NBER Working Paper Series

LABOR QUALITY, THE DEMAND FOR SKILLS, AND MARKET SELECTION

Sherwin Rosen*

Working Paper No. 162

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR
AND SOCIAL INSTITUTIONS

National Bureau of Economic Research, Inc.
204 Junipero Serra Boulevard, Stanford, CA 94305

January 1977

*University of Rochester and NBER. The author acknowledges with appreciation the financial support of the National Science Foundation and the Department of Health, Education and Welfare. He is indebted to Kamran Norman and Robert Willis for helpful discussions.

LABOR QUALITY, THE DEMAND FOR SKILLS, AND MARKET SELECTION

Sherwin Rosen
University of Rochester and NBER

I. NATURE OF THE PROBLEM

This paper investigates some alternative definitions of labor for productivity and demand analysis. The central issue requires introducing notions of "skill" into the production function and revolves around the nebulous concept of labor quality and the existence of meaningful aggregates of heterogeneous labor services. While capital aggregation has been discussed at length, far less attention has been devoted to labor aggregation, though the required structure is different enough to warrant independent development. A fundamental feature of labor is that the supply of potential services of any worker typically spans a much broader spectrum than those actually put to use in the market. Specialization and division of labor in the organization of work activities are ubiquitous; suggesting that labor markets might be usefully analyzed as marriage markets, matching workers to jobs. Central to this view is the fact that the distribution of potential skills among members of the labor force renders some individuals more capable of performing certain jobs than others. The work reported below illustrates some of these issues, based on the theory of optimal assignments (Koopmans and Beckmann), the Ricardian theory of comparative advantage (especially the development of Dorfman, Samuelson and Solow) and the theory of marriage (Becker).

Three approaches have guided empirical work in this field and this work is related to all of them:

(i) Economists working in the human capital tradition have maintained a distinction between "raw labor" and human capital, measuring the latter by a wage-weighted index of the distribution of education in the labor force, thereby in effect including education as a factor of production. This split obviously follows quantity/quality lines and appears to be a natural first-approximation toward measuring labor in efficiency units. However, it has not fully adressed the heterogeneity issue and substitution among various types of human capital. We begin to analyze some of those factors here.

(ii) Duality theory offers a wide variety of feasible empirical specifications for production functions involving many factors of production and empirical applications have disaggregated labor according to official skill and occupational categories. This approach relates various labor inputs to their functions as productive agents more closely than the human capital approach, but appears to be less than fundamental. First, the principles underlying official classifications are not transparent. Second, technological factors alone almost never determine the content of work activity and what is called a job. The bundling of work activities into packages labeled jobs and occupations is at least as much influenced by economic decisions affecting the organization of production as by the technology per se. There is no better example than Smith's pin factory. "Pin Maker" would represent an adequate occupational classification with the crudest kind of organization, whereas a more sophisticated organization would call for distinguishing among wire stretchers, point sharpeners and so on. Some examples of the endogenity of occupational classifications are shown below.

(iii) There has been some attempt (Welch) to reduce the dimensionality of labor inputs using an unobserved factor approach. This method assumes that labor services in production can be represented by a small set of latent, unobserved factors, with observed labor categories embodying these factors in alternative proportions (fixed "within" but varying "between"). For example, the latent factors might be strength and intelligence, and the observed labor categories might be several education groups cross-classified by race, sex and work experience. A market arbitrage argument is used to establish restrictions among wages in the observed categories. Assuming that any vector of unobserved factors can be achieved by linear bundles of observed categories, observed wages in each category must be similar linear combinations of the implicit prices of the unobserved components. The latter form a basis for reducing the dimensionality of the input space to the number of latent variables. A difficulty with this technique is that it may not survive aggregation over several goods with dissimilar latent factor technologies. Furthermore, the underlying hypothesis that workers embody fixed bundles of characteristics that are linearly combinable does not necessarily accord with the principles of comparative advantage and the virtues of specialization, so that some investigation of the practical limitations of the method is worthwhile.

The paper is organized as follows: Section II considers the organization of work activities in a simple fixed coefficient technology in the presence of comparative advantage among various classes of workers. Assuming that the number of independent productive activities exceeds the number of comparative advantage classes, an application of the envelope theorem shows the derivation from first principles of a neoclassical production function with input dimension (the number of workers of each type) smaller than the engineering technology (the number of activities). This is the basic result illustrating that occupational classifications depend on both the technology and the distribution of skills (factor supplies) in the working population, a fact that may be relevant to international and other cross-sectional differences in productivity and the demand for labor. The situation is reversed in section III, which treats the case where the number of worker classifications exceeds the number of production activities. In this case the micro-technology cannot be reduced below the basic set of work activities one starts with, and within these categories labor can be aggregated according to efficiency units. However, the nature of factor endowments in economies of this sort is rather different than in the neoclassical model, and leads to an output transformation function that has all the neoclassical properties. This result is reminiscent of an example of Houthakker (also, see Sato) who also obtained smooth neoclassical behavioral functions from underlying distributional phenomena. Section IV examines the characteristics--factor approach to labor aggregation and relates it to the results in section III, noting an inherent difficulty arising from selectivity of various ability groups of workers among work activities due to comparative advantage. In effect, the existence of rent destroys the possibility of simple linear aggregation. The point is also related to a general approach to income distribution originated by A.D. Roy and carried forward by Tinbergen, Mandelbrot, Houthakker and Sattinger. Finally, section V indicates some problems with applying the theory of marriage directly to labor demand. These issues become most interesting when there are incomplete markets that limit the gains from fully exploiting comparative advantage, due to transactions costs. The results are limited, but some examples show that any predictions concerning positive or negative assortive matching of workers depends not only on the correlation of talents among members of the work force, but also on the nature of technology and the distribution of demands for various outputs.

II. AN INDIRECT PRODUCTION FUNCTION

A basic result on the virtues of specialization and optimum job assignments in the presence of comparative advantage is most easily shown in the context of a simple engineering production function with fixed coefficients. Capital is ignored, without apology, in what follows. Let x be output and T_i be a production activity. The technology is given by

$$x = \min(T_1/a_1, T_2/a_2, \ldots, T_n/a_n) \tag{1}$$

where a_j is the activity input requirement per unit output. For concreteness think of (1) as the engineering production function for Smith's pin factory, with the T_j representing all the independent steps in the production process, such as drawing the wire, sharpening the points, and so forth. More generally, it is natural and convenient to associate T_j with independent "tasks," a collection of which constitutes a job. Let there be m types of workers. Workers of each type are differentiated by a capacity vector $(t_{1j}, t_{2j}, \ldots, t_{nj})$, $j=1,\ldots,m$, with t_{ij} indicating the maximum amount of activity i obtainable from workers of type j. These maximal amounts occur when a single activity is pursued full time. However, a worker's time may be divided among several activities. Assume that output in each activity is strictly proportional to the time devoted to it, with no further interactions from mixing activities. Thus the direct virtues of specialization due to indivisibilities, on-the-job learning and innovation discussed by Smith are ignored. Comparing workers j and i, worker j will be said to have a comparative advantage in task h relative to task k if $t_{hj}/t_{kj} > t_{hi}/t_{ki}$. Equivalently, worker i has a comparative advantage in task k. Comparative advantage is assumed to exist in all tasks and among all types, i.e.,

$$t_{hj}/t_{kj} \neq t_{hi}/t_{ki} \qquad (2)$$

for all pairs (h,k) and (j,i).

A familiar diagram illustrates the situation when there are two activities and two types. Two workers are shown in Figure 1, A and B, with A having comparative advantage in T_2 and B in T_1. An efficient assignment maximizes the activity levels attainable from a given labor force and the efficient frontier has two facets: Along the upper edge A is completely specialized to T_2 and B engages in both activities, since B is assigned to T_1 first due to comparative advantage there (fractional assignments will be the rule and not the exception in this problem). A is completely specialized to T_2 and B to T_1 at the corner. Only if the production requirement for T_1 exceeds t_{1B} is it efficient to allocate some fraction of A's time to T_1, and along the lower edge B is specialized to T_1 while A does both. All other assignments sacrifice activities. For example the most inefficient (the dotted line segments) assign inversely to comparative advantage (B specialized to T_2 and A fractionally assigned along the upper edge, etc). Of more interest is the case of no specialization, where A and B act as independent agents of production. Since the engineering technology dictates activities in fixed proportions, the maximum total activity levels in this case are the sums of A's and B's separate activity levels along arbitrary rays through the origin--the dashed line. It is inefficient because no comparative advantage is exploited. The gains from forming a production team and taking advantage of different talents by optimum assignment are shown by the shaded area. Evidently comparative advantage produces a superadditivity or synergy. This interactive effect captures a fundamental notion of "complementarity" in production.[1]

While this example is well known certain features have no counterpart in its use in trade theory and form a more fundamental basis for production theory. In particular, it is possible to derive a quasi-concave production function in terms of bodies of each type from first principles. Consider the problem

$$x = \max_{T_{ij}}[\min\{\sum_j T_{1j}/a_1, \sum_j T_{2j}/a_2, \ldots, \sum_j T_{nj}/a_n\}] \quad (3)$$

subject to

$$T_{1j}/t_{1j} + T_{2j}/t_{2j} + \ldots + T_{nj}/t_{nj} \leq N_j, \; j = 1,2,\ldots,m \quad (4)$$

where $\underline{N_j}$ is the number of workers of type j available. Then the envelope theorem proves that there exist nonnegative multipliers $(q_1,\ldots,q_m)$ and a quasi-concave function $x = F(N_1,N_2,\ldots,N_m)$ such that

$$x = F(N_1,\ldots,N_m) = \max_{T_{ij}}[\min\{\sum_j T_{1j}/a_1,\ldots, \sum_j T_{nj}/a_n\} + \sum_j q_j(N_j - \sum_i T_{ij}/t_{ij})] \quad (5)$$

The function F(N) is an efficient "indirect" production function.[2] Its derivatives when defined satisfy $\partial x/\partial N_i = q_i$ and represent induced marginal products of worker types rather than of production activities.[3] The theory holds for any $\underline{n}$ and $\underline{m}$, not necessarily of the same dimension, but gains considerable practical interest when the number of activities ($\underline{n}$) greatly exceeds the number of types of workers ($\underline{m}$).

