilon_i$ and $\epsilon_k$. Then (1.1) implies that personal productivity is $y = K(1+\lambda)^n$, with $K = \prod k_i$. Even if $\lambda$ is symmetric and normally distributed, the distribution of $K(1+\lambda)^n$ has a long right tail, and the highest earners account for a disproportionate share of total earnings.

An example brings the point close to home. This kind of model accounts for the highly skewed distribution of journal publications among members of an academic profession. Remarkably few scholars produce the lion's share of academic publications: the median number of publications in economics is one, while the mean is much larger (Lovell, 1973). The distribution of professional citations is even more concentrated (Leamer, 1981). Interpreting each $\epsilon_i$ as the probability of carrying through on the large number of steps needed to bring a research project to successful publication, scholars who are slightly better in each dimension are many times more likely to succeed overall.

When the $\epsilon_i$'s are interpreted as factors of production rather than as probabilities, it is natural to expect that some inputs are more important than others. A weighted specification is preferred, such as $y=\prod(\epsilon_i)^{\alpha_i}$, where the $\alpha_i$'s are fixed weights that affect marginal products. So long as $\sum\alpha_i > 1$, positive correlation among the $\epsilon_i$'s produces a long right tail in the y-distribution. A small common increment is magnified overall. However, if $\sum\alpha_i \leq 1$ there is no magnification and skew cannot arise in this way. There is a sense in which a form of increasing returns to scale produces the characteristic skew of earnings distributions.

### 1c Scale-of-Operations Effects and Superstars

There are attempts to study the explicit determinants of high earnings in terms of concepts such as power and span-of-control (Simon, 1957; Lydall, 1959) that are closely related to scale economies. What is at issue is a person's "scale of operations" (Mayer, 1960, Tuck, 1954, Reder, 1968). Most personal economic activity occurs on a small scale, but in some activities the size of a person's market can be noticeable, even at the scale of the economy. The term "scale of operations" comes from the tendency for extremely wealthy individuals, such as founders of large companies, to control vast amounts of capital and wealth. Scale economies account for this. For if person A can invest a dollar more efficiently than person B, person A can likely invest $1 billion better than person B and will end up controlling resources on a much larger scale. A small extra talent for control will command a large rent in the market equilibrium.

The executive labor market is a leading practical example of this effect. Control aspects of firms and management decisions contain "local public goods" elements because they affect the ways in which all resources within the organization are directed. Top level managerial decisions, such as what products to produce and how to produce and market them, affect the productivity of all resources employed by the firm. Just as a good soldier is not effective if fighting the wrong war, a talented and highly motivated worker is not socially productive if the organization is producing a good no one wants to purchase. Better decisions of this kind are worth a great deal when applied to a large organization. A person with a small edge in talent can have a very large marginal product.

The relevant theory describes how managerial talents are allocated to control functions in a market economy (Tuck, 1954; Rosen, 1982), e.g., how the executive labor market allocates talent. In essence, workers and the owners of the capital bid for the privilege of being managed by talented people. Thus, rents from superior capacity to make superior control decisions get transferred to managers. In equilibrium, more talented managers control larger or more valuable organizations. It is the principle reason why top level managers and executives in large companies earn such high salaries (Rosen, 1992) and are typically found in the top percentiles of the earnings distribution.

The scale of operations effect extends to a broader array of services that have local public goods qualities for other reasons. In many transactions, the marginal costs of serving another customer are small. Though different customers are served simultaneously, access is rationed by price. The services of intellectual property often have these properties, especially when rendered through mass media. Use of the service by one person does not greatly diminish availability others. An author who must personally tell his story to each person individually has much smaller chances of earning a high income than one who can, in effect, clone himself to many customers at a time by writing the story down once-and-for all and duplicating it, at negligible cost, to others as a book, cassette or compact disc. Similar considerations apply to entertainers, actors, television personalities and newscasters, musicians and composers, artists, intellectuals, software developers, and athletes (Rosen, 1981). Typically the chances of success in these endeavors are rather small, but the rewards of the successful can be enormous because relatively few people are needed to satisfy the entire market. The small chances of great success attract many entrants, and the "stars" who are successful ex post appear to dominate their fields. They earn enough to put them well into the upper tail of the earnings distribution (Frank and Cook, 1995 present many examples).

However, the upper tail is hardly dominated by stars, superstars and chief executive officers, because scale economies in the services they render means that few people are required to produce them. They simply are not numerous enough to make much of a dent in the upper tail, even in a large economy such as U.S., where the incomes and activities of such people receive so much attention and publicity. Highly successful professional practitioners in law, medicine and other professions occupy the upper tail much more frequently. Here, technology sharply constrains one's personal volume of transactions. Doctors can see a relatively small number of patients per day, and lawyers can manage only a few cases at a time. Instead, the scale-of-operations effect works through prices. The most capable lawyers tend to be assigned to the largest claims, where a slight edge in talent can have huge financial effects on the value of resources in dispute (Spurr, 1987). Large claims bid up the fees of the most successful lawyers. The wealthiest people in need of medical services outbid others for access to the most capable physicians.[1]

## 2 Selection Theory

In 1951, Roy sketched the basic structure of selection, but economists did not widely recognize the importance of his work until the 1970's. Tinbergen's (1959) version is more elaborate (and more obscure)[2], so we follow Roy (1951) here. The microeconomic content of "the Roy model," as it is sometimes called, is minimal. It is based on a choice problem of the simplest kind. Its analytic interest lies in its aggregate market implications. By describing how different individuals sort across job categories in the economy at large, the model enhances our understanding of the nature, determinants, and distribution of economic rents in labor markets. The model also illustrates how selection may generate asymmetric densities for earnings distributions even when the distribution of

[1]Alfred Marshall (1947) briefly introduced this point 100 years ago.

[2]Specifically, Tinbergen specifies that tastes for different kinds of work follow a specific form that has never been pursued in the literature. Tastes have been treated in the literature on equalizing differences, but not specifically in the occupational or job choice literature.

potential earnings in each job is symmetric.

## 2a The Roy Model

Consider an economy with n different types of jobs. Earnings are worker specific because worker productivity varies from job to job, and some workers are better at some jobs than others. Assume that supply and demand are equal in all job markets and examine the self-selection of workers to jobs in market equilibrium. Let $y_{ij}$ be the earnings capacity of worker i on job j. It is what worker i could earn if job j was chosen. Let $y_i$ be the earnings observed for worker i. Maintaining the economic hypothesis that self-interested workers choose jobs that maximize their earnings, observed earnings for worker i are

$$y_i = \max(y_{i1}, y_{i2}, \ldots, y_{in}) \tag{2.1}$$

The model is completed by specifying a joint probability distribution $f(y_{i1}, y_{i2}, \ldots, y_{in})$ for earnings prospects in the population at large. Roy (1951) assumed that f(y) is jointly normal. Mandelbrot (1962) assumed that f(.) follows the Pareto-Levy form. The observed earnings distribution, $g(y_i)$, is the transformation of f(y) implied by (2.1).

This simple model yields surprisingly complicated outcomes. The observed distribution g(y) is a mixture of conditional distributions resulting from maximization. Note that if n is large, (2.1) suggests the extreme value distribution (the distribution of the first order-statistic) as a possible stationary distribution. The extreme value theorem, like the central limit theorem, proves that the first order-statistic of independent and identically distributed parents tends to a unique limiting distribution, independent of how the parents are distributed. The stationary distribution is a double exponential. It is skewed and leptokurtic, characteristic of the upper tail of earnings distributions.

Though this idea is insightful, the extreme-value theorem cannot be applied to this problem because the component distributions of earning potential $y_{ij}$ in each job are not identically distributed (they have different means and variances) and may not be independent. It proves necessary to work through all the conditioning information to derive the observed distribution g(y). Mandelbrot (1962) shows that g(y) follows the Pareto-Levy distribution if f(.) is Pareto-Levy. In fact, this is the only case for which g(y) is in the same family as its parent f(.). Otherwise the conditional distributions which compose it are truncated, and g(y) is a complicated mixture that is not easily described in closed form.

The model is best illustrated when there are two jobs (n = 2). Suppose $f(y_1,y_2)$ is normal (or log-normal), with means $\mu_1$ and $\mu_2$, variances $\sigma_1^2$ and $\sigma_2^2$ and covariance $\sigma_{12}$. The correlation coefficient between earning capacities $y_1$ and $y_2$ is $\rho = \sigma_{12}/\sigma_1\sigma_2$. The assignment of workers to jobs is depicted in Figure 2.1. The probability contours depicted assume $\mu_1 > \mu_2$, $\sigma_1 > \sigma_2$ and $\rho > 0$. Each person in the population is a point in the $(y_1,y_2)$ plane. The probability mass there is proportional to the number of people in the population with those prospects. All workers along the 45° line ($y_1 = y_2$) find the two jobs equally attractive. Earnings in job 1 exceed those in job 2 for all people above the line, they choose job 1 and $y_i = y_{1i}$. Similarly, all people in the region below the 45° line maximize earnings by choosing job 2, so $y_i = y_{2i}$ for them. Therefore the observed distribution g(y) is the weighted sum of two truncated normals, one above the 45° line and the other below it. Maddala (1977) and Heckman and Honore (1990) present complete details of this model.

Economic rents enter this model both in the sense of relative talents ("producer" surplus or comparative advantage) and in the sense of absolute talents ("ability" rents and absolute advantage). Transforming a person's endowments to polar coordinates, the former is measured by the deviation of the angle of the vector ($y_{i1}$,$y_{i2}$) from the 45° line, and the latter by the length of the vector itself. Workers with prospects along the equal earnings line represent the extensive margin between job types. They receive zero surplus from their job choice. All others are inframarginal. Workers furthest from the 45° line (in either direction) receive the largest surplus from their choices, whether or not their earnings are large or small. The length of the ($y_{i1}$,$y_{i2}$) vector determines interpersonal earnings levels.

With a little stretch, the variances and covariance in the probability contours can be given an economic interpretation. For instance $\sigma_1 > \sigma_2$, implies a sense in which job 1 is more difficult than job 2. In this case, the outcome in job 2 is less dependent on who does the work. Pretty much the same outcome results no matter who does it. There is a greater range of possibilities in job 1. It offers more chances for talent to stand out. $\rho > 0$ means that a person whose earnings prospects are larger in job 1 is likely to have better than average prospects in job 2. For example, a one-factor structure of the type discussed above produces positive correlation between potential earnings in the two sectors and implies hierarchical sorting. In Figure 2.1 people with the highest earning prospects in both jobs tend to work in job 1. Those with lesser overall prospects tend to work in job 2. Notice that the relative variance condition is crucial to the order of the hierarchy. If $\sigma_1 < \sigma_2$ and $\rho > 0$, the probability contours cut the 45° line from above, and talented people are more often found in job 2.

The sign of the correlation coefficient is crucial to whether or not sorting tends to be hierarchical. When $\rho < 0$, a person who is good at one job is likely to be below average in the other. The ellipses in Figure 2.1 would be negatively sloped, and there would be positive selection in both jobs. Sorting is not hierarchical in that case, because absolute advantage does not dominate the determination of economic ability. Rather, people are sorted more by "comparative advantage." Sorting by comparative advantage is empirically supported for choice between high school and college education in some U.S. data (Willis and Rosen, 1979). People choosing to stop school upon high school graduation are more likely to be attracted to blue collar jobs and those continuing through college are more likely to be found in white collar jobs. Willis and Rosen provide evidence that, on average, those who choose blue collar jobs are actually more productive in those jobs than more educated white collar workers would have been had they chosen blue collar work. This result is plausible to the extent white and blue-collar jobs use much different kinds of talent and to the extent different kinds of talent are negatively correlated. In such circumstances, the relevant concept of ability for earnings is clearly multidimensional, and single factor measures like IQ scores may be quite misleading.

Computational complexities of estimating selection models rise quickly in n, so these models have only been used for broad classification of job types. However, Garen's (1984), who provides an approximation for selection bias in an educational choice model with a continuum of education levels, provides support for the multi-factor interpretation of comparative advantage sorting. Heckman and Sedlacek (1985) expand the typical bivariate discrete econometric model to more alternatives, and recent applications of their methods reveal increasingly positive covariance in the earnings potential of individuals across occupations in the U.S. data for the 1980's and 90's (Gould, 1996), a period when earnings inequality in the U.S. increased markedly. Perhaps sorting by comparative advantage has become less important and sorting by absolute advantage more important over time. For instance, there is an alleged connection between increasing wage inequality and the information-processing content and computer intensity of work (Krueger, 1993).

Another possibility is that complementaries among qualities of workers have become more important for job assignments lately (Kremer, 1993; 1996). Worker complementarities are present when increasing stratification of worker types into more homogeneous groups raises their total productivity. Such an effect might arise from in-divisibilities of work itself at the work-site. Increasing use of academic ability and other types of testing that sort people to schools tend to be inherently hierarchical and appear much more important now than in the past. However, much research remains to be done in this under-developed area.

## 2b Ability Sorting and Stratification

Figure 2.1 reveals that workers with similar earnings prospects choose the same type of job. When the number of job types n is large, it is possible to examine interesting details of the sorting and stratification of choices by thinking of talents in terms of a relatively small number of worker attributes that determine specific job productivity. Assume

$$y_{ij} = w_j \pi_{ij} \tag{2.2}$$

where $\pi_{ij}$ is the volume of the ith worker's output in the jth job, and $w_j$ is the price per unit of worker output in job j (think of $w_j$ as a piece rate). The choice criteria remains as in (2.1), so all pair-wise comparisons of $w_j/w_k$ with $\pi_{ik}/\pi_{ij}$ reveal a person's comparative advantage (Sattinger, 1975, 1993; Rosen, 1978). Furthermore, write

$$\pi_{ij} = \alpha_{1j}C_{1i} + \alpha_{2j}C_{2i} + \ldots + \alpha_{Kj}C_{Ki} \tag{2.3}$$

where the $C_{ki}$'s are individual worker characteristics such as mathematical, motor, and verbal skills, etc., and the

$\alpha_{kj}$'s are the marginal products of those traits for the specific job. These marginal products generally vary from job to job. For instance, verbal skills are of greater value to a lawyer than to a carpenter, but manual dexterity is of greater value to a carpenter than to a lawyer. Combining (2.2) and (2.3),

$$y_{ij} = \beta_{1j}C_{1i} + \beta_{2j}C_{2i} + .... + \beta_{vj}C_{vi} \tag{2.4}$$

where $\beta_{kj} = w_j\alpha_{kj}$ is the value of the marginal product of characteristic k in job j. Idiosyncratic random terms could be added to (2.2) - (2.3), but we ignore them here.

