NBER WORKING PAPER SERIES

THE EQUILIBRIUM APPROACH TO LABOR MARKETS

Sherwin Rosen

Working Paper No. 1165

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

July 1983

The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #1165
July 1983

The Equilibrium Approach to Labor Markets

ABSTRACT

This paper exposits the modern theory of equalizing differences, viewed as optimal assignments of workers to jobs. The basic ideas are first illustrated in a simple model with binary choices of work attributes. Multinominal choices are briefly considered after that. Empirical implications are stressed, with special emphasis on elements of selectivity and stratification by tastes and technology. Applications are sketched for certain aspects of the economics of discrimination, human capital, the value of safety and the theory of implicit contracts. Issues raised by assignment stratification according to worker traits and productivities are discussed, and the principle sorting model by comparative advantage is outlined. The implied valuation system on personal traits and its relationship to factor-analytic models, as well as selectivity issues in educational and occupational choice illustrate this aspect of the theory.

Sherwin Rosen
The University of Chicago
Department of Economics
1126 E. 59th Street
Chicago, IL 60637

(312) 753-1234

THE EQUILIBRIUM APPROACH TO LABOR MARKETS

Sherwin Rosen*

University of Chicago

I. Introduction

This paper elaborates variations on some themes of Adam Smith, particularly those in the first ten chapters of the Wealth of Nations. Those chapters are without parallel in economics and still define research frontier subjects in economics today. In reading those chapters one is struck by the timeless importance of the subject matter, by the quality of Smith's writing and the clarity of his exposition, and by the great scope and depth of his ideas. Two central themes stand out:

The first and most important for economic analysis as a whole is the discussion of gains from trade due to specialization and division of labor. Specialization exploits scale economies which multiply the fruits of labor resources. The resulting division of labor provides the basis for mutually advantageous exchanges of goods and services among economic agents. This is the sine qua non of decentralized competitive market organization, and is of course fundamental to the main theme of the work as a whole. It has been the main intellectual preoccupation of economists ever since. While the limitations of decentralized market allocation mechanisms are now well understood, there is small irony in the fact that Smith's own argument rests on scale economies that are not entirely consistent with competitive market organization of economic activities.

The second theme, and the one we will be most occupied with here, is the celebrated theory of equalizing differences. This is the basic equilibrium theory in labor economics. It rests on the proposition that wages paid to various types of labor and under various circumstances must equalize total advantages, both pecuniary and nonpecuniary, among them. It is interesting, but would take us too far afield, to track the fortunes of this theory over the years. Nevertheless, it is worth mentioning that the institutional approach to labor economics, which itself dominated the subject for most of its life, was far more concerned with anomolies and noncompeting group qualifications of Mill and Cairnes than with Smith's contributions themselves. Indeed, the theory of equalizing differences, which is fundamentally a theory of spatial equilibrium, found much greater direct use in agriculatural and urban economics, stemming from its development to those fields by von Thunen. There is no question, however, that whatever its status in the past, Smith's theory rests securely at the center of modern labor economics. His analysis of professional income was the first rigorous theory of human capital. His treatment of the influence of random fluctuations in income on occupational choice is closely related to the modern theory of implicit contracts. His treatment of consumption elements of work environments is essential for empirical understanding of the structure of wages. The work even contains a nascent statement of the principal and agent problem! The theory of labor supply, production theory, and some elements of economic demography are the only subjects of inquiry that are missing from Smith's treatment.

What follows is an analytical sketch of the theory of equalizing differences and some of its applications. These applications hardly are exhaustive, but rather illustrate the approach from studies in which I have had a hand and am most familiar with. I hope they convey the power, simplicity, and practical value of the theory. The basic model is sketched in Section II. It takes off on the treatment contained in Friedman and Kuznets, Income from Independent Professional Practice, but gives a more complete account of the nature of equilibrium and some of its implications. Section III outlines some applications. These include elements of the economics of discrimination; human capital; imputations of the value of work safety and the valuation of life in risk-benefit analysis; and the nature of implicit contracts and income risk. Section IV extends the model to valuations of personal traits and specific skills rather than job and work-environment characteristics. This is necessarily incomplete because several conceptual problems have not been resolved at this date. Implications for income distribution, and selectivity bias in occupational and educational choice are the examples chosen to illustrate these aspects of the theory.

II. The Theory of Equalizing Differences

The theory of equalizing differences is fundamentally a theory of valuation of job attributes. In this account I specialize the job attribute to a consumption item, so the problem is basically one of a tie-in in which the worker sells the services of his labor but simultaneously purchases the characteristics of his job, viewed as on-the-job consumption. In this model work attributes are fixed for a given work situation, but vary from job to job. The basic problem is to match each worker's preferences for on-the-job consumption versus market consumption to the proper work attribute; that is to say, to assign each worker to a firm which offers the desired job characteristics. A twin choice problem must be addressed: workers choose among job attributes by working for firms that offer the desired amounts; and firms choose which job attributes to offer. A competitive price mechanism guides these choices and takes the form of wage differentials -- equalizing differences -- on jobs with different attributes. The tie-in nature of attribute transactions makes clear that the problem is basically one of spatial equilibrium, assignment, and sorting.

Most of the basic ideas are illustrated in the simplest possible case. Consider a job characteristic D which is discrete and binary: D takes on two values, say 0 or 1. To be specific, suppose D indexes a disamenity such as airborne particulates at the workplace. Hence a job is either "dirty," in which case D = 1 (some particulates); or it is clean (no particulates), in which case D = 0. We simplify further by assuming that all workers are alike in terms of their basic underlying skills (this assumption will be relaxed later) and that the job attribute D does not in and of itself directly affect any worker's productivity. This means that D, or its absence, is a pure consumption good from a worker's point of view. Allowing for direct productivity interactions complicates the analysis, but does not affect the main conclusions. In these circumstances there must be two possible wages in the market: W_1 for jobs characterized by D = 1 and W_0 for jobs characterized by D = 0. The wage differential between them is the equalizing difference.

A. Worker Choice and Supply

A worker chooses the value of D that maximizes utility subject to a budget constraint. Preferences are described by a utility function u = U(C,D) defined over market consumption goods C and job consumption goods (attributes) D. As usual, U(C,D) represents the internal trade-offs between C and D that the worker would find acceptable. Since we are here interpreting D as a disamenity (a "bad" rather than a "good"), the map of U(C,D) in the C,D plane has convex indifference curves that are positively inclined rather than negatively inclined as in the standard case. (It is always possible to convert a disamenity problem into an amenity by an appropriate change in scale, e.g., consider the cleanliness of a job rather than its dirtiness, but the result is the same in either case.) In deciding on the choice of job D, the worker contemplates the fact that his labor services are exchanged for money which buys market consumption goods and that D is an undesirable by-product of the work environment chosen. It is clear that if the worker dislikes larger values of D, compensation in the form of additional market consumption is necessary for him to voluntarily choose it compared to a job with a smaller value of D. This is the common sense of the idea of equalizing differences. Preferences alone therefore predict that $W_1 \geq W_0$. Still, we are interested in the determinants of how large the difference between W_1 and W_0 might be and also what can be inferred from wage-attribute data.

When D is binary, the budget constraint reduces to two points.[1] Since $C = W_i$, for i = 0, 1, the available consumption possibilities in the (C,D) plane are $(W_1,1)$ and $(W_0,0)$. The worker chooses the point, and therefore the value of D, which lies on the highest indifference curve. Consequently, the choice is D = 1 if $U(W_1,1) > U(W_0,0)$ and is D = 0 if the inequality is reversed. The worker is indifferent between the two jobs if $U(W_1,1) = U(W_0,0)$. In fact, given W_0 it is possible to find a W_1 that would equalize utility between D = 1 and D = 0. That value of W_1 compared with W_0 is known as the reservation price. It represents the equalizing difference for that particular worker. Therefore an equivalent description of the solution to the worker's choice problem is to compare the actual, available market wage difference between the two jobs with this reservation price. If the difference in market wages exceeds the reservation wage, the worker chooses D = 1 because the additional market goods available from that choice more than compensates for the additional disagreeableness of the job. Similarly, the worker chooses D = 0 if the market wage differential is less than the reservation price.

To illustrate, suppose preferences are described by a one parameter family of curves $u = Ce^{-\rho D}$. The parameter ρ is a measure of the worker's distaste for D. The larger is ρ the greater the additional market compensation necessary to bribe the worker to undertake a job with a larger value of D. To analyze the choice problem, form an index function defined by $I = \log U(W_1,1)/U(W_0,0) = \log (W_1/W_0) - \rho$. The reservation wage is defined as the value of (W_1/W_0) that sets $I = 0$. Define the reservation wage as R. Then substituting into the above yields $R = e^{\rho}$, which is an increasing function of ρ. The worker chooses $D = 1$ if $I > 0$, or equivalently, if $(W_1/W_0) > e^{\rho}$. $D = 0$ is chosen if $I < 0$, of if $(W_1/W_0) < e^{\rho}$. Choice is made by a random device if $(W_1/W_0) = e^{\rho}$.

Aggregation over worker choices is required to describe market supply conditions to each type of job. In the example given, a worker is completely characterized by the taste parameter ρ. Different workers might have different preferences, and that is conveniently described by a distribution function over all workers. Denote that distribution by $g(\rho)d\rho$, which is readily transformed into a distribution of R, say $h(R)dR$. Since all workers for whom $R < (W_1/W_0)$ choose D = 1, the supply function of workers to jobs of type 1 must be the sum

$$(1) \qquad N_1^s = \int_{-\infty}^{W_1/W_0} h(R)dR \ .$$

The supply to jobs for which D = 0 is just the rest of the distribution, the integral of $h(R)$ from W_1/W_0 to the largest value of R.

Expressed in this way it is apparent that the market assignment of workers to job characteristics neatly partitions the distribution of tastes into two parts. The situation is illustrated in Figure 1. Consider a given market wage ratio (W_1/W_0) marked by the heavy vertical line in the figure. Then by the choice rule set forth above, the area under the distribution $h(R)$ to the left of W_1/W_0 is N_1^s, the number who supply their labor to dirty work. Similarly, the area in the right tail of the distribution is the number who supply their labor to clean jobs. The figure clearly shows that were, say (W_1/W_0) to increase, those people who were close to the margin of indifference at the old wage ratio would switch over and choose D = 1 rather than D = 0. Therefore, the supply curve of workers to dirty jobs is increasing in (W_1/W_0) and literally sweeps out the transformed distribution of preferences for D as captured in $h(R)$.

Were it possible to econometrically identify and estimate supply curve (1) or its counterpart N_0^s, it would be straightforward to impute the entire underlying distribution of preferences h(R) or g(ρ), conditional of course on the assumed form of the underlying utility function. However, it is usually not possible to do that because of data limitations. At this point most studies must rest content with estimating the equilibrium wage differential (W_1/W_0) in a cross section, itself not a trivial task. This identifies only one point on the distribution h(R), but still provides some information on the conditional first moments of h(R) or g(ρ).

