

NBER WORKING PAPER SERIES

THE DEMAND FOR LABOR IN
THE LONG RUN

Daniel S. Hamermesh

Working Paper No. 1297

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

March 1984

Professor of Economics, Michigan State University, and research associate, National Bureau of Economic Research. To be published in O. Ashenfelter and R. Layard, eds., Handbook of Labor Economics, North-Holland Press. Orley Ashenfelter, George Borjas, George Johnson, Richard Layard, Andrew Oswald and John Pencavel provided helpful comments. 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.

The Demand for Labor in the Long Run

ABSTRACT

The theory of the demand for labor is presented along with a catalog and critique of methods that are used to estimate the parameters that describe empirical labor-demand and substitution possibilities. A critical survey is presented of studies of own-price demand elasticities for labor as a whole and for workers categorized by demographic group, of substitution parameters among workers of different types, and of workers for capital. The main findings are: 1) The long-run constant-output demand elasticity for labor that is treated as homogeneous is between .15 and .5; 2) Own-price demand elasticities are higher for workers that have less general human capital embodied and them; 3) Skilled labor and physical capital are p-complements; and 4) More tentatively, youths and women are q-substitutes in production. The implications and importance for policy of these and other results are discussed. Suggestions for improving the literature and narrowing the range of knowledge of the underlying parameters, especially by concentrating more on disaggregated and even microeconomic data, are presented.

Daniel S. Hamermesh
Department of Economics
Michigan State University
East Lansing, Michigan 48824

(517) 355-7349

I. Introduction

The demand for labor in the long run should be important to labor economists for a variety of reasons. So long as the supply of labor to an occupation, industry or area is not perfectly elastic in the long run, the nature of demand for labor in that subsector interacts with the shape of the supply function to determine the level of wages. As in the market for a commodity, so too in the market for labor the demand is an integral determinant of the price of what is exchanged.

In many cases economists are interested in the demand for labor for its own sake rather than for its effects on wage determination. In some instances, e.g., in unionized employment or where the supply of labor to a subsector is perfectly elastic, the wage can be viewed as unaffected by labor demand. In such cases knowledge of wage elasticities of labor demand allow one to infer the effects of exogenous changes in wage rates on the amount of labor employers seek to use. The impact of changes in the price of one type of labor on its employment and on the employment of other types of labor (cross-price effects) can be discovered using estimates of labor-demand relations alone. Alternatively, one can in many instances assume that the employment of workers of a particular type is fixed (and determined solely by the completely inelastic supply of such workers to the market). In those cases the demand for their labor determines the wage rate they are paid. Knowledge of the shape of the labor-demand function enables one to infer how exogenous changes in supply (due perhaps to changes in the demographic mix of the labor force or to shifts in suppliers' preferences for entering different occupations) affect the wage rate of workers in the group whose supply has shifted and in other groups too (cross-quantity effects).

Economists interested in policy questions should be concerned with issues of labor demand. The effects of any policy that changes factor prices faced by employers will depend on the structure of labor demand. Thus to predict the impact of wage subsidies, payroll tax changes, investment tax credits, etc., one must have satisfactory estimates of underlying parameters. Similarly, the impact on wages of policies such as skills training or population control that change the demographic or human-capital mix of the labor force can be assessed only if one knows the underlying structure of substitution relations among groups of workers.

Bearing in mind throughout that the purpose of studying the demand for labor is to understand how exogenous changes will affect the employment and/or wage rates of a group or groups of workers, we begin this essay by examining the theory of labor demand. The theoretical discussion is divided into two parts: Demand for labor in the two-factor case, and demand in the multi-factor case. In each part we first derive the results generally, then proceed to specific functional forms. In Sections IV and V we point out the issues involved in estimating labor-demand relations for one type of homogenous labor, and then summarize the state of knowledge in this area. Sections VI and VII perform the same tasks for the demand for labor of several types.

The focus throughout is on the relations between exogenous wage changes and the determination of employment, and between exogenous changes in inelastically supplied labor and the structure of relative wages. We ignore the possibility that firms may not maximize profits or minimize costs, and assume throughout that employers are perfect competitors in both product and labor markets. While this latter assumption may be incorrect, the analysis applies mutatis mutandis to employers who have some product-market power. Most important, we focus only on the long-run, or static theory of labor

demand, and thus only on the long-run effects of exogenous changes in wage rates or labor supply. The dynamics of labor demand, particularly the role of adjustment costs and the distinction between the amount of labor used and its intensity of use (employment versus hours per period), are ignored (and left to Nickell's essay in this volume). Most lags in the adjustment of labor demand to its long-run equilibrium do not appear to be very long (Hamermesh, 1980); the slow adjustment of relative wages to exogenous shocks appears due mostly to lags in suppliers' decisions about training and mobility. That being the case, the theory of labor demand in the long run, and the estimates of parameters describing that demand, are useful in answering questions of interest to policy-makers and others who are interested in the near-term effects of various changes in the labor market.

II. Two Factors--The Theory

While the theoretical results on labor demand can be generalized to N factor inputs, many useful insights into the theory can be gained by examining the demand for homogeneous labor when there is only one cooperating factor, usually assumed to be capital services. Since much of the terminology of labor demand applies in the two-factor case, concentrating on it also has some pedagogical advantages. Also, many of the specific forms for the production and cost functions from which labor-demand functions are derived were initially developed for the two-factor case and make a good deal more economic sense applied to only two factors than generalized to several. The presentation here and in Section III goes through some derivations, but our aim is to provide a theoretical outline to link to empirical work. More complexity can be found in Varian (1978); still more is available in the essays in Fuss-McFadden (1978).

Assume that production exhibits constant returns to scale, as described by F , such that:

$$(1) \quad Y = F(L, K), \quad F_i^D > 0, \quad F_{ii} < 0, \quad F_{ij} > 0,$$

where Y is output, and K and L are homogenous capital and labor inputs respectively. A firm that maximizes profits subject to a limit on costs will set the marginal value product of each factor equal to its price:

$$(2a) \quad F_L - \lambda w = 0;$$

$$(2b) \quad F_K - \lambda r = 0,$$

where w and r are the exogenous prices of labor and capital services respectively, λ is a Lagrangean multiplier showing the extra profit generated by relaxing the cost constraint, and we assume the price of output is unity. The firm will also operate under the cost constraint:

$$(2c) \quad C^0 - wL - rK = 0.$$

The ratio of (2a) to (2b) is the familiar statement that the marginal rate of technical substitution equals the factor-price ratio for a profit-maximizing firm.

Allen (1938, p.341) defines the elasticity of substitution between the services of capital and labor as the effect of a change in relative factor prices on relative inputs of the two factors, holding output constant. (Alternatively, it is the effect of a change in the marginal rate of technical substitution on the ratio of factor inputs, defined as an elasticity.) In this two-factor linear homogeneous case it is (see Allen (1938) pp. 342-343):

$$(3) \quad \sigma = \frac{d \ln K/L}{d \ln (w/r)} = \frac{d \ln \left(\frac{K}{L} \right)}{d \ln \left(\frac{F_L}{F_K} \right)} = \frac{F_L F_K}{Y F_{LK}}$$

The own-wage elasticity of labor demand at a constant output and constant r is [Allen (1938) pp. 372-373]:

$$(4a) \quad \eta_{LK} = [1-s]\sigma > 0$$

where $s = wL/Y$, the share of labor in total revenue. Intuitively, the constant-output elasticity of labor demand is smaller for a given technology (σ)

when labor's share is greater because there is relatively less capital toward which to substitute when the wage rises. The cross-elasticity of demand (for capital services) is:

$$(4b) \quad \eta_{LK} = [1-s]\sigma > 0$$

[What is the intuition on the inclusion of $1 - s$ in (4b)?].

Both (4a) and (4b) reflect only substitution along an isoquant. When the wage rate increases, the cost of producing a given output rises; and the price of the product will rise, reducing the quantity of output sold. The scale effect depends on the (absolute value) of the elasticity of product demand, η , and on the share of labor in total costs (which determines the percentage increase in price). Thus to (4a) and (4b) the scale effects must be added, so that:

$$(4a') \quad \eta'_{LL} = - [1-s]\sigma - s\eta ,$$

and

$$(4b') \quad \eta'_{LK} = [1-s] [\sigma - \eta]$$

The results here and in (4a) and (4b) are the most important in the theory of labor demand. They will be proved below using the cost-function approach.

Both (4a) and (4a') are useful, depending on the assumptions one wishes to make about the problem under study. Certainly in an individual firm or particular industry, which can expand or contract as the wage it must pay changes, scale effects on employment demand are relevant. For an entire economy, in which output may be assumed constant at full employment, (4a) and

(4b) are the correct measures of the long-run effect of changes in the wage rate on factor demand.

All of these measures assume that both factors are supplied elastically to the firm. If they are not, the increase in employment implicit in (4a') when the wage decreases cannot be complete: The labor that is demanded may not be available; and the additional capital services whose presence raises the marginal product of labor ($F_{LK} > 0$) also may not be. In such cases the demand elasticities are reduced (see Hicks (1964, Appendix)). Though such cases may be important, we ignore them in this essay (though we do deal with the polar case in which the wage depends upon the level of exogenous employment).

An alternative approach makes use of cost minimization subject to an output constraint. Total cost is the sum of products of the profit-maximizing input demands and the factor prices. It can be written as:

$$(5) \quad C = C(w,r,Y), \quad C_i > 0, \quad C_{ij} > 0, \quad i,j = w,r,$$

since the profit-maximizing input demands were themselves functions of input prices, the level of output, and technology. By Shephard's lemma (see Varian (1978, p. 32)) the firm's demand for labor and capital at a fixed output Y can be recovered from the cost function (5) as:

$$(6a) \quad L^* = C_w$$

and

$$(6b) \quad K^* = C_r .$$

Intuitively, the cost-minimizing firm uses inputs in a ratio equal to their marginal effects on costs. The forms (6) are particularly useful for estimation purposes, as they specify the inputs directly as functions of the factor prices and output.