Construction of an indirect production function in the 2 x 2 case follows directly from Figure 1 and illustrates the general method. Consider the problem of producing one unit of $\underline{x}$ with alternative numbers of types of workers. If only type $\underline{A}$ workers are available, then $a_1/t_{1A} + a_2/t_{2A}$ of them are required. Similarly, if only $\underline{B}$ workers are available, then $a_1/t_{1B} + a_2/t_{2B}$ of them are necessary. Suppose $t_{2A}/t_{1A} > t_{2B}/t_{1B}$ as before. Then at the specialization point in Figure 1, $\underline{B}$ is assigned to $\underline{T}_1$ and $\underline{A}$ to $\underline{T}_2$, and if $N_A = a_2/t_{2A}$ and $N_B = a_1/t_{1B}$ one unit of $\underline{x}$ is produced. These three possibilities are shown in Figure 2. However, Figure 1 shows that any linear mixtures of $\underline{A}$'s and $\underline{B}$'s between adjacent points in Figure 2 are efficient because they follow optimum assignments: The two connected line segments in Figure 2 are dual to the activity possibility frontier, with the upper edge corresponding to specialization of B (the lower face in Figure 1) and the lower one corresponding to specialization of A. In fact the connected lines in Figure 2 represent the efficient unit isoquant--the level set of $F(N_A,N_B)$ in this case--and serve as a perfectly adequate basis for a production function.

There are inefficient production functions as well. For example, corresponding to no specialization in Figure 1 would be the straight line connecting the intercepts of Figure 2. Here $\underline{N}_A$ and $\underline{N}_B$ appear as perfect substitutes with efficiency units measured by total product per worker. Corresponding to the perverse assignment in Figure 1 are the lines connecting the intercepts and the cross in Figure 2. This argument immediately shows that the efficient indirect production function is quasi-concave: It exhibits <u>substitution</u> even though the underlying factors are in some sense complements. The marginal rate of substitution is t_{1B}/t_{1A} in the branch where $\underline{B}$ is specialized and is t_{2B}/t_{2A} in the branch where A is specialized. These are determined by factor skill endowments only and are independent of the technology. Imperfect substitution around the corner does depend on the technology, but its extent also depends on the "distance" between skill endowments of worker types. A different configuration of input requirements, a_i, changes the location of the corner. Whatever its location, curvature around the corner depends on the extent of comparative advantage $t_{2A}/t_{1A} \div t_{2B}/t_{1B}$.

The experiment resulting in Figure 2 fixed the unit isoquant in the (T_1,T_2) plane and varied the number of workers of each type to maintain output, always efficiently assigning worders to activities along the way. These variations alter the intercepts and location of the corner of the efficient frontier in Figure 1, but leave the slope of each facet unchanged. Where the activity isoquant touches a facet and not a corner of the frontier, A and B workers have one activity in common and it is this commonality that determines the marginal rate of substitution between them in the efficient indirect production function. The same experiment applies when the underlying engineering technology admits some substitution among activities, though the induced substitution among worker types is somewhat tempered. Figure 3 illustrates the experiment in such a case. Variations in N_A and N_B beyond certain limits result in unique tangents between the unit isoquant and the activity possibility frontier at points c and d (these degenerate to a single point in the fixed coefficient case). N_A and N_B share one activity at these points, which establishes a unique MRS between them as before. Between these limits the isoquant touches the frontier at a corner, the A's and the B's share no activities and the indirect MRS between N_A and N_B merely follows the direct MRS between T_1 and T_2. The indirect MRS is trapped by t_{1B}/t_{1A} and t_{2B}/t_{2A} and varies smoothly in between. Hence an isoquant of the indirect production function would appear just as it does in Figure 2, except with a rounded corner. Whatever the nature of substitution among activities, the indirect production function displays more substitution than the engineering technology because of the substitution possibilities in the assignment of workers to jobs in addition to direct substitution among activities.

The argument is readily extended to n activities and involves the same two steps of first finding efficient assignments of workers to activities to construct the efficient activities possibility set, and then varying the number of workers of each type, maintaining efficient assignments and mapping out the indirect production function.

The efficient activity possibility set is found by solving an artificial maximum problem (see Dorfman, Samuelson and Solow). Define a set of shadow prices p_i one for each activity. In context these might be thought of as piece rates. Maximize the value of production activities subject to the capacities of each type of worker and determine how these assignments are related to the implicit prices. There is a simple analytic solution in the n x 2 case. Consider the problem

$$V = \max_{T_{ij}}\{p_1(T_{1A} + T_{1B}) + p_2(T_{2A} + T_{2B}) + \dots + p_n(T_{nA} + T_{nB})\} \qquad (6)$$

Subject to

$$\begin{aligned} T_{1A}/t_{1A} + T_{2A}/t_{2A} + \dots + T_{nA}/t_{nA} &\leq N_A \\ T_{1B}/t_{1B} + T_{2B}/t_{2B} + \dots + T_{nB}/t_{nB} &\leq N_B \end{aligned} \qquad (7)$$

The dual problem requires choosing shadow prices on worker types to minimize

$$V = \min_{r}\{r_A N_A + r_B N_B\} \tag{8}$$

subject to

$$\left.\begin{aligned} r_A/t_{iA} &\geq p_i \\ r_B/t_{iB} &\geq p_i \end{aligned}\right\} \quad i = 1,\ldots,n \tag{9}$$

r_A and r_B in (8) and (9) may be interpreted as wage rates of each type. Thus (6) maximizes production value, while (8) minimizes cost. Since the primal problem involves two constraints, the solution to (8) requires that exactly two constraints in (9) are binding, so that there are n^2 possible solutions. Of these, $\underline{n}$ solutions are immediate. They require both $r_A = t_{iA}p_i$ and $r_B = t_{iB}p_i$ for $i = 1,\ldots,n$, and correspond to the case where $\underline{p_i}$ is so large that both $\underline{A}$'s and $\underline{B}$'s work only at task $\underline{T_i}$ (the intercepts in fig. 1). However, the solution to (6) may involve two noncongruent activities, say $\underline{T_i}$ and $\underline{T_j}$, illustrated by the interior corner solution in Figure 1. Suppose $\underline{A}$ has comparative advantage at $\underline{T_i}$ and $\underline{B}$ has comparative advantage at $\underline{T_j}$. Consider the solution

$$r_A/t_{iA} = p_i \text{ and } r_B/t_{jB} = p_j \tag{10}$$

so that $\underline{A}$ is assigned to $\underline{T_i}$ and $\underline{B}$ to $\underline{T_j}$. Then it must also be true from (9) that

$$r_A/t_{jA} > p_j \text{ and } r_B/t_{iB} > p_i \tag{11}$$

Equations (10) and (11) imply

$$t_{iA}p_i > t_{jA}p_j \text{ and } t_{jB}p_j > t_{iB}p_i \tag{12}$$

or

$$t_{jB}/t_{iB} > p_i/p_j > t_{jA}/t_{iA} \tag{13}$$

which holds true for some p_i/p_j, given the comparative advantage assumption. If B had been assigned to T_i and A to T_j, the same logic leads to the inequality

$$t_{jB}/t_{iB} < p_i/p_j < t_{jA}/t_{iA} \tag{14}$$

which cannot possibly hold true given A's comparative advantage on T_i, and is therefore nonoptimal. Continuing in this way for all (i,j) pairs ($i \neq j$) it is seen that $n(n-1)/2$ possible assignments of workers to activities must be nonoptimal; and only $[n + n(n-1)/2]$ of the possible n^2 solutions form a basis for the efficient activity possibility set.