Consider the distribution of R in Figure 1 conditional on D = 1 having been chosen. Then we know from the definition of a conditional expectation that the average value of R (and therefore of ρ) for those who choose D = 1 must be less than the observed value of (W_1/W_0). It also must be less than or equal to the unconditional expectation of R in the entire population. That is to say, the observed value of log (W_1/W_0) is an upper bound estimate of the conditional mean of ρ for D = 1. By a parallel argument, it is a lower bound estimate of the conditional mean value of ρ for those who choose D = 0. If we have prior information that the variance of h(R) is small, if individuals have very similar tastes, then the estimated value of W_1/W_0 tells us much more about those tastes than if the variance is large. If we furthermore have reason to believe that most workers do not work on D = 1 jobs so that N_1^s is a small fraction of available workers in the market, then we can confidently predict that the observed value of log (W_1/W_0) is a lower bound on the average value of ρ in the population as a whole. While not providing complete information about preferences, these inequalities are very useful in many practical applications of the theory. Note that the choice rule and market assignment in Figure 1 sorts people to jobs in a systematic way and that the people who are observed in each category form a censored sample of the population as a whole. This is perhaps the most simple and fundamental example of selectivity in labor economics. It should be apparent that it is a very general implication of the theory of revealed preference and applies to virtually all economic choices. Another way of saying this is that there are economic rents inherent in these choices, since most people who make a given choice would continue to make it even if wages were somewhat different. It is only when all persons have the same tastes that rents disappear and that observed wages index the equalizing difference for the whole population, for the average person in the market as well as for the marginal one.

B. Firm Choice and Demand

If most workers dislike D we know from the analysis above that W_1 generally must exceed W_0. But higher wages on D = 1 jobs must be supported somehow and what sustains them must be larger productivity among firms who find it optimal to offer that kind of work. Clearly if that was not the case we would never observe any jobs for which D = 1. All jobs would offer D = 0 and would be preferred by everyone. Hence we conclude that in some general sense attributes that are disamenable must be productive. It does not follow though that job attributes which are amenable must be counterproductive. Rather, if the two coexist, disamenities must exhibit a productivity advantage to the firms that offer them. This is the basis for ascertaining which types of jobs a given firm chooses to supply to the labor market. For this purpose, and analogous to its role in workers' utility functions, D must enter as an argument into a firm's production function. The choice of D is then made by comparing the enhanced productivity of a given labor force when $D = 1$ with the additional cost of labor due to the fact that W_1 rather than W_0 must be paid to each worker. If the additional value of productivity is larger than the additional wage bill the firm chooses $D = 1$; while, if the productivity effect is small, dirty work does not generate sufficient revenue to cover the extra labor costs, and the firm cleans up its technology. It installs ventilation equipment and uses other resources for cleaning up its environment and offers $D = 0$.

To illustrate this in the simplest possible way, consider a fixed coefficient technology with production function $x = \alpha_1 n$ if $D = 1$ and $x = \alpha_0 n$ if $D = 0$. Here x is output, n is labor input, and α_i are fixed coefficients with $\alpha_1 > \alpha_0$. The last condition reflects the fact that productivity is larger if $D = 1$ is chosen, that some otherwise productive labor must be used up in the cleaning process if $D = 0$ were chosen instead. With this technology unit cost is W_i/α_i, for $i = 1,0$. The firm chooses the value of D which minimizes unit costs. Define another index function $I^* = (W_1/\alpha_1) \div (W_0/\alpha_0) = (W_1/W_0)(\alpha_0/\alpha_1) \equiv (W_1/W_0)(1/\beta)$. Then $D = 1$ or 0 as $I^* < 1$ or as $(W_1/W_0) \lessgtr \beta$. If $I^* < 1$ the productivity effect outweighs the added labor cost; and if $I > 1$ added productivity is insufficiently large to compensate for added labor costs. In this case we have that β is itself the firm's reservation wage ratio, the firm's equalizing difference. If the market wage ratio happens to equal β the firm is indifferent about its choice of D.

To obtain the market demand for workers in type $D = 1$ jobs it is necessary to aggregate among firms that choose to offer them: Suppose firms differ in their technology ratio β and that $f(\beta)d\beta$ describes the distribution of β among firms. Then since $\beta > (W_1/W_0)$ describes firms who choose $D = 1$, the number of such workers demanded is[2]

$$(2) \qquad N_1^d = \int_{W_1/W_0}^{\infty} f(\beta)d\beta \ .$$

The partition of firms among $D = 1$ and $D = 0$ is displayed in Figure 2. Equation (2) defines a curve in the $N,(W_1/W_2)$ plane that is negatively inclined. Now all those to the right of the heavy vertical line choose $D = 1$ and all those to the left choose $D = 0$. As (W_1/W_0) falls the market demand sweeps out the distribution of β among firms from right to left and the demand for workers in dirty jobs increases. The same arguments about conditional moments as in Figure 1 apply to this distribution and will not be repeated. Suffice it to say that a single observed value of the market wage ratio measures the equalizing difference for firms on the margin of choice only and marginal and average firms depart in systematic ways unless all firms have the same technologies.

It is common in economics today to maintain the assumption that all firms in a given industry have access to the same technology, so some comment on the rationale for a nondegenerate distribution of β is warranted. The most important point is that the construction in Figure 2 is not necessarily confined to a given industry. Variance in β is produced by interindustry differences in technologies. Some production processes are inherently dirtier than others. While there is no technology that cannot use resources to clean up its work environment, it is undoubtedly more costly to do so in some industries than in others; e.g., think of coal mining versus insurance. The second point is that there is nothing sacred about the assumption of identical technologies in a given industry, though it is sometimes convenient. Indeed, many aspects of firms, such as differences in their size in the same industry, are simply not consistent with identical technologies. Whatever firm-specific factors cause these differences may well interact in nonseparable ways with job attribute-output tradeoffs, even in the same industry.

C. Market Equilibrium

The joint solution to market supply and demand functions (1) and (2) determines the wage ratio (W_1/W_0) that is observed in the market.[3] Since supply price is increasing and demand price is decreasing, an equilibrium in which some workers are observed in both types of jobs must be unique, if it exists. The equilibrium wage ratio is determined so that the partitions in Figures 1 and 2 are conformable, so that all workers seeking $D = 1$ jobs are able to find them and all firms seeking workers on such jobs can fill them. It is interesting to note that the equilibrium assignment of workers to firms exhibits a negative assortive matching property. Workers with lower than average values of ρ are matched to firms with higher than average values of β. Workers who have the least distaste for the disamenity work in firms for whom the disamenity is the most productive. The equilibrium assignment matches the proper worker type to the proper firm type, analogous to a marriage market. This kind of allocation is characteristic of all spatial equilibrium problems, but the question of who works for whom is especially important in the labor market. Finally, it is also obvious that the equilibrium assignments and choices are Pareto optimal when information is complete.

D. Generalization

While binary choices takes us a long way toward understanding the issues raised by equalizing differences and what can be inferred from the data, it remains true that most job attributes exhibit much more variation. Thus to continue along with the problem above, jobs differ greatly in the degree of cleanliness offered. A straightforward generalization of the model is to consider multinomial choice rather than binary choice. I used the index function approach to suggest a probit analog: more choices would require a multinomial probit or logit approach. Thus let D take on k possible values, with $k \geq 2$. Since D is ordered, let larger values of k index correspondingly large values of D. Then k distinct markets must be considered. The competitive wage in the kth market is W_k and the budget constraint for a worker is represented by k distinct points (W_j, D_j), for $j = 1, 2, \ldots, k$. The worker chooses that value of j which maximizes utility. While conceptually straightforward the problem is difficult to analyze for general utility functions because a computational algorithm that makes pairwise comparisons between all possible choices is required. It is easy to see that the optimal choice depends not only on local curvature properties of preferences, but on global curvature as well. Nonetheless, it is clear that the ordering property of the optimal assignment by tastes and relative costs shown in Figures 1 and 2 are more or less preserved. Thus with suitably regular parameterizations of preferences, the distributions are partitioned into at most k ordered regions. Workers with the largest values of ρ are assigned to the smallest values of j and firms whose cleaning costs are largest are assigned to the largest values of j. The negative assortive matching feature of market equilibrium is thereby generalized.

A marginal analysis serves to illustrate this very nicely when k is so large and D is sufficiently divisible that there are an infinite number of choices for all practical purposes.[4] Then D may be represented as a continuous variate, measured say in parts per million particulates. There remains a wage associated with every value of D, so now income possibilities for a worker are represented by a continuous function W(D), which is nondecreasing if D is a disamenity. The worker maximizes utility subject to C = W(D): therefore D is chosen to maximize u = U(W(D),D). A maximum is characterized by the marginal condition $-U_D/U_C = W'(D)$. Here U_D/U_C is the marginal rate of substitution between D and consumption goods and is negative if D is disamenable. Notice the slight variance from a standard constrained maximum problem in that the gradient of W(D) is the correct (marginal) price in the optimization calculation, not W(D) itself. Notice also that W(D) need not be linear, so the marginal price W'(D) may vary with D. The solution is represented as a proper spatial equilibrium in Figure 3. The curves labeled θ^1 and θ^2 are (C,D) indifference curves for two different types of workers. Worker 1 exhibits a greater distaste for D and chooses a smaller value in equilibrium.

A similar development, wherein D shifts production possibilities rather than tastes, is available for firms. I omit it here. A summary of the solution is also depicted in Figure 3 by profit indifference curves in the (W,D) plane, labeled ϕ^1 and ϕ^2 for two different types of firms. ϕ^1 type firms find it easier to provide clean workplaces than ϕ^2

type firms do and therefore choose to offer smaller amounts of D to the market. The equilibrium assignment allocates worker taste types to firm technology types in a systematic manner. That a profit indifference curve "kisses" a worker's indifference curve at the equilibrium assignment best summarizes the marriage aspects of the problem solved in the implicit market for job attributes.

Figure 3 well illustrates the revealed preferences, sorting aspects of the equilibrium assignment and shows what can be inferred from the observed wage-attribute schedule W(D). For example, it is apparent that the gradient W'(D) identifies the marginal rate of substitution only for workers and firms who happen to choose that particular value of D. Still, when workers are approximately identical in their preferences then W(D) identifies an indifference curve and W'(D) measures the marginal rate of substitution all over the map. Similarly, were firms identical rather than workers, W(D) would coincide with a profit indifference curve and its gradient function would closely approximate the marginal cost function for achieving smaller values of D. When firms and workers are both heterogeneous, the data are censored and selected by the optimal assignment. Thus for example, the difference in wages between D_1 and D_2 is an underestimate of the equalizing difference required for type 1 persons -- that is why type 1's are located at D_1 rather than at D_2; and it is an overestimate for type 2 persons, who evidently found the wage premium sufficiently large to more than buy-off their distastes and who chose D_2 instead of D_1. Hence, if W(D) is estimated over its upper range, for the largest values of D, it could be confidently predicted that the gradient W'(D) in that range underestimates the average person's marginal rate of substitution, because their intrinsic distaste for D was much larger, by revealed preference. Comparable statements can be made about firms.