Using equations (6) and the result that $C(w,r,Y) = YC(w,r,1)$ if Y is linear homogeneous, the elasticity of substitution can be derived (see Sato-Koizumi, 1973) as:

$$(7) \quad \sigma = \frac{C}{C} \frac{C_{wr}}{C_w C_r} .$$

Note that the elasticity of substitution derived from a cost function looks strikingly similar to that derived from a production function. Obviously they are equal, suggesting that the form one chooses to measure σ should be dictated by convenience.

The factor-demand elasticities can be computed as:

$$(8a) \quad \eta_{LL} = -[1-m]\sigma$$

and

$$(8b) \quad \eta_{LK} = [1-m]\sigma,$$

where m is the share of labor in total costs. Since by assumption factors are paid their marginal products, and the production and cost functions are linear homogeneous, $m=s$, and (8a) and (8b) are equivalent to (4a) and (4b).

We are now in a position to prove (4a') easily following Dixit (1976, p. 79). If we continue to assume constant returns to scale, we can reasonably treat the firm as an industry and write industry factor demand as:

$$(6a') \quad L = YC_w$$

and

$$(6b') \quad K = YC_r .$$

Under competition firms equate price, p , to marginal and average cost:

$$p = C.$$

Noting that if markets clear, so that output equals industry demand $D(p)$, we obtain:

$$\partial L / \partial w = Y C_{ww} + D'(p) C_w^2 .$$

Because C is linear homogeneous, $C_{ww} = \frac{-r}{w} C_{wr}$. Substituting for C_{ww} , then from (7) for C_{wr} , and then for C_w and C_r from (6a') and (6b'):

$$\partial L / \partial w = \frac{rK}{Y} \frac{\sigma L}{wC} + \frac{D'(p)L^2}{Y^2} .$$

To put this into the form of an elasticity, multiply both sides by pw/pL , and remember that $p = C$:

$$\eta_{LL} = \frac{-rK}{pY} \sigma + \frac{pD'(p)}{Y} \frac{wL}{pY} = -[1-s]\sigma - s\eta,$$

by the definition of factor shares under linear homogeneity.

The production or cost functions can also be used to define some concepts that are extremely useful when examining markets in which real factor prices are flexible and endogenous, but factor supplies are fixed (and, because of the flexibility of input prices, are fully employed). The converse of asking, as we have, what happens to the single firm's choice of inputs in response to an exogenous shift in a factor price is to ask what happens to factor prices in response to an exogenous change in factor supply. Define the elasticity of complementarity as the percentage responsiveness of relative factor prices to a one-percent change in factor inputs:

$$(9) \quad c = \frac{\partial \ln w/r}{\partial \ln K/L} .$$

This is just the inverse of the definition of σ .

Thus:

$$(10) \quad c = \frac{1}{\sigma} = \frac{C_w C_r}{C C_{wr}} = \frac{Y F_{LK}}{F_L F_K} .$$

In the two-factor case in which the production technology is linear homogeneous, one can find the elasticities of substitution and of complementarity equally simply from production and cost functions; and, having found one of them, the other is immediately available.

Given constant marginal costs, the elasticities of factor price (of the wage rate and the price of capital services) are defined as:

$$(11a) \quad \epsilon_{ww} = - [1-m] c$$

and

$$(11b) \quad \epsilon_{rw} = [1-m]c .$$

Equation (11a) states that the percentage decrease in the wage rate necessary to accommodate an increase in labor supply with no change in the marginal cost of the product is smaller when the share of labor in total costs is larger (because labor's contribution to costs -- a decrease -- must be fully offset by a rise in capital's contribution in order to meet the condition that marginal cost be held constant).

Consider now some examples of specific production and cost functions.

A. Cobb-Douglas technology

The production function is:

$$(12) \quad Y = L^\alpha K^{1-\alpha} ,$$

where α is a parameter; marginal products are:

$$(13a) \quad \frac{\partial Y}{\partial L} = \alpha \frac{Y}{L} ,$$

and

$$(13b) \quad \frac{\partial Y}{\partial K} = [1-\alpha] \frac{Y}{K} .$$

Since the ratio of (13a) to (13b) is $\frac{w}{r}$ if the firm is maximizing profits, taking logarithms and differentiating with respect to $\ln \frac{w}{r}$ yields $\sigma = 1$.

Equations (4a) and (4b) imply:

$$\eta_{LL} = - [1-\alpha] ,$$

$$\eta_{LK} = 1 - \alpha .$$

Minimizing total costs subject to (12), one can derive (Varian (1978, p. 15)) the demand functions for L and K, and thus the cost function. The latter reduces to:

$$(14) \quad C(w,r,Y) = Z w^\alpha r^{1-\alpha} Y ,$$

where Z is a constant. Using Shephard's lemma, one can again derive:

$$(15) \quad L/K = \frac{\alpha}{1-\alpha} \frac{r}{w} ,$$

Taking logs, the calculation that $\sigma = 1$ follows immediately. It is also clear from (15) that $c = 1$.

B. Constant elasticity of substitution technology

The linear homogeneous production function is:

$$(16) \quad Y = [\alpha L^\rho + (1-\alpha)K^\rho]^{\frac{1}{\rho}} ,$$

where α and ρ are parameters. Marginal products are $\frac{1}{\rho}$:

$$(17a) \quad \frac{\partial Y}{\partial L} = \alpha \left(\frac{Y}{L}\right)^{1-\rho} ,$$

and

$$(17b) \quad \frac{\partial Y}{\partial K} = (1-\alpha) \left(\frac{Y}{K}\right)^{1-\rho} .$$

Setting the ratio of (17a) to (17b) equal to the factor-price ratio, taking logarithms, differentiating with respect to $\ln \frac{w}{r}$, and making $\sigma > 0$, yields:

$$(18) \quad \frac{-\partial \ln \frac{L}{K}}{\partial \ln \frac{w}{r}} = \sigma = 1/[1-\rho] .$$

The CES is sufficiently general that any value of $\rho < 1$ is admissible, and the relationship (18) can be used to estimate σ .

Among special cases are: 1) The Cobb-Douglas function ($\rho = 0$, as should be clear from (18)); 2) The linear function ($\rho = 1$), in which L and K are perfect substitutes. (Go back to (3), and note that if $\rho = 1$, so that (16) is linear and $F_{LK} = 0$, $\sigma = \infty$.) 3) The Leontief function ($\rho = -\infty$), in which output is the minimum function $Y = \min \{L, K\}$, and the inputs are not substitutable at all.^{2/} The constant-output factor-demand elasticities in each case follow immediately from the definitions and the recognition that α is labor's share of revenue if the factors are paid their marginal products.

The CES cost function can be derived (Ferguson (1969, p. 167)) as:

$$C = Y [\alpha^\sigma w^{1-\sigma} + [1-\alpha]^\sigma r^{1-\sigma}]^{\frac{1}{1-\sigma}},$$

where, as before, $\sigma = 1/[1-\rho] > 0$. The demand for labor is:

$$(19) \quad L = \frac{\partial C}{\partial w} = \alpha^\sigma w^{-\alpha} Y.$$

Taking the ratio of (19) to the demand for K, the elasticity of substitution can again be shown to be σ .

In both of these examples it is very straightforward to derive c first, then derive σ as its inverse. It is worth noting for later examples and for the multifactor case that c is more easily derived from equations (17) and the factor-price ratio (since w/r , the outcome, appears alone), than from (19) and the demand for capital. σ is more readily derived from the cost function, since the ratio L/K appears alone. Obviously in the two-factor case the simple relation (10) allows one to obtain c or σ from the other; but the ease of obtaining c or σ initially differs depending on which function one starts with, a difference that is magnified in the multifactor case.

Two other specific functional forms, the generalized Leontief form of Diewert (1971) and the translog form (Christensen et al (1973)), are second-order approximations to arbitrary cost or production functions. Each has the advantage over the CES function in the two-factor case that σ (or c) is not restricted to be constant, but instead depends on the values of the factor inputs or prices. In each case we examine here only the cost function.

C. Generalized Leontief

$$(20) \quad C = Y \{ a_{11} w + 2 a_{12} w^{.5} r^{.5} + a_{22} r \} ,$$

where the a_{ij} are parameters. Applying Shephard's lemma to (20) for each input, and taking the ratios:

$$(21) \quad \frac{L}{K} = \frac{a_{11} + a_{12} \left(\frac{w}{r}\right)^{-1/2}}{a_{22} + a_{12} \left(\frac{w}{r}\right)^{1/2}} .$$

As is easily seen from (21), in general $\sigma \left(- \partial \ln \left(\frac{L}{K} \right) / \partial \ln \left(\frac{w}{r} \right) \right)$ depends on all three parameters and the ratio $\frac{w}{r}$. Under restrictive assumptions (20) reduces to some of the examples we have already discussed. If $a_{12} = 0$, it becomes a Leontief function (since the ratio L/K is fixed). If $a_{11} = a_{22}$, it becomes a Cobb-Douglas type function.

D. Translog

$$(22) \quad \ln C = \ln Y + a_0 + a_1 \ln w + .5b_1 [\ln w]^2 + b_2 \ln w \ln r + .5b_3 [\ln r]^2 + [1-a_1] \ln r ,$$

where the a_i and b_i are parameters. Applying Shephard's lemma to each input, and taking the ratios:

$$(23) \quad \frac{L}{K} = \frac{r}{w} \frac{a_1 + b_1 \ln w + b_2 \ln r}{[1-a_1] + b_2 \ln w + b_3 \ln r} .$$

Again σ depends on all parameters and both factor prices. Under specific circumstances ($b_i = 0$ for all i), the cost function reduces to a Cobb-Douglas technology.

Both the generalized Leontief and translog functions may be useful for empirical work (see below), even when written out as in (20) and (22). Each has the virtue of allowing flexibility and containing some simpler forms as special cases. That suggests that they should supplant the Cobb-Douglas and CES functions even for empirical work involving just two inputs.

Throughout this Section we have assumed the production and cost functions are linear homogeneous. This also implies they are homothetic: Factor demand is such that the ratio of factor inputs is independent of scale at each factor-price ratio. This assumption may not always make sense. For example, large firms may be better able to function with a more capital-intensive process at given w and r than are small firms.