Finding the efficient frontier requires determining prices $\{p_i\}$ that imply degenerate solutions to (6). The 3x2 case illustrate the method.[4] For the sake of example, suppose $t_{2A}/t_{1A} > t_{2B}/t_{1B}$, $t_{2A}/t_{3A} > t_{2B}/t_{3B}$ and $t_{3A}/t_{1A} > t_{3B}/t_{1B}$. Let V_{ij} represent the maximum value of activities in (6) and (8) when type A workers are optimally assigned to T_i and B workers are assigned to T_j. The argument leading to (13) shows that V_{12}, V_{13} and V_{32} cannot be optimum for this problem. Array the other possibilities in a matrix, with row $\underline{i}$ representing the assignment of A's to T_i and row $\underline{j}$ representing the assignment of B's to T_j. If V_{ij} does not rest on optimal assignments, its value is left blank in the matrix. In the case under consideration

$$\{V_{ij}\} = \begin{bmatrix} p_1t_{1A}N_A + p_1t_{1B}N_B & \text{———} & \text{———} \\ p_2t_{2A}N_A + p_1t_{1B}N_B & p_2t_{2A}N_A + p_2t_{2B}N_B & p_2t_{2A}N_A + p_3t_{3B}N_B \\ p_3t_{3A}N_A + p_1t_{1B}N_B & \text{———} & p_3t_{3A}N_A + p_3t_{3B}N_B \end{bmatrix}. \quad (15)$$

The fully degenerate solutions to (6) or (8) imply equality among various combinations of the elements of V_{ij}. They can be found as follows. Consider the assignment of $\underline{A}$ to T_1 and $\underline{B}$ to T_1, which is optimal if $\underline{p}_1$ is large enough. Then $V_{11} = p_1t_{1A}N_A + p_1t_{1B}N_B$. Given some arbitrary values of $(\underline{p}_1, \underline{p}_2, \underline{p}_3)$ consistent with the optimality of V_{11}, find the minimum possible value of $\underline{p}_1$ that maintains V_{11} optimal. Notice in (15) that only V_{11}, V_{21} and V_{31} are altered as $\underline{p}_1$ varies. A fully degenerate solution corresponding to a facet of the efficient frontier therefore implies $V_{11} = V_{21} = V_{31}$, or the equivalent restrictions $p_2/p_1 = t_{1A}/t_{2A}$ and $p_3/p_1 = t_{3A}/t_{1A}$. Now consider the element V_{21}. Here two experiments are required, since V_{21} is a function of both $\underline{p}_1$ and $\underline{p}_2$. First, find the minimum value of $\underline{p}_1$ consistent with the optimality of V_{21}. This also implies $V_{11} = V_{21} = V_{31}$, or the same restriction on the $\underline{p}$'s as before. Second, change $\underline{p}_2$ and find its minimum consistent with the assumed optimality of V_{21}. Since only V_{21}, V_{22} and V_{23} are changed when p_2 varies, another fully degenerate solution corresponding to another facet of the efficient frontier implies $V_{21} = V_{22} = V_{23}$, or the restrictions $p_2/p_1 = t_{1B}/t_{2B}$, and $p_2/p_3 = t_{3B}/t_{2B}$. Proceeding in this fashion for all the elements of V_{ij} leads to only one more restriction, namely $p_3/p_1 = t_{1B}/t_{3B}$ and $p_2/p_3 = t_{3A}/t_{2A}$. Hence the efficient frontier has exactly three facets, one for each restriction. It is shown in Figure 4 (c.f. Whitin). The restriction implied by $V_{11} = V_{21} = V_{31}$ corresponds to the lower triangular facet. Here B is completely specialized to T_1, while A's engage in all three activities. The restriction implied by $V_{21} = V_{22} = V_{33}$ corresponds to the upper triangular facet; there A is completely specialized to T_2 and B engages in all three activities. Finally, the remaining restriction, $V_{31} = V_{23} = V_{33}$ corresponds to the quadralateral, where A is assigned to activities T_2 and T_3, while B is assigned to T_1 and T_3. Notice that within each facet, A and B have precisely one activity in common.[5] The argument readily extends to the $n \times 2$ case, and the efficient activities polyhedron has the same number of facets as activities, so long as (2) holds.

Construction of an indirect production function in the 3-activities case is shown in Figure 5. Each connected line segment corresponds to a facet of the efficient frontier, with the marginal rate of substitution determined by the commonality of activities there. Evidently the argument above shows that if there are n activities the indirect isoquant consists of n connected line setments. Clearly, as n grows large the indirect production function gets arbitrarily smooth and neoclassical.

It remains to be shown that the indirect production function is supported by a price system. That this is so follows from the analysis of (6) and (7). Suppose workers are paid in proportion to their activity output at unit rewards p_i, and individually choose production activities to maximize their incomes. For example, each worker of type A chooses T_{iA} to maximize

$$p_1T_{1A} + p_2T_{2A} + \dots + p_nT_{nA}, \tag{16}$$

subject to

$$T_{1A}/t_{1A} + T_{2A}/t_{2A} + \dots + T_{nA}/t_{nA} \leq 1, \tag{17}$$

and similarly for workers of type B. But the solution to (16) is exactly the same as the solution to (6). Therefore, the market solution is efficient. The rest follows from the fact that the market shadow prices or piece rates p_i are prescribed by the efficient activities frontier, as shown above. Consequently, the factor price frontier of the indirect production function (5) in terms of wage rates, r_j, consists of n isolated points corresponding to each facet of the efficient activity frontier, each one of which is dual to a facet of an isoquant. For example, in the 3 x 2 example discussed above, the factor price frontier in the (r_A, r_B) plane lies along the three rays $r_A/r_B = t_{1A}/t_{1B}$, $r_A/r_B = t_{2A}/t_{2B}$, and $r_A/r_B = t_{3A}/t_{3B}$. The first ray corresponds to the lower triangular facet in Figure 4. Here a person of type A is indifferent among all three activities, while B's definitely prefer activity 1. Hence $r_A = p_1t_{kA} = p_2t_{2A} = p_3t_{3A}$ and $r_B = p_1t_{1B}$, and so forth for all three facets.[6] Finally, the construction of the indirect isoquant explicitly shows that the marginal rate of substitution along each of its facets equals the wage ratio there. Consequently, maximizing the indirect production function subject to a wage-cost constraint on bodies truly does maximize output. Any indeterminacy on the interior of a line segment is eliminated because N_a and N_B are given.

We have exhibited a mechanism for incorporating the concepts of skill into neoclassical production theory via the assignment of workers to jobs, and shown how an observable indirect production function is mapped out by variations in factor supply conditions. A number of consequences follow:

(i) The construction points out some potential pitfalls in using published occupational classes for productivity analysis, since the definition of an occupation in terms of a collection of work activities is not invariant and endogenously depends on factor supplies. Workers with similar skill endowments will be found on different work assignments, since differences in factor supplies locate them on different facets of the indirect production function (e.g., the upper, lower, or middle branches of Figure 5). This difficulty is tempered insofar as occupational classifications index "capacities" rather than actual endogenous job assignments. The stated principles of occupational classification are not encouraging in this regard, but examination of the actual classifications suggests increasing logical difficulties with the level of disaggregation.

(ii) If worker types are tentatively identified with observed economic-demographic categories such as years of schooling, the above shows precisely how education enters production, thus providing a link between supplies of human capital and less well-analyzed demands for them. Since differences in capacities among worker types automatically induces substitution in the indirect production function, fixed wage weighted indexes of labor input for total factor productivity indexes are subject to well known substitution bias that can be easily understood in terms of this framework. A more subtle difficulty emerges if the capacity vectors among the various observed categories change over time, as seems likely. Unless these changes approximate uniformity over all classes, the indirect production function is shifted nonhomogeneously, and both total factor productivity and the "systematic" part of the production function are altered through time. The latter shows up as "biased technical change." It seems probable that variable weight Divisia type indexes advocated by Jorgenson and Griliches might eliminate part of the problem. But it is by no means apparent that they solve it completely.

iii. For cross-sectional analysis of production and demand, one must be reasonably confident that the observed "capacity" classifications are more or less uniform across data points. If not, price-quantity observations are not mapped out of a common structure. It is not difficult to derive examples whereby estimated substitution possibilities may be either greater or less than the true possibilities when productivity differences within classes over the observations exist but are not statistically controlled. In particular, one might expect this problem to be most severe in the case of international comparisons. Simple though it is, a virtue of this framework is in clarifying the meaning of "international differences in technology." It is possible and plausible that there exist no differences in direct engineering production functions, yet differences among indirect observable production functions arise if there are differences in capacity endowments among the relevant populations. The often observed fact that a factory in the U.S. is more productive than its identical twin in a less developed country is partially due to the fact that the work assignments inherent in the design of capital equipment are optimal for the U.S. labor force, but are definitely suboptimal for the foreign labor force because of differences in the extent and distributions of worker comparative advantage.