Applying the choice criteria in (2.2) to (2.4) reveals how people with similar C-endowments tend to choose the same type of job. Figure 2.2 illustrates the equilibrium stratification for 5 job types and two characteristics ($n = 5$, $v = 2$). Given job j, all people whose characteristics lie along the downward sloping line $C_1 = Y^*/\beta_{1j} - (\beta_{2j}/\beta_{1j})C_2$ would earn the same amount of money Y* if they chose to work on it. Draw such a line in the $(C_1,C_2)$ plane for every job type and consider the envelope of the entire family of lines. This is the negatively inclined piecewise linear curve in the figure, after ordering job numbers by $\beta_{2j}/\beta_{1j}$. Rays from the origin through the corners of the envelope define regions where workers endowed with such traits maximize their incomes by choosing jobs 1, 2,..., or 5. These partitions are completely deterministic when there is no randomness in (2.3). If specific random errors (noises) are present in (2.3), the partitions in Figure 2.2 represent assignments in expected value. The partitions of the space of characteristics in Figure 2.2 are open cones when the relations in (2.3) are homogeneous. However, if (2.3) has constant terms, the partitions can be closed and are much more complicated to describe.

Workers tend to choose jobs that highly value the personal characteristics they possess in relative abundance. The slope of each linear segment of the envelope in Figure 2.2 is the ratio of marginal products $\alpha_{2j}/\alpha_{1j}$ in (1.3). As illustrated, characteristic $C_1$ is highly productive in job 1 and $C_2$ is highly productive in job 5. Workers with relatively greater endowments of $C_1$ tend to choose job 1 and those with relatively greater endowments of $C_2$ tend to choose job 5. For instance, if $C_1$ is verbal skill and $C_2$ is manual dexterity, job 1 might be lawyer and job 5 might be carpenter. Both characteristics are equally important for productivity in job 3 and it tends to attract people with more balanced endowments. In this model, large earnings require large personal endowments of the C''s. Since it is much less probable to have large endowments of many characteristics than of only a few, high earners would tend to be more frequently found in occupations such as 1 and 5, which weigh only one characteristic strongly (Mandelbrot, 1962).

Prices, $w_j$, also affect the size of the partitions in Figure 2.2. For instance, an increase in $w_j$ reduces the intercept of the line that defines the facet and expands the width of the cone for which occupation j is optimal. Some persons otherwise would have chosen neighboring occupations $j\pm1$ now choose occupation j. The supply price of workers to jobs is always rising.

## 2c Factor Price Equalization?

The selection model contrasts with the application to labor markets of the Lancasterian combinable characteristics (hedonic) approach to product differentiation. Both place economic structure on a statistical factor-analytic model of personal earnings. In the hedonic approach, each worker supplies certain factors of production to employers, like strength and intelligence, etc.. Equating demand and supply of each factor in an implicit aggregate factor market establishes market equilibrium prices and fixes the factor-loadings on the statistical structure. A worker's labor income is a fixed weighted sum of personal factor endowments (plus noise). If the number of factors is relatively small, the earnings distribution is compacted into space of equally small dimensions. In effect, there are as many basic types of workers as underlying factors, and the earnings of every worker can be expressed as a linear combination of basic types (Welch, 1969)

This method rests on the crucial assumption that the total amounts of worker characteristics employed by a firm or sector of the economy are combinable and serve as its factors of production, independent of how the totals are obtained. It is a multivariate generalization of the old "efficiency units" assumption; e.g., a firm hiring two workers with half as many characteristics is the same as hiring one worker with twice as many. In essence, workers can unbundle their skills. Workers are allowed to allocate one type of skill to a given task while simultaneously allocating other skills to different tasks. In this form, the model is isomorphic to the theory of international trade, and the main question for its empirical relevance is whether or not there is "factor price equalization" across firms and industries. Unique prices imply that factor loadings are the same for all workers, but if factor price equalization fails, the factor loadings are job specific, and the dimensionality economies afforded by the hedonic approach are lost.

Taking this model on its own terms, factor price equalization in labor markets is improbable. It fails if specialized factors are used in some firms but not in others. Further, because workers skills are embodied, they cannot sell their skills in separate markets. An agent cannot sell brains to one employer while selling braun to another. This restriction limits the scope of arbitrage activities required to ensure unique factor prices in the labor market. The failure of factor price equalization makes the selection model in Figure 2.2 more relevant (Rosen, 1983a; Heckman and Scheinkman, 1987).

## 3 Information, Learning and Sorting

In the models discussed to this point, agents know their own productive capacities. We now review models where both workers and firms are uncertain about worker endowments. Both learn about the expected productivity of particular workers through successive observations of their performance. These models are dynamic extensions of selection and have implications for how the size distribution of earnings within a cohort evolves over time.

In pure statistical-decision models, there is no real allocative feedback from observations of current productivity to the level of future productivity. Agents do not use information from the record of previous output when making subsequent allocation decisions. Information only influences wage rates. Since each observation of a person's output is an error-ridden measure of productivity, labor market competition implies that workers are paid their expected marginal products, given the information available. Thus wages for a given worker evolve as information accumulates over time. If the record is long enough, earnings in a cohort converge to the stationary distributions described in the stochastic models of section 1.

More economically interesting information models introduce feedbacks from the data to future allocations, along the lines of the selection models of section 2. In sorting models, workers learn about their comparative advantage in various jobs and can switch jobs over time to better exploit their most valuable traits. In matching models, workers and firms gradually learn how productively they match with each other. If a realization is unfavorable, the relationship is terminated, and each side seeks a new match. Because workers keep good matches but abandon bad matches, wages grow, on average, as workers gain more information through work experience. Further, because workers leave poor matches, the wage distribution for a cohort becomes truncated over time, and wage distributions for experienced workers tend to be skewed relative to the distribution of wages among new workers.

Both matching and sorting models capture an important facet of individual careers. Without information about their relative aptitudes for different types of work, workers cannot make occupation choices that maximize their earnings. Thus, workers receive two forms of compensation from work experience. They receive their current earnings, and they receive information about their skills. This information affects future earnings because it informs future job choices.

Analyzing dynamic "Roy Models" requires putting more detailed structures on individual choices, and these structures make it more difficult to study the full market equilibrium aspects of the problem. Little work has been done on how individual decisions based on gaining better personal information link up to overall market behavior. For instance, most learning models implicitly assume infinitely elastic demands for all types of skills. An exception is MacDonald (1982), who nests a learning model inside a supply and demand framework. MacDonald illustrates how learning generates sorting of workers across tasks in a general equilibrium context. The sorting process in his model also generates aggregate wage growth for a cohort and skewness in the size distribution of wages.

### 3a The Basic Learning Model

Assume that two individual traits, $X_i$ and $\theta_i$, affect personal productivity, $y_{it}$. All agents observe $X_i$. No one observes $\theta_i$ directly (workers have no private information). In each period, nature randomly assigns workers to firms. They are paid their expected marginal product as assessed at the beginning of the period, given $X_i$ and their productivity history, $y_{it-1}, y_{it-2}, .. y_0$. The realization of worker output in each period is subject to an i.i.d. shock. Thus, output is a noisy measure from which to infer $\theta_i$. Repeated observations of past performance allow inferences to become sharper over time. At each point, the worker's output $y_{it}$ is an independent draw from the distribution $G(y_{it}|X_i,\theta_i)$.

Farber and Gibbons (1996) study how the distribution of wages evolves for a given cohort of workers in these circumstances. This problem is greatly simplified by the fact that true productivity does not vary over time. Consider a group of workers who share the same value of $X_i$. Since $y_{it}$ is a random sample out of a fixed distribution, the mean wage for this group of workers does not change as the workers age. There is no cohort wage growth in large samples, and the mean wage in any period among these workers is always

$$E(w_{it}|X_i) = E(y_{it}|X_i) = E[(y_{it}|X_i,y_{it-1},y_{it-2},..y_0)|X_i] \tag{3.1}$$

However, from well-known results on signal extraction in statistical decision theory, the variance of wages within each $X_i$ class, and hence overall, increases with cohort experience. Simple as it is, this model produces the fundamental and important result that the distribution of wages is more compressed among young cohorts because there is not so much information to differentiate among them. Wage dispersion increases as work records reveal each worker's productivity more precisely.

Information models are more interesting (and more difficult) when agents use acquired information to make real allocative decisions. The literature treats two kinds of problems. In one, workers use information to decide to invest in specific skills depending on what they learn about themselves (Jovanovic, 1979b). The implications for earnings distributions are perhaps not distinctive enough from the generic human capital models considered in section 4 to warrant discussion here. In the other, matching and sorting models, workers optimize work choice decisions based on their cumulated work record. This kind of learning does not change skills per se. Rather, it changes perceptions of what skills one has and how they may be used more effectively.

The basic economic idea in information theory is that agents anticipate a priori how information gained currently and in the future will affect their subsequent decisions. For instance taking a low paying, low productivity job might have value in providing lots of information about the best sequence of subsequent jobs. The ability to change one's decision as information accumulates lends option value to choices. But this is very difficult to characterize analytically, and the earnings distribution literature has made little use of option values. The structure of most work in labor economics minimizes this aspect of the problem.[3]

---

[3] Miller (1984) is an exception. Miller shows how workers sort across occupations when jobs differ with respect to both current expected returns and information content. In Miller's model, experience in a given job provides no information about expected productivity in other jobs, and given this form of independence across jobs, he is able to use results from the literature on Dynamic Allocation Indices (or Gittins indices) to describe a tractable solution to the model. In models without this type of independence, tractable solutions are unlikely unless the information content of all jobs is the same.

3b Sorting and Information about Comparative Advantage

Consider the following problem. There are N workers and J possible jobs. In each period output, $y_{ijt}$, of worker i on job j is

$$y_{ijt} = h(\theta_i, X_i, Z_j, \varepsilon_{ijt}) \tag{3.2}$$

where $\theta_i$ are unobserved traits, and $X_i$ are observed traits as before, $Z_j$ is a set of job attributes, and $\varepsilon_{ijt}$ is an idiosyncratic i.i.d. shock drawn from a known distribution $G(\varepsilon)$. The function h(.) is known, and outputs, $y_{ijt}$, are observed publicly. The "person effect" $\theta_i$ is an i.i.d. draw from a known distribution $F(\theta)$. Wage setting goes as before, and workers and firms are risk neutral. In the case where job characteristics are known, wages are given by

$$w_{ijt} = E(y_{ijt} \mid X_i, Z_j, y_{ijt-1}, .. y_{ij1}) \tag{3.3}$$

Workers choose the sequence of jobs that maximizes the present discounted value of lifetime earnings, $\Sigma\beta^t w_{ijt}$, where $\beta$ is the discount factor.

The solution is trivial when all traits of jobs and individuals are known: pick the job with the highest expected output and never switch. However, if some traits of jobs or individuals are unknown, then workers may sort across jobs based on their observed record of productivity in various employments. Sorting models describe environments where workers learn about their unobserved traits, $\theta_i$, when these traits have different values on jobs with different Z's. Information helps allocate workers to these jobs more efficiently over time. By contrast, all information in matching models is match-specific and has no external value about the distributions of potential rewards associated with other jobs.

Matching models are easier to analyze, but sorting models may be more realistic. Nonetheless, both approaches tell the same basic story. As workers learn about their comparative advantages in different types of employment, or about the quality of their current job match, they sort away from things that they do poorly. This implies that, for a given cohort, average wages grow with experience. Further, in the sorting model, the truncation of distributions associated with selection generates skewness in the distribution of wages.

The sorting model of Gibbons, Katz, and Lemieux (1996) nicely illustrates the main ideas while finessing many technical dynamic programming difficulties.[4] Here, the function h(.) in 3.2 follows

$$y_{ijt} = d_j + c_j(\theta_i + \varepsilon_{ijt}) \tag{3.4}$$

---

[4] See Gibbons and Katz (1992) for a similar but less general sorting model. Kim (1998) and Ross, Taubman, and Wachter also provide variants of this type of model.

As in section 2, this is a one-factor structure on "worker ability" $\theta_i$, with factor loadings differing across job types. Here, the $Z_j$'s in (3.2) are two parameters, $c_j$ and $d_j$. Workers have one trait, $\theta_i$. We ignore observed individual characteristics, $X_i$. A known pair $\{c_j, d_j\}$ characterizes each job.

Coexistence of all types of jobs requires that if $c_j > c_{j'}$, then $d_{j'} > d_j$. Figure 3.1 illustrates three jobs. Job 1 requires no skill. It pays a wage that is independent of any assessment of worker ability. Job 2 offers positive returns to ability, and job 3 offers the highest returns to ability. The shocks, $\varepsilon_{ijt}$, come from a standard normal distribution. Individual ability, $\theta_i$, also comes from a normal distribution with mean $\mu_\theta$ and standard deviation $\sigma_\theta$. Notice from the figure that if agents know $\theta_i$, the analysis reduces to the deterministic factor model in section 2, and sorting is strictly hierarchical: people with the highest values of $\theta_i$ choose job type 3 and those with the lowest values choose job type 1.

Hierarchial sorting simplifies the problem.[5] The problem is further simplified because the specified technology ensures that the information content of all jobs is the same and therefore does not affect job choice. Agents base posterior beliefs about the distribution of a worker's value of $\theta_i$ on the statistic

$$V_{ijt} = \theta_i + \varepsilon_{ijt} = (y_{ijt} - d_j)/c_j \tag{3.5}$$

Since the distribution of $V_{ijt}$ is invariant across jobs j, the worker's optimal policy is to pick the job that offers the highest wage in each period. This is given by

$$\max_j \; w_{ijt} = c_j \mu_{\theta it} + d_j \tag{3.6}$$

where $\mu_{\theta it}$ is the posterior mean of $\theta_i$ based on information available up to time t.

Consider a single cohort of workers. Assume these workers initially have the same expected ability, in the sense that their $\theta$'s come from the same distribution. The optimal policy described in (3.6) generates three implications for the evolution of the distribution of wages for this cohort. First, the distribution of wages becomes more dispersed as the cohort gains experience. Second, as workers sort across jobs in response to the information they receive, the average wage for the cohort as a whole increases with experience. Third, sorting across jobs generates skewness in the distribution of wages.

To understand how the sorting of workers across jobs creates wage growth in this model, note that since

[5] As noted in section 2, if workers differ with respect to more than one skill, the presence of constant terms in equation (3.4) complicates sorting partitions.

both $\theta_i$ and $\varepsilon_{ijt}$ are normally distributed, $\mu_{\theta it}$ is normally distributed across workers for $t > 1$. Further, the mean and median of $\mu_{\theta it}$ equals $\mu_\theta$ in each period. Consider the job choice of a worker with $\mu_{\theta it} = \mu_\theta$ in Figure 3.1. This worker chooses $j = 2$. If this were the only job in the economy, wages would be determined by

$$w_{it} = c_2\mu_{\theta it} + d_2, \tag{3.7}$$

and the variance of wages would increase with t, but the mean would remain unchanged at the starting wage, $w_{start} = c_2\mu_\theta + d_2$. This is just a version of the pure statistical information model discussed in the previous subsection.