In the general case nonhomotheticity means that the production function cannot be written as:

$$Y = G(F[L,K]),$$

where G is monotonic and F is linear homogeneous. Alternatively, the cost function cannot be expressed as (Varian (1978, p. 49)): $C(w,r,Y) = C^1(Y) C^2(w,r)$; i.e., output is not separable from factor prices. Some special cases are useful for estimation; and a nonhomothetic CES-type function (Sato (1977)) and translog form (Berndt-Khaled (1979)) have been used.

III. Several Factors — The Theory

Mathematically the theory of demand for several factors of production is just a generalization of the theory of demand for two factors presented in the last section. Empirically, though, the generalization requires the researcher to examine a related aspect of factor demand that is not present when the set of inputs is classified into only two distinct aggregates. The issue is illustrated when one considers a three-factor world, for example, three types of labor, L_1 , L_2 and L_3 . One could assume that production is

characterized by:

$$(24) \quad Y = F(G(L_1, L_2), L_3) ,$$

where F and G are two-factor production functions of the kind we discussed above. The difficulty with (24) is that the aggregation of L_1 and L_2 by the function G is a completely arbitrary description of technology. Far better to devise some method that allows this particular aggregation to be a subcase whose validity can be tested. This problem, one of separability of some factors from other(s), provides the major reason why labor economists must be interested in multi-factor labor demand. As an example, it means that one should not, as has been done by, for example, Dougherty (1972), combine pairs of labor subaggregates by hierarchies of two-factor CES functions.

Intuitively this is because changes in the amount of one type of labor in a particular subaggregate could affect the ease of substitution between two groups of labor that are arbitrarily included in another subaggregate. If so, one will draw incorrect inferences about the ease of substitution between the latter two factors (and about the cross-price demand elasticities).

Consider a firm (industry, labor market, economy) using N factors of production, X_1, \dots, X_N . Let the production function be:

$$(25) \quad Y = f(X_1, \dots, X_N), \quad f_i > 0, \quad f_{ii} < 0 .$$

Then the associated cost function, based on the demands for X_1, \dots, X_N , is:

$$(26) \quad C = g(w_1, \dots, w_N, Y), \quad g_i > 0,$$

where the w_i are the input prices. As in the two-factor case:

$$(27) \quad f_i - \lambda w_i = 0 , \quad i = 1, \dots, N;$$

and, using the cost function:

$$(28) \quad X_i - \mu g_i = 0, \quad i = 1, \dots, N,$$

where λ and μ are Lagrangian multipliers.

The technological parameters can be defined using either the equilibrium conditions based on the production function ((25) and (27)) or those based on the cost function ((26) and (28)). Allen (1938) used f to define the partial elasticity of substitution, the percentage effect of a change in w_i/w_j on X_i/X_j holding output and other input prices constant, as:

$$(29) \quad \sigma_{ij} = \frac{Y}{X_i X_j} \frac{F_{ij}}{|F|},$$

where:

$$F = \begin{vmatrix} 0 & f_1 & \dots & \dots & f_n \\ \vdots & & & f_{ij} & \\ \vdots & & & & \\ f_N & & & & f_{NN} \end{vmatrix},$$

the bordered Hessian determinant of the equilibrium conditions (25) and (27), and F_{ij} is the cofactor of f_{ij} in F .

The definition in (29) is quite messy. An alternative definition based on the cost function is:

$$(30) \quad \sigma_{ij} = \frac{C}{g_i g_j} \frac{g_{ij}}{g_i g_j}.$$

(Note the similarity to the definition of σ in (7) in the two-factor case. Note also that the definition in (30) requires knowledge only of a few derivatives of (26), unlike that of (29), which requires a complete description of the production function.)

If one differentiates the system (25) and (27) totally, the comparative-static equations are:

$$(31) \quad [F] \begin{bmatrix} d\lambda/\lambda \\ dX_1 \\ \vdots \\ dX_N \end{bmatrix} = \begin{bmatrix} dY \\ dw_1/\lambda \\ \vdots \\ dw_N/\lambda \end{bmatrix} .$$

Holding Y and all other w_k constant:

$$(32) \quad \partial X_i / \partial w_j = \frac{F_{ij}}{\lambda |F|} .$$

Multiplying the numerator and denominator of (32) by $w_j X_i X_j Y$:

$$(33) \quad \partial \ln X_i / \partial \ln w_j = \eta_{ij} = \frac{f_{ij} X_j}{Y} \cdot \sigma_{ij} = s_j \sigma_{ij} ,$$

where the last equality results from the assumptions that factors are paid their marginal products and f is linear homogeneous.^{3/} The η_{ij} , factor demand elasticities, can, of course, be calculated more readily using the definition of σ_{ij} based on (26).

Since $\eta_{ii} < 0$ (and thus $\sigma_{ii} < 0$), and since $\sum_j \eta_{ij} = 0$ (by the zero-degree homogeneity of factor demands in all factor prices), it must be the case that at least one $\eta_{ij} > 0$, $j \neq i$. But (and what makes the multifactor case interesting) some of the η_{ij} may be negative for $j \neq i$.

The partial elasticity of complementarity between two factors is defined using the production function as:

$$(34) \quad c_{ij} = \frac{Y f_{ij}}{f_i f_j} .$$

(Here the definition is just a generalization of (10).) The c_{ij} show the percentage effect on w_i/w_j of a change in the input ratio X_i/X_j , holding marginal cost and other input quantities constant.

The c_{ij} can also be defined from the cost function (from the system of equations (26) and (28)) in a way exactly analogous to the definition of σ_{ij} from the production function:

$$(35) \quad c_{ij} = \frac{C}{w_i w_j} \frac{G_{ij}}{|G|},$$

where $|G|$ is the determinant of the bordered Hessian matrix that results from totally differentiating (26) and (28), and G_{ij} is the cofactor of g_{ij} in that matrix (see Sato-Koizumi (1973, p. 48)). Note that unlike the two-factor case, in which $c = 1/\sigma$, $c_{ij} \neq 1/\sigma_{ij}$.

The result of totally differentiating (26) and (28) under the assumption that G is linear homogeneous is:

$$(36) \quad [G] \begin{bmatrix} dY/Y \\ dw_1 \\ \vdots \\ d\dot{w}_N \end{bmatrix} = \begin{bmatrix} Yd\mu \\ dX_1 \\ \vdots \\ dX_N \end{bmatrix}.$$

Solving in (26) for $\partial w_i / \partial X_j$:

$$(37) \quad \partial w_i / \partial X_j = \frac{G_{ij}}{|G|}.$$

Multiply both numerator and denominator in (37) by $C w_i w_j X_j$ to get:

$$(38) \quad \partial \ln w_i / \partial \ln X_j = \epsilon_{ij} = s_j c_{ij},$$

the partial elasticity of factor price i with respect to a change in the quantity X_j .

Since $\epsilon_{ii} = s_i c_{ii} < 0$, and $\sum_j s_j c_{ij} = 0$, $c_{ij} > 0$ for at least some factors. It is quite possible, though, that there are factors for which $\epsilon_{ij} < 0$ for some $j \neq i$, i.e., for which an exogenous increase in the quantity of input j reduces the price of input i at a constant marginal cost.

The partial elasticities of demand and of factor prices can be used to classify pairs of factor inputs. Using the ϵ_{ij} , inputs i and j are said to be q-complements if $\epsilon_{ij} = s_j c_{ij} > 0$, q-substitutes if $\epsilon_{ij} < 0$. (Note that it is possible for all input pairs (i,j) to be q-complements.) Using the η_{ij} , inputs i and j are said to be p-complements if $\eta_{ij} = s_j \sigma_{ij} < 0$, p-substitutes if $\eta_{ij} > 0$. (Note that it is possible for all input pairs (i,j) to be p-substitutes.) If there are only two inputs, they must be q-complements and p-substitutes.⁴

The use of these definitions should be clear, but some examples may demonstrate it better. If skilled and unskilled labor are p-substitutes, one may infer that a rise in the price of skilled labor, perhaps resulting from an increase in the ceiling on payroll taxes, will increase the mix of unskilled workers in production. These two factors may also be q-complements. If so, an increase in the number of skilled workers (perhaps resulting from increased awareness of the nonpecuniary benefits of acquiring a college education) will raise the wage of unskilled workers by increasing their relative scarcity.

The concepts developed in this section can be illustrated by a number of the specific functional forms that have been used in the literature to estimate production/cost relations describing several inputs.

A. Multifactor Cobb-Douglas and CES functions

These are just logical extensions of the two-factor cases. The N-factor Cobb-Douglas cost function can be written:

$$(39) \quad C = \prod_1 w_1^{\alpha_1}, \quad \sum \alpha_1 = 1 .$$

Each $\sigma_{ij} = 1$ (just apply (30) to (39)), making this function quite uninteresting in applications where one wishes to discover the extent of p-substitutability or examine how substitution between X_i and X_j is affected by the amount of X_k used. That $c_{ij} = 1$ can be readily derived from a generalization of the argument in (13)-(15).

The N-factor CES production function is:

$$(40) \quad Y = [\sum \beta_1 X_1^\rho]^{1/\rho}, \quad \sum \beta_1 = 1 .$$

As with the N-factor Cobb-Douglas function, the technological parameters are not interesting:

$$c_{ij} = 1 - \rho \text{ for all } i \neq j .$$

The degree of substitution within each pair of factors is restricted to be identical.

A slightly more interesting case is that of the two-level CES function containing M groups of inputs, each of which contains N_i individual inputs:

$$(41) \quad Y = \left\{ \left[\sum_1^{N_1} \alpha_1 X_1^{\rho_1} \right]^{\frac{v}{\rho_1}} + \dots + \left[\sum_{N_{M-1}}^{N_M} \alpha_k X_k^{\rho_N} \right]^{\frac{v}{\rho_N}} \right\}^{\frac{1}{v}}, \quad \sum_1^{N_M} \alpha_i = 1,$$

where the ρ_j and v are parameters to be estimated. Equation (41) is the same as (40), except that groups of factors aggregated by CES subfunctions are themselves aggregated by a CES function with parameter v . For factors within the same subaggregate:

$$c_{ij} = 1 - \rho_k, \quad k = 1, \dots, M .$$

For factors in different subgroups, $c_{ij} = 1 - v$. While (41) is less

restrictive than (40), it still imposes the assumption that the ease of substitution is the same between all pairs of factors not in the same subgroup; and it also imposes separability--substitution within a subgroup is unaffected by the amount of inputs from other subgroups.