iv. This construction helps understand the well known phenomena of skill bumping associated with layoffs and business cycles (see Reder). So long as layoffs are not proportional across worker types, short run employment declines imply corresponding reassignments of work activities among employed workers, a kind of short-run substitution effect as it were. Any differential adjustment costs that insulate some groups from transitory demand shocks has the effect of flip-flopping work assignments among various branches of the indirect production function isoquant. If in addition the fixed factors (capital) embody work routines that are optimal for longer term, permanent conditions, the short-term observations may even switch over to inefficient regions of the indirect production function, implying systematic variations in measured labor productivity that are consistent with observed phenomena. Extending Stigler's idea of flexibility to this problem suggests less specialization and rigidity of work routines in cyclically sensitive industries to minimize these short-run inefficiencies.

v. That workers' skills are developed over the course of their labor market and work experiences is central to the theory of human capital and implies a progression of work assignments over the life cycle. Thus life cycle variation in the capacity vectors $\underline{t}_{ij}$ is to be expected and the indirect production function suggests imperfect substitution among age groups or cohorts, a fact that is difficult to understand in the usual neoclassical model in terms of efficiency units. It follows that the size of a cohort may have an effect on relative cohort earnings throughout its life cycle. Being a member of a relatively large birth cohort may haunt one for a whole lifetime, as in Easterlin's relative income-fertility hypothesis. This also would appear to explain why alternative vintages of graduates in some professional markets (notably the present academic market) fare relatively better or worse when general demand conditions are changed. It also suggests a common cohort variance component in wage rates that renders the covariance between starting salary and lifetime income larger than one might otherwise suppose, and thus makes initial salary a better predictor of new entry than the standard permanent-transitory decomposition would suggest.

III. PRODUCTION POSSIBILITIES

We turn now to the case where the number of worker categories exceeds the number of activities. Consider technologies using two activities and denote the efficient activities frontier by $F(T_1,T_2) = 0$. Now add a third type, C, to the construction of Figure 1, with comparative advantage somewhere between A and B: $t_{2A}/t_{1A} > t_{2C}/t_{1C} > t_{2B}/t_{1B}$. Then the activities possibility set would have three facets instead of two, with slopes corresponding to these ratios. As still more types are added, more facets appear, filling in the corners of Figure 1 and smoothing out $F(T_1,T_2) = 0$. Its limiting behavior can be derived by making use of the result of (13) and (14) that a price system and free choice implies efficient assignments. Discussion is confined to the two activity case, which generalizes immediately to $\underline{n}$ activities.

Consider a randomly chosen person with capacity vector (t_1,t_2). Then his potential income is $p_1T_1 + p_2T_2$ and is constrained by $T_1/t_1 + T_2/t_2 \leq 1$. Potential income in the (T_1,T_2) plane is described by a family of straight lines with slope $-p_1/p_2$. The time-resource constraint is also described by a straight line with slope $-t_2/t_1$. Devoting full time to activity 2 maximizes income if $t_2/t_1 > p_1/p_2$ and similarly for activity 1 if the inequality goes in the opposite direction. The worker is indifferent among activities if the ratio of his capacities equals the price ratio. Thus, there is a convenient ordering of choice by workers according to comparative advantage: Pick an arbitrary price ratio, $\rho \equiv p_1/p_2$. Then on the efficient frontier all workers with comparative advantage in excess of ρ are optimally assigned to activity 2 and actually choose it, while all those with comparative advantage less than ρ are assigned to activity 1.

As the number of types of workers becomes indefinitely large the pairs t_1 and t_2 are conveniently described by a density function $f(t_1,t_2)$, with each worker belonging to a point in the (t_1,t_2) plane. The activity possibilities frontier is defined parametrically by the conditional expectations

$$T_1 = \int_0^\infty \int_0^{\rho t_1} t_1 f(t_1,t_2)dt_1dt_2 = g(\rho) \tag{18}$$

$$T_2 = \int_0^\infty \int_{\rho t_1}^\infty t_2 f(t_1,t_2)dt_1dt_2 = h(\rho) \tag{19}$$

and is "swept out" of the density $f(t_1,t_2)$ as ρ varies from zero (everyone assigned to T_2) to infinity (everyone assigned to T_1). Furthermore, differentiating (18) and (19) yields

$$g'(\rho) = dT_1/d\rho = \int_0^\infty t_1 f(t_1,\rho t_1)dt_1 \tag{20}$$

$$h'(\rho) = dT_2/d\rho = \int_0^\infty - \rho t_1 f(t_1,\rho t_1)dt_1 \tag{21}$$

which implies that the slope of the efficient frontier is the shadow price ratio itself:

$$dT_2/dT_1 = h'(\rho)/g'(\rho) = -\rho. \tag{22}$$

Further application of the implicit function theorem demonstrates that $d^2T_2/dT_1^2 \le 0$, so that $F(T_1,T_2) = 0$ is quasi-concave.

Example: Let the distribution of t_1 and t_2 in the labor force be independent, strongly Paretian. Then $f(t_1,t_2) = M\alpha^2 t_1^{-\alpha-1}\, t_2^{-\alpha-1}$ for $t_1 \geq 1$, $t_2 \geq 1$, where M is the number of workers and α is a fixed parameter greater than unity. Applying (18) and (19) gives[7]

$$\left.\begin{aligned} T_1 &= k\lambda\rho^{\alpha-1} \\ T_2 &= k(1-\rho^{\alpha}) + k\lambda\rho^{\alpha} \end{aligned}\right\} \quad \rho \leq 1$$

and

$$\left.\begin{aligned} T_1 &= k(1-\rho^{\alpha}) + \lambda k\rho^{-\alpha} \\ T_2 &= k\lambda\rho^{-\alpha+1} \end{aligned}\right\} \quad \rho \geq 1$$

with $k = M\alpha/\alpha-1$, $\lambda = \alpha/2\alpha-1$, which yields

$$T_2 = k[1 + (\lambda-1)(T_1/k\lambda)^{\alpha/\alpha-1}] \quad \text{for} \quad T_1 \leq k\lambda \leq T_2$$

$$T_1 = k[1 + (\lambda-1)(T_2/k\lambda)^{\alpha/\alpha-1}] \quad \text{for} \quad T_1 \geq k\lambda \geq T_2 \, .$$

The activity frontier is composed of two symmetric branches that link up with a common slope of unity at $T_1 = T_2 = k\lambda$.

The shape of $F(T_1,T_2)$ is determined wholly by the parameter α. As α gets large, workers' capacities become increasingly concentrated on (1,1), they become more nearly alike, and $F(T_1,T_2) = 0$ approaches a straight line. Clearly the construction of Figure 1 and its extension to many types reveals that curvature in the activity possibility frontier depends on worker diversity in comparative advantage, and diversity is minimized when α is large. The behavior of F as $\underline{\alpha}$ approaches unity is more complicated because the unconditional means of t_1 and t_2 become unbounded. However, the logic about diversity suggests that $F(T_1,T_2)$ should become more like a square, with a sharply rounded corner at $T_1 = T_2$, and that is indeed the case. For $\underline{\alpha}$ between these limits $F(T_1,T_2)$ has a shape rather like a portion of an ellipse characteristic of the way production sets are typically drawn in economics. For example, for $\alpha = 2$ it appears like a flattened quarter circle.

Evidently the envelope theorem can be applied when $\underline{m} > \underline{n}$, but it is not terribly interesting to do so because the engineering technology already is of minimum dimension. However, just as increasing the number of activities smoothed out the indirect production function in section II, a similar result applies here to production possibility sets in the economy at large. Consider two goods with technologies

$$X_1 = \min\{T_1/a_{11}, T_2/a_{21}\} \qquad X_2 = \min\{T_1/a_{12}, T_2/a_{22}\}.$$

The production possibilities set maximizes X_2 given X_1 and $F(T_1,T_2) = 0$. Its construction is shown in Figures 6 and 7. In spite of the assumption of fixed coefficient technologies, the macro production set is smooth and all workers are fully employed. Here is another example of smooth neo-classical macro behavior deriving from underlying distributional micro phenomena that forms an adequate basis for macro equilibrium theory, but is rather different from the usual construction. Notice how the activity set of Figure 6 maps into the production set in Figure 7 by the addition of two vectors (each representing an allocation of activities to each good) summing to a point on $F(T_1,T_2)$. Each division of output is supported by a unique shadow price ratio ρ and a corresponding division of the labor force among activities; and the effect of the distribution of output and product demand conditions on the distribution of earnings is immediately revealed. Also, comparative advantage in goods production clearly depends on the distribution of talents in the labor force, a natural and manageable extension of the Heckscher-Ohlin kind of results to "many factors."

IV. A LATENT FACTOR APPROACH

An alternative methodology for labor aggregation has been proposed by Welch. The technology is specified in terms of unobserved latent factors or "characteristics", with observed labor categories embodying the latent characteristics in alternative proportions. It is further assumed that any given factor requirement can be achieved by linear combinations of worker types, which implies a kind of perfect substitution among types depending on the characteristics they embody. If all worker types are to be fully employed, relative wage rates cannot deviate from the fixed technical rates of substitution between them. Further, if the number of latent factors is less than the number of worker categories, a basis for aggregating observed classes to a smaller dimensionality has been found.