III. Applications

A. Value of Safety[5]

Recent years have seen an explosion of interest in this topic that pervades many aspects of environmental legislation, workplace safety regulation, food and drug safety, consumer safety, and so on. It is now well understood that proper cost/benefit analysis of alternative policies requires both an estimate of the magnitude of risks involved and some valuation of the additional safety that might be provided by the policy. Following general economic practice, the appropriate valuation of risk is the willingness to pay to reduce it. Let V measure this sum for any given person. It is defined as the marginal rate of substitution between consumption and mortality risk and is sometimes labeled the "value of life." To motivate that terminology, consider the following conceptual experiment. Think of a large group of N people who contemplate a project that would reduce morality risk by 1/N. Then each would be prepared to pay approximately V/N for the project and they would collectively pay $N(V/N) = V$. Since the project reduces mortality by 1/N and N people are involved, approximately one statistical life is saved; hence these people are prepared to pay V for one statistical life.

Thoroughgoing analysis of safety has been hampered by lack of direct measures of valuations V. As is usually the case in economics, it must be inferred from actual behavior of persons in risky situations. A basis for inference is provided by the common observation that people do in fact voluntarily undertake many risks in their everyday lives and do so by weighing the perceived costs and benefits of their actions. Nowhere is this so apparent as in the labor market, where we observe many jobs with substantial risks to health and longevity that pay correspondingly large wages. This is a straightforward application of the theory of equalizing differences: if workers find health risks distasteful, jobs that involve considerable perceived risks must bribe workers to accept them by paying a wage premium. The observed wage premium, in conjunction with the size of the risk therefore provides a possibility for inferring V from the risk premium.

Consider a worker with von-Neuman-Morgenstern utility $(1-q)U(C)$, where q is the risk of a job and C is consumption. It is readily verified that the marginal rate of substitution between q and C is $U(C)/(1-q)U'(C) \equiv V$. Suppose the worker has an opportunity to work in jobs of various risk q which pay wages $W(q)$, with $W'(q) > 0$. The worker chooses q and C to maximize expected utility subject to the constraint $W(q) = C$. Substituting into the utility function and differentiating with respect to q yields the marginal condition $V = W'(q)$. Therefore the wage gradient provides an estimate of the marginal value of life V at q. The analysis in Section II applies virtually intact, with q replacing D -- see Figure 3. Only if all workers had the same preferences would it be true that $W(q)$ would cover a unique indifference curve. More generally, workers have different tastes. For many reasons some are more risk averse and have higher values of V than others. Then revealed choices suggest that workers found on riskier jobs have lower values of V than those who work on safe jobs. However, if we can find the wage premium on very risky jobs that should serve at least as a lower bound estimate of the average value of V in the population as a whole.

Econometric estimates of W(q) are obtained by regression or related methods. Required data are wage rates, risks workers are exposed to, and measures of personal characteristics such as schooling, experience, and other variables that are known to affect wages and which serve as statistical controls. Two types of risk measures are available for this purpose: occupational and industry risks. Both are obtained from accident statistics collected by the federal government and from life insurance company records. They are matched to earnings and related data available from census survey records. All studies undertaken so far have shown that the empirical wage-risk gradient is positive and prove the feasiblity of the approach. Having said that, however, there is far less agreement from study to study on the magnitude of the gradient and therefore on the size of V. Studies using occupational risk data provide estimates of V that are systematically smaller than studies using industrial risk data. The former estimates are in the vicinity of $500,000 (in 1983 dollars), whereas the latter estimates range as high as $2M or more. The reason for these substantial differences in the estimates has not yet been resolved, but probably lies in the crudeness of the risk measures available. It is interesting to note that estimates inferred from observed risk choice behavior outside of the market, such as cigarette smoking, tend to corroborate these figures, wide as their range may be.

B. Economics of Discrimination

It is in this area that the theory of equalizing differences has found its earliest and most widespread use. Discrimination is viewed as arising from tastes or distastes of association with identifiable groups in the workplace or other environments. Such preferences, which in some contexts may be viewed as socially illegitimate, effectively serve to tax members of despised groups and subsidize members of favored groups. The theory of tax incidence may then be applied to predict the distribution and size of wage differentials among workers.[6] Since this theory and the many empirical studies that support it are so well known, I have chosen a less familiar example drawn from the market for public school teachers.[7] This example is not only interesting in its own right, but considerably broadens the scope of the theory of equalizing differences and raises issues that apply to many other labor markets as well.

We seek to study the implicit valuation of student attributes by teachers, particularly the racial composition of the student body within a school. For example, how much additional pay, if any, is required to entice a white teacher to work in a school with mainly black students? Answers to questions such as these have obvious relevance for estimating real educational costs indexes necessary to implement Equal Educational Opportunity policies. The analytical issues raised by this problem involve a nontrivial extension of the theory which has much broader applicabiilty. While teachers may have well defined preferences for schools and students of various characteristics, it is also true that schools may well have distinct preferences for various types of teachers and <u>their</u> attributes. The matching problem is therefore much more complicated than was indicated in Section II.

Denote school characteristics by the vector S and teacher characteristics by the vector T. A teacher endowed with a particular value of T searches out a school with the desired value of S, given the wage prospects available. Similarly, a school is endowed with a particular value of S and searches for teachers with desired characteristics T because teaching effectiveness may differ among persons with different traits for a particular composition of the student body. A match occurs when desired values of T and S are conformable with each other. It is particularly interesting that the matching problem gives rise to possibilities for trade refusal. A given teacher may desire to work at a particular school because it offers a preferred wage and student characteristic configuration. But the school may not be willing to hire him if he does not possess desirable teaching attributes T relative to someone else. Similarly, a school may desire to hire a particular teacher, but may not offer the value of S necessary to attract him. The equilibrium concept therefore must be extended to cover the joint space (S,T), which implies that the equilibrium pricing mechanism is defined over both sets of variables: W(S,T) is the market clearing wage for any feasible S,T combination.

A teacher's utility function is defined over market consumption C and school attributes, as before: $u = U(C,S)$. Choice of S is found by maximizing U subject to the constraint $C = W(S,T)$, given the teacher's particular value of T, leading to the marginal condition $-W_S(S,T) = U_S/U_T$. Conditioning the choice on T is necessary for feasibility, given the definition of W(S,T). Therefore the S subgradient of the observed wage-attribute function measures the marginal valuation of S for those teachers who were able to choose it. Revealed preference-selectivity bias again applies for persons who are not located at that particular margin.

A school's choice of teachers is made on the basis of the effects of teacher traits T on educational output, represented in the educational production function $E = F(T,S)$, where E is educational value-added per student. Notice that T and S strongly interact in production, if the effectiveness of a teacher of given traits varies according to the characteristics of the school and students to which he is assigned. School administrators serve as agents for parents and choose teachers with traits that minimize costs given E and S (or equivalently that maximize E given costs and S). This leads to the marginal condition $W_T(S,T) = \lambda F_T(T,S)$, where λ is the marginal cost of E. That is, the school chooses teacher attributes such that their marginal cost is proportional to their marginal product, all conditional on the student characteristics that the school is endowed with. Therefore the T-subgradient of the observed wage-attributes function estimates marginal productivity of T for schools who were able to hire those persons. The selectivity bias argument again applies to schools who are located at yet other margins.

Empirical work on this problem has concentrated on estimating the function W(S,T) in cross-section data. This requires information on wages paid to teachers, on the student and neighborhood characteristics of the schools they work in, and productivity attributes of teachers. The basic unit of observation is a school-teacher pair, and wages of teachers are regressed on empirical proxies for S and T. While several such studies have been made, the one I am most familiar with used a national random sample of schools from 1965 survey data. When S is summarized in a single statistic, the racial composition of students measured by proportion black, it is found that white teachers prefer to teach in schools with mainly black students. The average compensating differential was $6 per percentage point black students for white teachers (1965 $). It was $2 per percentage point white for black teachers. This suggests that it would be necessary to compensate a white teacher at least $600 to move from an all-white school to an all-black school; whereas a black teacher would need to be compensated at least $200 to move from an all-black school to an all-white one. Experimentation with other school and student characteristic variables indicates that these differentials reflect much more than racial preferences per se. These indicators include measures of student ability; attendance, truancy and disciplinary problems; college-going preferences of students; and neighborhood characteristics. Regression coefficients on these variables typically reveal that teachers prefer to teach in schools located in more amenable neighborhoods with more able and better motivated students.[8] They are willing to pay someting for these opportunities in the form of wage reductions. In American society today it is an unfortunate fact that student racial composition is highly correlated with these other attributes of students and schools. The correlation is sufficiently large that it is not possible to disentangle the separate influences of each dimension of S. An index on the entire vector is the best that can be done to summarize the data, because schools are very strongly stratified by race and other school-student attributes. Nonetheless, the sorting implications of the basic model were strongly confirmed. The average white teacher in this sample required additional compensation of more than $400 to teach in schools with the characteristics of the average black teacher. Similarly, the average black teacher required additional compensation of at least $300 to work in schools with the average characteristics of those in which white teachers were found.

C. Human Capital

The theory of human capital has been very important to labor economics in the last two decades. Its implications for the distribution of income and inequality in economic life are profound. The major outline of the subject is readily found in the Wealth of Nations, when Smith notes that occupations requiring time and money expenditures on training must pay larger wages to compensate both for that expense and for the briefer duration of labor market productivity implied by it. As is now well known, observed earnings differentials between schooling levels provides a basis for imputing rates of return to education. However, these concepts have much greater generality to learning opportunities and skill acquisition in the labor market as well as to the analysis of education per se. It is these largely informal, learning-by-doing aspects on which the following account focuses.

It is a common observation that most specific job skills are learned from work activities themselves. Formal schooling paves the way, both by setting down a body of general knowledge and principles as well as teaching students how to learn. But even in the case of professional training there is no perfect substitute for apprenticeship, that is, for work experience itself. These ideas can be captured in the following way.[9] Think of a job as a tied package of work and learning: a worker simultaneously sells the services of his skills and jointly buys the opportunity to augment these skills. Learning potential is a by-product of the work environment itself. It is tied to a specific work activity but varies from activity to activity and from job to job. Some provide more learning opportunities and some provide less. Therein lies a margin of choice, for both workers and firms.