B. Generalized Leontief

The cost function, an expanded version of (20), is:

$$(42) \quad C = Y \sum_{ij} a_{ij} w_i^{.5} w_j^{.5}, \quad a_{ij} = a_{ji} .$$

The technological parameters can be estimated from:

$$(43) \quad X_i = a_{ii} + \sum_j a_{ij} [w_j/w_i]^{.5}, \quad i = 1, \dots, N.$$

The partial elasticities of substitution are:

$$\sigma_{ij} = \frac{a_{ij}}{2[X_i X_j s_i s_j]^{.5}},$$

and

$$\sigma_{ii} = \frac{a_{ii} - X_i}{2 X_i s_i} .$$

To derive the σ_{ij} from this functional form, one need only know those parameters that involve factors i and j .⁵ A production function analogous to (42) can be used to derive the c_{ij} easily (and the σ_{ij} with great effort!).

C. Translog

In general the translog cost function is:

$$(44) \quad \ln C = \ln Y + a_0 + \sum_i a_i \ln w_i + .5 \sum_i \sum_j b_{ij} \ln w_i \ln w_j ,$$

with

$$(45) \quad \sum_i a_i = 1; \quad b_{ij} = b_{ji} ; \quad \sum_i b_{ij} = 0 \text{ for all } j.$$

The first and third equalities in (45) result from the assumption that C is

linear homogeneous in the w_i (proportionate increases in the w_i raise costs proportionately). By Shephard's lemma:

$$(46) \quad \frac{\partial \ln C}{\partial \ln w_i} = X_i w_i / C = s_i, \quad i = 1, \dots, N,$$

where both sides of the factor demand equation have been multiplied by w_i/C , and we have assumed factors receive their marginal products.

The reason for writing (46) as it is rather than as a set of factor-demand functions is that, while the latter are nonlinear in the parameters,

(46) is linear:

$$(47) \quad s_i = a_i + \sum_{j=1}^N b_{ij} \ln w_j, \quad i = 1, \dots, N.$$

The partial elasticities of substitution are:

$$\sigma_{ij} = [b_{ij} + s_i s_j] / s_i s_j, \quad i \neq j;$$

$$\sigma_{ii} = [b_{ii} + s_i^2 - s_i] / s_i^2.$$

The σ_{ij} can also be calculated from a translog production specification, but to do so requires using (29), and thus the determinant of what could be a large matrix. The production form is useful, though, to derive partial elasticities of complementarity.

These functional forms and the associated production functions are all summarized in Table 1 for the multifactor case. (Though the Cobb-Douglas and CES should not be used when there are more than two factors, I present them here to allow their use in the two-factor case.) The relative merits of and problems with the alternative cost and production tableaux are discussed in the next sections.

TABLE 1
SUMMARY OF FUNCTIONAL FORMS

Theoretical Forms

Estimating Forms and
Demand Elasticities

1. Cobb-Douglas

a. Cost

$$C = Y^a \prod w_i^{\alpha_i}; a = 1 \text{ if CRS}$$

$$\ln C/Y = \sum \alpha_i \ln w_i$$

b. Production

$$Y = \prod X_i^{\beta_i}; \sum \beta_i = 1 \text{ if CRS}$$

$$\ln Y = \sum \beta_i \ln X_i;$$

$$\eta_{11} = [1 - \beta_1]; \eta_{1i}$$

2. CES

a. Cost

$$C = Y^a \left[\sum_i \alpha_i w_i^{\sigma} \right]^{\frac{1}{1-\sigma}}, a = 1 \text{ if CRS}$$

$$\ln X_i = a_0 + \sigma \ln w_i + a \ln Y;$$

$$\eta_{11} = s_i \sigma;$$

b. Production

$$Y = \left[\sum_i \beta_i X_i^{\rho} \right]^{\frac{1}{\rho}}, b = 1 \text{ if CRS}$$

Little use

3. Generalized Leontief

a. Cost

$$C = YEE \alpha_{ij} w_i^{.5} w_j^{.5},$$

$$a_{ij} = a_{ji}$$

$$X_i = a_{11} + \sum_j a_{ij} [w_j/w_i]^{.5}, \quad i = 1, \dots, N;$$

$$\eta_{ij} = \frac{s_j a_{11}}{2[X_i X_j s_i s_j]^{.5}}$$

$$\eta_{11} = \frac{[a_{11} - X_1]}{2X_1}$$

b. Production

$$Y = EEB_{ij} X_i^{.5} X_j^{.5},$$

$$b_{ij} = b_{ji}$$

$$w_i = b_{11} + \sum_j b_{ij} [X_j X_i]^{.5}, \quad i = 1, \dots, N$$

$$\epsilon_{ij} = \frac{s_j b_{11}}{2[w_i w_j s_i s_j]^{.5}}$$

$$\epsilon_{11} = \frac{b_{11} - w_1}{2w_1}$$

4. Translog

a. Cost

$$\ln C/Y = a_0 + \sum \alpha_i \ln w_i + .5EEb_{ij} \ln w_j$$

$$b_{ij} = b_{ji}$$

$$s_i = a_i + \sum_j b_{ij} \ln w_j; \quad i = 1, \dots, N$$

$$\eta_{ij} = [b_{ij} + s_i s_j]/s_i$$

$$\eta_{11} = [b_{11} + s_1^2 - s_1]/s_1$$

b. Production

$$\ln Y = \alpha_0 + \sum \alpha_i \ln X_i + .5EE\beta_{ij} \ln X_i \ln X_j$$

$$\beta_{ij} = \beta_{ji}$$

$$s_i = \alpha_i + \sum_j \beta_{ij} \ln X_j; \quad i = 1, \dots, N$$

$$\epsilon_{ij} = [\beta_{ij} + s_i s_j]/s_i$$

$$\epsilon_{11} = [\beta_{11} + s_1^2 - s_1]/s_1$$

IV. Homogeneous Labor--Estimation and Empirical Issues

In this section we deal with the problems involved in estimating the demand for homogeneous labor. We examine how one estimates the demand parameters under the assumption that all units of labor are identical. The parameters of interest, the labor-demand elasticity and the cross-price and substitution elasticities, have been produced both in the two- and the multifactor cases. We discuss both issues of how the estimating equations are to be specified, and how they are to be estimated and the results interpreted.

A. Specification

The first approach to estimation relies on the production or cost function "directly." In the case of the Cobb-Douglas function this method produces the distribution parameters. (If, for example, data on factor prices are unavailable, these parameter estimates are necessary to compute the factor-demand elasticities. If data on shares can be computed, there is no reason to estimate such a function.) Estimating a CES function directly is, as inspection of (16) shows, not easy, so the direct approach does not apply here. The generalized Leontief and translog approximations can be estimated directly (either in their cost or production function forms). Though little work has relied upon this approach, it is quite feasible in the two-factor case. In the multifactor case the problem of multicollinearity ($N+1$ terms involving each factor of production are included in the translog approximation, N in the generalized Leontief approximation) becomes severe (but see Hansen *et al* (1975)). With more than one other factor included, direct estimation should not be done unless one arbitrarily imposes a multifactor Cobb-Douglas technology.

The second approach uses labor-demand conditions, either from the marginal productivity condition (2a) or the Shephard condition (6a). In the simplest case, a CES function, this means estimating an equation like:

$$(48) \quad \ln L = a_0 + \sigma \ln w_L + a_1 \ln Y,$$

where the a_i are parameters, with $a_1=1$ if the production function is characterized by constant returns to scale.⁶ (Indeed, estimating (48) without constraining a_1 to equal one is the standard way of testing for constant returns to scale while estimating the labor-demand equation.) In the generalized Leontief and translog cases the amount of labor demanded is a nonlinear function of the factor prices, which makes these approaches inconvenient.

In the multifactor case the labor-demand approach involves the estimation of an equation like:

$$(49) \quad \ln L = \sum b_i \ln w_i + a_1 \ln Y, \quad \sum b_i = 0,$$

where one can test for constant returns to scale ($a_1=1$). Clearly, (49) should be viewed as part of a complete system of factor-demand equations; if data on all factor quantities are available, a complete system should be estimated. If not, though, (49) will provide all the necessary estimates, for:

$$\delta \ln L / \delta \ln w_i = [s_i / s_L] \delta \ln X_i / \delta \ln w_L.$$

The multifactor labor-demand approach provides a useful way of testing whether the condition that the demand for labor be homogeneous of degree zero in factor prices holds, and whether it is homogeneous of degree one in output. A similar approach can be used to examine a wage equation specified as a linear function of the logarithms of all factor quantities.

Yet a third approach may be called the relative factor demand method. In the two-factor CES case this just involves estimation of (18), with $\ln L/K$ as a dependent variable, from which the demand elasticities can be calculated. Some research has used this method, but none has used (21) or (23) directly.

The relative factor-demand method should not be used in the multifactor case, for it involves the estimation of all pairs of equations like (18), in the CES case, or like (21) and (23) in the more general cases. While there is nothing inherently wrong with this approach, it prevents the imposition of the restrictions that factor demand be homogeneous of degree zero in all factor prices. Since that restriction is a postulate of the theory, the specification that prevents the researcher from imposing or at least testing it does not seem desirable.

The fourth approach is to estimate the demand for labor as a part of a system of equations based upon one of the approximations, like the generalized Leontief or translog forms, that we discussed in Section III. Even in the two-factor case a single equation like (47) for $i=L$ could be used, with the only parameters to be estimated being the constant term and the slope on $\ln(w_L/w_j)$ (since the homogeneity restrictions make an equation for the other factor redundant and the coefficients on $\ln w_L$ and $\ln w_j$ equal and of opposite sign). In the case of several factors homogeneous labor becomes one of the factors in a system of $N-1$ equations. These are the share equations for the translog approximation, or equations (43) for the generalized Leontief approximation.