With more than one job, workers still begin their careers in $j = 2$, and all receive the same first period wage. But now, $\mu_{\theta it}$ is not the only wage determinant that changes over time because workers have the opportunity to change jobs. Those who get favorable information and remain on $j = 2$ see their wage go up. Those who get unfavorable information and remain on $j = 2$ see their wage fall. However, the ability to change jobs convexifies the mapping from ability to wages, because those who learn they are much more capable move to $j=3$, while those learning that their value of $\theta_i$ is small switch to $j = 1$, where ability is not so important. Those who switch from $j = 2$ to either $j = 1$ or $j = 3$ earn strictly more than they would have earned if they stuck with $j = 2$. Thus, the expected wage for a cohort of experienced workers subsequently exceeds the starting wage. Convexity of the market payoff to ability also imparts skew to the wage distribution. Since the median wage is constant, the mean wage growth that results from sorting creates a growing gap between the mean and median wage.

The same logic applies in a model with an arbitrary number of jobs. As long as some workers sort away from their starting job, average wages will grow on average with worker experience, and the distribution of wages will be skewed. In addition, since the variance of expected ability $\mu_{\theta it}$ increases monotonically with t, the fraction of workers who have switched switch to new jobs increases monotonically with t. This implies that the gap between the mean and median wage of a cohort increases with experience. The truncation of distributions with sorting makes it difficult to say how the variance of wages changes with cohort experience. However, it can be shown that interdecile range measures of dispersion increase with t.[6]

To put models of this kind into perspective, bear in mind that much learning about individual talents (and preferences) occurs in school, prior to labor market entry. This is one of the reasons why all workers do not start work at the same wage on the same job. Instead, workers begin their careers on a range of jobs and at different wages. The point is that both of these ranges expand as workers gain experience. This model provides a useful description of how work experience yields information that affects the sorting of workers across jobs and how

---

[6]For example, the 90-10 ratio increase with t because the variance of $\mu_{\theta it}$ increases. As workers gain experience, both the 90th and the 10th percentiles of the $\mu_{\theta it}$ distribution are moving farther away from the median, $\mu_\theta$, which is constant over time. Since wages are a strictly monotonic function of $\mu_{\theta it}$, the 90th and 10th percentiles of the wage distribution must also be moving farther away from the median wage, $w_{start}$.

sorting affects the evolution of the distribution of wages for a given cohort of workers.

3c Matching Models

In matching models, all learning is idiosyncratic to a particular worker-job match. Experience in one job provides no information about a worker's potential productivity in any other jobs. The following sketch draws heavily on Jovanovic's (1979a) model, which in turn resembles a search model. Worker i's productivity in job j at time t equals

$$y_{ijt} = \eta_{ij} + \varepsilon_{ijt.} \qquad (2.9)$$

Agents observe $y_{ijt}$, but do observe its components. $\eta_{ij}$ is the expected value of the match between worker i and job j, and $\varepsilon_{ijt}$ is an iid realization of a normally distributed shock. The matches, $\eta_{ij}$, are i.i.d. draws from a normal distribution with mean $\mu_\eta$ and standard deviation $\sigma_\eta$. They are specific to workers and jobs jointly, not to either one separately as in sorting models. Again, workers are paid at the beginning of each period, before they produce output, and agents are risk neutral. There are enough firms competing for workers that the labor market is competitive: the wage for worker i in job j equals the expected value of the match, $\eta_{ij}$, given past outputs.

As in search theory, the worker's optimum policy is to choose a reservation wage and switch jobs if the expected value of the current match, given all available information, falls below it.[7] The reservation wage is a function of $\mu_\eta$, $\sigma_\eta$, and tenure in the current job, $\tau$. A job change in this model is another drawing of $\eta_{ij}$. Average wages for a cohort grow with experience because workers and firms have the option to abandon bad matches. Further, the reservation wage policy ensures that, in the limit, the distribution of wages is truncated. The left tail of the distribution is cut off at the reservation wage for a worker who knows $\eta_{ij}$.

3d Learning and General Equilibrium

Most sorting and matching theory is partial equilibrium and does not analyze how the number of people engaged in an activity affect its returns. MacDonald (1982) embeds a learning model in a market equilibrium context, and his results reinforce the predictions of basic sorting models. Wages grow, on average, with worker experience, and the sorting process tends to generate skewness in the distribution of wages.

MacDonald adapts the comparative advantage structure of Sattinger (1975) and Rosen (1978) to a learning environment in which workers learn their abilities over time. The structure of production is such that individual workers perform a large number of tasks arranged along a continuum, indexed by s; $0 \leq s \leq 1$. There are two types

[7]Similarly, it is optimal for the firm to adopt a reservation policy such that if the worker's productivity falls below a threshold, it is best to layoff the worker and draw another sample. The reservation values of firms and workers are consistent with each other. There is no meaningful distinction between quits and layoffs here.

of workers, A and B, but neither firms nor workers know the type of a specific person. Workers of types A and B differ in their ability to perform the tasks. Type A workers have comparative advantage at one end of the spectrum, and type B workers have comparative advantage at the other end.

A fraction $\delta$ of all workers are actually type A workers. At the end of a period, each worker exogenously receives a piece of public information; an independent realization of a random variable X, where X takes one of two values. X=a, or X=b. The probability that X=a is $\theta \geq ½$ if the worker is an A. Likewise, the probability that X=b is also $\theta$ if the worker is a B. Given this information technology, if a cohort of workers has worked for N periods, each worker in the cohort belongs to one of 2N+1 information classes, where an information class is a group of workers with the same posterior probability of being of type A. At any point in time, the supply of workers in a given information class is determined by the age distribution in the population. Workers work for a fixed number of periods, N+1.

MacDonald illustrates the determination of firm demands for workers from each information class in two steps. Given any set of workers from various information classes, the firm must determine the best way to assign workers to tasks. The solution to the assignment problem involves a set of critical tasks that define intervals on the task continuum, with one interval per information class. Because comparative advantage is ordered when there are only two types, the critical task intervals are ordered. Information classes are matched with these intervals according to the probabilities that workers in the class are of type A. These information classes correspond to conventional factors of production. Given an array of market wages for workers of various information classes, the firm must also choose the combination of optimally assigned workers that minimizes the cost of producing output. The solution yields the cost function and the associated compensated factor demands. Markets clear when the demand for workers in each information class equals their availability in the population. Overlapping cohorts lead to a steady state equilibrium.

Equilibrium wages are positively correlated with worker experience in this model. As the market acquires more information about a given cohort of workers, these workers are, on average, assigned to tasks that more fully exploit their true comparative advantage, like the sorting model described above. However, the result is reinforced in this model by supply conditions. In this economy, young workers are found in information classes that contain relatively large stocks of workers, both within each cohort and in the labor market as a whole. Thus young workers have many substitutes, which constitutes a second reason that young workers, on average, earn low wages.

For most interesting parameter specifications, the size distribution of wages in this model is skewed. If the skill endowments are such that neither A nor B has an absolute advantage at all tasks, the distribution of wages is always skewed. Further, in the benchmark case $\delta=1/2$, mean wages exceed the median wage, regardless of initial skill differences between types A and B. Here, the median is constant, and mean wages grow with cohort age as workers are matched with tasks that better exploit their comparative advantage. However, if type A workers are more productive in all tasks, and $\delta \neq ½$, then the distribution of wages may or may not be skewed.

## 3e Summary

Learning and information models add a dynamic dimension to the literature on assignment models. In both cases, wage distributions are skewed relative to distributions of wage offers because workers sort to jobs where they have a comparative advantage. However, in information models, workers learn their comparative advantage only with time and experience. This implies aggregate wage growth for a cohort, and often implies that the distribution of wages for a cohort becomes more skewed as the cohort ages.

## 4 Human Capital

In The Wealth of Nations, Adam Smith wrote,

> "When any expensive piece of machine is erected, the extraordinary work to be performed by it before it is worn out, it must be expected, will replace the capital laid out upon it, with at least the ordinary rate of profits. A man educated at the expence of much labor and time ... may be compared to one of those expensive machines. The work which he learns to perform, it must be expected, over and above the wages of common labour, will replace to him the whole expence of his education with at least the ordinary rate of profits of an equally valuable capital."

This is a central idea in modern human capital theory. Some wage inequality is a consequence of the fact that employments differ greatly with respect to the "difficulty and expence of learning the business." Because wages must compensate workers for the cost of learning their trade, wages would differ across employments even in a world where all workers possess the same initial skill endowments. The monetary value of the effort and time required to learn neurosurgery is great. If the wages of brain surgeons did not greatly exceed the wages of workers who learn their craft easily and quickly, we would observe a dire shortage of brain surgeons.

The two previous sections highlight how the market assigns people with heterogenous endowments to heterogeneous tasks and how assignments change as markets reveal information about workers' true endowments. Many models of human capital investment also highlight the role of heterogenous endowments, but the most important insight of the human capital approach is that patterns of investment in training help determine differences in productive capacities across workers. Here, investment costs drive earnings differences across groups of workers who have different levels of education, experience, or professional training.

### 4a Compensation for Investments in Training

Mincer (1958) is among the first modern economists to formalize Adam Smith's ideas concerning wage inequality and occupational training requirements. Suppose a given worker who seeks to maximize his lifetime earnings is faced with two possible careers. The first requires d periods of training and pays a wage of $W_0$ per period. The second requires training for d+s periods and pays a skilled wage of $W_1$ per period. Training requires no direct expense. The worker lives for n periods and discounts the future at rate r. A worker is indifferent between the two occupations if they offer the same lifetime earnings, or

$$\int_{d}^{n} W_0 e^{-rt} dt = \int_{d+s}^{n} W_1 e^{-rt} dt \tag{4-1}$$

The ratio $W_1/W_0$ that leaves the worker indifferent between the two occupations is given by

$$k(d,s,n) = \frac{W_1}{W_0} = \frac{e^{-rn} - e^{-rd}}{e^{-rn} - e^{-r(d+s)}} \tag{4-2}$$

Therefore k is the relative supply price of the schooling-intensive occupation. It is increasing in s and decreasing in n. Though schooling has no direct cost, an increase in s postpones entry into the labor market and increases foregone earnings: k must increase for the worker to remain indifferent. Further, since an increase in work life, n, implies a longer working period in which to recover foregone earnings, k decreases.[8] Finally, k is an increasing function of d. Relative wage ratios and the rate of return to additional schooling should be larger for persons with 14 years of schooling than for those with 10 years, because when life is finite and independent of investment, a higher relative wage premium is required to cover the costs of schooling that occurs later in life. These last two effects are small when n is large.

As n goes to infinity, k(d,s,n) ceases to be a function of d or n. In the limit (Mincer, 1974)

$$\ln W_1 = \ln W_0 + rs \tag{4-3}$$

This is the clearest statement of how wages compensate for added training costs. Let s be the additional schooling required to perform job 1 instead of job 0. A worker is indifferent between the two jobs if the percentage difference between the wage for job 1 and the wage for job 0 is roughly equal to the product of r and s. If a group of workers, who are homogeneous with respect to talent and tastes, occupy a variety of occupations that require different levels of training or schooling, the distribution of wages for trained workers in this group is skewed relative to the distribution of schooling. For example, if the distribution of schooling is normal, the distribution of wages is log normal.

Education and training means that wage inequality is not exclusively driven by heterogenous endowments

---

[8] In societies where people work until roughly the end of their lives, increasing expected life span is associated with decreases in the relative supply price to training-intensive occupations. This effect has been important historically.

and rents. Wage inequality results when some occupations require more training than others. Wage differences also must equalize the net advantages of work in other ways (see the survey in Rosen, 1986). For instance, some jobs are more onerous or dangerous than others. Some offer more earnings stability than others. Though these are important factors affecting measured inequality indexes, the scope of the subject is too broad to be included here. However, education and training deserve special emphasis for this survey because they are so important in determining personal productivity.

Mincer's approach is valuable because it illustrates an equilibrium condition that must be satisfied when workers acquire skills by investment. But, it does not tell us which workers invest in training and under what circumstances. We next review models of optimal investment in human capital to shed more light on the mapping between the distribution of individual traits and the distributions of wages and earnings.

For the theories reviewed up to this point, distinguishing between wage rates and total earnings has been inconsequential. In endowment models and learning models, workers decide where to allocate their time, but, by assumption, workers allocate all of their time to some type of production. Further, in Mincer's original paper, while there is a training period, it is distinct from work, and once training is over, workers spend all their time producing goods. By contrast, in most human capital models, workers decide how to blend different activities at each point in time. Some models address the division of time between training and production. Others also address the consumption of leisure. In all cases, earnings, which equal incomes per unit of time, need not equal wages, which equal incomes per unit of time spent in goods production.

Having said this, we discuss the implications of human capital models in terms of actual earnings not wage rates or potential earnings. Links between the predictions of human capital models and empirical regularities in labor market data are more clear in earnings data than in wage data. Existing data often provide straightforward measures of actual earnings, but this is not always the case with hourly wage rates.[9]

### 4b Optimal Investment

Among the first and most well known attempts to formulate a model that captures the implications of optimal investment in human capital for the distribution of labor earnings is Becker (1967),[10] who developed the useful graphical representation in Figure 4.1. In Becker's model, workers supply labor inelastically and seek to maximize the present value of their lifetime earnings. Great simplification is achieved by reducing lifetime human

---

[9]For example, consider a college student who works for only three months in the summer and reports annual earnings of $5,000. Researchers commonly conclude that the student's summer earnings imply an annual wage rate or potential annual earnings of $20,000. Among other things, this conclusion rests on the assumption that the student receives no training on the summer job. If the student is working in an apprentice program or an internship, his actual earnings may grossly understate his earning potential which is based on the wage he could receive in a job that involved no training.