As is generally known, a worker's incentives for capital accumulation (learning) are largest at young ages. Hence young workers are typically assigned to those jobs and work activities for which learning potential is largest. The optimal human capital investment program is implemented by a sequence of assignments in which workers systematically move across work activities and jobs that offer successively smaller learning opportunities. Thus, the optimal program implies a systematic pattern of job mobility and "promotions" with experience. Firms accommodate this by structuring work activities in various ways to provide greater or smaller learning options. While some learning invariably is jointly supplied with all work activity, prospects for altering learning potential arises from reallocating experienced workers' time away from direct production and toward instructing inexperienced personnel. This is costly, because marketable output is foregone. Thus, a firm can be viewed as jointly producing both marketable output and training output, summarized by a production possibilities frontier between the two. Training services are directly sold to existing employees. These sales are implicit in wage reductions of workers who undertake training. In this case the equalizing differential is defined over the learning opportunity connected with some activity.

To see what this implies, index the training potential of a work activity by a latent variable I. Let p(I) represent the market equalizing difference, with P increasing in I and $P(0) = 0$. P is the foregone earnings paid by a worker if assigned to activity I. Let the worker be endowed with skill k, which rents for unit price R. Then the worker's observed earnings are y and $y = RK - P(I)$. This illustrates the tie-in. The worker sells services of value Rk but buys back a learning opportunity worth P(I). The worker demands learning opportunities because they enhance future skills and rewards. The model is closed by specifying a relationship between the effect of choice of I on $\dot{k} = dk/dt$. This technological constraint is $\dot{k} = g(I)$, where g is an increasing function. Inverting g, expressing I as a function of $\dot{k}$ and substituting into the definition of earnings yields $y = Rk - G(\dot{k})$, where G is an increasing convex function. The worker chooses a time valued function I(t) and therefore $\dot{k}(t)$ to maximize the present value of earnings over working life. The economics of the choice problem is illustrated in Figure 4. The smooth curve shown as an envelope is the function $y = Rk - G(\dot{k})$ conditioned on the current value of k. The step functions that this curve envelopes represent alternative learning opportunities I, with I increasing as $\dot{k}$ increases and which cost more in terms of P(I). Choice of a larger value of I is costly because current income falls. The potential return is a larger value of k in the future which shifts income-investment opportunity locus upward and to the right and expands future choices.

For example, consider a simple case where $\dot{k} = \gamma I$, with γ interpreted as a learning efficiency parameter; and where P(I) is a simple quadratic, $P(I) - I^2/2$. Then $y(t) = Rk(t) - [\dot{k}(t)/\gamma]^2/2$. The discounted value of human wealth is $\int_0^N y(t)e^{-rt}dt$, where N is the length of working life and r is the discount rate. In this simple case the program that maximizes human wealth (subject to an initial stock $k(0) = k_0$) equates marginal cost of investment to its marginal return. Marginal cost is simply the slope of the income-investment possibilities curve in Figure 4, or $\dot{k}(t)/\gamma^2$ in this case. The discounted marginal return of a unit of skill is the rental that will be obtained from it over its useful life. At time t this is nothing other than $(R/r)(1 - e^{-r(N-t)}) = Q(t)$, the present value of an annuity paying R for (N-t) periods. Notice that Q(t) is decreasing and concave in t. Along the optimum trajectory we therefore have $\dot{k} = \gamma^2 Q(t)$: the worker's rate of learning is largest at young ages and monotonically falls over the life cycle. The sequence of learning options that implement the optimal policy is given by $I(t) = \gamma Q(t)$. Young workers are assigned to positions with the largest learning opportunities and are successively promoted to "higher" levels as their skills increase.

An interesting selectivity aspect of this problem arises if γ is thought of as a fixed effect that varies from person to person in the population. Some persons may be more efficient in converting a given learning opportunity into useful marketable skills. A more complicated problem would specify an interaction between learning ability and previously acquired knowledge, as well as with inherent ability. Whatever the source of these differences, the formulas above for $\dot{k}$ and I reveal that workers with larger values of γ accumulate more human capital and are assigned to jobs with greater learning opportunities at each age. Greater learning efficiency reduces the real price of investment to the more able, and they purchase greater amounts. This may be an important source of income inequality in the population as a whole, because human wealth is increasing in γ and observed wage differences are not completely equalizing across the population at large.

This theory implies a corresponding theory of promotions within an organization and possibly a "stepping stone" theory of job mobility among firms. For example, it is not hard to imagine that some firms might exhibit comparative advantage in producing learning opportunities. If so, they would cater mostly to young workers and provide a source of supply of experienced workers to firms which have relative advantages in other lines of production. Little empirical work has been done so far along these lines.[10] Instead most empirical studies have concentrated on observable life cycle earnings. The basis for this is easily seen in the simple example. If the expression for $\dot{k}$ and the implied life cycle trajectory for k is substituted into the definition of y, a closed form solution for y(t) is obtained. It is easily imagined, and turns out to be true, that the implied functional form of y(t) provides information on underlying parameters such as γ and r. The few studies that have been made suggest that workers with more formal schooling are more efficient learners. They also suggest that their depreciation and obsolescence rates on human capital investment are larger than those with less schooling, implying another obvious source of selection and assignment of workers among different types of work activities.[11]

D. Implicit Contracts

Recent research on implicit contracts extends the idea of equalizing differences to unemployment risk. An implicit contract is a mutual understanding between workers and employers in which the firm is given wide latitude to make decisions concerning employment and layoff status and hours of work of its employees at its own discretion and and as circumstances arise. The agreement is implicit because myriad unforeseen contingencies make contracting costs so large that formal contracts are uneconomic. Employers decisions serve their firm's self interest in any given situation. Were these decisions otherwise unconstrained, they might involve a degree of worker exploitation. In most analyses sufficient common information and potential mobility of workers across firms insures that an employer's decisions are constrained to achieve a minimal level of expected utility by its workers; viz., the level expected at other firms.[12]

Figure 3 still serves as an organizing device for this class of problems. Assuming that all workers are alike, the equalizing difference function, defined say on the risk of layoff, maps out the representative worker's indifference curve, and constrains a firm's choice, which is described by the usual tangency condition. There is an equivalent dual representation of the problem. Instead of treating the implicit contract as maximizing profits subject to a utility constraint, think of the contract as maximizing worker utility subject to a profit constraint; in fact with free entry of firms, profit is constrained to be zero. Competition for workers among firms guarantees that contractural features make workers as well off as possible. I follow the dual approach here.

Early work on compensating differentials for unemployment risk took the worker's objective to be expected income maximization.[13] Expected earnings are the wage while working times the probability of employment, and the latter is 1.0 minus the probability of unemployment. Therefore, jobs offering high unemployment prospects must pay higher wages in order to attract workers and equalize expected utility among them. Implicit contract theory shows that this approach may be seriously misleading when various aspects of insurance and risk sharing are considered.

A sketch of the basic idea is illustrated by a simple example. Imagine a worker in some activity where the value of production x is stochastic. Let x be distributed by the known probability law G(x). The worker is permanently attached to the activity, but has the option of not working in any period if his productivity (the realized value of x) is small enough. Let the nonmarket value of the worker's time be k, and assume that k is nonstochastic. Thus k is the value of leisure or of home production. Let u(.) denote the worker's concave utility function of consumption. Finally, let $\delta(x)$ denote an employment indicator function such that $\delta(x) = 1$ when the worker is employed in market production x and $\delta(x) = 0$ when the worker is unemployed (or employed in the nonmarket sector), "producing" k.

I propose to analyze the problem in two steps. First consider a worker-firm that doesn't trade with anyone else. Then examine the gains from trade through risk shifting and insurance.

The worker's autarky expected utility is

$$\bar{U} = \int_k^\infty u(x)dG + u(k)G(k) \tag{3}$$

$$= (1-G(k))E(u(x)|\delta = 1) + G(k)u(k)$$

$$= (1-\theta)E(u(x)|\delta = 1) + \theta u(k)$$

where $\theta \equiv G(k)$ is the unemployment rate and $E(u|\delta = 1)$ is a conditional expectation. Expression (3) provides the simple intuition for compensating differentials on unemployment risk. Consider two activities, one with a greater unemployment rate than the other. Then, since the employment rule implies that $E(u(x)|\delta = 1) > u(k)$, the activity with the larger value of θ

has smaller expected utility at equal average consumption when employed. A wage premium is necessary to attract any workers. That something might be awry with this logic is suggested by the fact that θ and $E(u(x)|\delta = 1)$ are both endogenously determined, presumably by such things as the riskiness of the distribution $G(x)$. This is most easily shown when risk is diversifiable through insurance.

Consider now a situation where there are many firms, each facing statistically independent risks which can be fully diversified by pooling. Availability of insurance breaks the personal link between consumption and production. It allows a worker and firm to form a relationship in which the firm pays $w(x)$ to employed workers and unemployment compensation $\bar{w}(x)$ to its unemployed members, both possibly state dependent. This is achieved by shifting productivity risk from workers to risk-neutral employers or to the market at large by portfolio diversification. Full risk shifting is equivalent to an actuarially fair insurance arrangement and must satisfy a budget constraint: expected income to the insurance agent equals expected outgo. Thus the implicit contract solves the following constrained maximum problem

$$(4)\qquad \bar{U} = \int[\delta(x)u(w(x)) + (1 - \delta(x))u(\bar{w}(x) + k)]dG$$

$$+ \lambda\int[\delta(x)x - \delta(x)w(x) - (1 - \delta(x))\bar{w}(x)]dG$$

with respect to functions $\delta(x)$, $w(x)$ and $\bar{w}(x)$.

State-by-state differentiation of the policy functions in (4) reveals the following. The equilibrium contract equalizes consumption in all employed states: $w(x) = w$ for all x for which $\delta(x) = 1$. Similarly, unemployment compensation equalizes consumption in all unemployed states: $\bar{w}(x) = \bar{w}$ for all x where $\delta(x) = 0$. In fact fair insurance implies complete insurance in all states independent of x: $w = \bar{w} + k$, and the worker's consumption is guaranteed independent of employment status. These are all familiar consequences of risk aversion and are sometimes thought to imply wage rigidity in the optimal contract. However, that is slightly mistaken. It is more accurate to say that consumption rigidity is implied, akin to the permanent income hypothesis. Finally, the optimum employment policy is efficient as before and has $\delta(x) = 1$ when $x \geq k$ and $\delta(x) = 0$ when $x < k$. Complete insurance disassociates productive efficiency from distribution.

These properties of the optimum contract allow us to write maximum expected utility as

$$\bar{U} = u(\bar{w} + k) + \lambda[\int_k^\infty (x - k)dG - \bar{w}] \tag{5}$$

$$= u(\bar{w} + k) + \lambda[(1 - G(k))(E(x|x \geq k) - k) - \bar{w}]$$

from which comparative statics can be derived. Consider an increase in risk in the sense of a mean preserving spread in the distribution G(x). Increasing risk changes expected utility in proportion to its effect on $\bar{w}$, and expression (5) suggests that the effect of such a change on $\bar{w}$ depends on how $E(x|x \geq k)$ is affected. We know from definition that a conditional mean in increasing in the spread of its parent distribution. Therefore we conclude that increasing risk raises expected utility because it increases $\bar{w}$. The logic of this paradoxical result, that greater risk is preferred to less risk, is analogous to option pricing formulas in finance. The availability of a nonmarket "production" alternative truncates the distribution of employed states from below. Increasing risk allows the firm to be more selective in its employment policy, employing workers with greater frequency in the most productive states, increasing average output and supporting larger w and $\bar{w}$ payments.