Throughout the discussion in this section we have dealt only with methods of estimating the constant-output labor-demand elasticity. As we indicated in Section II, in the short run, or for individual firms, sectors or industries, a change in the price of labor will induce a change in output (especially if a

small industry is the unit of observation). The effect of the output change can be measured indirectly or directly. The indirect approach simply takes some extraneous estimate of the demand elasticity for the product of the industry, and uses (4a') to derive a labor-demand elasticity that includes the scale effect. A direct approach would estimate equations like (48) or (49) but with output (Y) deleted.

B. Measurement and Interpretation Issues

There are many data considerations in estimating elasticities involving labor demand; we concentrate here only on problems concerning the measurement of L and w. The simpler issue is the choice of a measure of the quantity L. In the literature the alternatives have mostly been total employment and total hours. Clearly, if workers are homogeneous, working the same hours per time period, the choice is irrelevant. If they are heterogeneous along the single dimension of hours worked per time period, using number of workers to represent the quantity of labor will lead to biases if hours per worker are correlated with factor prices or output. In studies using cross-section data, in which there may be substantial heterogeneity among plants, firms or industries in hours per worker, this consideration suggests that total hours be used instead of employment. In time-series data (on which most of the estimates of demand elasticities for homogeneous labor are based) the choice is probably not important, since there is relatively little variation in hours per worker over time. However, if one is also interested in the dynamics of labor demand, the choice is crucial, for there are significant differences in the rates at which employment and hours adjust to exogenous shocks (see Nickell in this volume).

The choice of a measure of the price of labor is more difficult. Most of the published data from developed countries are on average hourly earnings or

average wage rates. A few countries publish data on compensation (employers' payments for fringes and wages per hour on the payroll). While most of the studies of the demand for homogeneous labor use one of the first two measures, none of these three is satisfactory. There are two problems: 1) Variations in the measured price of labor may be the spurious result of shifts in the distribution of employment or hours among subaggregates with different labor costs, or of changes in the amount of hours worked at premium pay; and 2) Data on the cost of adding one worker (or one hour of labor services) to the payroll for one hour of actual work are not available.'

The first problem can be solved in studies of labor demand in the United States using the adjusted earnings series covering most of the postwar period for the private nonfarm economy. The second problem is soluble (except for labor costs resulting from inputs into training) for studies of the United States labor market beginning in 1977 by the Employment Cost Index that the Bureau of Labor Statistics has produced. Clearly, future work using aggregate data should rely upon that index. That the distinction is important is shown in Hamermesh (1983), in which a measure of labor cost per hour worked is developed and shown to lead to substantially higher own-price demand elasticities than do average hourly earnings or average compensation measures.

The second measurement issue is what variables if any should be treated as exogenous. Ideally the production or cost function, or labor-demand equation, will be embedded in an identified model including a labor supply relation. In such a case methods for estimating a system of equations are appropriate, and the problem is obviated: Both the price and quantity of labor may be treated as endogenous. If a complete system cannot be specified, one may have sufficient variables that are not in the equation based on the cost or production function and that can be used to produce an instrument for the

endogenous right-hand side variable. However, given the difficulty of specifying a labor supply relation in the aggregate data on which most studies of labor demand are based, it seems unlikely that a good set of variables can be found.

The choice usually boils down to whether price or quantity can be viewed as exogenous in the problem under study. In studies based on small units--plants, firms, or perhaps even geographical areas--one might well argue that supply curves to those units are nearly horizontal in the long run. If so, the wage rate may be treated as exogenous; and estimates of cost functions, labor-demand equations, or share equations based on factor prices are appropriate (for they include the wage instead of the quantity of labor as an independent variable). In studies using aggregate data this assumption has not been considered valid since Malthusian notions of labor supply were abandoned. If, as many observers believe, the supply of labor to the economy is quite inelastic even in the long run, demand parameters are best estimated using specifications that treat the quantity of labor as exogenous; production functions and variants of second-order approximations that include factor quantities as regressors should be used.

Since in reality it is unlikely that the supply of labor to the units being studied is completely elastic or inelastic, any choice other than estimating production parameters within a complete system including supply is unsatisfactory. However, since supply relations have not been estimated satisfactorily except in certain sets of cross-section and panel data, one is left to make the appropriate choice based on one's beliefs about the likely elasticity of supply to the units, the availability and quality of data, and whether factor-demand elasticities or elasticities of factor prices are of interest.

V. Homogeneous Labor--Results and Problems

A. Results with Output Constant and Wages Exogenous

Remembering that the chief parameter of interest in analyzing the demand for homogeneous labor is the constant-output own-price elasticity of demand, let us consider a number of studies that have produced estimates of this parameter.⁷ I have divided the studies into two main types depending on the specification of the equations estimated: Labor-demand studies and production or cost-function studies. All of the latter use either a CES production function or a translog cost function. In the translog cost functions labor is specified as one of several factors of production (with energy, the focus of interest in these studies, included as one of the other factors).

In Table 2 I list the classification of the available studies of the constant-output long-run demand elasticity for labor. The estimates are of the absolute value of the own-price elasticity of demand for homogeneous labor. (The studies listed in Part I.A are based on relationships like (48); since the values of s_L are unavailable for the particular samples, I present the estimates of $\eta_{LL}/(1-s_L)=\sigma$.) The estimates in the studies based on a marginal productivity condition imply a measure of the responsiveness of demand that is quite consistent with constant-output demand elasticities holding other factor prices constant of between .2 and .4 (assuming the share of labor is 2/3, and noticing that the range of most of the estimates is .67 - 1.09). Only Black-Kelejian (1970) and Drazen *et al* (1984) among those studies using this approach produce estimates that imply a constant-output demand elasticity holding other factor prices constant that is well below this range. The latter may be an outlier because of the difficulties with the wage data for some of the countries; why the estimates in the former are so low is

TABLE 2

STUDIES OF THE AGGREGATE EMPLOYMENT-WAGE ELASTICITY

Author and Source	Data and Industry Coverage	η_{LL}
I. Labor Demand Studies		
A. Marginal productivity condition on labor (estimates of $\eta_{LL}/[1-s]$)		
Black-Kelejian [1970]	Private nonfarm, quarterly, 1948-65	.36
Dhrymes [1969]	Private hours, qtrly., 1948-60	.75
Drazen et al [1984]	Manuf. hours, qtrly., 10 OECD countries, mostly 1961-1980	.21 ^a
Hamermesh [1983]	Private nonfarm, qtrly., based on labor cost, 1955-1978	.47
Liu-Hwa [1974]	Private hours, monthly, 1961-71	.67
Lucas-Rapping [1970]	Production hours, annual, 1930-65	1.09
Rosen-Quandt [1978]	Private production hours, annual 1930-73	.98
B. Labor demand with price of capital		
Chow-Moore [1972]	Private hours, qtrly., 1948:IV-1967	.37 ^b
Clark-Freeman [1980]	Manuf. qtrly., 1950-76: Employment	.33
	Hours	.51
Nadiri [1968]	Manuf. qtrly., 1947-64: Employment	.15
	Hours	.19
Nickell [1981]	Manuf. qtrly., 1958-74, United Kingdom (materials prices)	.19
Tinsley [1971]	Private nonfarm, qtrly., 1954-65: Employment	.04 ^b
	Hours	.06 ^b
C. Interrelated factor demand		
Coen-Hickman [1970]	Private hours, annual 1924-40, 1949-65	.18
Nadiri-Rosen [1974]	Manuf. employment, qtrly., 1948-65: Production	-.11
	Nonproduction	.14
Schott [1978]	British industry, annual, 1948-1970: Employment	.82
	Hours	.25

II. Production and Cost Function Studies

A. CES production functions

Brown-deGani [1963]	Private nonfarm hours, annual, 1933-58	.47
David-van de Klundert [1965]	Private hours, annual, 1899-1960	.32
McKinnon [1963]	2-digit SIC manuf., annual, 1947-58	.29 ^c

B. Translog cost functions

Berndt-Khaled [1979]	Manuf., annual, 1947-71; capital, labor, energy and materials:	
	Homogeneous, neutral techn. change	.46
	Nonhomothetic, nonneutral techn. change	.17
Magnus [1979]	Enterprise sector, annual, 1950-76, Netherlands; capital, labor and energy	.30 ^b
Morrison-Berndt [1981]	Manuf., annual, 1952-71; capital, labor energy and materials	.35
Pindyck [1979]	10 OECD countries, annual, 1963-73; capital, labor and energy	.43 ^a

^a Simple average of country estimates.

^b Estimates calculated at the sample end-point.

unclear.

Studies included under I.B. in the table in most cases specify the price of capital services in a labor-demand equation that can be viewed as part of a complete system of demand equations. The estimates have the virtue that the own-price demand elasticity is simply the coefficient of $\ln w_L$ in the equation containing $\ln L$ as the dependent variable. The estimates are substantially lower than those produced in studies in I.A. that include only the wage rate. However, when one remembers that the estimates in I.A. are of the elasticity of substitution, the two sets of estimates are in the same fairly narrow range. Only the estimates based on interrelated factor demand (Part I.C. in the table) are below the range implied by the estimates in I.A. and I.B. Clark-Freeman (1980) have shown that measures of the price of capital services are much more variable than measures of wages or earnings (presumably reflecting at least in part errors of measurement). Studies of interrelated factor demand, by estimating labor and capital demand simultaneously, inherently base the estimated labor-demand elasticities in part on the responsiveness of the demand for capital to what is likely to be a poorly measured price of capital. This view suggests the studies in Part I.C. of the table probably shed little light on the demand parameters of interest.

Among the cost and production function studies listed under Part II of Table 2 there is a remarkable degree of similarity in the implied constant-output labor-demand elasticity. Given the diversity of specifications, sample periods and units that are studied, the extent of agreement is astounding. These studies produce estimates that are roughly in agreement with those listed under I.A. and II.B. Again, whether one takes information on other factor prices into account or not seems to make little difference for the estimates of the labor demand elasticity. All that is

required is that one interpret one's results carefully, relating the parameter estimates back to the elasticity one is trying to estimate.