[10]This was later reprinted in Becker (1975).

capital investment to a static problem. Assume that human capital investment activities require complementary purchased inputs (e.g., books and teachers), as well as the student's time and that each investment activity combines time and purchased inputs in fixed proportions. Returns on successive investments diminish at some point because the remaining work life is shortened and because the worker's opportunity cost of time increases with previous investments in human capital.[11]

The horizontal axis in Figure 4.1 measures total lifetime expenditure on inputs purchased for human capital production. The downward sloping "demand" curve represents the marginal internal return on these expenditures. It equals $(PV_1 - PV_0 - C_1) / C_1$, where $PV_1$ and $PV_0$ are the present values of lifetime earnings with and without the marginal investment in human capital, and $C_1$ is the dollar value of purchased inputs for the marginal investment.[12] Note that because $(PV_1 - PV_0)$ is the difference in the lifetime present values of two earnings streams, the marginal return in Figure 4.1 is a net rate of return on the financial cost of investment, $C_1$. This rate of return incorporates the opportunity cost of foregone earnings.

The upward sloping "supply" curve traces out the opportunity cost of purchased inputs in human capital production. In most industrialized countries, elementary and secondary schooling are approximately free in terms of tuition costs, but investments in on-the-job training or post-secondary schooling require other inputs that families must pay for with their resources. For wealthy agents, the opportunity cost of funds spent on these purchased inputs is the market rate of interest. For poor agents, the price of these funds is the interest rate at which they can borrow to finance their human capital investments. The supply curve of funds for poorer agents is generally upward sloping but may have horizontal segments. The person's lifetime investment is determined where demand equals supply.

Figure 4.1 pushes a number of important details into the background. Nonetheless, it is a useful general description of how ability and family wealth interact to determine the distribution of lifetime earnings. In a rough and ready way, the "demand" and "supply" representations in the figure correspond to the concepts of "nature and nurture" that dominate popular debates on inequality. The relevant concept of ability here is associated with returns from human capital investments. Worker A is more able than worker B if A's returns from particular human capital investments are always higher than B's returns. The demand curve for human capital by worker A is shifted up relative to worker B, and given the same supply conditions, worker A invests more than worker B. The effects of family circumstances and capital market imperfection work through the supply curve. Students with wealthier parents can finance their investments on better terms. Their supply curves are farther to the right than those of students from poorer backgrounds. Holding ability constant, students from wealthier families invest more.

---

[11]There may be regions of increasing returns if capital accumulation facilitates future learning. Below, we discuss assumptions concerning the way in which existing human capital affects the creation of new human capital. These assumptions are key to pinning down exactly how human capital theory links the distribution of talent and the distribution of earnings.

[12]These present values are calculated for all individuals and investments using a common market interest rate.

Superimposing the demand and supply functions of all individuals on the same graph yields a cloud of points that described the entire distribution of life-investment outcomes. For instance, the distribution of lifetime earnings tends to be more skewed when people have different talents (demand curves) and the same opportunities for financing investments (identical supply curves) than when people have identical talents but differ in their access to funds. Becker calls the first case "equal opportunity," because the terms of investment are independent of personal financial circumstances and family wealth. Those with large total investments in human capital also have high returns at the margin. But if all agents have the same ability and face different costs of funds, family wealth dominates investment outcomes: those with large total investments have low marginal returns because their marginal costs are low. Policies, like free education, that promote equal opportunity reduce the dispersion in supply curves and allow people to exploit their endowed talents to best advantage.

This model has limited empirical content for understanding the details of observed earnings distributions because it deals with unobservable human capital concepts and does not impose enough structure to pin down the precise mapping from human capital wealth to observed earnings. This difficulty is transparent when one compares the model to the pure equalizing difference model of Mincer (1958). In Mincer's model, all persons have the same talent and tastes, and therefore, the present value of lifetime earnings is the same for everyone. Although there is considerable cross-section earnings inequality, there is no lifetime inequality. While available data provide evidence that lifetime earnings do vary greatly across individuals, Becker's model provides no clear guidance concerning what portion of cross-section earnings inequality represents compensating differences on investment costs and what portion represent ability rents.

We need more specific models of human capital accumulation in order to precisely map talents and opportunities into life-cycle earnings profiles. The leading model of this kind is Ben-Porath's (1967) analysis of human capital investments as an intertemporal optimization problem. Workers seek to maximize the discounted present value of their lifetime earnings by allocating their time between investments in human capital and work and by choosing the optimal path of purchased investment goods over the life-cycle. The worker's problem is

$$\max_{s(.),i(.)} \int_0^T (w(v) - P\,i(v))\, e^{-rv}\, dv \tag{4-4}$$

$$s.t. \quad w(t) = R(1-s(t))k(t), \quad k'(t) = \beta(s(t)k(t))^{\gamma_1} i(t)^{\gamma_2}$$

where w(t) is the wage at t, s(t) is the fraction of time spent training, k(t) is the stock of human capital, i(t) is purchased inputs used to produce human capital, R is the rental rate on human capital, P is price of purchased inputs, and r is the interest rate. k(t) is "efficiency units" of human capital. Analysts typically assume that the equilibrium rental rate R and the price of purchased inputs P are constant over time. The expression for k'(t) is the educational production function.

This is a standard control problem. The details of solution depend on the form of the production function. Ben Porath's Cobb-Douglas form simplifies the problem immensely. In this model, the gross returns to using time either for direct production in the market or for investment in the production of human capital are increasing in the current stock of human capital. The technology described in (4.4) makes these effects exactly offsetting, and renders the optimal investment path k'(t) independent of k(t). This property makes it possible to get closed form expressions for w(t) and k(t).

In the Ben Porath framework, human capital k(t) is used to produce both output and additional human capital. Therefore, one might treat the initial stock of human capital, k(0), as an index of initial talent or ability. However, given this definition of ability, the Ben Porath model does not yield the prediction that able workers invest more resources in human capital production. It is straightforward to show that both efficiency units of time investment, s(t)k(t), and levels of purchased investment goods, i(t), are independent of the current stock of human capital, k(t). Workers with higher initial stocks of human capital have higher lifetime earnings, but they do not make larger investments in human capital production. Here, initial capital stocks are pure rents.

Becker's notion of ability is better captured by $\beta$, the scale parameter in equation (4.4). Consider two workers with identical human capital stocks, k(t), but different values of $\beta$. Assume these worker make identical investments of purchased inputs i(t) and time s(t). The costs of these investments are the same in terms of forgone earnings and purchased inputs, but the high-$\beta$ worker receives a greater return in terms of additional human capital, k'(t). The worker with a higher $\beta$ learns more efficiently, and it is easy to show that workers with more learning ability have higher lifetime disposable earnings. Further, for $t < T$, the more able worker invests more financial resources, Pi(t), and more effective units of human capital, s(t)k(t), in the production of human capital. Holding resource costs constant, workers who learn efficiently invest more in human capital and receive greater total returns from their investments. Thus, the ability to learn is a possible demand shifter in Becker's model above.[13]

Lifetime accumulation of human capital also has implications for the evolution of the distribution of earnings for a particular cohort. Figure 4.2 presents life-cycle earnings profiles for three workers in a Ben Porath model with the same initial stocks of human capital but different learning efficiencies.[14] The dispersion in earnings

[13]The strong Markovian properties of Ben-Porath's functional specification afford so many analytical conveniences that its use has become routinized. Alternatives are seldom considered. While convenient and simple, it is not necessarily descriptively accurate. For instance, the marginal costs of investment may be decreasing in k(t). Then people who know more find it cheaper to invest. Initial stocks k(0) as well as production function parameters affects optimum investment profiles and the affects of ability are more complex to describe. See Rosen (1976) for an example of this type. See Schultz(1975), Welch(1975)and Foster and Rosenzweig (1996) for extended discussion and evidence on how ability to learn affects productivity.

[14]The three values for $\beta$ are .05, .10, and .15. The other parameters are $\gamma_1$=.2, $\gamma_2$=.075, r=.03, and R/P=4. The initial capital stock is normalized to one. The relative price of human capital roughly equals the ratio of mean annual earnings for full-year workers who are twenty years old, male, and high school graduates ($16,300) to the mean cost of four quarters of college at a state university for the year 1994 ($4,021). The earnings data come from the March, 1995 Current Population Surveys, and the tuition data come from The Digest of Educational Statistics (1996).

for a given cohort is greater for mature workers than for young workers. Further, the increase in cohort earnings dispersion begins at a point well into the workers' careers. In fact, in this example, dispersion decreases over the first few years and then increases dramatically after the more able workers, who make large initial investments in training, catch up. Mincer (1975) first identified this effect.

This U-shaped relationship between cohort earnings variance and cohort age is an important theme in the literature on human capital. If able workers who learn efficiently make large investments in training, the variance of observed wages for a cohort of young workers greatly understates the variance of lifetime earnings for the cohort. Further, if the relationship between learning ability and training investments is strong enough, able workers actually begin their careers earning less than their less able counterparts. In this scenario, the variance of earnings for a cohort initially decreases with age as efficient learners workers make fewer training investments and begin to catch up to the earnings levels of less able workers. At some point, the earnings of able workers overtake the earnings of less able workers. Experience-earnings profiles in the cohort fan-out and the variance of earnings increases thereafter.

When workers vary in their ability to acquire skills, the location of young workers in the distribution of current earnings for their cohort misstates their position in the distribution of lifetime earnings because greater levels of investment depress their current earnings relative to less able workers.[15] Workers who learn efficiently have both relatively steep earnings profiles and relatively high lifetime earnings. In contrast, among workers who differ only with respect to their initial stocks of human capital, k(0), earnings profiles are parallel over the life-cycle,[16] and a worker's position in the distribution of current earnings is a perfect indicator of position in the distribution of lifetime earnings (Lillard, 1977; Lillard and Willis (1978), Weiss and Lillard (1978),Weiss, 1986).

It is difficult to analyze human capital market imperfections in optimal investment theory. Here "equal opportunity" corresponds to an environment where every worker can borrow and lend at the same interest rate, r, and everyone can purchase inputs for human capital production at the same price, P. Capital market imperfections break down the separation between maximization of wealth and of utility. But one can represent imperfections by exogenous differences in costs and returns for human capital investments. These differences reflect "noncompeting groups" and affect not only the shape of the cross-section earnings distribution but also how the distribution of

---

There are no available estimates of $\beta$, $\gamma_1$, and $\gamma_2$. However, Haley (1975) does provide estimates from a model without purchased inputs. Here, the parameters are chosen in order to get initial investments of time and purchased resources that roughly correspond to three investment levels: full-time college, part-time college or vocational training, and a modest level of on the job training. See Von Weizacker (1993) for an complete exposition of similar models and similar simulations exercises concerning the evolution of the distribution of earnings for a cohort.

[15]For example, medical students appear "poor" on current earnings accounts, but are wealthy on human capital accounts.

[16]This result holds under the assumption that human capital does not depreciate. In a model with skill depreciation, earnings profiles get closer together over the work life.

earnings for a given cohort evolves over time.

Here are some examples.

i) the level of government subsidy for both secondary and post-secondary education varies across countries and across regions within countries. Or education quality may differ among groups in the population. These can be represented by interpersonal differences in P. Investment in human capital varies directly with R/P, the ratio of the rental rate on human capital to the price of purchased inputs in human capital production. Persons facing low values of P invest heavily at the beginning of their careers and receive high earnings later in life.

(ii) Optimal human capital investment models usually are based on the idea of mutually exclusive time allocations between work and training. However, there is an equivalent alternative set up where work and learning are jointly produced on the job. In this learning-by-doing framework, learning opportunities differ across jobs and pay different wages as an equalizing difference. Workers choose the sequence of jobs that implements their optimal investment program (Rosen, 1972). Labor market discrimination and other barriers sometimes restrict access of some people to jobs with greater learning opportunities and better chances for advancement. This is equivalent to an increase in the marginal costs of learning, represented by a decrease in $\beta$ or an increase in P. Either way, those affected reduced their investments, and consequently, their experience-earnings profiles are flatter.

(iii) Labor market discrimination can also cause rental rates (the R's) to differ across groups. This too reduces the marginal return to investment and restricts human capital accumulation incentives, again resulting in smaller earnings levels and flatter earnings profiles.[17]

## 4c Labor Supply

A substantial amount of earnings inequality is directly attributable to interpersonal differences in hours of work supplied to the market. This issue is important for empirical work and also bears on theory. The human capital models considered above treat labor supply as inelastic, independent of investment activities. However, one expects interactions because workers choose their investments and labor supply jointly. For instance, unless human capital investments equally improve the productivity of a worker's time in both market and nonmarket uses, workers who expect to work more have greater incentives to invest because greater utilization of skills increases the rate of return to investment. Conversely, those who invest more face higher costs of leisure at the margin and may substitute against nonmarket activities. Forces that determine the quantity and timing of leisure consumption can affect both the size distribution of earnings and the evolution of the distribution of earnings for a particular cohort.

Blinder (1974) was among the first to investigate these issues. He restricts his analyses to the pure effects of hours choices because he treats the path of wages as exogenous. Still, his work serves as a useful starting point.

---

[17]The natural extension of these models is to introduce explicit intergenerational linkages. These issues of intergenerational mobility are crucial to the study of economic equality and are treated separately by Piketty in this volume.

An agent's utility is a time- and goods-separable function of consumption and leisure. Workers borrow and lend at a fixed rate of interest, r and face an exogenous life-cycle pattern of wages, w(t). Each person maximizes utility by choosing optimal paths of consumption c(t) and leisure l(t) given their initial wealth, M(0), and a concave bequest function, B[M(T)]:

(4.5)
$$\max_{c(t),l(t)} \int_0^T e^{-\rho t}[U(c(t))+V(l(t))]dt + B(M(T))$$
$$s.t.\ c(t) + s(t) = rM(t) + w(t)(1-l(t)),\ M'(t) = s(t),$$
$$M(0) = M_0,\ 0 \le l(t) \le 1$$

where s(t) is savings at t, and ρ is the rate of time preference. To illustrate the solution assume U(c(t)) = log(c(t)), and V(l(t)) = ϕlog(l(t)) and ρ = r for all workers. The first order condition for l(t) is

(4.6)
$$l(t) = \frac{\phi c(t)}{w(t)}$$

When ρ = r, c(t) = c(0) is a constant, and earnings are given by

(4.7)
$$E(t) = w(t) - \phi c(0)$$

so actual earnings profiles, E(t), and potential earnings profiles, w(t)---what the worker would earn if working flat out---are parallel. Because consumption c(0) is an increasing function of lifetime earning potential, the gap between lifetime earnings and lifetime earning potential increases with earning potential. Thus, when the path of potential earnings is exogenous and leisure is a normal good, the consumption of leisure mitigates lifetime earnings inequality. Blinder (1974) demonstrates a similar result concerning overall income inequality. Workers who begin life with large initial asset endowments, $M_0$, consume more leisure over their lifetime. Non-earned income is a substitute for earnings.