Let the production function be $X = G(Z)$, where Z_i are the unobserved latent factors, with $i = 1,\ldots,\underline{v}$. There are $\underline{m}$ classes of workers, with $\underline{m} > \underline{v}$, and each class is completely described by an endowment vector $(Z_{1j}, Z_{2j},\ldots,Z_{vj})$, $j = 1,\ldots,\underline{m}$. $\underline{N_j}$ is the number of workers of type $\underline{j}$. The total endowment of factor $\underline{i}$ is $\overline{Z}_i = \sum_j Z_{ij}N_j$ and full employment output is determined by $\overline{X} = G(\overline{Z})$. At full employment output, the marginal product of the $\underline{i}$th latent factor defines an implicit price for it, $\underline{W_i}$, up to a factor of proportionality: $G_i(\overline{Z}) = bW_i$, where $\underline{b}$ is a normalizing constant, and the income, $\underline{y_i}$, of a member of the ith class must be

$$W_1Z_{1i} + W_2Z_{2i} + \ldots + W_vZ_{vi} = y_i \qquad \text{for } i = 1,\ldots,n \qquad (23)$$

In addition, the linear combinability assumption implies that the latent characteristic vector embodied in any group can be built up by a linear combination of exactly $\underline{v}$ other groups, e.g., there exist constants c_{ij} not necessarily all positive such that[8]

$$c_{1j}Z_{11} + c_{2j}Z_{12} + \ldots + c_{vj}Z_{1v} = Z_{1j}$$

$$c_{1j}Z_{21} + c_{2j}Z_{22} + \ldots + c_{vj}Z_{2v} = Z_{2j} \qquad (24)$$

$$\ldots\ldots\ldots\ldots \qquad j = v+1,\ldots,m$$

$$c_{1j}Z_{v1} + c_{2j}Z_{v2} + \ldots + c_{vj}Z_{vv} = Z_{vj}$$

Finally, by the market equilibrium property (23) of a common implicit price for each latent factor, (24) implies

$$y_j = c_{1j}y_1 + c_{2j}y_2 + \ldots + c_{vj}y_v \qquad j = v+1,\ldots,m \qquad (25)$$

There are $\underline{m}$-$\underline{v}$ linear restrictions on the observed wages of the $\underline{m}$ groups, or only $\underline{v}$ independent sources of variation in observed wages. These independent categories are the basis for aggregating the $\underline{m}$ observed categories to $\underline{v}$ linear combinations.

For empirical implementation one uses cross-section data (subeconomies) with the total endowments $\bar{Z}$ differing across observations, maintaining the assumption of a common latent structure Z_{ij}. For example, suppose the observations are mean earnings of various education-experience classes across states (indexed by $\underline{k}$). Assume that $Z_{ijk} = Z_{ij}$ for all i,j,k, but let N_{jk} vary across states. Variation in $\overline{Z}_i$ arises from differences in N_{jk} and provokes differences in implicit prices W_{ik} among the states. (23) becomes

$$y_{jk} = W_{1k}Z_{1j} + W_{2k}Z_{2j} + \ldots + W_{vk}Z_{vj} \qquad j = 1,\ldots,n \qquad (26)$$

and may be viewed as a factor-analytic statistical model, where the implicit prices W_{ik} represent $\underline{v}$ latent factors and the Z's represent factor "loadings." Notice that (26) is not identified. However, the assumption $Z_{ijk} = Z_{ij}$ implies that the same linear restrictions (24) apply to all observational units. Hence (25) becomes

$$y_{jk} = c_{1j}y_{1k} + c_{2j}y_{2k} + \ldots + c_{vj}y_{vk} \qquad j = v+1,\ldots,n \quad (27)$$

and may be viewed as a system of $\underline{m-v}$ regression equations in $\underline{v}$ "independent" variables. The c's are regression coefficients. Given a specification of exactly v underlying factors, (27) may be estimated by regression methods, $c_{1v},\ldots, c_{1j},\ldots,c_{vj}$ are identified for all j, and have the ready interpretation of fixed marginal rates of substitution among types. Note that (26) is not constant across observations because differing endowments cause corresponding differences in implicit prices W_{ik}. The method works because (27) is independent of endowments, due to the repackaging assumptions imposed on the latent factors. This suggests a test for the number of unobserved factors: If (27) is correct and there are exactly $\underline{v}$ factors, then adding observed numbers of workers N_{jk} to the regressions should have no explanatory power. It is by no means clear how such an hypothesis is nested however.

This model is correct and ingenious, given its assumptions. Still it is interesting to see what happens when the assumption of a single homogenous good is relaxed. For illustrative purposes, assume two goods with different technologies in two latent factors. Figure 8 depicts the situation in terms of a modified box diagram. The dimensions of the box are given by the total endowments ($\bar{Z}_1\bar{Z}_2$) in the subeconomy (observational unit) and the contract locus is shown as usual by the smooth curve. However, not all points on the contract curve are feasible because workers come in fixed bundles of latent factors. The diamond shaped area within the box represents the set of feasible allocations and its edges represent the boundary of the parallelogram sums of vectors of fixed endowments of each type of worker. The diamond has as many facets as the number of classes of workers (m). The area between its edges and the edges of the box cannot be obtained from feasible linear combinations of worker types. Suppose output demand conditions result in an equilibrium at a point on the interior (such as a). Then the mutual tangency of isoquants there establishes unique implicit factor prices W_i as before, and (26) and (27) remain valid. However, output demand conditions may result in an equilibrium along an edge (such as point b). Now there is no longer a single price of factors. Instead there are two sets of implicit factor prices, one for each industry. Seemingly, this difficulty can be handled empirically by extending the observations to industrial classifications within regions or states. This should work because the transformation from (24) to (27) washes out factor price differences among both regions and industries due to the fact that the same linear packaging restrictions among workers apply to all industries and regions. Letting l denote industry, the hypothesis $Z_{ijkl} = Z_{ij}$ must be maintained for this method to be valid.

In other words, it must be assumed that there exist no industry-specific skills, or no latent factors specific to industries. At this point it could be argued, and maybe correctly, that (27) still applies in the presence of industry-specific skills if the estimation is confined within industries (i.e., a model such as (27) for each industry, allowing the c's to vary among industries).

The basic difference between this model and one presented above lies in the packaging assumptions. (24) implies statements such as "eight elementary school graduates plus three college graduates are productively equivalent to 13 high school graduates." These kinds of restrictions make no sense in the models presented here, because workers are selected into productive activities according to their comparative advantage and there exists a great deal of economic rent in earnings, contrary to (25). In those models, productivity is comparable among workers who engage in the same activity, but not between activities because different work activities require and make use of different talents. Presumably Billy Rose's manual dexterity played a small role in his activities as an impressario. Indeed, section II demonstrates precisely how imperfect substitution derives from the existence of comparative advantage and specialization. These factors play no role whatsoever in (24) and (27).

A latent factor, characteristics model is still available for these models, however. It takes quite a different form from (26) and has some interesting features for analyzing the distribution of earnings. Consider the model of section III, and recall that the capacities of each worker were taken as exogenous (though perhaps affected by prior schooling decisions, family background, etc.). We can go one step back in this process. Consider the following linear specification:

$$t_i = b_{oi} + b_{1i}Z_1 + b_{2i}Z_2 + \ldots + b_{vi}Z_v \quad i = 1,\ldots,\underline{n} \qquad (28)$$

(28) may be considered to be a "production function", where a worker's skill capacity in activity $\underline{i}$ is determined by $\underline{v}$ latent factors or characteristics. The "marginal productivity" b_{ji} of the jth factor in producing the ith skill is allowed to vary among skills and each worker is described by a point in the space of Z. The general equilibrium in the economy, including product demand conditions, determines a set of implicit prices for each activity as above. Potential earnings in activity $\underline{i}$ are $p_i t_i$, and

$$y_i = p_i t_i = p_i b_{oi} + p_i b_{1i} Z_1 + \ldots. + p_i b_{vi} Z_v \qquad (29)$$

Now (29) bears a similarity to (25), but its interpretation is altogether different. It is useful to think of (29) as having a factor-analytic structure, but here the Z's are the latent factors, while the pb's are factor loadings that are similar to prices, just the reverse of (25). Finally, assume the latent factors are distributed according to some joint probability density $\theta(Z_1,\ldots,Z_v)$. In multivariate analysis it is customary to assume that θ is log normal, but that is too restrictive for our purposes.