It must be stressed that the risk-is-good argument applies only to those shocks that are sufficiently transient and independent to be diversified. In that case permanent wealth of a worker is not affected by any particular realization of x. This suggests that equalizing differences would only be observed for anticipated undiversifiable risk because those do imply permanent changes in wealth and therefore in consumption prospects.

Empirical work on this problem is in a surprisingly elementary state and the few studies that have been done found small wage premiums for unemployment risk. However, most investigators have not clearly distinguished between diversifiable and nondiversifiable risk, so all returns are not in. Nevertheless there are some other broad implications of these models that are consistent with observation. First, the theory implies that consumption should be smoothed relative to stochastic realizations. Permanent income studies of consumption broadly confirm this prediction. Second, the model implies systematic sorting and assignment of different types of workers to risk classes. Those workers whose nonmarket uses of time are most valuable should bear a proportionately greater share of unemployment risks because it is socially less costly for them to do so. The incidence of unemployment among various socioeconomic groups is broadly consistent with this prediction.

IV. Selection, Assignment and Productivity

This section sketches how a competitive labor market assigns workers of different talents to alternative tasks and occupations. The equalizing difference function is now defined over productivity attributes of workers and is closely related to the Ricardian theory of rent (consumption attributes of jobs are ignored in this discussion). Just as different climate, irrigation, and soil nutrient attributes of agricultural land affect productivity and therefore value, so too do different ability and other attributes of workers affect their valuations in a competitive labor market. One of the problems in Section III showed how this might be addressed by introducing worker traits directly into production functions. Then a variant of Figure 3 applies with W(.) defined over worker traits. The equilibrium wage-attributes function is an envelope across bid functions of firms for attributes. Workers are systematically sorted and stratified to jobs according to their productivity attributes and are matched to firms according to the intensity with which attributes affect production of various goods and services.

Such an approach has two limitations: it restricts analysis to those situations where production functions can be expressed directly in terms of worker characteristics; and it is easily manageable only when each firm specializes its choice of personnel to a single type of worker. The former is not always possible or meaningful, and the latter is seldom strictly observed. I therefore outline an alternative model in which the valuation system on characteristics is the indirect outcome of market assignments, in the manner of the conventional theory of rent. This model is better articulated with standard analytical methods, particularly those used in the theory of international trade, and is also more fruitful for analyzing the distribution of earnings.[14]

Consider a simple economy with n goods and production functions $x_i = F^i(T_{1i},T_{2i})$, where x_i is output of good i = 1, ..., n and T_{1i} and T_{2i} are total input of two labor factors in the production of x_i, each measured in efficiency units. The T's may be given a variety of interpretations depending on the particular problem to be analyzed, but for present purposes are two distinct occupations, which themselves are (exogenously) defined as collections of productive tasks and work activities. Notice that this formulation preserves a certain additivity: It is the sum of productive inputs among all workers employed in the production of a particular good that matters, not how this sum is distributed over different types of workers in a production unit. Another way of saying it is that there is a form of perfect quantity-quality substitution among workers in production.

Ignoring hours of work decisions for simplicity, a worker is completely described by two numbers (t_1,t_2) where t_j, j = 1,2 represents a worker's productive efficiency in job j, his contribution to T_j if full time is allocated to that task. I assume for this exposition that t_1 and t_2 are endowed for each worker, but differ among them. Then the absolute scale of t_1 and t_2 measures a worker's absolute advantage in each activity and the ratio $r = t_2/t_1$ is an index of comparative advantage on activity 2 (equivalently, 1/r indexes comparative advantage in activity 1). Define a worker type by an index v of the comparative advantage ratio, $r = r(v)$, with v defined on the unit interval. Then v can be chosen so that r is ordered from largest to smallest. I assume that there are such a large number of different types of workers in the sense of v that r(v) is continuous for all practical purposes, so that $dr/dv < 0$. The total supply of worker talents is described in either of two equivalent ways.

One is as a distribution by type v, say n(v)dv, where n is appropriately scaled to account for differences within type v for absolute advantage or differences in absolute efficiency units. Another is by a joint distribution on the random vector (t_1,t_2) across the entire working population, say $g(t_1,t_2)dt_1dt_2$. We wish to study how the market assigns workers to activities, how it partitions these distributions, and what those partitions imply about observable variables.[15]

It is clear that the optimum assignment of workers to tasks follows the principle of comparative advantage. An economy-wide task possibility-frontier in the T_1,T_2 plane is found by maximizing T_2 for any given level of T_1 subject to the constraints implied by either n(v) or $g(t_1,t_2)$. This is done in exactly the same way that the world production possibility frontier is derived in the theory of international trade with many countries. Each worker specializes in the activity in which he exhibits comparative advantage and potential skill in the activity to which he is not assigned remains latent and unutilized. The assignment is ordered on v: a given point on the task possibility frontier is supported by a critical value v^* such that persons with $v > v^*$ are assigned to T_2, those with $v < v^*$ are assigned to T_1, and those for whom $v = v^*$ are assigned by a random device. The frontier itself completely describes relevant factor endowments in the economy, which in turn define the conditions of supply of all goods x_i -- the production possibility set of outputs in the economy at large. Finally, demand conditions for goods and services determine the general equilibrium, including the assignments of workers to various firms and industries and the precise menu of goods actually produced.

Since it is not immediately relevant to the purposes at hand, I omit details of how the complete equilibrium in the economy is determined. What is crucial is that this equilibrium is supported by a maximum problem and a competitive labor market. General equilibrium determines linear prices p_1 and p_2 on the inputs T_1 and T_2 that maximize the total value of inputs in the economy. A plane tangent to the task possibility frontier defines these prices, and its gradient can be thought of as piece rates for personal output in each activity. Each worker faces these prices parametrically and decides whether to supply his labor to T_1 or T_2. Given the worker's endowed values of t_1 and t_2, the alternative that maximizes income is chosen. Income potential in activity j is simply $y_j = p_j t_j$, for $j = 1,2$. Define $p = p_1/p_2$ as the relative price. Then maximization of earnings is equivalently described by the rule: choose T_1 or T_2 as $r(v) \gtrless p$. The equality $r(v^*) = p$ defines the critical value of v^* discussed above.

An alternative description of the optimum assignment is represented as a linear partition of the joint distribution $g(t_1,t_2)$ shown in Figure 5. Each person is uniquely described by a point in the t_1,t_2 plane. The ellipses in Figure 5 show the probability contours in the overall working population, the level sets of $g(t_1,t_2)$. The linear function $t_2 = pt_1$ serves as the dividing line in the optimal and market assignment of workers to activities. All those whose skill endowment point lies above the line choose activity 2 because their incomes are largest there. All those whose skill endowment point lies below the line choose activity 1. Notice that as p rises, a margin of workers (it would be represented by a wedge-shaped area in Figure 5) find it optimal to reassign themselves to activity 1 rather than 2 because their relative income prospects change. Notice also that for small changes in p most people are inframarginal and continue using the same skill. This is very similar in spirit to what happened in the model of Section II.

The contours in Figure 5 are drawn on the assumption that t_1 and t_2 are positively correlated in the population, so that a person who is very productive in activity 2 is also likely to be very productive in activity 1 as well. Positive correlation signals the presence of absolute advantage and suggests a factor structure interpretation of t_j which places substantial weight on a single dimension, such as general intelligence or I.Q. However, it is entirely possible that the correlation is negative. Then the ellipses would have been negatively inclined, and a person who was very good at one activity would more than likely be worse than average in the other. In that case absolute advantage is not important. A factor structure interpretation of t_j suggests at least two distinct factors in which skills in one activity load heavily on one factor and skills in the other activity load heavily on a distinctly different factor. The ellipses in Figure 5 have been drawn so that the marginal variance of t_2 in $g(t_1,t_2)$ exceeds the marginal variance ot t_1. This may be interpreted in terms of the inherent difficulty of performing the two activities. Smaller variance in activity 1 suggests that just about everyone achieves more or less the same amount of useful output if they devote themselves to that activity. In that sense activity 2 in Figure 5 is inherently more difficult and offers greater scope for talent and ability to stand out and make its mark.

Given the shape of $g(t_1,t_2)$ market selectivity implies some interesting productivity calculations in response to relative price changes. For example, suppose t_1 and t_2 are positively correlated and p increases. Then the average personal productivity of people observed in both occupations would tend to rise: the average person observed in T_2 initially exhibits absolute advantage in both, and some of these persons switch over to T_1. On the other hand, were the correlation between latent skills negative, an increase in p raises average productivity of workers remaining in T_2, because those finding it advantageous to switch from T_2 to T_1 are drawn from the lower tail of the conditional (marginal) t_2 distribution. Those making the switch are of lower average productivity in T_1 than workers who initially chose it, so average personal productivity observed in T_1 decreases.

Equivalent statements can be made about the observed distribution of earnings in each activity. These statements are readily obtained from the fact that both the latent and actual income distributions are simple linear transforms of the distribution of $g(t_1,t_2)$ due to the linearity of prices p_1 and p_2. This provides a very powerful basis for analyzing the cross-sectional distribution of earnings, both between activities and among them, from the relationship $y = \max(y_1,y_2)$, where y is observed earnings independent of activity choice and y_j is earning opportunities of a person in activity j. Space limitations preclude a complete development of those ideas here. Suffice it to say that the actual choices are made optimally. This implies selectivity bias in between-activity earnings comparisons. Average earnings observed by those who chose T_1 are unlikely to be an unbiased estimate of average earnings prospects available to those who chose T_2 had they somehow been assigned to T_1 instead. The converse also applies. Again, these statements are very similar in spirit to the sorting implications of revealed preference discussed in Section II.

V. Ability Sorting: Applications

Two applications illustrate the ideas in Section IV. The first is theoretical and the second is empirical.

A. Sorting by Latent Characteristics.[16]

Given the general flavor of this essay, it is perhaps noteworthy that the general outline of equilibrium in Section IV made little or no reference to worker characteristics or traits. That deficiency is remedied here. It is shown that the selectivity aspects of optimal assignments imply corresponding sorting of workers by those characterisitcs which influence abilities and comparative advantage.

The number t_j is a direct measure of a worker's talent in activity j. The vector of talents and prices p_j were seen to be all that was necessary to study market assignments and income distribution. Further development is possible by writing the determinants of personal talent in an activity in terms of another set of latent factors (in the sense of factor analysis in statistics), according to "production functions" $t_j = \Theta^j(Z_1, Z_2, \ldots, Z_m)$. Here $Z_1, \ldots, Z_m$ are a relatively small set of latent factors which determine ability in a given activity, such as physical strength, manual dexterity, verbal abilities, analytical abilities, and so forth. As written, the Z's influence potential output in all activities, though the marginal product of any Z may differ among activities, as shown by the index on production function Θ^j. This is clearest in the case of a proper factor model, when Θ^j is linear as in Z_i: $t_j = \sum_i \theta_{ij} Z_j$ where θ_{ij} are constants.