If all workers have the same tastes, the positive relationship between wealth and consumption of leisure ameliorates overall income inequality due to individual differences in asset income and earning potential. However, all workers do not have the same tastes, and tastes for leisure may well be correlated with individual wage rates and assets. Blinder uses simulations to gauge what sources of individual heterogeneity contribute significantly to the observed inequality in labor incomes. His results suggest that individual differences in wage profiles are the most important factor in determining earnings inequality, but there is also an interaction between tastes and wages. If wages and preferences for leisure are negatively correlated, realized lifetime earnings inequality may be much greater than the underlying inequality in potential earnings. This finding is important because we expect tastes to influence earnings capacities. In Blinder's framework, correlations between potential earnings and tastes are simply correlations between worker endowments. However, in human capital models, earnings capacity is, in part, determined by tastes. Because workers with weak preferences for leisure forfeit relatively little in utility terms when they work, they have additional incentive to invest in market skills. In short, we expect future earnings capacity to vary inversely with current tastes for leisure.

A complete life cycle model adds human capital investment to problem (4.5) and yields an analysis of three uses of time: investment, s(t), leisure, l(t), and work time (1 - s(t) - l(t)).[18] We maintain the Ben-Porath technology for human capital investment in (4.4), but ignore bequests and purchased inputs, and write the instantaneous utility function as u(c(t),l(t)) . Individual earnings capacities evolve as workers make investments in human capital. Assuming that the rate of time-preference equals the rate of interest, first order conditions for c(t), l(t) and k(t) satisfy

$$u_c(c(t),l(t)) = \lambda \tag{4.8}$$

$$u_l(c(t),l(t)) = \lambda R K(t) \tag{4.9}$$

$$\frac{R(t)[s(t)k(t)]^{1-\gamma}}{\beta\gamma} = \int_t^T R(\tau)(1-l(t))\exp(-r(\tau-t))\,d\tau \tag{4.10}$$

The right hand side of (4.10) is the discounted marginal value of a unit of current investment. λ is the multiplier associated with the lifetime budget constraint. The left-hand side is the marginal cost of additional human capital, k'(t), in terms of forgone current earnings.

Equations (4.8) and (4.9) are the standard intertemporal substitution conditions, with earning capacity, R(t)k(t), replacing the wage rate. The multiplier λ equals exp(rT). It is clear from (4.9) that the marginal utility of leisure is inversely proportional to R(t)k(t) along the optimum path (over the life cycle). If the marginal utility of leisure is independent of current consumption, this implies that leisure varies inversely with earnings capacity over the life cycle. As in the standard Ben Porath model, time spent investing, s(t), declines monotonically with age. Thus, when we examine a cross-section of workers, part of the observed difference in earnings between experienced and inexperienced workers may reflect the fact that experienced workers not only enjoy higher earnings capacity because of previous human capital investments but also spend less time investing and consuming leisure. A similar result holds in Blinder's framework if one assumes that earnings capacity rises over time. Human capital models

[18] Examples include Heckman (1976), Ryder, Stafford, and Stephan (1976), and Blinder and Weiss (1976). Weiss (1986) provides a survey of life-cycle human capital models.

generate rising earnings capacities endogenously.

Further, while the model we describe above yields few general comparative dynamics results that are of interest here, the model does provide a framework for discussing how differences in tastes for leisure across individuals may contribute to differences in the evolution of life-cycle earnings profiles. Take the special case where R(t) is constant for all t, assume that preferences are separable over time and goods, and assume interior solutions exist for all $t < T$. Now, consider two workers with identical skill endowments but different tastes for leisure. Because worker A has a stronger preference for leisure, he begins his career by devoting more time to leisure than worker B. Worker B always invests more in terms of efficiency units of human capital, s(t)k(t), because his weak tastes for leisure increase the returns from investments in market human capital. Therefore, over time, the earning capacity of B grows apart from the earning capacity of A. Since B always has higher earning capacity and weaker tastes for leisure, B always consumes less leisure than A.

In sum, if we ignore corner solutions, A always consumes more leisure than B. But, one cannot say that A always works less because A may spend less time training. As the workers age, differences in leisure consumption become more pronounced because earnings capacities continue to grow apart. However, differences in training must eventually diminish because training tends to zero for both workers as $t \to T$. Thus, differences in tastes for leisure lead to life-cycle patterns of leisure and training that eventually magnify earnings differences among experienced workers. Experienced workers with weak tastes for leisure have relatively high earnings capacities because of previous investments. For experienced workers, the combination of weak tastes for leisure and high earnings capacities reduces current leisure consumption and, at some point, must translate into both higher earnings and more labor supply.

While the exact life-cycle relationship between tastes for leisure and labor supply is not clear, it is clear that, holding ability constant, those with weaker tastes for leisure have higher discounted lifetime earnings.[19] Further, human capital models demonstrate that this result is more than a link between labor supply and tastes for leisure. There are two components to the link between tastes and earnings. Given earnings capacity, workers with weak tastes for leisure work more. But earnings capacity is not independent of tastes. Weak tastes for leisure enhance investment and therefore raise earnings capacity.

Adding labor supply considerations to human capital investment models extends the economic implications of these models. People with either weaker tastes for leisure (inferior opportunities for work outside the market) or lower costs of human capital production invest less. Thus economic development, decreasing fertility, and increasing labor force participation of women increase the demand for human capital investments and augment the future market supply of skill. What remains for future research is a better understanding of the demand for skilled labor and how such shifts in the supply of skill affect the equilibrium rental rate on human capital.

---

[19] If two workers have the same opportunities, the one with weaker tastes for leisure will enjoy more goods consumption. This requires more higher lifetime earnings.

## 5 Insurance, Agency and Earnings

If firms could pay workers strictly on the basis of their own production, most incentive and labor relations problems between employers and employees would disappear. Pay would be contingent on a mutually acceptable outcome. Few labor contracts work this way because individual output is costly to assess at each point in time. Payments of salaries and wage rates for a worker's effort and time are the norm, and firms loosely structure the detailed duties and obligations of employees. Workers typically have some degree of control and discretion over how they perform their work. Thus, workers have opportunities to pursue their own interests instead of their employer's interest. Examples are shirking and taking malfeasant actions, such as theft of goods from the firm and accepting payments from outside contractors.

The economic theory of agency analyzes such problems and describes internal compensation mechanisms that resolve conflicts and align the interests of employer and employee. The literature on agency examines the scope and limitations of different payment schemes, such as piece rates, for eliciting desired worker behavior, but our interest in agency theory is restricted to its role in understanding the distribution of compensation in labor markets. In this regard, agency models provide an important contrast to the models reviewed earlier in this survey. In most theories of earnings distributions, competition ensures that individual earnings reflect individual productivity, and the chief aim of theory is to explain the distribution of individual productivity. The agency literature reverses things: methods of compensation affect the allocation of worker effort across tasks and therefore determine the distribution of individual productivity. Moreover, observed wage rates at any point in time may not reflect personal productivity at that point.

A related literature focuses on a different information problem. When firms can observe productivity, risk averse workers who are not privately informed of their personal productivity do not want to be paid strictly according to their ex post output. They want insurance against random factors that affect productivity but are beyond their control. This demand for insurance shapes compensation policies by weakening the links between pay and performance.

Below we discuss implications of both agency models and models of earnings insurance for the distribution of earnings. We begin by describing the insurance components of wage payments. This is the starkest possible departure between personal pay and personal productivity: pooling risks introduces voluntary redistributions of income and tends to compress wage distributions. Next, we consider agency models where firms use policies that increase wage dispersion in order to elicit effort. Tournament models of competition for promotion and models of bonding and deferred compensation loosen the links between current productivity and current earnings. Although insurance models describe forces that compress earnings inequality, both yield an asymmetric distribution that is truncated on the left.

We conclude with an examination of efficiency wage models that address monitoring problems. In these models, firms use wage premiums to elicit certain desired aspects of employee performance, but other nonwage aspects of pay, such as performance bonds are not present and labor markets do not clear. These models do not have clear implications for the overall distribution of earnings, but they do provide a possible explanation for observed differences in mean wages across sectors of the economy.

5.1 Attitudes toward Risk

Worker preferences alone can give rise to voluntary redistribution in the presence of uncertainty. Friedman (1953) is the first to address earnings inequality in such terms (see also Kanbur, 1979). Risk averse workers have incentives to equalize their incomes when earnings have stochastic components. If ex post outcomes contain random elements and no one has private information, a group of ex ante identical risk averse workers who know the probability distribution of outcomes have incentives to pool the risks and insure each other. All could agree ex ante to a binding contract in which everyone contributed their ex post earnings to a common fund. All would receive the ex post mean, and income would be completely equal even though the distribution of ex post productivity might be dispersed.[20]

Full ability-insurance is not widely observed in the labor market as a whole, due to moral hazard and adverse selection. People who turn out to be successful would be reluctant to share their windfalls with the less fortunate, and even if they did, they would have little incentive to develop and exploit their skills. Further, people who thought themselves more talented ex ante would be reluctant to participate in such a scheme in the first instance.[21]

Harris and Holmstrom (1982) provide an interesting model of earnings insurance that is an elaboration of the simple learning model described in section 3.1. The record of production gradually reveals a person's productivity, and there is market competition for workers. If workers and firms could write lifetime employment contracts, risk neutral firms could provide complete earnings insurance for workers. However, in the absence of such contracts, labor mobility limits the ability of firms to provide insurance. Ex post, highly productive workers have an incentive to leave a firm that pays each worker according to the expected productivity of all its workers. However,

---

[20]Friedman himself investigated the possibility that this kind of insurance is not observed because people are risk-loving rather than risk averse. Workers with Friedman-Savage preferences (risk-loving in a certain range) have incentives to increase the dispersion of income by gambling. Recently Bergstrom (1986) and others (see Freeman, 1996 and references therein) have shown that Friedman-Savage risk preferences can result when the marginal utility of material goods consumption interacts with the type of work a person does or with other indivisible life-circumstances, such as place or residence. However, these type of risk preferences need not be addressed by labor market mechanisms. Purely monetary side gambles are sufficient. After winners and losers settle these side bets, they can make their occupational and other indivisible choices deterministically (Rosen, 1996), just like the models in sections 1 - 4.

[21]There is a certain awkwardness in the pure theory of insurance as applied to labor market contracts. Why should the arrangements flow through firms rather than being managed by third party insurance companies? One possible answer is that the firm is better able to control the worker's work performance, on which the insurance is contingent.

even in an environment where workers are free to change firms, employers can provide partial insurance by matching outside offers and guaranteeing that a worker's earnings never fall, no matter what the subsequent record reveals about personal productivity. It does this by guaranteeing that each worker will always receive at least the starting wage.

Since the firm rationally expects to pay higher wages to successful workers later on, the starting wage must be less than the unconditional mean productivity in the group. But risk averse workers are willing to accept a lower starting wage for the privilege of being insured against subsequent wage cuts should their productivity record prove unfavorable. The difference between the guaranteed starting wage and mean productivity of the entry cohort represents an up-front insurance premium payed by workers necessary to finance the insurance. Likewise, when a worker who performs well receives a new wage offer, the offer is less than the worker's expected productivity given the new information. The worker's ability is still uncertain, and the worker is willing to accept a new wage below expected productivity in exchange for a guarantee that the new wage will never fall in the future. As the worker becomes more experienced, there is less uncertainty about his true ability, and he demands less insurance. Therefore, as a cohort of workers gains experience, those receiving wage increases pay smaller premiums for earnings insurance.

The predictions of this model are similar, but stronger than those of the sorting model in section 3.2. In the sorting model, wages become more disperse as the market learns more about individual abilities. Over time, average wages rise and the distribution of wages for a cohort spreads asymmetrically. Those who learn they possess lower levels of ability mitigate their losses by switching to jobs where output is less sensitive to ability, while those who learn the opposite magnify their gains by switching to jobs that more fully exploit their talent. In the Harris and Holmstrom framework, learning also creates wage growth and asymmetric wage dispersion because the distribution is always truncated at the starting wage. But in addition, wages of individuals never fall, as they would in the pure sorting model. There is a kind of "downward wage rigidity."

In general, worker demands for insurance compress wages relative to the distribution of individual productivities, but, because workers and firms cannot write complete insurance contracts, the resulting compression may be one sided. Models of effort elicitation also generate asymmetric wage distributions, but here the emphasis is on stretching the right tail of the distribution not truncating the left tail.

## 5.2 Effort Incentives and Compensation in Internal Labor Markets

Wages differ substantially across job titles within firms. Not only are there substantial pay distinctions between rank and grade, but empirical work shows that wage jumps associated with promotions account for a high proportion of a worker's wage growth within firms, especially in white-collar jobs (Baker, Gibbs and Holmstrom, 1995). At first blush, this is hard to reconcile with standard productivity theory. Though a person's job activities usually change after a promotion, one's inherent skills and productivity do not jump on promotion day. Evidently, firms tie wages to jobs as well people.

Job transfers of existing personnel and promotions from within can be thought of as an internal labor market. A military organization, where outside recruitment occurs only at the lowest ranks and all higher level positions are filled by internal promotions, is the most familiar example. Other organizations recruit at all levels but fill most vacant positions through internal transfers. Private law firms promote associates to partnership status, academic departments grant tenure and change the rank of professors, and large firms draw many of their executive staff from lower ranking positions in the same firm.

The indivisibility of job assignments is the main reason for internal recruitment of personnel: there can only be one chief executive officer at a time. Promotion policies allow firms to learn worker capacities for more important jobs by observing performance on less important ones; and the workers gain specific knowledge of the firm's operations and procedures that enhance future productivity in such positions. Issues of testing and sorting naturally arise in the assignment of persons to positions in internal labor markets. Firms promote workers who surpass a certain threshold of performance. In setting these thresholds, firms face a trade-off between how high to set the bar and the wage differential between categories. Pay differences between grades increase the incentives for workers to perform, while higher standards decrease incentives.

Let x be a worker's productivity on the current job, with $x = \mu + \epsilon$, where $\mu$ is the worker's effort and $\epsilon$ is a random variable. $C(\mu)$ is the cost to the workers associated with effort level $\mu$. Normalizing the price of x at unity, the efficient choice of $\mu$ for a risk neutral worker occurs when the marginal cost of effort equals its marginal return, or $C'(\mu) = 1$. The probability of passing the test[22] is calculated from the distribution of $\epsilon$ as $p = p(\mu,s)$, where s is the pass threshold measured in units of x, and $p_\mu > 0$ and $p_s < 0$. The firm promotes those who pass to a job that pays wage $W_1$, and those who fail go to a lower paying job $W_2$.[23]

The expected return to the worker is

$$(5.1) \qquad p(\mu;s)(W_1 - W_2) + W_2 - C(\mu)$$

The worker puts forth effort to satisfy the first order condition

$$(5.2) \qquad p_\mu(\mu;s)(W_1 - W_2) = C'(\mu)$$

---

[22]If the cumulative distribution of $\epsilon$ is $G(\epsilon)$ and the worker is promoted if $s > x$ then $p(\mu;s) = 1 - G(s-\mu)$. Note that this is not necessarily concave in $\mu$ so first order conditions must be carefully checked for optimality. Those complexities are ignored here.