Obviously there is no one correct estimate of the constant-output elasticity of demand for homogeneous labor in the aggregate. The true value of the parameter will change over time as the underlying technology changes, and will differ among economies due to differences in technologies. However, a reading of the estimates in Table 2 suggests that, in developed economies in the late twentieth century, the aggregate long-run constant-output labor-demand elasticity lies roughly in the range .15 - .50. While this range is fairly wide, it does at least put some limits on the claims one might make for the ability of, for example, wage subsidies to change the relative labor intensity of production at a fixed rate of output. These limits suggest that the huge empirical literature summarized here should narrow the debate over what the likely effects would be of any change imposed on the economy that affects the demand for labor.

An examination of these empirical studies and a consideration of the problems of specification indicates that the labor-demand elasticity can be obtained from a marginal-productivity condition, from a system of factor-demand equations, from a labor-demand equation that includes other factor prices, or from a system of equations that produces estimates of the partial elasticities of substitution among several factors of production. Often data on other factor prices will not be so readily available as the wage rate. The lack of differences we have noted between studies that include other factor prices and those that do not suggest the effort devoted to obtaining series on those other prices will not result in major changes in the estimates of the labor-demand elasticity.

B. Varying Output or Endogenous Wages

While our major interest is in the constant-output labor-demand elasticity, it is may be worth asking a short-run question: What is the elasticity when output can vary, that is, what is a reasonable value for η' in (4a')? The responses to changes in wage rates under these assumptions are obviously of special interest to those concerned with short-run macroeconomic problems. One recent study, Symons-Layard (1983), examined demand functions for six large OECD economies in which only factor prices, not output, were included as independent variables. The estimates range from .4 to 2.6, with four of the six being greater than one. These relatively large estimates suggest, as one should expect from comparing (4a) and (4a'), that there is more scope for an imposed rise in real wages to reduce employment when one assumes output can vary.

The discussion thus far has dealt with the demand for homogeneous labor in the aggregate. Nearly all the studies summarized treat factor prices, including the wage rate, as exogenous. Yet, as we noted in Section IV, this assumption is strictly correct only if the elasticity of labor supply is infinite, which hardly seems correct in those studies based upon data from entire economies. (It is unlikely that the private nonfarm sector can elicit more labor from households without any increase in the market price of time.) The remarkable similarity of the results discussed in this Section may merely arise from the authors' use of methods that are similar, but essentially incorrect, and that fail to provide a proper test of the theory of labor demand. Studies based on units of observation to which the supply of labor can be claimed to be truly exogenous thus provide a clearer test of the predictions of the theory of labor demand.

Estimates of labor-demand elasticities for small industries, for workers within a narrowly-defined occupation, for workers within small geographical areas, or even within individual establishments, are less likely to be fraught

with problems of simultaneous-equations bias than are the macro time-series that underlie the studies summarized in Table 2. Unfortunately, relatively little attention has been paid to this problem; but those studies that have treated less aggregated data describing the demand homogeneous labor are summarized in Table 3. The estimates of the constant-output labor-demand elasticities are quite similar to those summarized in Table 2. This suggests that the estimated elasticities that seem to confirm the central prediction of the theory of labor demand are not entirely an artifact produced by using aggregate data.

One might claim that even these units of observation are not the establishments or firms upon which the theory is based. It is true that, in contrast to the myriad studies of labor supply behavior based observations on households, there is a shocking absence of research on the empirical microeconomics of labor demand. Thus the most appropriate tests of the predictions of the theory have yet to be made. For those skeptical even of the results in Table 3 that are based on data describing occupations or industries, an additional confirmation of the theory is provided by analyses of the effects of the minimum wage. An overwhelming body of evidence (see the summary in Brown *et al.* (1982)) indicates that imposed, and thus exogenous, changes in minimum wages induce reductions in the employment of workers in those groups whose market wages are near the minimum.

VI. Heterogeneous Labor--Estimation and Empirical Issues

Most of the methods for specifying and estimating models involving several types of labor carry over from the discussion of homogeneous labor in the previous section. Yet because one is generally interested in many more parameters than in the case of homogeneous labor, there are several

TABLE 3

INDUSTRY STUDIES OF LABOR DEMAND

<u>Author and Source</u>	<u>Data and Industry Coverage</u>	<u>η_{LL}</u>
Ashenfelter and Ehrenberg (1975)	State and local government activities, states, 1958-69	.67 ^a
Field-Grebenstein (1980)	2-digit SIC manuf., annual, 1947-58	.29 ^a
Freeman (1975)	U.S., university faculty, 1920-1970	.26
Hoperoft-Symons (1983)	U.K. road haulage, 1953-80, capital stock held constant	.49
Lovell (1973)	2-digit SIC manuf., states, 1958	.37 ^a
McKinnon (1963)	2-digit SIC manuf., annual, 1947-58	.29 ^a
Sosin-Fairchild (1984)	770 Latin American firms, 1970-74	.20
Waud (1968)	2-digit SIC manuf., qtrly., 1954-64	1.03 ^a

^aWeighted average of estimates, using employment weights.

considerations that do not arise in that case.

A. Specification

If one assumes that there are only two types of labor, and that they are separable from nonlabor inputs, the discussion in the previous section applies and the ways of estimating substitutability between the two factors should be apparent. (But see below for some problems that arise in this case.) In most instances, though, the problem at hand involves estimating the degree of substitutability among several types of labor (and among them and other factors). In that case, as the discussion in Section III should make clear, the restrictive Cobb-Douglas and CES forms will not be appropriate to answer the questions under study except under highly unlikely circumstances.

Two alternatives are possible, with the choice depending on the availability of data: 1) A complete system of factor-demand equations, essentially a series of N equations with the L_i , $i=1\dots N$, as dependent variables, and the same set of independent variables as in (49); and 2) A system of equations based on one of the flexible approximations to a production or cost function, e.g., the generalized Leontief or translog forms, such as are shown in Table 1. (Whether one specifies these systems with factor prices or quantities as independent variables is another issue, which we discuss below.) Each of these approaches requires data on all factor prices and quantities. Each of the approaches using the flexible forms allows the ready inference of the partial elasticities of substitution (or of complementarity) as well as the factor-demand (factor-price) elasticities.

As in the case of homogeneous labor, one would ideally specify factor demands simultaneously with factor supplies and be able to estimate a model that obviates the need to consider whether factor prices or quantities are to

be considered exogenous. However, if it is difficult to specify such a model involving homogeneous labor, it seems impossible to do so for a model that includes several types of workers. Accordingly, one must be able to argue that supplies of each type of labor are either completely inelastic or completely elastic in response to exogenous changes in demand.⁸

No satisfactory choice appears to have been made in the studies that have estimated substitution among several types of labor. For example, consider a study that seeks to examine the extent of substitutability among adult women, adult men, youths and capital. It seems reasonable to treat the quantity of adult men in the work force as exogenous, and increasingly also for adult women, but that assumption hardly makes sense for youths whose labor supply appears to be quite elastic. (The supply elasticity of capital is also a problem.) That being the case, the absence of an appropriate set of variables from which to form instruments for the wage or labor quantities used means one must accept some misspecification whether one chooses to treat wages or quantities as exogenous.

As another example, one might argue that the supplies of blue- and white-collar labor to the economy are highly elastic in the long run; but it is unlikely, given the heterogeneity among workers' abilities, that these supplies are completely elastic. Even if one believes they are, the long run over which they are infinitely elastic is probably longer than the quarter or year that forms the basic unit of observation of time-series studies that focus on this disaggregation of the work force. That being the case, there is no clear-cut choice dictated by theory alone about whether wages or quantities should be treated as exogenous in this example either.

The problem is not solved by estimating the cost or production parameters using aggregated cross-section data. For example, the persistence of regional

wage differentials unexplained by apparent differences in amenities suggests that one cannot claim that labor of all types is supplied perfectly elastically to geographical areas. Thus using data on metropolitan areas or other geographical subunits does not guarantee that factor prices can be considered exogenous. The same problem arises when data on industries are used: Insofar as industries use industry-specific skills, the supply of labor to the industry could well be upward-sloping in the long run. The only satisfactory solution, one that has not been tried in practice, is to use data on firms or establishments as the units of observation.

In practice the best guide to the choice between treating wages or quantities as exogenous is the link between this choice and the researcher's own priors on the supply elasticities of the factors whose demand is being examined (and thus how the misspecification that is induced can be minimized). In the example involving adult females, adult males and youths the overwhelming shares of output are accounted for by the first two groups, whose supply of effort is relatively inelastic. That being so, treating factor quantities as exogenous is probably the better choice. This also means that one should focus the analysis on the elasticities of complementarity and of factor prices, which are estimated more readily using production rather than cost functions (see Section III).

B. Measurement and Interpretation Issues

Whether labor subaggregates are separable from capital, or whether some groups within the labor force are separable in production from others, is of central importance in empirical work estimating substitution among heterogeneous workers. Consider first the issue of separability of labor subaggregates from capital. In many cases the available data provide no way of obtaining a measure of the price or quantity of capital services. Even if

such data are available, they may be measured with much greater error than the data on wage rates or employment in each labor subgroup. If the errors of measurement are small, one might well argue the Cambridge position that the notion of trying to aggregate the capital stock in an economy, or even in a labor market, is senseless. That being the case, one must be sure that labor is separable from capital when one estimates substitution relations among labor subgroups in the absence of a measure of capital price or quantity. Otherwise, the estimates of labor-labor substitution will be biased.

A similar problem arises when one concentrates on substitution among several subgroups in the labor force and assumes that they are separable from the rest of labor. (For example, Welch-Cunningham (1978) examine substitution among three groups of young workers disaggregated by age under the assumption that the σ_{ij} of each for adult workers are identical.) The estimates of the σ_{ij} (or c_{ij}) between the pairs of labor subgroups being studied will generally be biased. The separability of the labor subgroups from capital should always be tested rather than imposed if the data permit.