From here there are two methods of proceeding:

(i) Use (29) to transform the distribution of latent factors θ to the distribution of potential income, $\phi(y_1,\ldots,y_n)$. It is to be emphasized that ϕ is a distribution of earning <u>potential</u>. To get the distribution of actual earnings, employ the choice rule of sections II and III that individuals choose activities that maximize their income. Then the distribution of observed earnings in activity $\underline{i}$ is conditioned on the fact that the earnings of people

people found in it is greater than they could have achieved in other activities: The distribution of observed earnings, $\hat{y}_i$, in activity i follows the conditional density

$$\xi^i(\hat{y}_i) = \Pr(y_i | y_i > y_1 \ldots, y_i > y_{i-1}, y_i > y_{i+1}, \ldots, y_i > y_n)$$

$$= \int_0^{y_i} \cdots \int_0^{y_i} \phi(g) \prod_{j \neq i} dy_j / \Pr(y_i > y_1, \ldots, y_i > y_n) \quad (30)$$

This kind of model was first elaborated by A. D. Roy. The main point is that observed earnings distributions are truncations from the distribution of earnings potential.[9] Comparison of observed means across activities lead to ubiquitous "selectivity bias", just as in the controversies about "ability bias"in return to schooling computations. The extent and nature of these biases evidently depends on patterns of correlation in the distribution of earnings potential (see Roy; also Lewis, Maddala and Heckman for more modern development). Roughly speaking, strong positive correlations among the y_i lead to greater selectivity bias. Finally, the overall distribution of observed earnings is a weighted sum of these truncated distributions (see Houthakker for a simple instructive example).

(ii) Alternatively, and equivalently, one may partition the factor space (Z) into acceptance regions, where max $\{y_1, y_2, \ldots, y_n\} = y_i$ in region i. This is the approach of Mandelbrot, and brings out the selection aspects of observed earnings more clearly.[10] These acceptance regions are cones[11] completely determined by the relations (29). For example, suppose (29) load very heavily on a single factor, with the effects of other factors being negligible. This factor could be thought of as some kind of general ability or human capital, which transforms differently into specific productive activity capacities. Then the sorting of workers to activities tends to be severely stratified, with the most capable workers found in the most remunerative jobs, the next most capable found in the next most remunerative jobs and so forth. The income distributions across activities follows a similar rank order, because the single factor induces large positive correlations in the distribution of potential earnings $\theta(y)$. The same hierarchical stratification would occur in the presence of many factors if there were sufficiently strong positive correlations among them. However, casual observations suggest a much more complex pattern of loadings and correlations, in which case observed assignments need not follow any clear hierarchical patterns. Nevertheless, the partition of $\underline{Z}$ into acceptance sets implies that people with characteristics within well defined limits are found in each activity. The selection phenomena remains, whatever the distribution of factors.

V. A NOTE ON THE THEORY OF MARRIAGE

We have been able to go some distance in analyzing part of Smith's famous proposition on the division of labor, based entirely on specialization and comparative advantage. All results above are independent of size; and the price system is efficient and achieves the optimal degree of specialization, given the resource constraint, for any scale of operations. Thus comparative advantage itself provides little insight into the more profound and difficult part of Smith's proposition linking specialization with the extent of the market. The reason for this lies in the linearity assumptions and absence of indivisibilities imposed on the present model, which imply that total service flows within each of the various activities are all that matter, independently of the distribution of embodied talents among members of the labor force. Under these circumstances the implicit prices of services efficiently clear all markets and depend only on total skill potential in the economy, not on how skills come prepackaged in indivisible bodies.[12] These assumptions appear to be reasonable approximations for those activities, such as pin making and other production work that have fairly clear-cut and identifiable outputs. However, they would appear to fail for many other activities where output is less clearly defined: Two mediocre economists or doctors or lawyers do not necessarily add up to some multiple of a "good" one. In terms of the marriage market analogy, the price system described above does indeed optimally select workers to jobs, but the distribution of workers across firms and the question of "who works with whom" is simply irrelevant. In this sense super-polygamous marriages in the labor market are pareto optimal, given the assumptions!

It is clear that any theory of selection of worker quality matched to firm quality must confront the indivisibility issue at ground zero. The result must be to make firms play a much more active role in the sorting process than has been described above. In addition to Smith's elegant discussion, more recent work on the theory of the firm (Alchian and Demsetz) and the theory of signaling (Stiglitz) suggests a possible route by way of technical externalities and true joint production in the underlying engineering technology, thus giving an incentive for entrepreneurial activity in assembling an optimal production team and an optimal distribution of talents within the firm. However, the discussion above points to another fundamental source of indivisibility and that is in the operation of markets themselves (Stigler). At a low level of output, the fixed cost of marketing specialized talents in point sharpening, wire stretching and so forth cannot be covered because these markets would be too thinly traded. If so, then there are clear gains from entrepreneurial activity assembling an optimal work force outside the market mechanism.

The following example illustrates some of the resulting complexities. Consider a fixed coefficients production process in two work activities, T_1 and T_2, and let there be four workers, A, B, C, and D, each described by capacity vectors as before. We know that a marriage of the four is efficient and instead look for some second-best solutions due to market failure. The restriction that workers must be assigned to two-person production teams is arbitrarily imposed, thus allowing use of the assignment algorithm of Koopmans and Beckmann or the marriage market solution proposed by Becker. One possibility is shown in Figure 9, where there is a clear rank order of comparative advantage and no absolute advantage. The second-best activity frontiers are built up by forming all possible pairwise frontiers, as in Figure 1, and then adding them up along all possible rays through the origin.[13]

As shown, a negative sorting by comparative advantage is second best for all technologies. The efficiency loss from incomplete markets is proportional to the radial distance between the efficient frontier and the negatively sorted frontier, and is close to zero for some technologies. That negative sorting is not an automatic result of comparative advantage is shown in Figure 10. Here the negative assortment frontier crosses the positive assortment frontier and the second-best solution is the envelope of the two. With this distribution of talent, negative sorting is best for some technologies and positive sorting is best for others. Again, the efficiency loss is negligible for some technologies. Note that there is a curious asymmetry about these second-best solutions: They require four prices, one for each worker, which combine to more than two implicit prices for activities. The costs of assembly are greater because more information is required on this account. Yet the differential public goods aspects of each kind of marketing system may sustain the second-best solution even if its total resource costs exceed operating markets in skills rather than in people.