[23]There is no mobility here. Workers and firms agree on a lifetime employment contract that gives the workers their reservation level of utility. If workers are free to leave in the second period, firm face additional participation constraints in choosing their wage structure.

Comparative statics and the second order condition imply that $\mu$ is increasing in $(W_1 - W_2)$ and decreasing in s, so combinations of the promotion bonus, $(W_1 - W_2)$, and the height of the bar, s, can be found that achieve the efficiency condition $C'(\mu) = 1$. Firms pay a cost to set their standard, and these costs are increasing and convex. With homogenous firms and workers, competition by firms for workers ensures an efficient equilibrium combination $[(W_1 - W_2), s]$.

$$pW_1 + (1-p)W_2 = E(x) = \mu \tag{5.3}$$

Given this combination, workers receive their opportunity cost in terms of expected lifetime utility, and firms earn zero profits.

It is interesting is that although the average wage equals average productivity. The actual wage almost never equals ex post realized productivity. For instance, the probability that a person's realized value of x equals either $W_1$ or $W_2$ is zero, if $\epsilon$ is continuous. Here, wage dispersion does not match up with ex post personal productivity. Rather, wage dispersion elicits ex post productivity.

In many instances, it may straightforward to rank workers on an ordinal performance scale but difficult or impossible for firms to rank workers on a cardinal scale. Even when it is clear that worker A is better than worker B, firms may not know how much better. Under these circumstances, firms make promotion decisions based on relative performance, such as head to head competition among lower ranking workers vying for higher level positions. The noise introduced by common-error components in measuring individual performance can be eliminated by comparing contestants with each other. The result is as a "tournament" (Lazear and Rosen, 1981; Nalebuff and Stiglitz, 1983, Green and Stokey, 198; McLaughlin) or some other form of relative performance evaluation (Holmstrom, 1982).

In tournament theory, contestants vie against each other rather than against a fixed standard. This introduces strategic considerations to the problem above, because the probability of promotion for each player depends on what others are doing. The efforts of all other contenders replaces s in the probability of winning function, and we must analyze wages as outcomes of a noncooperative game. Nonetheless, prizes exist that produce the socially efficient level of effort among ex ante homogeneous players.

The tournament literature provides other interesting results concerning the distribution of earnings within firms. For instance, one can show that more contestants competing for a fixed number of promotions adversely affects incentives (O'Keefe, et. al, 1984). Thus, the age and experience distributions within a firm affect its internal wage policy. Furthermore, analysis of sequential contests in promotion ladders reveals another reason why compensation is extraordinarily high in top level positions. For in tournaments where contestants are eliminated from further contention after failing, the difference in rewards between adjacent ranks provides only part of the incentives to perform. There is substantial option value in the probability of contending for even larger prizes in better jobs down the line. But in the last rounds of the tournament, that option plays out. Maintaining performance incentives at this crucial stage requires that the option value be replaced by an extraordinary jump in wages at the top of the organizational chart (Rosen, 1986). Again, wages do not necessarily equal ex post personal productivity of the person who wins the prize but do affect the productivity of others in the organization.

5.3 Agency and Performance Bonds

Contests or tournaments involving workers constitute incentive schemes that magnify earnings dispersion within a firm for a particular cohort of workers. A given cohort of workers enters a contest and over time, their earnings grow apart as some win and some lose. However, more standard models of agency problems address earnings inequality across cohorts. These models highlight the role of performance bonds and deferred compensation as monitoring devices.

The basic solution to agency problems is a performance bond, in which a worker is put in a position to lose something if shirking or malfeasance are detected by the firm. The threat of dismissal alone does not provide sufficient incentives if an alternative job with the same pay (utility) can be readily found. Further, paying wages in excess of opportunity costs just to prevent shirking essentially creates rents for workers fortunate enough to land such jobs. In agency theory proper, other components of pay are adjusted to eliminate these rents and achieve market clearing. These take the form of payments at the front (bonds) and back (pensions) of the contract. Section 5.4 discusses efficiency-wage models, where these extra market-clearing adjustments are not admissible.

Workers in positions to take malfeasant actions against the firm can be deterred by sufficient monitoring (Alchian and Demsetz, 1972), but, like police and lawyers, monitoring is expensive. A self-enforcing payment mechanism that induces workers not to take advantage of the firm may be less expensive and achieve the same result. Obviously this requires penalizing malfeasant behavior (the "stick" approach) or rewarding good behavior (the "carrot" approach), so that workers find it in their best self interest to behave as the firm desires them to behave (Lazear, 1995). Ross (1973) and Becker and Stigler (1974) established that an up-front performance bond and back loaded pension accomplish this.

Following Becker and Stigler (1974), consider the last period of job which pays a wage $W_n$ and offers a chance at malfeasance worth x to the worker. Knowing that malfeasance can occur, the firm installs an imperfect monitor that detects malfeasance with probability p. The firm fires the worker if it detects malfeasance, and the worker may take a job elsewhere that pays $V_n$. The worker's expected return to malfeasance is $(1-p)(W_n+x) + pV_n$. The return to honesty is $W_n$ itself. Equating the two gives the minimum honesty-inducing wage

$$W_n = V_n + [(1-p)/p]x \qquad (5.4)$$

Honesty requires paying a wage premium of $[(1-p)/p]x$ that varies inversely with the probability of being caught.

Consider now a two-period job in which the wage paid in (5.3 ) is the second period wage, and thus, workers do not shirk in the second period. What wage is required to deter shirking in the first period? The present value of earnings in a job where the worker is deterred from malfeasance in both periods is $W_{n-1} + W_n/(1 + r)$, where r is the discount rate. If the firm dismisses all malfeasant workers it detects in the first period, the expected present value of malfeasance is $(1-p)[W_{n-1} + x + W_n/(1+r)] + p[V_{n-1} + V_n/(1+r)]$. Equating present values gives the minimum payment needed to deter malfeasance in the first period as

$$W_{n-1} = V_{n-1} + [(1-p)/p][r/(1+r)]x \qquad (5.5)$$

Equation (5.5) generalizes to all contracts lasting an arbitrary number of periods.

Notice the difference between (5.4 ) and (5.5). The wage premium needed to deter malfeasance is largest in the final period of the contract because the threat of termination has less deterrent value the shorter the remaining horizon. But, in every period of the contract, $W_t > V_t$. There are rents in every period, and the expected present value of the job strictly exceeds that of the alternative. The difference in present values works out to be $[(1-p)/p]x$. To equalize present values and avoid job rationing, workers are obliged to put up front money of $[(1-p)/p]x$. Shirking may be eliminated without job rationing under the following scheme: (i) workers post a performance bond, (ii) wages at each period equal opportunity wages plus interest on the bond, and (iii) deferred pay at the end of the contract equals the bond itself.

There is an important general point in this example. Current rates alone are insufficient to achieve efficient incentive alignment in many agency problems. In this example, front and back-loaded payments, as well as wages are needed. In a sense, one price, the wage rate, has to do too many tasks (Lazear, 1995). The firm needs to attract applicants for the job and provide performance incentives once workers take the job. Generally more than one price is needed to accomplish more than one economic function. The bond creates a "price," as it were, for malfeasance.

Pensions and deferred pay represent significant components of total compensation in many jobs where malfeasance is of obvious concern. Examples include the military, police, judges, politicians and some civil servants. Of course pensions are of growing importance in total compensation for other reasons having to do with the tax advantages of saving for retirement in this way. Nonetheless, most private pension plans that are based on a person's previous wage record, greatly penalize departures from the firm prior to the standard retirement age and serve a deterrence function in this respect. The data do not support the up-front bonding aspects of the solution. Explicit performance bonds are seldom observed in labor markets. If x is large and p is small, large bonds might be necessary but infeasible for liquidity constrained workers. Further, the theory sketched above does not address the possible reluctance to post performance bonds when firms have the opportunity to malfeasantly claim malfeasance by workers and seize their bonds

Lazear (1979) developed an important extension of the basic model by noting that deterrence does not require strict equality between the values of malfeasance and honesty in each period. Inequalities also work. In Lazear's model there is no explicit up front bond. Instead workers gradually post the equivalent of a bond by working for less than their full productivity to the firm when they are young. The firm breaks even because it pays them more than their direct productivity when they are older. In a sense, workers are building an implicit "equity" or "partnership" position in the early years and receiving "stakeholder" returns in the form of wage premiums and pensions in later years. In this model, wage-experience profiles are upward sloping for incentive reasons. Thus, the distribution of wages within a firm is more disperse than it would be in an environment without monitoring concerns[24].

Akerlof and Katz (1989) point out that the implicit bonding scheme proposed by Lazear does not work at the beginning of careers. At each point in time, the worker weighs the benefits of shirking against its cost----the expected value of the implicit bond that the workers loses if detected. At the beginning of the employment relationship, the value of the implicit bond is zero, and the worker has nothing to lose by shirking. Implicit bonding by workers requires some mechanism that gets around this "start up" problem. For instance, the firm may offer entry level jobs where output is easily observed, with workers posting bond by being paid piece-rates below their marginal product. After posting the bond, the worker can then be promoted to more complex jobs where output and effort are not directly observed. For workers who begin their careers with few skills, minimum wage laws or other employment regulations may frustrate attempts to post bonds in this manner. Nonetheless, there are other degrees of freedom. Firms may engage in intensive direct monitoring until workers develop the human capital necessary to post a bond while working at the minimum wage. If bonds cannot be posted implicitly, upward sloping wage profiles still serve as an incentive device, but the market does not clear. Efficiency wage models explore these issues.

## 5.4 Efficiency Wage Models

Efficiency wage models come in several flavors, but most share the common feature that workers are not allowed to post bonds. We devote most of our attention to the shirking model.[25] The model of Bulow and Summers (1986) uses inter-industry differences in monitoring technologies to explain how wage levels of observationally

[24]In addition, the model provides a rationale for mandatory retirement. When the contract reaches a termination date set at the beginning of the contract, the wage being paid at that time are greater than the worker's opportunity cost, and the worker "prefers" to keep working. The firm must terminate the worker by mandatory retirement at that point else the ex ante scheme would go insolvent. This is one of the only economic theories of mandatory retirement that has been proposed.

[25]Shapiro and Stiglitz (1985) also provide a well known shirking model. Some efficiency wage models address asymmetric information about worker quality , turnover costs or the nutritional needs of workers in undeveloped countries. In all these cases, up front payments are not allowed and ex ante rents and nonprice rationing of jobs creates unemployment. Bonds are not so crucial in efficiency wage models based on the sociological concept of voluntary gift exchange.

similar workers differ across sectors of the economy in a cross-section. Firms in the "primary" sector are, by assumption, unable to effectively monitor their workers. Direct expenditures on supervision do not change the probability of catching a worker who is shirking. Firms in the "secondary" sector are able to monitor their workers perfectly are zero cost.

In the primary sector, firms catch workers who shirk with probability p. Primary sector firms cannot change p, and by assumption, workers cannot post performance bonds either implicitly or explicitly. Firms fire shirkers if they catch them, but firms have only one other means to deter shirking. They can make dismissal more costly by paying wages above the going wage rate in the secondary sector.[26] The wage premium required to prevent shirking is an increasing function of total employment in the primary sector because a large primary sector makes it relatively easy for a dismissed worker to find a new primary sector job. This no-shirk function is depicted in Figure 5.1. The intersection between it and the demand for primary sector workers determines sectoral employment and the wage premium enjoyed by primary sector workers. The model generates an equilibrium where identical workers in different sectors of the economy earn different wages. The relative demand for primary sector output determines both primary sector employment and the magnitude of the inter-sector wage gap.

Wage levels differ considerably across the industries of modern economies. Controlling for observed differences among workers in various industries eliminates some of the interindustry wage variability, but much remains.[27] This theory offers an explanation for the variance remaining after standardizing for measured skills. It is easy to imagine a generalization of the model where w* represents the value of home production, and there are N no shirk conditions that correspond to N different industries with various degrees of monitoring problems. Such a model would generate a pattern of differences in wages across industries that reflect differences in monitoring technologies across various workplaces rather than inherent differences in worker skill. However, as an empirical matter unobserved differences in skills of workers across industries appear to account for at least part of these wage differentials.[28]

Efficiency wage models generate a distribution of wages across sectors of the economy that reflects heterogeneity across firms with respect to monitoring technologies. In efficiency wage models, workers are not allowed to post bonds, and because this option is shut down, firms use wage premiums as a mechanism to provide performance incentives. Complete market agency models involve either implicit or explicit performance bonds, and over a worker's entire tenure with a firm, he earns lifetime wages equal to his opportunity cost in present values. It is difficult to separate these models empirically because available data provides little information about

---

[26]The result is similar to the one presented above in our discussion of Becker and Stigler (1974). However, the Bulow and Summers model is static and does not address how time horizons interact with the threat of dismissal.

[27]See Krueger and Summers (1988).

[28]See Gibbons, Katz, and LeMieux (1997) and Gibbons and Katz (1992).

monitoring technologies. However, previous studies have not demonstrated a link between monitoring practices and inter-industry wage premiums.[29]

6 CONCLUSION

We began by noting that earnings distributions are always skewed with the right tail being much longer than the left. Selection and sorting models illustrate that when workers choose the job that best suites their skills, observed earnings do not include many of the bad matches in the left tails of potential earnings distributions for specific jobs. Insurance models describe how partial insurance contracts truncate wage distributions on the left, while tournament models illustrate how firms stretch the right tail in order to provide incentives. Human capital models may generate skewed distributions because individual endowments like learning ability, productive capacity and tastes for leisure may interact to generate considerable heterogeneity in investment behavior. However, human capital models stand out because they tie observed differences in earnings across education and occupation groups to observed differences in skill investment.

The literature on scale economies and span of control describes a link between technology and the distribution of earnings. Because technology determines scales of operation, it affects the deamnd for skill and the distribution of ability rents. However, most of the models presented here generate earnings distributions from supply decisions. They illustrate how decisions concerning the allocation of talents, time and energy generate and sustain earnings inequality. Selection models illustrate how workers supply their endowments to takss that generate the highest earnings. Human capital models describe skill formation and demonstrate that inequality is necessary when different types of work require different levels of skill investment. Models of the investment process illustrate how talent and financial resources affect investment decisions and therefore demonstrate how human capital accumulation links initial endowments, family background, and realized inequality. Agency models focus on the supply of effort and describe wage inequality within firms as an incentive device.