A worker is endowed with particular values of attributes Z and is completely described by a point on the vector $(Z_1, \ldots, Z_m)$. All workers together are described by a distribution over $(Z_1,\ldots,Z_m)$, which in conjunction with the production functions relating the Z's to t_j's lead to the joint density $g(t_1,t_2)$ extensively utilized above. As a consequence the equilibrium analysis of Section IV applies intact, and the pricing system that supports it induces an implicit pricing scheme on the Z's through the relations $y_j = p_j t_j$. Substituting the production relations into these expressions in the linear case results in $y_j = \sum_i \phi_{ij} Z_i$, where $\phi_{ij} = p_j \theta_{ij}$ can be regarded as the implicit prices of Z_i in activity j. Notice that the implicit prices ϕ_{ij} on any given characteristic Z_i are distinct across activities. There are no exploitable arbitrage opportunities that equalize implicit prices across different types of jobs in the economy. This is a fundamental implication of the assumption that the Z's themselves are not direct objects of choice by firms and do not directly enter into a firm's production function. Viewing the linear relations between the Z's and y_j's as a factor structure suggests interpreting ϕ_{ij} as a "factor loading." However, the analogy is not complete because factor loadings vary from job to job (i.e., across j).

Nevertheless, systematic sorting of activities according to Z is strongly implied by the model. For example, consider Figure 6, which is specialized to two underlying factors (m = 2) and four activities (j = 1,...,4). Given ϕ_{1j} and ϕ_{2j}, the dashed intersecting lines show those combinations of Z_1 and Z_2 that result in a dollar's worth of earnings in activity j, for j = 1,...,4. A person with an endowment at point A makes the most money in activity 1 and is therefore assigned to that activity. The rays through the origin are defined by intersections of consecutive dashed lines, and it follows from homogeneity of the linear Θ^j production functions that all individuals whose endowment lies between the vertical axis and the first (steepest) ray optimally choose activity 1, by exactly the same logic. The rays therefore define the optimal partition of the joint distribution of the Z's across activities. There is systematic sorting of characteristics and worker types to activities. Notice that the relative slopes of the dashed lines in Figure 6 depend on relative factor prices ϕ_{ij}. Thus income prospects in activity 1 weight attribute Z_2 much more heavily than Z_1 and vice-versa for activity 4. Activities 2 and 3 weigh both attributes more equally. Those individuals who are heavily endowed with one attribute are more likely to be found in the extreme activities (1 or 4), while those with more balanced endowments are more likely to be found in the middle activities (2 or 3). Again, systematic selection of persons by latent characteristics is the rule rather than the exception.

B. Educational Selection[17]

The rate of return to schooling is the fundamental concept in the economics of education. The basic idea is simple. Suppose a person takes $\underline{s}$ years of schooling and enters the market thereafter. Let V(s) represent the discounted present value, as of some early age, of future earnings prospects given that $\underline{s}$ years are completed. Here earnings include income during the schooling period (with deductions for direct costs) as well as during periods of full participation in the work force. The person's incremental rate of return to schooling is the gradient V'(s). A major empirical task is to calculate the function V(s). This is complicated by the fact that we never observe a path not taken, and only one value of $\underline{s}$ is realized for any given person. Therefore, V(s) must be imputed by a "counterfactual" comparison of income streams between people who stopped school at different levels, raising possibilities of bias due to interpersonal differences in returns and costs of schooling.

Public policy interest in these calculations rests on cost-benefit analysis of interventions, such as subsidies, which induce young people to undertake more schooling than they otherwise would. If we were confident that actual schooling choices were randomly determined, then imputing V(s) from interpersonal differences in earnings would map out the average social rate of return. However, there are strong reasons for believing that school completion choices are systematic: rational decision requires choosing s to equate the marginal rate of return with the personal rate of discount. Still if family wealth (interpersonal differences in discount factors) was the only constraint on individual choice, interpersonal estimates of V(s) would yield excellent approximations of average schooling prospects for a random person in the population. This is not the case if private rates of return systematically differ among people.

Early investigations of schooling returns clearly recognized possible problems of "ability bias," that those who continue school are likely to be more able in some sense than those who stop at lower levels. Thus selection and admissions policies of colleges are contingent on adequate high school performance, and colleges themselves are highly stratified by abilities of students. Nevertheless, most studies have adopted a very narrow view of ability-as-IQ which, by the logic of section IV, implicitly restricts attention to questions of absolute advantage and ignores comparative advantage. Since different levels and types of schooling are closely associated with different occupations and work activities, it seems evident that a priori restrictions of absolute advantage or disadvantage are too confining for the problem. Would a person who was very successful as a lawyer have been a very successful plumber as well (and vice-versa)? Or is it more likely that the verbal skills that make for successful lawyering would not have as much value in the plumbing profession, for which mechanical abilities are more important? In short, it seems likely that the constructions in Figures 5 and 6 apply to these choices and that individuals are sorted across occupations and school completion levels by latent characteristics (talents) which produce comparative as well as absolute advantage.

The apparatus of section IV can be applied directly to this problem when school completion levels are discrete rather than continuous. Thus let V_i be the present discounted value of earnings of a person for schooling level i, with $i = 1, \ldots, n$. Then the population at large is described by a random vector $(V_1, V_2, \ldots, V_n)$ in the sense of the distribution g(.) used above. Each person chooses the value of i which maximizes V, that is the distributions of V and i actually observed are generated by the rule $V = \max_i (V_1, \ldots, V_n)$ for each person, leading to partitions much like Figures 5 and 6.

My study with Willis examined the choice of continuing schooling beyond high school compared with stopping after high school graduation, a special case in which $n = 2$. This restriction of the choice set was dictated by limitations of available econometric technology which currently is best suited to binomial rather than multinomial choices. We found substantial evidence of ability sorting in these two classes, though necessary adjustments to simple rate of return calculations were relatively small. More interestingly, when observed earnings patterns in the two classes were purged of selection bias there was strong indication of the presence of comparative advantage and negative correlation in underlying talents. Those who entered college would have earned less had they stopped upon high school graduation than those who actually stopped. Had high school completers continued on to college, they would have earned less than those who actually found it in their best interests to enter college. We also discovered that prospective personal financial gains from college entrance were important determinants of the decision to attend, though family background and costs were important considerations as well. Studies of this sort have to be replicated on a broader scale and on a variety of data sources to get a more complete picture of the practical importance of these effects.

VI. Conclusions

I hope my exposition and examples illustrate both the simplicity and broad range of applicability of equalizing differences in labor markets. Much more work remains to be done. Existing analysis of valuation of worker traits and characteristics and the assignment of workers to firms seems to me to be a weak sister in this enterprise. A main function of labor markets is to solve a type of marriage problem, to match a particular worker to a particular position in a particular firm. While the framework I set forth above does this in a general way, it ignores productivity interactions among workers within firms, how a specific worker fits a niche in the enterprise and becomes a proper member of a productive team. A good deal of the search and turnover activity we observe among young workers probably is attributable to this kind of matching. While much interesting recent research on matching has been done, most of it is partial equilibrium and conceals the larger picture of assignments in the market as a whole. Finally, we are a long way toward understanding how disequilibrium affects mobility decisions and how movements provoked by disequilibrium can be distinguished empirically from equilibrium mobility that naturally arises over the life-cycle through skill and job upgrading and reassignment.

FOOTNOTES

*I am indebted to Nasser Saidi for the opportunity to present this material and for his substantial contributions over the years to many of the ideas in this paper. The National Science Foundation provided financial support. As will be apparent, my bibliography is highly selected, but more complete citations are found in the references listed.

[1]The presence of nonearned income modifies the budget constraint in an obvious way and is therefore omitted. In that case, nonzero income effects in preferences imply systematic sorting by wealth, along the lines indicated below for differences in taste.

[2]Determination of firm size and related questions are ignored because they are not directly relevant for the problem at hand: the distribution $f(\beta)$ incorporates the exogenous distribution of firms by size, but each firm's scale decision could be incorporated without great difficulty.

[3]This example has been carefully chosen so that the wage ratio enters both supply and demand, and is not general. More typically the wage difference is relevant for workers, while the ratio is sufficient for firms. Clearly, no great issues of principal are involved here.

[4]It is straightforward to extend this to multivariate (vector) attributes: see Rosen [1974] for detailed elaboration. Mas-Colell [1975] proves an existence theorem for models of this type.

[5]My exposition is based on Rosen [1981] and Thaler and Rosen [1975]. Useful surveys of this and other issues are found in Linnerooth [1979] and Jones-Lee [1976].

[6]See especially Becker [1957]. Also, Arrow [1972].

[7]I draw mainly on the study by Antos and Rosen [1975]. Several other studies along similar lines appear in the literature; e.g., Chambers [1978].

[8]Existence of substantial wage differentials between public and private schools confirms this interpretation. Racial mixing of students may produce social tension that creates nonlinearities in the wage regression, but this remains to be investigated.

[9]See Rosen [1972] for elaboration.

[10]MacDonald [1982] develops a major extension in which learning about workers' latent talents allows more efficient work assignments as experience is accumulated. See also Ross, Taubman and Wachter [1981].

[11]These studies are surveyed in Rosen [1977].

[12]The basic references are Azariadis [1975] and Baily [1974]. My account follows my own work (Rosen [1983]) and Mortensen [1983].

[13]See Hall [1972] for example. Harris and Todaro [1970] present an especially interesting application to underdeveloped countries. Abowd and Ashenfelter [1981] pursue this line to its logical limit.

[14]Roy [1952] pioneered this approach, though see the related effort by Tinbergen [1959]. That work was rediscovered and elaborated in recent years by Rosen [1978] and Sattinger [1980]. Heckman and Sedlacek [1981] develop econometric methods suitable for these models.

[15]The distribution of latent talents g(.) is the fundamental primative and readily generalizes to an arbitrary number of work activities. The

ordered distribution n(v) does not usefully generalize to more than two skills, but has certain expository virtues.

[16]The ideas in this section were first formulated in a neglected article by Mandelbrot [1962]. Welch [1969] takes a much different approach in which the Z's are direct factors of production in place of the T's.

[17]This discussion is based on Willis and Rosen [1978].

REFERENCES

Abowd, J. and O. Ashenfelter, "Anticipated Unemployment, Temporary Layoffs and Compensating Differentials," in S. Rosen (ed.), Studies in Labor Markets: Chicago, 1981.

Antos, J. and S. Rosen, "Discrimination in the Market for Public School Teachers," Journal of Econometrics, 1975.

Arrow, K., "Models of Discrimination," in A. H. Pascal, (ed.), Racial Discrimination in Economic Life: Lexington, 1972.