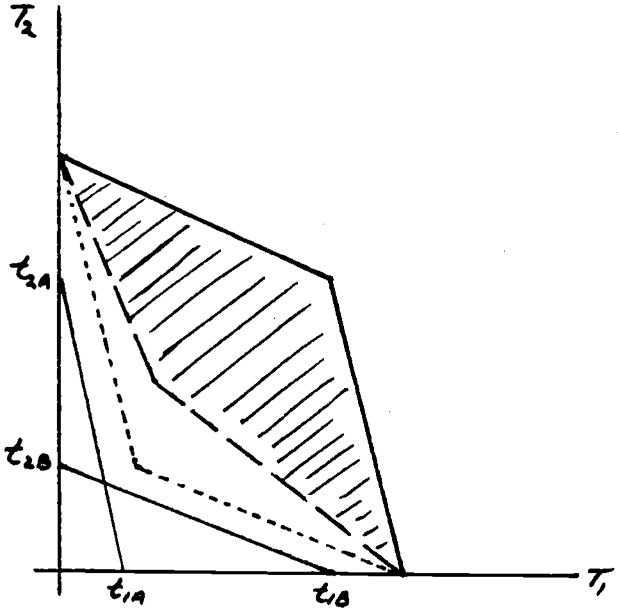

Fig. 1

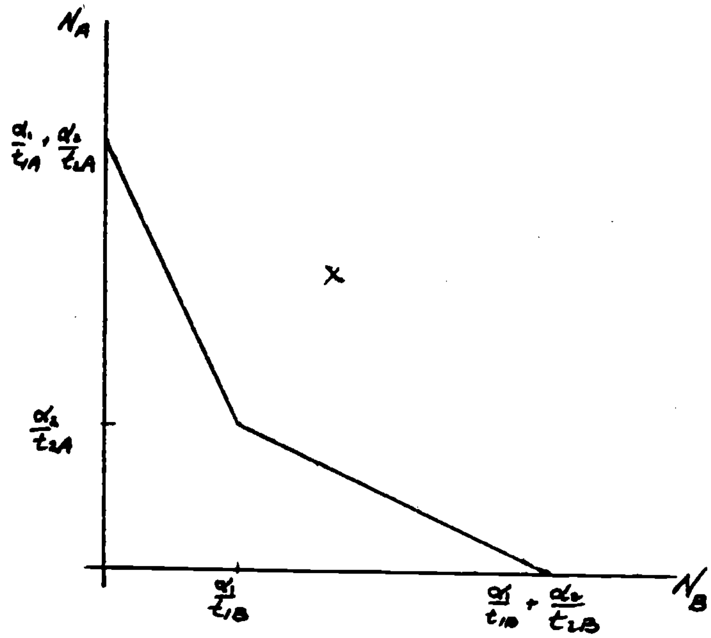

Fig. 2

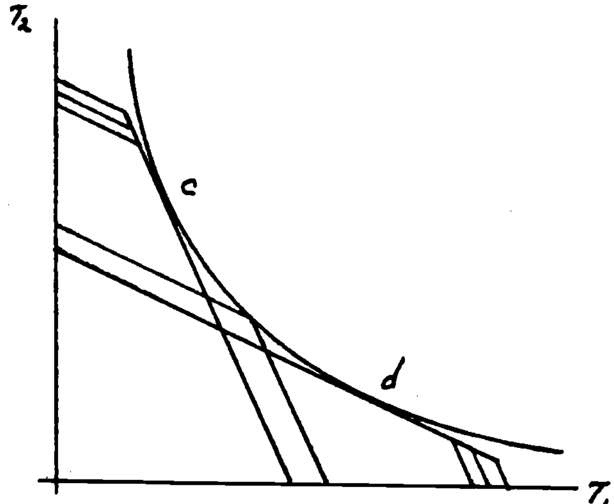

Fig. 3

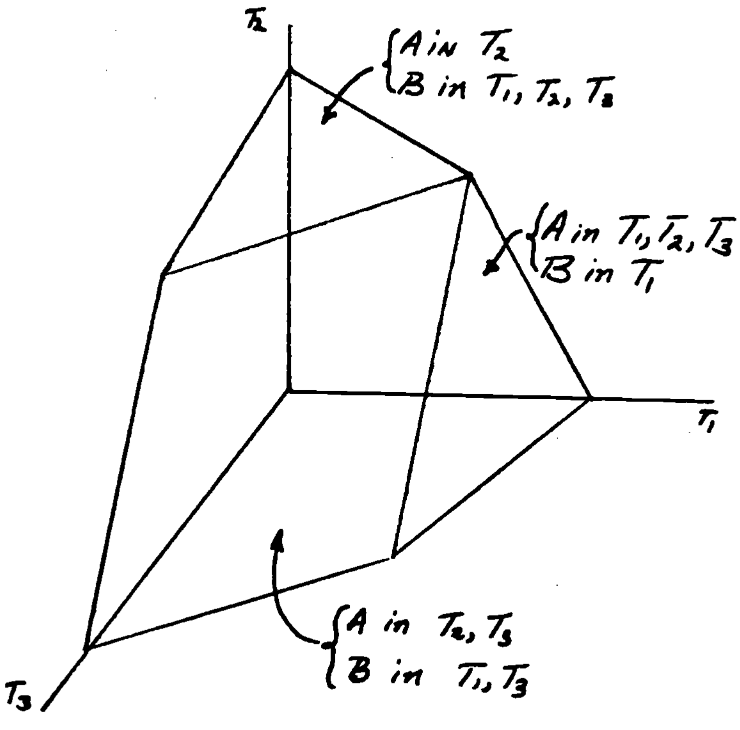

Fig. 4

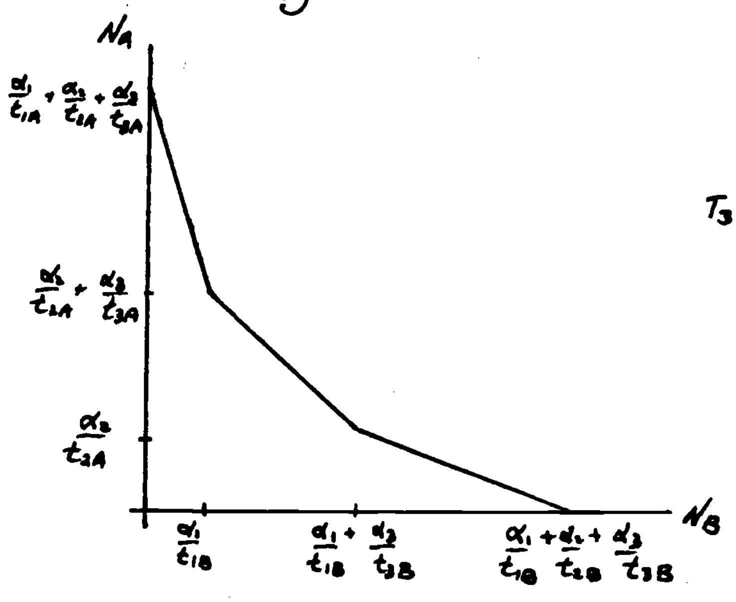

Fig. 5

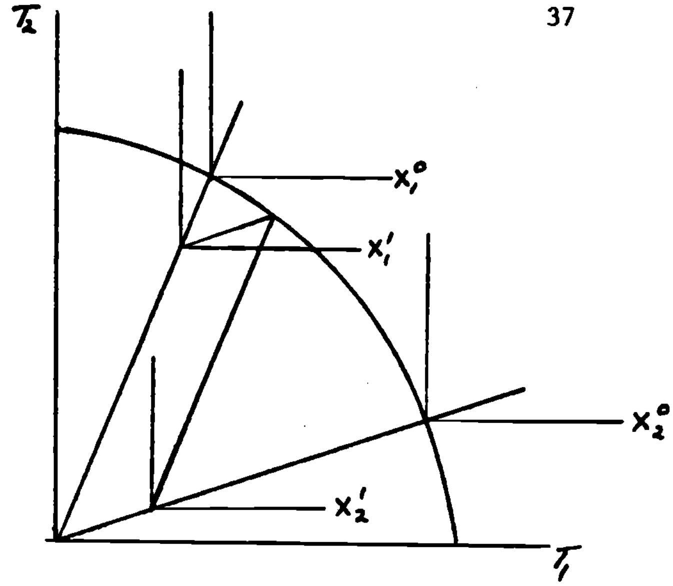

FIG. 6

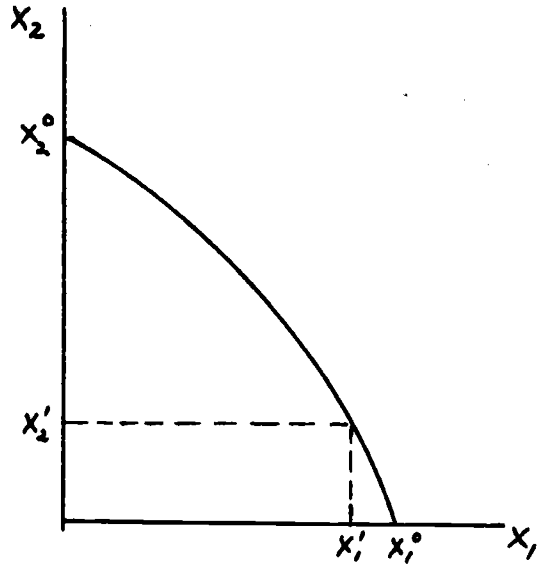

Fig. 7

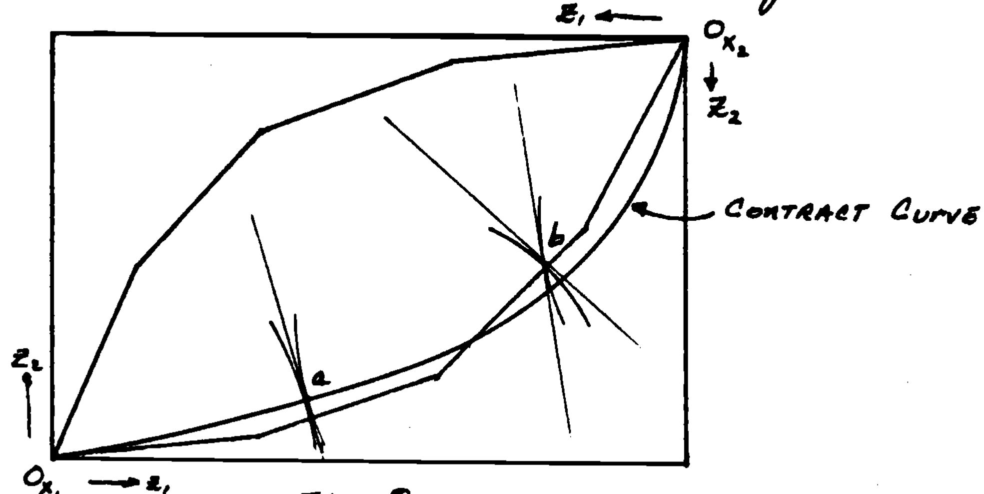

Fig 8

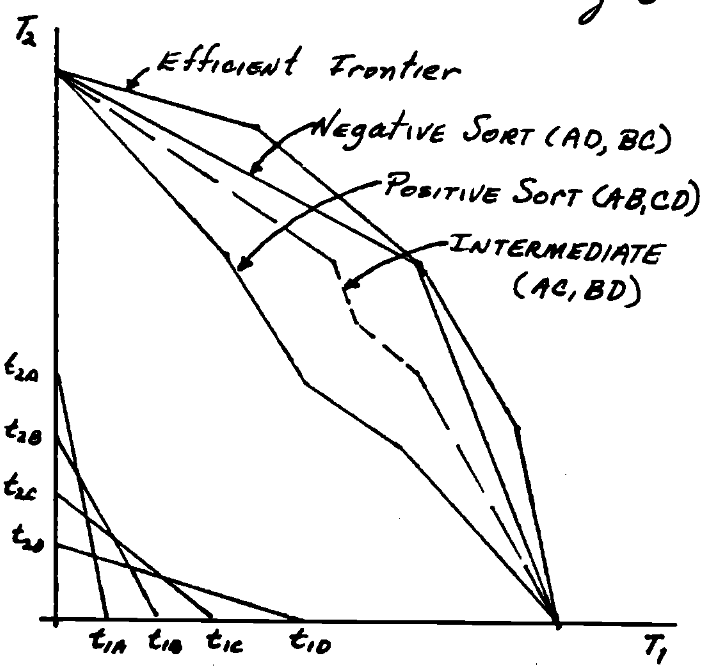

Fig. 9

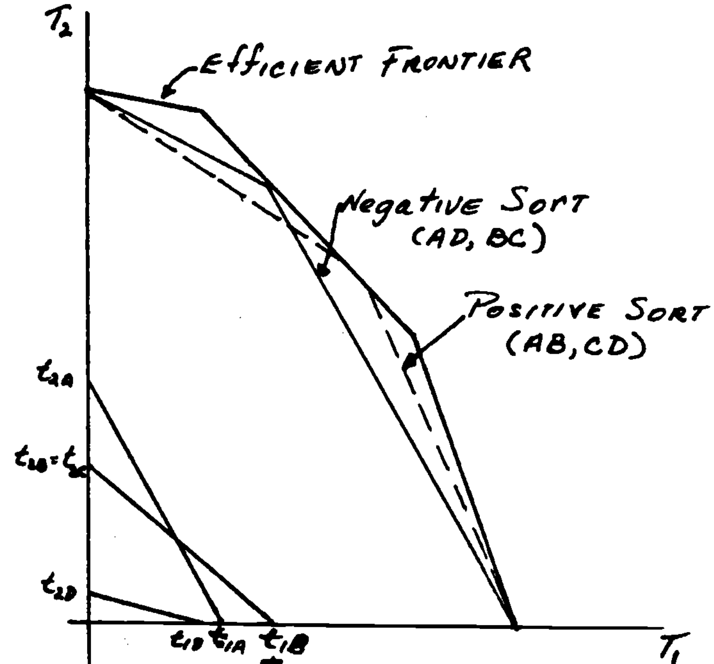

Fig. 10

FOOTNOTES

1. As simple as it is, this variant of the familiar gains from trade argument seems to be unrecognized in this connection. It shows how specialization without indivisibility leads to something that looks like an economy of scale, though it is not a scale economy in the usual sense, since all the production functions derived from it below are linear. Ames has attempted to build up a Smithian technology by a series of nested neoclassical production functions. However, if all the nested functions exhibit constant returns, so must the aggregate function and the synergistic effects exhibited here do not arise.

2. The following is not unlike the usual construction of neoclassical isoquants in a linear programming framework (e.g., see Baumol), because they are both applications of the same envelope theorem. The novelty here lies in the nature of the constraints and in the substantive conception of the problem.

3. In the nature of the case, constraints (4) are binding at an optimum. Also, only left or right-hand derivatives may be defined for some $\underline{N}$ (see below).

4. (6) and (7) of course have a solution for any $\underline{n}x\underline{m}$ dimension, and the envelope theorem always guarantees existence of an indirect production function. However, analyzing its properties requires finding the efficient activity possibility frontier and I have only been able to do so in the $\underline{n}$x2 base. The reason is that the assignment restrictions such as (13) and (14) are

considerably more complex when m exceeds 2. For example, see Jones for the nxn case. It is clear that the parametric method used here is not very useful for $m > 2$.

5. If the comparative advantage assumptions are different from those specified, there will be a different pattern of blank entries in (15). Nevertheless, the same argument shows that there are three facets on the efficient polyheron, though their orientation will be different from those in Figure 4. Also, if there is no comparative advantage in some activities, the number of facets may be less than 3. This same point applies to the nx2 case.

6. Recall that a point on the factor price frontier maps to a flat on the isoquant and that a flat on the factor price frontier goes to a corner of an isoquant. To find the absolute values of the r's on the factor price frontier, use the restriction $\Sigma p_j \alpha_j = 1$ to solve for the p's along each facet of the activity frontier, and use the choice-indifference restrictions to solve for r_A and r_B. With n activities and complete comparative advantage, the factor price frontier obviously lies along n rays.

7. The integration must be split into two regimes because of the lower bounds on t_1 and t_2 in the Pareto distribution. The choice of an example based on Pareto's law is not fortuitous, but follows Houthakker's well known example, to which this model bears a distinct family resemblance. Closed form solutions for $F(T_1,T_2)$ are not available for other simple distributions such as the exponential or the lognormal.

8. I have arbitrarily expressed these restrictions in terms of the first v types, though any v distinct groups will do.

9. I had originally thought that these kinds of selection phenomena could themselves account for observed skewness of earnings distributions with symmetrical distributuions of underlying factors. One can construct examples where this is so, but it need not be true in general. The reason is that the correlations among the Z's may imply truncations in some activities from both above and below; and even if truncation occurs in the lower tail for all activities, the configuration of implicit prices p_i may affect the within-activity means in ways that do not produce the simple kind of skewness observed in overall earnings distributions.