Even if the labor subaggregates are separable from capital (or, if they are not separable, the biases induced by assuming separability are small), a problem of interpretation arises. Assume, for example, that the true production function is:

$$Y = F(K, G[L_1, L_2]),$$

where the function G aggregates the two types of labor. Estimates based on:

$$(50) \quad L = G(L_1, L_2),$$

implicitly measure substitution along an isoquant that holds L , but not necessarily Y constant. Thus the factor-demand elasticities computed from

(50) are not constant-output demand elasticities (see Berndt (1980) for a discussion of this). They are gross elasticities; constant-output labor demand elasticities will differ from these, for any rise in the price of, say, L_1 , will induce a reduction in L (because the price of aggregate labor has fallen). If, for example, the L -constant demand elasticity for L_1 is η^*_{11} , the constant-output demand elasticity will be:

$$(51) \quad \eta_{11} = \eta^*_{11} + s_1 \eta_{LL},$$

where η_{LL} is the constant-output elasticity of demand for all labor (see Berndt-Wood (1979)). In general:

$$\eta_{ij} = \eta^*_{ij} + s_j \eta_{LL}.$$

The true (constant-output) demand elasticity is more negative (greater in absolute value) than the gross elasticity, η^*_{11} ; and the true cross-price demand elasticities are more negative than those based on estimates of substitution using (50) as the underlying production relation.

Assuming the labor subgroups can be treated as separable from capital, there is nothing wrong with the estimates of factor-demand (or factor-price elasticities in the dual case). However, they are not the usual elasticities, and should be adjusted accordingly. Otherwise, one will underestimate own-price demand elasticities, and infer that the types of labor are greater p-substitutes than in fact they are.

Another consideration is the choice of a disaggregation of the work force. Much of the early empirical work (through the middle 1970s) focussed on the distinction between production and nonproduction workers. This was dictated partly by the ready availability of time-series data on this disaggregation, partly by the belief that this distinction represented a

comparison of skilled and unskilled workers. Recent work by labor economists has recognized that differences in skill (embodied human capital) between production and nonproduction workers are not very great. Also, most of the policy issues on which studies of labor demand can have a bearing involve labor subgroups disaggregated according to other criteria. Thus most of the recent work has disaggregated the work force by age, by race or ethnicity, by sex, or by these criteria in various combinations.

Given one's interest in substitution among particular groups of labor necessitates the aggregation of workers who differ along other dimensions that are of less interest to the researcher. Care should be exercised, though, that the aggregations decided upon make sense, in that substitution toward other groups is the same for all workers within a subaggregate.⁹ In practice this means that, wherever possible given the limitations of the data being used, one should test for the consistency of aggregating workers into larger groups. For example, if one is concerned about substitution among males, females and capital, one should if possible test whether the substitution between young men and females (or capital) is the same as that for older men.

The problem of deciding which disaggregation to use and the larger difficulty of deciding what we mean by a "skill" have led to efforts to circumvent the decision by defining a set of characteristics of the workers. In this view (see Welch (1969) and Rosen (1983)) each worker embodies a set of characteristics (by analogy to Lancasterian models of the demand for goods).¹⁰ This approach has the appeal of avoiding the aggregation of what may be very dissimilar workers into a particular group; instead, it "lets the data tell" what the appropriate skill categories are, in a manner similar to factor analysis. One of its difficulties is that it has not as yet been developed enough that the powerful restrictions of production theory can be imposed on

estimates using this approach. Also, for many issues that attract public interest the arbitrary disaggregations of workers by age, race, sex, etc. are of substantial importance.

VII. Heterogeneous Labor--Results and Problems

A summary of the parameters of interest in the studies that have examined heterogeneous labor disaggregated by occupation is shown in Table 4.¹¹ Perhaps the most consistent finding is that nonproduction workers (presumably skilled labor) are less easily substitutable for physical capital than are production workers (unskilled labor). Indeed, a number of the studies find that nonproduction workers and physical capital are p-complements. This supports Rosen's (1968) and Griliches' (1969) initial results on the capital-skill complementarity hypothesis. This finding has major implications for the employment effects of such policies as accelerated depreciation, investment tax credits and other attempts to stimulate investment in physical capital, suggesting that they will increase the demand for skilled relative to unskilled labor.

Although not uniformly observed in all the studies tabulated, in most the demand elasticity for nonproduction workers is lower than that for production workers. This difference reflects what seems to be a consistent result among studies examining all the disaggregations of the labor force: The own-price demand elasticity is lower, the greater is the amount of human capital embodied in the average worker in the particular class of labor. Thus skill *per se* ties employers to workers by making labor demand less sensitive to exogenous changes in wage rates.

One would like to draw some inferences about the ease of substitution of

TABLE 4

STUDIES OF SUBSTITUTION OF PRODUCTION
AND NONPRODUCTION WORKERS

Study	Data and Method	σ_{BK}	σ_{WK}	σ_{BW}	η_{BB}	η_{WW}
I. Capital Excluded						
A. Cost Functions						
Freeman-Medoff (1982)	Manufacturing plants, 1968, 1970, and 1972, detailed industry dummy variables; CES Union Nonunion	.19	.28			
B. Production Functions						
Berndt-Christensen (1974b)	Manufacturing, 1929-68; translog, 1968 elasticities		4.9	-1.63	-2.87	
Dougherty (1972)	States, Census of Population, 1960; CES		4.1			
II. Capital Included						
A. Cost Functions						
Berndt-White (1978)	Manufacturing, 1947-71; translog, 1971 elasticities	.91	1.09	3.70	-1.23	-.72
Clark-Freeman (1977)	Manufacturing, 1950-76; translog, mean elasticities	2.10	-1.98	.91	-.58	-.22
Dennis-Smith (1978)	2-digit manufacturing 1952-73; translog, mean elasticities	.14	.38	-.05		

Denny-Fuss (1977)	Manufacturing, 1929-68; translog.	1.50	-.91	2.06	
Freemar-Medoff (1982)	1968 elasticities Pooled states and 2-digit manufacturing industries, 1972; translog	.53 .90	-.02 1.02	-.24 .76	-.12 -.43
Grant (1979)	Union .94 Nonunion SMSAs, Census of Population, 1970; translog	.47	.08 .14	.52	-.32
Kesselman et al. (1977)	Professionals & managers Sales & clericals Manufacturing, 1962-71; translog	.46			-.18
Woodbury (1978)	1971 elasticities Manufacturing, 1929-71; translog	1.28	-.48	.49	-.34
	1971 elasticities			-.70	-.52
B. Production Functions					
Berndt-Christensen (1974b)	Manufacturing, 1929-68; translog	2.92	-1.94	5.51	-2.59
Chiswick (1978)	1968 elasticities States, Census of Population, 1910 and 1920				
Denny-Fuss (1977)	Manufacturing; CES Professionals vs. others Manufacturing, 1929-68; translog		2.5		
Hansen et al. b (1975)	1968 elasticities 3- & 4-digit industries, Census of Manufactures, 1967, translog; ^a highest quartile of plants lowest quartile of plants	2.86	-1.88	4.76	
			6.0 2.0	-1.3 -1.5	

^a Estimates are medians of parameters for individual industries.

^b Ranked by value added per manhour. Estimates are medians of parameters for individual industries.

white- for blue-collar labor, and about the absolute size of the demand elasticities for each. Unfortunately, there appears to be very little agreement among the studies on these issues. Examining the table more closely, though, one notices that the estimated demand and substitution elasticities are generally higher in those studies that base them on estimates of production functions. Since inferring these parameters from production functions requires inversion of an entire matrix of parameter estimates (see equation (29)), they will be affected by errors in any of the parameters estimated. While there is no reason to expect biases, the accumulation of errors is also to be avoided. For that reason the cost-function estimates are likely to be more reliable. The estimates shown in Parts I.A and II.A are better ones to use to draw inferences about the extent of substitution among these three factors. Using them, the demand elasticities for the broad categories, white- and blue-collar labor, seem to be roughly the same magnitude as the estimates of the demand elasticity for homogeneous labor that we discussed in Section V.

Only a few studies have disaggregated the labor force by educational attainment. Among them Grant (1979) finds that the own-price demand elasticity declines the more education is embodied in the group of workers. (This is consistent with the results on the relation of the elasticity to the skill level that we noted above.) Grant and others, including Welch (1970) and Johnson (1970), find that college and high-school graduates are p-substitutes. (These latter two studies, which estimate pairwise CES relations, are less reliable because they did not allow the level of other factor inputs to affect the measured extent of substitution within a pair of inputs. Essentially they estimate relative factor demand for many pairs of factors.) All the studies estimate the extent of substitution, and the own-price demand elasticities, to be roughly on the order of those found

between white- and blue-collar workers in Table 4.

The disaggregations of labor used in the studies discussed above are clear-cut. In the more recent research a large variety of disaggregations, mostly involving age and/or race and/or sex, have been used. This diversity makes it rather difficult to draw many firm conclusions from the findings because of the relative lack of replication. In Table 5 I list the results of these studies, separating them by whether they estimate substitution elasticities or elasticities of complementarity. Among the former several results appear consistently among the studies. The estimated demand elasticities (and, though they are not shown in the table, the substitution elasticities) are much larger when produced using methods that treat factor quantities as exogenous. This result parallels what we observed in Table 4; even though quantities may be exogenous, deriving any substitution elasticity from estimates based on this assumption requires estimates of all the production parameters. That requirement may induce large errors when one or more of the parameters is estimated imprecisely.

The estimates of the factor-demand elasticities vary greatly among the studies. (Indeed, in Merrilees (1982) some are positive, for reasons that are not clear; but their sign casts doubt on all of Merrilees' results.) However, the demand elasticity for adult men is generally lower than that for other groups of workers. This result is another reflection of the apparently general inverse relationship between a group's average skill level and the elasticity of demand for its labor. The final generalization from the studies listed in Part I of Table 5 is that in most of the disaggregations each factor is a p-substitute for the others.