Azariadis, C., "Implicit Contracts and Underemployment Equilibria," Journal of Political Economy, 1975.

Baily, M., "Wages and Employment Under Uncertain Demand," Review of Economic Studies, 1974.

Becker, G., The Economics of Discrimination: Chicago, 1957.

Chambers, J., "Educational Cost Differentials and the Allocation of State Aid for Education," Journal of Human Resources, 1978.

Hall, R., "Why Is the Unemployment Rate So High at Full Employment?" Brookings Papers on Economic Activity, 1970.

Harris, J. and M. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," American Economic Review, 1970.

Heckman, J. and G. Sedlacek, "The Impact of the Minimum Wage on Employment and Earnings of Workers in South Carolina," University of Chicago, 1981.

Jones-Lee, M., The Value of Life: An Economic Analysis: London, 1976.

Linnerooth, J., "The Value of Human Life: A Review of the Models," Economic Inquiry, 1979.

MacDonald, G., "Information in Production," Econometrica, 1982.

Mandelbrot, B., "Paretian Distributions and Income Maximization," Quarterly Journal of Economics, 1962.

Mas-Colell, A., "A Model of Equilibrium with Differentiated Commodities," Journal of Mathematical Economics, 1975.

Mortensen, D., "A Welfare Analysis of Unemployment Insurance: Variations on Second Best Themes," Carnegie-Rochester Papers on Public Policy, 1983.

Rosen, S., "Learning and Experience in the Labor Market," Journal of Human Resources, 1972.

________, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, 1974.

________, "Human Capital: A Survey of Empirical Research," in R. Ehrenberg (ed.), Research in Labor Markets: Greenwich, 1977.

________, "Substitution and Division of Labor," Economica, 1978.

________, "Valuing Health Risk," American Economic Review, 1978.

________, "Unemployment and Insurance," Carnegie-Rochester Papers on Public Policy, 1983.

Ross, S., P. Taubman and M. Wachter, "Learning by Observing and the Distribution of Wages," in S. Rosen (ed.), Studies in Labor Markets: Chicago, 1981

Roy, A. D., "Some Thoughts on the Distribution of Earnings," Oxford Economic Papers, 1951.

Sattinger, M., Capital and the Distribution of Labor Earnings: Amsterdam, 1980.

Thaler, R. and S. Rosen, "The Value of Saving a Life: Evidence from the Labor Market," in N. Terleckyj (ed.), Household Production and Consumption: New York, 1975.

Tinbergen, J., "On the Theory of Income Distribution," in Selected Papers: Amsterdam, 1959.

Welch, F., "Linear Synthesis of Skill Distribution," Journal of Human Resources, 1969.

Willis, R. and S. Rosen, "Education and Self-Selection," Journal of Political Economy, 1979.