10. Mandelbrot begins with a formulation such as (29) and specifies θ to be (weak) Paretian, though no underlying economic structure of the model is elaborated. Such a generating structure has been established above. What seems odd about (29) at first glance and without a theoretical structure is that prices of factors vary from activity to activity, a result that has been amply justified above.

11. This is easily seen when the equations in (29) are homogeneous ($b_{0i} = 0$) and there are two factors. Then the acceptance sets are open cones defined by rays through the origin. In the nonhomogeneous case, the cones may be closed and the partitions are less easy to compute.

12. This discussion abstracts from assortments of workers across firms according to nonpecuniary aspects of work and other consumption values offered by firms. Such a theory is available (Rosen) and could be grafted onto this one without too much difficulty. The presence of income effects would seem to imply observed talent stratification of workers across firms, but that is beside the point here.

13. A point on each pair's frontier defines an element of K-B's payoff matrix and their solution maximizes the sum of payoffs. There is sufficient structure on this problem to derive a graphical solution. In the presence of a continuous distribution of talents, some additional insight might be obtained by setting up an optimum control problem, but I have not been able to do so.

REFERENCES

Alchian, A. A., and H. Demsetz, "Production, Information Costs, and Economic Organization," American Economic Review, 1972.

Ames, E., "That's No Paradox, It's a Pin Factory," Purdue University, 1965.

Baumol, W., Economic Theory and Operations Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1965.

Becker, G., "A Theory of Marriage," in T. W. Schultz (ed.), The Economics of the Family, Chicago, University of Chicago Press, 1974.

Dorfman, R., P. Samuelson, and R. Solow, Linear Programming and Economic Analysis, McGraw-Hill, New York, 1958.

Easterlin, R., "Toward a Socio-Economic Theory of Fertility: A Survey of Recent Research on Economic Factors in American Fertility," in S. J. Behrman, L. Corsa and R. Freedman (eds.), Fertility and Family Planning, A World View, University of Michigan Press, Ann Arbor, 1969.

Heckman, J., "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables," University of Chicago, 1976.

Houthakker, H., "The Size Distribution of Labor Incomes Derived from the Distribution of Attitudes," in W. Sellekaerts (ed.), Econometrics and Economic Theory: Essays in Honor of Jan Tinbergen, Macmillan, New York, 1975.

____________, "Pareto Distributions and the Cobb-Douglas Production Function in Demand Analysis," Review of Economic Studies, 1955-56.

Jones, R., "Comparative Advantage and the Theory of Tariffs: A Multi-Country, Multi-Commodity Model," Review of Economic Studies, 1961.

Jorgenson, D. W., and Z. Griliches, "The Explanation of Productivity Change," Review of Economic Studies, 1967.

Koopmans, T. C., and M. Beckmann, "Assignment Problems and the Location of Economic Activities," Econometrica, 1957.

Lewis, H. G., "Comments on Selectivity Biases in Wage Comparisons," Journal of Political Economy, 1974.

Maddala, G. S., "Self-Selectivity Problems in Econometric Models," University of Florida, 1976.

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

Reder, M., "Wage Differentials: Theory and Measurement," in Aspects of Labor Economics, NBER, Princeton University Press, 1962.

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, 1972.

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

Sato, K., Production Functions and Aggregation, North-Holland, Amsterdam, 1975.

Sattinger, M., "Comparative Advantage and the Distribution of Earnings and Abilities," Econometrica, 1975.

Stigler, G., "Production and Distribution in the Short Run," Journal of Political Economy, 1939.

____________, "The Division of Labor is Limited by the Extent of the Market," Journal of Political Economy, 1951.

Stiglitz, J., "The Theory of Screening, Education and the Distribution of Income," American Economic Review, 1975.

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

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

Whitin, T., "Classical Theory, Graham's Theory, and Linear Programming in International Trade," Quarterly Journal of Economics, 1953.