However, with the possible exception of the scale economies literature, these models say little about the determinants of labor demand or the elasticities of demands for different types of skill. This is a striking omission because much recent research explores potential "demand driven" explanations for the recent rise in wage inequality observed in many Western countries. A large empirical literature debates competing claims about the nature of labor demand and its response to changes in technology or trade barriers. We do not explore these issues here because a related chapter by Lawrence Katz addresses these issues in detail.[30] But, we do note that most theoretical

---

[29] See Neal (1993) and Leonard (1987).

[30] Because the models described here do not address the elasticities of demand for various types of workers, they say little about how demographic changes affect the wage structure through changes in the relative supplies of various skills. Murphy and Welch (1992) and Katz and Murphy (1992) provide evidence concerning the relative importance

(continued...)

work on the distribution of labor earnings does not provide clear insights concerning the forces that shift or shape the demands for various types of market skills.

[30] (...continued)
of shifts in labor demand and supply.

## References

Akerlof, George A. (1982) "Labor Contracts as Partial Gift Exchange", *Quarterly Journal of Economics*, XLVII, no. 4, 543-569.

Akerlof, George A. and Katz, Lawrence (1989) "Workers' Trust Funds and the Logic of Wage Profiles", *Quarterly Journal of Economics*, August, 525-536.

Alchian, Armen (1972) "Production, Information Costs, and Economic Organization", *American Economic Review*, 62, no.5, Dec., 777-795.

Atchison, John and Brown, J. A. C. (1957) *The Lognormal Distribution, with special reference to its uses in economics*, Cambridge, University Press.

Baker, George, Gibbs, Micheal, and Holmstrom, Bengt (1994) "The Internal Economics of the firm: Evidence from Personnel Data", *Quarterly Journal of Economics*, 109, 881-919.

Becker, Gary S. (1967) "Human Capital and the Personal Distribution of Income", *W.S. Woytinsky Lecture, No. 1*, University of Michigan.

Becker, Gary S. (1975) *Human Capital, A Theoretical and Empirical Analysis with Special Reference to Education*, 2nd ed., University of Chicago Press: Chicago and London.

Becker, G. and Stigler, G. (1974) "Law Enforcement, Malfeasance and Compensation of the Enforcer", *Journal of Legal Studies*, 3, no.1, 1-18.

Ben-Porath, Yoram (1967) "The Production of Human Capital and the Life Cycle of Earnings", *Journal of Political Economy*, 75, 352-365.

Bergstrom, Theodore (1986) "Soldiers of Fortune?", in Heller and Starr (eds.), *Essays in Honor of Kenneth Arrow, vol. 2, Equilibrium Analysis*, New York: Cambridge University Press, 57-80.

Blinder, Alan S. and Weiss Yoram (1976) " Human Capital and Labor Supply: A Synthesis", *Journal of Political Economy* 84, no. 3, June, 449-472.

Blinder, Alan S. (1974) *Toward an Economic Theory of Income Distribution*, The MIT press, Cambridge, Massachusetts, and London, England.

Bowley, Arthur Lyon (1915) *Livelihood and Poverty*, London, G. Bell and sons, ltd.

Bulow, Jeremy, and Summers, Lawrence (1986) "A Theory of Dual Labor Markets with Applications to Industrial Policy, Discrimination, and Keynesian Unemployment", *Journal of Labor Economics*, 4, 376-414.

Champernowne, D.G. (1953) "A Model of Income Distribution", *Economic Journal* 63, 318-351.

Champernowne, D.G. (1973) *The Distribution of Income between Persons*, Cambridge [Eng.] University Press.

Farber, Henry S. and Gibbons, Robert (1996) "Learning and Wage Dynamics", *Quarterly Journal of Economics*, 111, 1007-1048.

Foster, Andrew D., and Rosenzweig, Mark R. (1995) "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture", *Journal of Political Economy*, 103, no. 6, 1176-1209.

Frank, Robert H. and Cook, Philip J. (1995) *The Winner-Take-All Society : How More and More Americans Compete for Ever Fewer and Bigger Prizes, Encouraging Economic Waste, Income Inequality, and an Impoverished Cultural Life*, New York: Free Press.

Friedman, Milton (1953) "Choice, Chance, and the Personal Distribution of Income", *Journal of Political Economy*, vol. 61, 4, 277-290.

Garen, John (1984) "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable", *Econometrica* 52, 1199-1218..

Ghez, G. and Becker G. (1975) *The Allocation of Time and Goods Over the Life Cycle*, NBER.

Gibbons, Robert, Katz, Lawerence and Lemieux, Thomas (1997) " Learning, Comparative Advantage, and the Inter-Industry Wage Structure", mimeo Universite de Montreal.

Gibbons, Robert and Katz, Lawerence (1992) "Does Unmeasured Ability Explain Inter-Industry Wage Differ ences?", *Review of Economics Studies*, 59, 515-535.

Gibrat, Robert (1931) *On Economic Inequalities*, Paris, Sirey, chap. 5-7, 62-90.

Green, Jerry and Stokey, Nancy (1983) "A Comparison of Tournaments and Contracts", *Journal of Political Economy*, 91, 349-364.

Grubb, David B. (1985) "Ability and Power over Production in the Distribution of Earnings", *Review of Economics and Statistics* 67, 188-94..

Harris, Milton and Bengt, Holstrom (1982) "A Theory of Wage Dynamics", *Review of Economic Studies* 49, July, 315-33.

Haley, William J. (1973) "Human Capital: The Choice Between Investment and Income", *American Economic Review* 6, December, 929-44.

Heckman, James (1976) "A Life-Cycle Model of Earnings, Learning, and Consumption", *Journal of Political Economy* 84, no.4, Part 2, August, S11-44.

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

Heckman, James and Scheinkman, Jose (1987) "The Importance of Bundling in a Gorman-Lancaster Model of Earnings", *Review of Economic Studies* 54, 243-55.

Heckman, James and Sedlacek, Guilherme (1985) "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market", *Journal of Political Economy* 93, 1077-1125.

Holmstrom, Bengt (1982a) "Managerial incentive problems-A Dynamic Perspective", in *Essays in Economics and Management in Honor of Lars Wahlbeck*, Helsinki: Swedish School of Economics.

Holmstrom, Bengt (1982b) "Moral Hazard in Teams", *Bell Journal of Economics*, 13, 324-340.

Jovanovic, Boyan (1979a) "Job Matching and the Theory of Turnover", *Journal of Political Economy*, 87, Oct., 972-990.

Jovanovic, Boyan (1979b) "Firm-Specific Capital ad Turnover", *Journal of Political Economy*, 87, Dec., 1246-1260.

Kalecki, Michael (1945) "On the Gibrat Distribution", *Econometrica* 13, 161-170.

Kanbur, S. M. (1979) "Of Risk Taking and the Personal Distribution of Income", *Journal of Political Economy*, 87, no. 4, Aug., 769-797.

Katz, Lawrence F. and Murphy, Kevin M. (1992) "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", Quarterly Journal of Economics, February, 35-78.

Kim, Dae Il (1998) "Reinterpreting Industry Premiums: Match Specific Productivity", forthcoming, *Journal of Labor Economics*.

Kremer, Michael (1997) "How Much Does Sorting Increase Inequality?", *Quarterly Journal of Economics*, 112, 115-39.

Kremer, Michael (1993) "The O-Ring Theory of Economic Development", *Quarterly Journal of Economics*, 108, 551-75.

Krueger, Alan B. (1993) "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989", *Quarterly Journal of Economics*, 107, 33-60.

Krueger, A. B., and Summers, L. H. (1988) "Efficiency Wages and the Inter-Industry Wage Structure", *Econometrica*, 56, no. 2, 259-294.

Lazear, Edward P. (1979) "Why is There Mandatory Retirement?", *Journal of Political Economy*, 87, 1261-1264.

Lazear, Edward P. and Rosen, Sherwin (1981) "Rank-order Tournaments as Optimum Labor Contracts", *Journal of Political Economy*, 89, 841-864.

Leamer, Edward (1981) "The Hit Parade of Economics Articles", in E. Tower, editor, *Economics Reading Lists, Course Outlines, Exams, Puzzles and Problems*, Volume 14, July, Eno River Press series.

Lebergott, Stanley (1959) "The Shape of the Income Distribution", *American Economic Review*, 49, 328-47.

Leonard, Jonathan S. (1987) "Carrots and Sticks: Pay, Supervision, and Turnover", *Journal of Labor Economics*, 5, Oct., 136-152.

Lillard, Lee, A. and Willis, Robert J. (1978) "Dynamic aspects of Earning Mobility", *Econometrica* 46, no. 5, 985-1012.

Lillard, Lee A. (1977) "Inequality: Earnings vs. Human Wealth", *American Economic Review*, 67, March, 42-53.

Lovell, Michael C. (1973) "The Production of Economic Literature: An Interpretation", *Journal of Economic Literature*, 11, 27- 55.

Lydall, Harold F. (1959) "The Distribution of Employment Incomes", *Econometrica*, 27, 110-115.

Lydall, Harold F. (1968) *The Structure of Earnings*, London: Oxford University Press.

Maddala, G.S. (1977) *Self-Selectivity Problems in Econometric Models*, in P.R. Krishnaia (ed), Applications in Statistics Amsterdam: North-Holland.

Mandelbrot, Benoit (1960) "The Pareto-Levy Law and the Distribution of Income", *International Economic Review*, vol. 1, 2, 79-106.

Mandelbrot, Benoit (1962) "Paretian Distributions and Income Maximization", *Quarterly Journal of Economics*, 77, 57-85.

Marshall, Alfred (1947) *Principles of Economics: 8th edition*, New York: Macmillan.

Mayer, Thomas (1960) "The Distribution of Ability and Earnings", *Review of Economics and Statistics*, 42, 189-195.

McLaughlin, Kenneth (1988) "Aspects of Tournament Models: a Survey", *Labor Economics*, 9, 225- 256.

Miller, Herman P. (1955) *The Income of the American People*, New York: Wiley.

Miller, Robert A. (1984) "Job Matching and Occupational Choice", *Journal of Political Economy*, 92, no. 6, 1086-1120.

Mincer, Jacob (1958) "Investment in Human Capital and Personal Income Distribution", *Journal of Political Economy*, vol. LXVI, 4, August, 281-302.

Mincer, Jacob (1974) *Schooling, Experience and Earnings*, New York: Columbia University Press for National Bureau of Economic Research.

Mincer, Jacob (1970) "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach", *Journal of Economic Literature*, vol. 8, no. 1, March, 1-26.

Moore, Henry L. (1911) *Laws of Wages*, New York: Macmillan.

Murphy, Kevin J. (1986) "Incentives, Learning, and Compensation: a Theoretical and Empirical Investigation of Managerial Labor Contracts", *Rand Journal of Economics*, vol.17, 1, Spring, 59-76.

Murphy, Kevin J. and Welch, Finis (1992) "The Structure of Wages", Quarterly Journal of Economics, February, 285-326.

Nalebuff, Barry J. and Stiglitz, Joseph (1983) "Prizes and Incentives: Toward a General Theory of Compensation and Competition", *Bell Journal of Economics*, 14, 21-43.

Neal, Derek (1993) "Supervision and Wages across Industries", *Review of Economics and Statistics*, August, 409-417.

Paglin, Morton (1975) "The Measurement and Trend of Inequality: A Basic Revision", *American Economic Review*, 65, 598-609.

Pareto, Vilfredo (1897) *Cours d'Economie Politique*, Vol II, Lausanne: Rouge & Cie.

Pen, Jan (1971) *Income Distribution: Facts, Theories, Policies*, New York: Praeger.

Reder, Melvin, (1968) "The Size Distribution of Earnings", *The Distribution of National Income*, Marchal and Ducros eds., chap. 21, 583-617.

Reder, Melvin (1969) "A Partial Survey of the Income Size Distribution", in Lee Soltow, ed., *Six Papers on the Size Distribution of Income*, New York: NBER, distributed by Columbia University Press.

Rosen, Sherwin (1972) "Learning and Experience in the Labor Market", *Journal of Human Resources*, 7, no. 3, 326-342.

Rosen, Sherwin (1976) "A Theory of Life Earnings", *Journal of Political Economy*, 84, no. 4, Part 2, August, S45-67.

Rosen, Sherwin (1978) "Substitution and Division of Labor", *Economica*, 45, 235-50.

Rosen, Sherwin (1981) "The Economics of Superstars", *American Economic Review*, 71, 845-58.

Rosen, Sherwin (1982) "Authority, Control, and the Distribution of Earnings", *Bell Journal of Economics*, 13, 311-23.

Rosen, Sherwin (1983a) "A Note on Aggregation of Skills and Labor Quality", *Journal of Human Resources*, 18, 425-31.

Rosen, Sherwin (1983b) "Specialization and Human Capital", *Journal of Labor Economics*, 5, 3-49.

Rosen, Sherwin (1986) "The Theory of Equalizing Differences", in Ashenfelter, Orley and Layard, Richard, (eds.), *Handbook of Labor Economics*, 1, Oxford and Tokyo: North-Holland, 641-692.

Rosen, Sherwin (1992) "Contracts and the Market for Executives" in Werin, Lars; Wijkander, Hans, eds., *Contract Economics*, Cambridge, Mass. and Oxford: Blackwell, 181-211.

Rosen, Sherwin (1997) "Manufactured Inequality", *Journal of Labor Economics*, 15, Apr.,189-196.

Ross, Stephen A. (1973) "The Economic Theory of Agency: The Principal's Problem", *American Economic Review*, 63, 134-139.

Ross, Stephen, Taubman, Paul and Wachter, Micheal (1981) "Learning by Observing and the Distribution of Wages", in Rosen, Sherwin (ed.), *Studies in Labor Markets*, 359-386.

Roy, A.D. (1950) "The Distribution of Earnings and of Individual Output", *Economic Journal*, 60, 489-505.

Roy, A.D. (1951) "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers* (N.S.), 3, 135-46.

Rutherford, R.S.G. (1955) "Income Distributions: A New Model", *Econometrica*, 23, 277-294.

Ryder, Harl E., Stafford, Frank P. and Stephan, Paula E. (1976) "Labor, Leisure and Training Over the Life Cycle", *International Economic Review*, vol. 17, 3, 651-674.

Sattinger, Michael (1993) "Assignment Models of the Distribution of Earnings", *Journal of Economic Literature*, 31, 831-80.

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

Schultz, Theodore W. (1975) "The Value of the Ability to Deal with Disequilibria", *Journal of Economic Literature*, 13, September, 827-46.