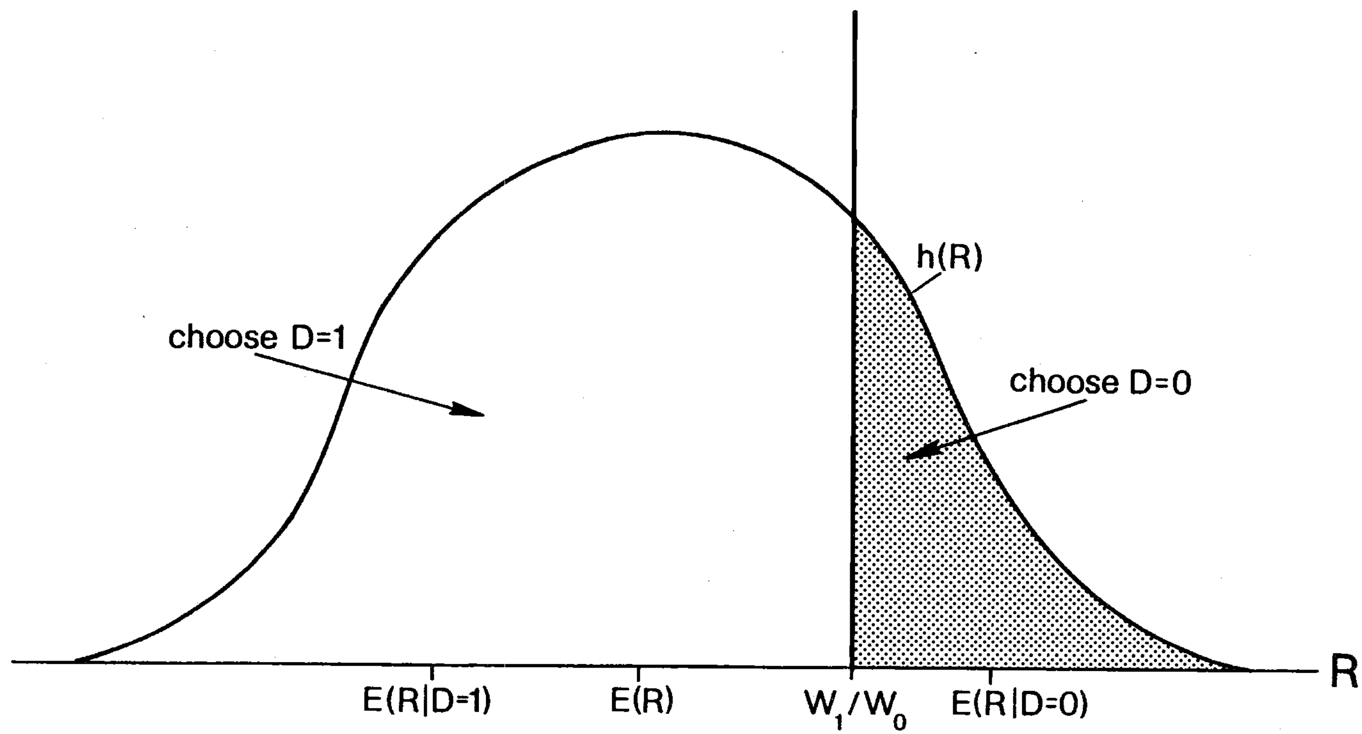

Figure 1

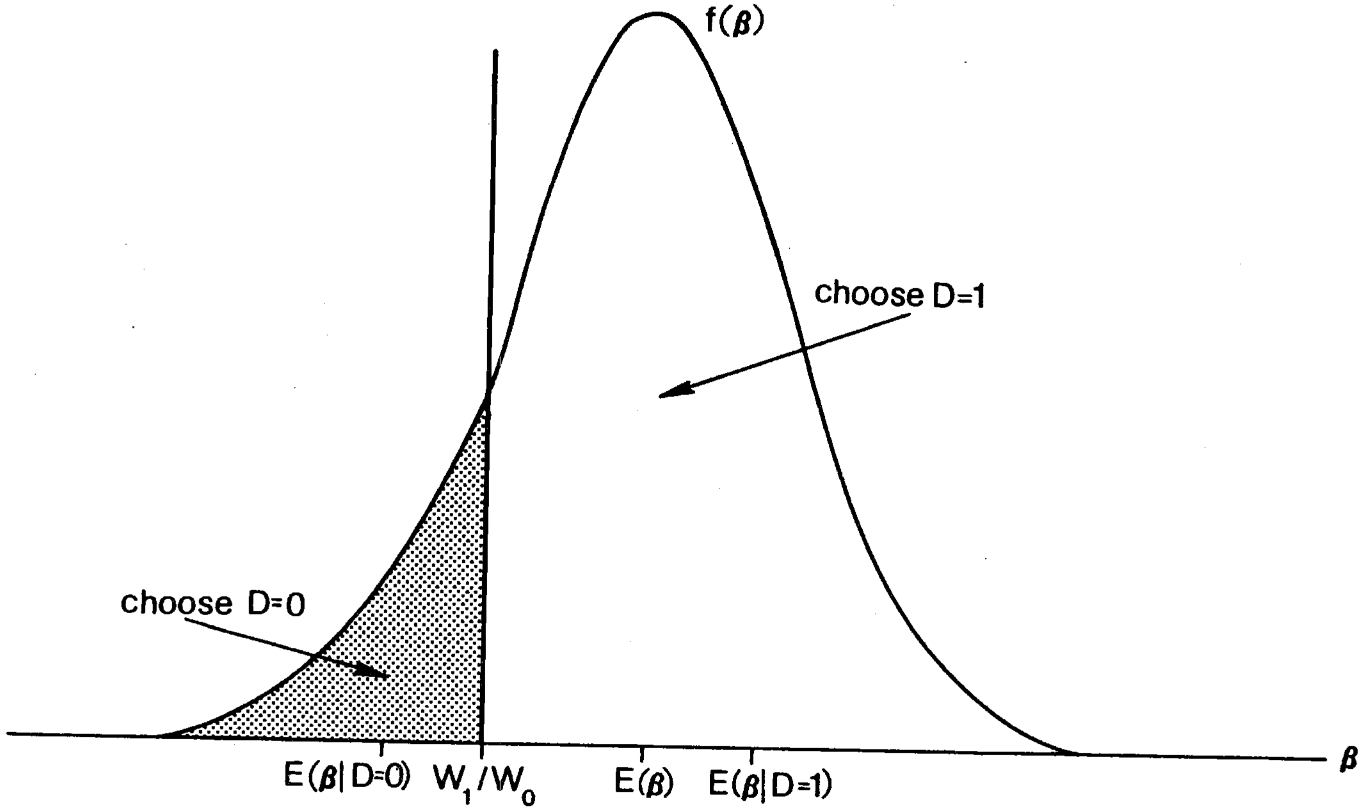

Figure 2

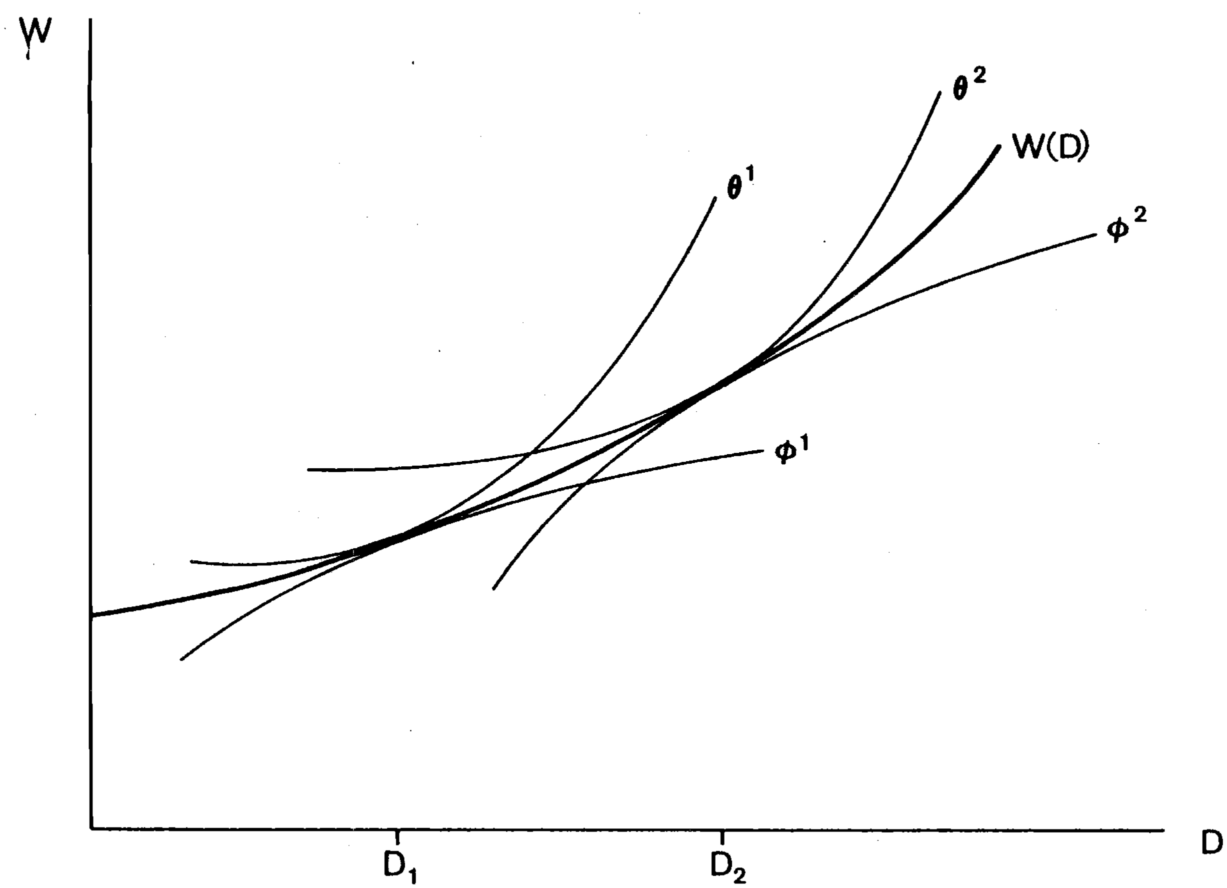

Figure 3

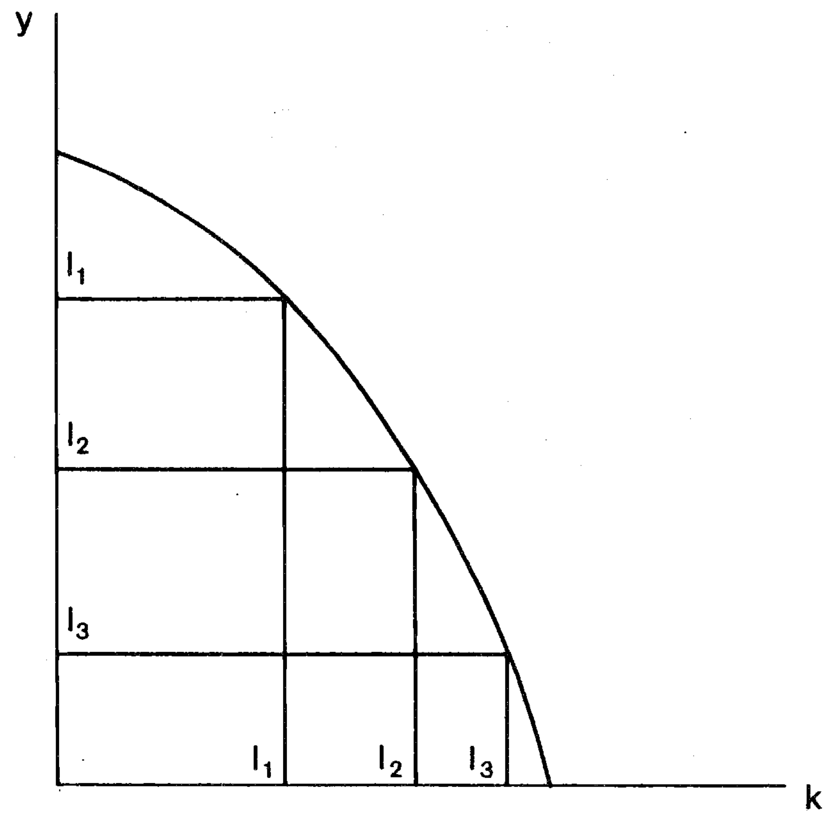

Figure 4

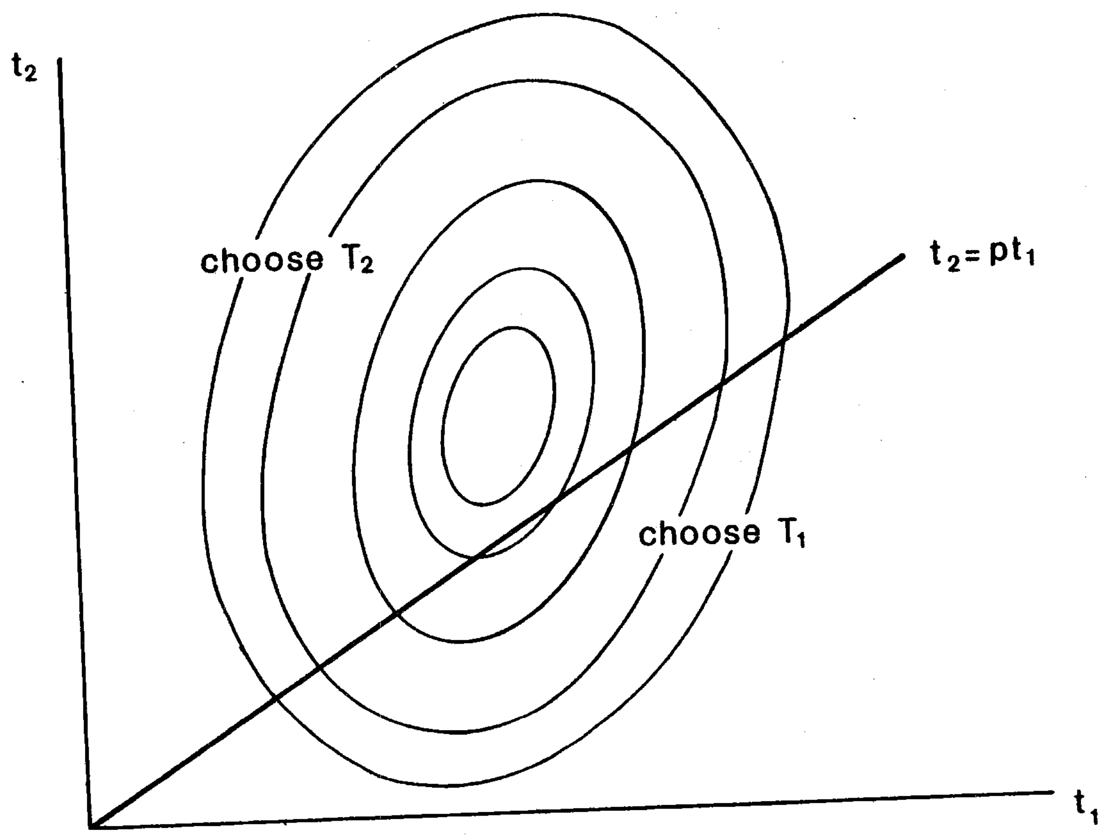

Figure 5

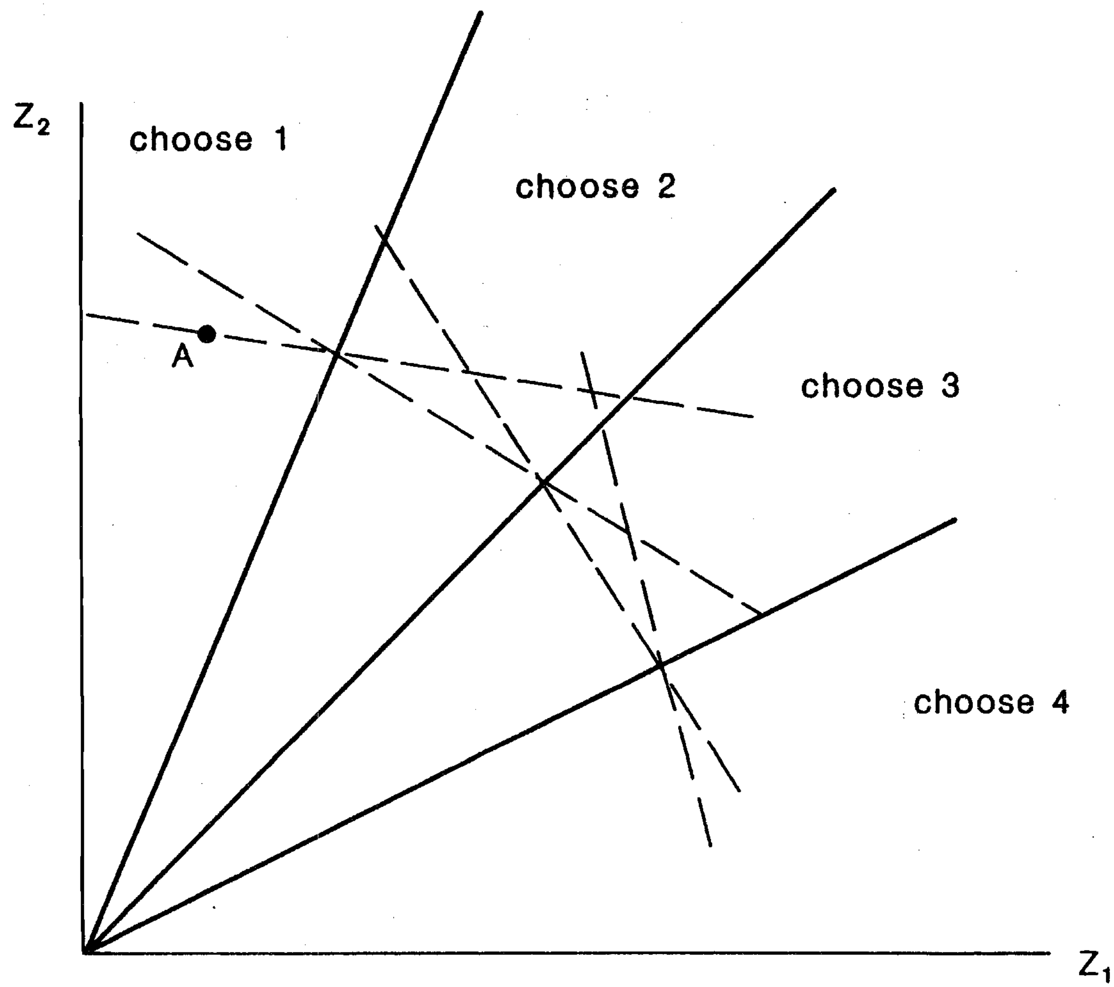

Figure 6