Shapiro, C. and Stiglitz, J. E. (1985) "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review*, 74, no. 3, 433-444.

Simon, Herbert A. (1955) "On a Class of Skew Distributions", *Biometrica*, 42.

Simon, Herbert A. (1957) "The Compensation of Executives", *Sociometry* 20, 32-35.

Singh, S.K. and Maddala, G.S. (1976) "A Function for Size Distributions of Income", *Econometrica* 44, 963-70.

Spurr, Stephen J. (1987) "How the Market Solves an Assignment Problem: The Matching of Lawyers with Legal Claims", *Journal of Labor Economics*, 5, no.4, Oct., 502-532.

Staele, Hans (1943) "Ability, Wages and Income", *Review of Economics and Statistics*, 25, 77-87.

Steindl, Joseph (1968) "Size Distributions in Economics", *International Encyclopedia of Social Sciences*, vol. 14, Macmillan and Free Press, New York, 294-300.

Tinbergen, Jan (1959) "On the Theory of Income Distribution", *Selected Papers* Amsterdam: North-Holland.

Tuck, R.H. (1954) *An Essay on the Economic Theory of Rank*, Oxford: Blackwell.

Willis, Robert J. and Rosen, Sherwin (1979) "Education and Self-Selection", *Journal of Political Economy*, 87, part 2, 7-36.

Weiss, Andrew (1990) *Efficiency wages: Models of unemployment, layoffs, and wage Dispersion*, Princeton: Princeton University Press.

Weiss, Yoram and Lillard, Lee A. (1978) "Experience, Vintage, and Time Effects in the Growth of Earnings: American Scientists, 1960-1970", *Journal of Political Economy*, 86, no. 3, 427-448.

Weiss, Yoram (1986) "The Determination of Life Cycle Earnings", in Ashenfelter, Orley and Layard, Richard, (eds.), *Handbook of Labor Economics*, 1, Oxford and Tokyo: North-Holland, 603-640.

Weizsacker, Robert K. von (1993) *A Theory of Earnings Distribution*, Cambridge University Press.

Welch, Finis (1969) "Linear Synthesis of Skill Distribution", *Journal of Human Resources*, 4, 311-327.

Welch, Finis (1970) "Education in Production", *Journal of Political Economy*, 78, no.1, 35-59.

Welch, Finis (1979) "Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust", *Journal of Political Economy*, 87, no. 5, October, S65-98.

Wold, Herman O. A. and Whittle, P. (1957) "A Model Exploring the Pareto Distribution of Wealth", *Econometrica*, 25, 591-595.

Figure 2.1

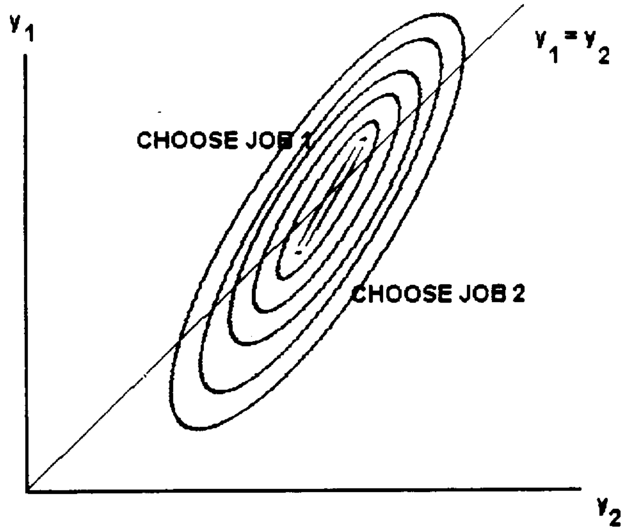
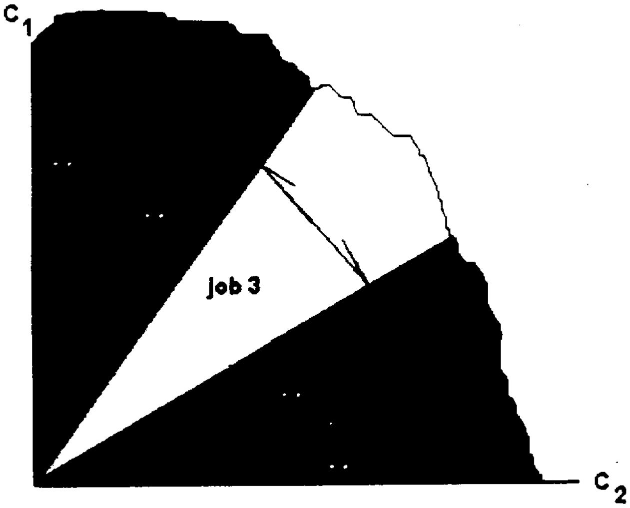
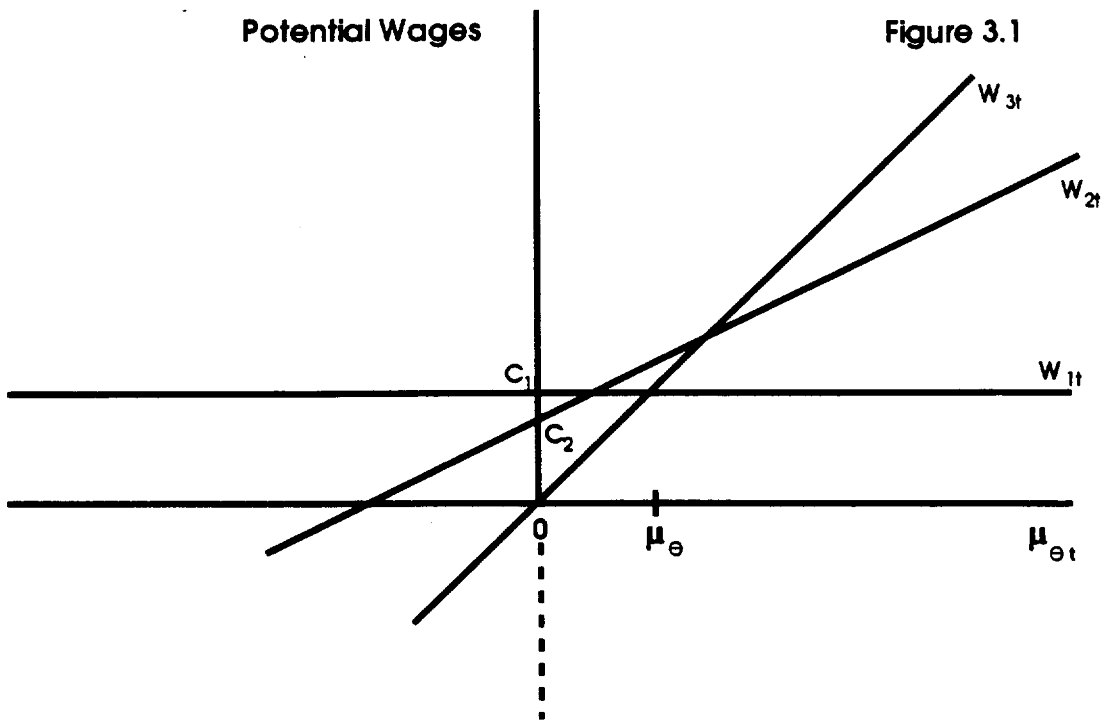


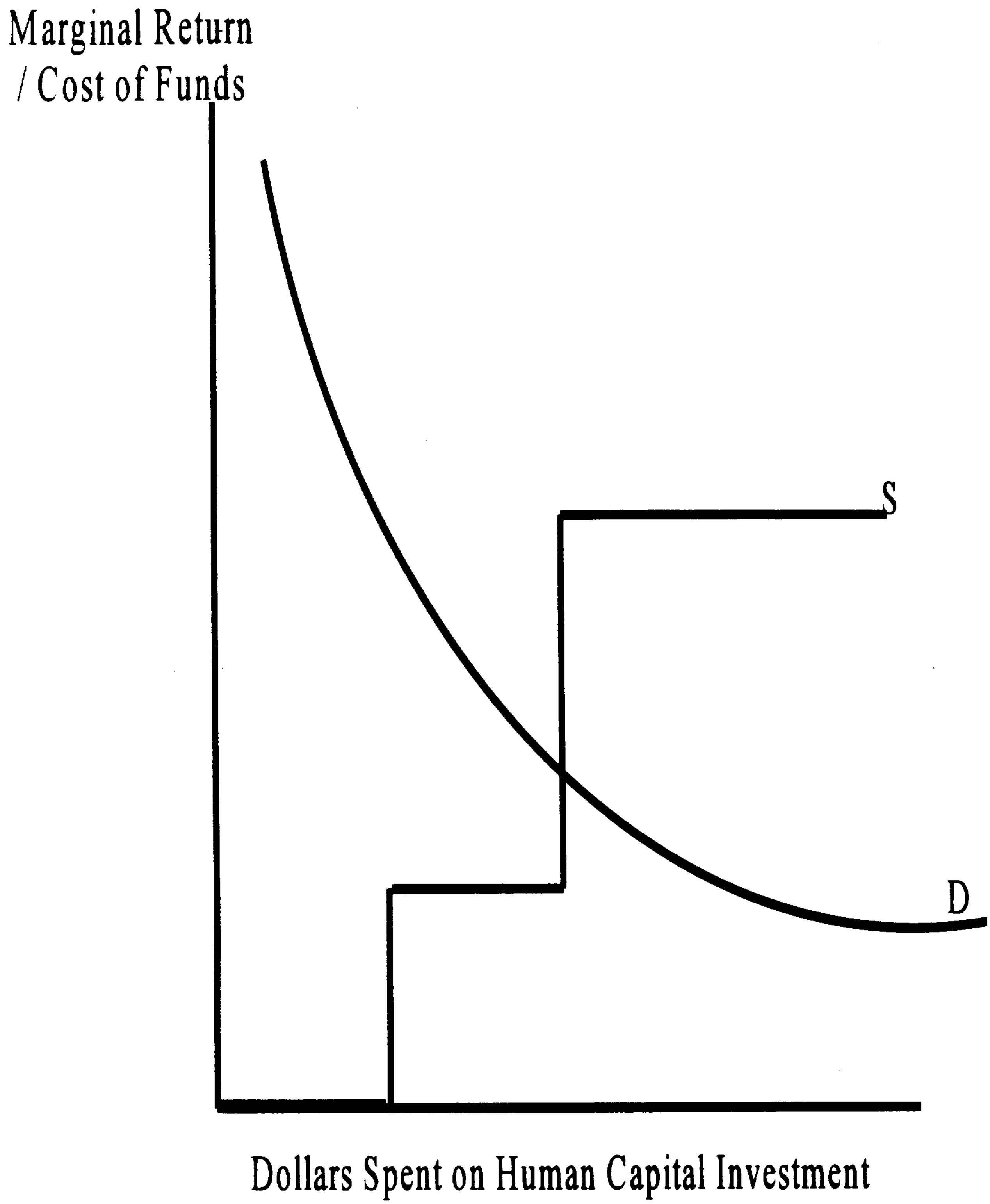
Figure 2.2





Potential Wages
Figure 3.1
W 3t
W 2t
W 1t
C1
C2
0
μθ
μθt

Figure 4.1



Marginal Return
/ Cost of Funds
S
D
Dollars Spent on Human Capital Investment

**Figure 4.2**

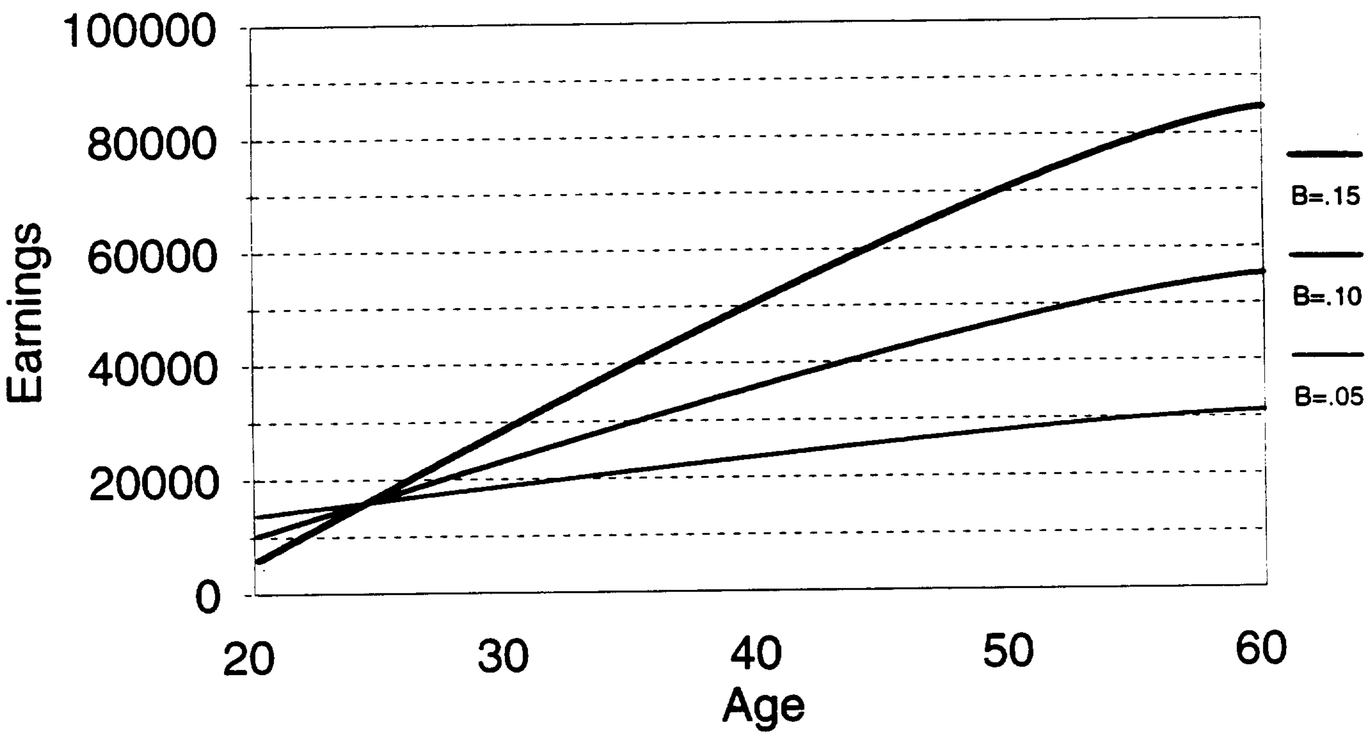

100000
80000
60000
40000
20000
0
Earnings
20
30
40
50
60
Age
B=.15
B=.10
B=.05

Figure 5.1