As we noted in Section VI, the elasticity of supply should guide the choice about whether to treat wages or quantities as exogenous. In the case of disaggregating by age and sex, treating quantities as exogenous and

TABLE 5

STUDIES OF SUBSTITUTION AMONG AGE AND SEX GROUPS

Category	Study	Data and Method	Types of Labor	σ_{ij}	η_{ii}
I. Substitution and Demand Elasticities					
Wages exogenous Cross-section	Welch- Cunningham (1978) CES	States, Census of Population, 1970; Teenage Labor	A. Capital Excluded		
			14-15, 16-17, 18-19	All are > 0	
Pooled cross-section time series	Government of Australia (1983) factor-demand equations	17 Australian industries, 1976-81; M 21+ F 21+	A. Capital Excluded		
			M < 21 F < 21	All are > 0	-1.80 -4.58
Time-series	Johnson- Blakemore (1979)	Entire U.S. economy, 1970-77; CES	A. Capital Excluded		
			Average for 14 age-sex groups	1.43	
Time-series	Layard (1982)	British manufacturing, 1949-69; translog	A. Capital Excluded		
			M < 21 F < 18 M 21+ F 18+	All are > 0 except F < 18 vs. F 18+	-1.25 -.31 -.35 -1.59
Wages exogenous Time-series	Merrilees (1982) equations	Canada, entire economy, 1957-78; factor-demand Adult males Adult females	B. Capital Included		
			Young males Young females adult male wages are < 0	Mixed, but all involving -.07 .11	.56 -.44

Hamermesh (1982)	Entire economy, 1955-75; translog, mean elasticities	14-24 > 0	- .59 - .01
Quantities exogenous Cross-section	Grant (1979) SMSAs, Census of Population, 1970; 25-44 translog 45+	14-24 > 0	-9.68
Time-series	Anderson (1977) Manufacturing, 1947-72; translog, 1972 elasticities 45+	16-24 > 0	-7.14
II. Elasticities of Complementarity and Factor Prices (Quantities exogenous)		ϵ_{ij}	ϵ_{ii}

A. Capital Excluded

Cross-section	Borjas (1983b) Entire U.S. economy, microdata 1975; Hispanics Generalized Leontief	Blacks > 0 Whites	All are - .64 ^a - .001 ^a
---------------	---	-------------------------	--

B. Capital Included

Cross-section	Borjas (1983a) Census of Population, 1970; Females Generalized Leontief Hispanic migrants White nonmigrants White migrants	Black males those involving Hispanic nonmigrants involving Hispanics	All > 0 except all 2.90 ^a females, and some -11.98 - .03 1.02 ^a
	Grant-Hamermesh (1981) SMSAs, Census of Pop- ulation 1970; translog White M25+ White F25+	Youths Blacks 25+	All > 0 except Youths vs. F 25+ - .13 - .19
	Grossman (1982) SMSAs, Census of Pop- ulation 1970; translog Foreign born	Natives Second generation	All are < 0 - .20 - .03

Pooled cross-section, time-series	Berger (1983)	States, U.S., 1967-74; translog, mean elasticities	M, 0-15 yrs. school, 0-14 experience	All are < 0	- .51a
			M, 16+ school, 0-14 experience	except young	-3.45a
Time-series	Freeman (1979)	Entire U.S. economy, 1950-74; translog, mean elasticities	M, 0-15 school, 15+ experience	vs. old	- .80a
			M, 16+ school, 15+ experience	college grads	-1.48a
			F		- .29a
			M 20-34	Only	- .38
			M 35-64	M 35-64 vs. F is > 0	- .49
			F		- .71

a Own-quantity elasticities of complementarity.

deriving elasticities of complementarity and factor price is the better choice (in the absence of a well-specified model of the supply of each type of labor) if data on large geographical units are used. (Clearly, if data on a small industry or even individual establishments are used, wages should be treated as exogenous. One's belief in the validity of the theory of labor demand should be strengthened by the results of those three studies--Rosen (1968), O'Connell (1972) and King (1980)--that use these small units and find the expected negative own-price elasticities for workers in narrowly-defined occupations.) The studies presented in Part II of Table 5 treat quantities as exogenous and estimate these elasticities for a variety of disaggregations of the labor force. As such they give a better indication of the substitution possibilities within the labor force disaggregated by age, race and sex than do those listed in Part I.

In all the studies the elasticities of factor prices are fairly low. (Given the small share of output accounted for by most of the inputs, the elasticities implied by Borjas' studies and by Berger (1983) are also quite low.) They suggest that the labor market can accommodate an exogenous change in relative labor supply without much change in relative wages.¹² No generalizations about the relative magnitudes of the elasticities are possible from the studies currently available.

One intriguing result occurs in all four studies (Borjas (1983a), Grant-Hamermesh (1981), Berger (1983) and Freeman (1979)) that examine the issue. Adult women are q-substitutes for young workers. Borjas (1983a) also disaggregates the black male work force by age and finds that most of the q-substitutability is between women and young black men. This finding suggests that the remarkably rapid growth in the relative size of the female labor force that has occurred in many industrialized countries, including the

United States, Canada and Sweden, in the past twenty years has contributed to a decline in the equilibrium relative wage rate for young workers. To the extent that relative wages cannot adjust because of real wage floors, and thus permanent unemployment, the assumptions needed to produce estimates of q -substitutability are incorrect. However, so long as adjustment eventually occurs, these cross-section estimates can be used to infer that the growth of the female labor force has also contributed to the high rate of youth unemployment in these countries during this time.

Among the studies discussed in this Section only a few have tested for the separability of labor from capital (and thus shed light on whether estimates of the (gross) elasticities of demand or of factor prices obtained when capital is excluded are biased). Berndt-Christensen (1974a) and Denny-Fuss (1977) examine this issue using the production-worker, nonproduction-worker disaggregation; and Grant-Hamermesh (1981) disaggregate the labor force by age, race and sex. All three studies conclude that the separability of labor from capital is not supported by the data. The findings suggest that the inclusion of the quantity or price of capital services is necessary to derive unbiased estimates of production and cost parameters even between subgroups in the labor force. The extent of the biases induced by assuming separability has not been examined, though Borjas (1983a) indicates that the a_{ij} involving labor-force subgroups change little when capital is excluded from a generalized Leontief system.¹³

There has also been very little effort made to examine whether the particular disaggregations used are correct in assuming that workers included within a subgroup are equally substitutable for workers in other subgroups. This absence is due partly to the difficulties of obtaining data on large numbers of narrowly-defined groups of workers. However, the evidence (see

Grant-Hamermesh (1981)) suggesting that it is incorrect to aggregate subgroups of workers into still larger subgroups should induce greater care in future research in this area.

VIII. Conclusions

Research into the demand for labor over the past 50 years has focussed on depicting demand in a decreasingly restrictive way as the outcome of employers' attempts at cost minimization or profit maximization. The outcome of this trend to date is a means of characterizing demand for N factors of production in a way that allows for complete flexibility in the degree of substitution within any pair of factors; for that flexibility to depend on the firm's output level; and for flexibility in the specification of returns to scale in production. Not only is the theory completely general: We have today the means to describe production relations empirically in a completely general manner.

Perhaps the main advantage of this increased generality is that it allows us to test whether some of the simpler specifications of labor demand describe the data well. Thus the many studies analyzed in Section V suggest that the Cobb-Douglas function is not a very severe departure from reality in describing production relations between homogeneous labor and physical capital. So too, returns to scale in production functions involving homogeneous labor do not seem to differ too greatly from one.

The major advance of the last 15 years has been the ability to estimate substitution within several pairs of inputs. While such estimation is really in its childhood (partly because of the wide range of interesting choices about how to disaggregate the labor force), some results are already fairly solid. 1) Skill (human capital) and physical capital are p-complements in production; at a fixed output employers will expand their use of skilled labor

when the price of capital services declines. 2) The demand for skill is also less elastic than the demand for raw labor; thus we find that the demand for more educated or more highly trained workers is less elastic than that for other workers. 3) No matter what the disaggregation, labor is not separable in production from physical capital. This finding implies that estimates of substitution among groups within the labor force should be based on models that include either the price or quantity of capital services. 4) Finally, though it is less solid a result than the other three, there is an accumulation of evidence that adult women are q-substitutes for young workers.

The theory and estimation techniques we have outlined provide many ways to estimate the degree of factor substitution and the responsiveness of factor demand (prices) to changes in factor prices (quantities). Though the appropriate specification depends upon one's beliefs about the behavior of the agents in the particular labor market, several guidelines for the analysis arise from this discussion. Where at all possible, the specification should allow the researcher sufficient flexibility to test whether simpler specifications are applicable. Where the data are available, physical capital should be included as a factor of production in the analysis along with the various types of labor.

Despite the substantial advances that have been made in analyzing the demand for labor, a remarkable amount is still unknown. We still understand very little about the absolute magnitudes of elasticities of demand, or elasticities of factor prices, for various labor-force groups. So too, the ease of substitution among groups is only now beginning to be analyzed.

More important than these lacunae in our understanding of labor demand, though, are problems induced by the failure to account for the interaction of substitution parameters with parameters describing the supply of labor-force

groups. Those relatively few studies that have estimated demand relations using highly disaggregated data corroborated the basic predictions of the theory of labor demand. However, there has been far too little work that has accounted for the possibility of simultaneity between wages and quantities of labor. Since we have seen how important the specification of labor supply is to deriving estimates of production parameters, the joint estimation of substitution parameters and labor supply should be an area that will lead to substantial advances in understanding the demand for labor. Alternatively, more research is needed that estimates demand relations using data on individual firms or establishments as units of observation.

REFERENCES

- Allen, R.G.D. Mathematical Analysis for Economists. London: Macmillan, 1938.
- Anderson, J. "Labor Force Age Structure Changes and Relative Wages." Unpublished paper, Harvard University, 1977.
- Arrow, K., Chenery, H., Minhas, B. and Solow, R. "Capital-Labor Substitution and Economic Efficiency." Review of Economics and Statistics 43 (1961): 225-250.
- Ashenfelter, O. and R. Ehrenberg. "The Demand for Labor in the Public Sector." In Labor in the Public and Nonprofit Sectors. Ed. D. Hamermesh. Princeton, NJ: Princeton University Press, 1975.
- Berger, M. "Changes in Labor Force Composition and Male Earnings: A Production Approach." Journal of Human Resources 18 (1983): 177-196.
- Berndt, E. "Modelling the Simultaneous Demand for Factors of Production." In The Economics of the Labor Market. Eds. Z. Hornstein et al. London: HMSO, 1980.
- _____ and Christensen, L. "Testing for the Existence of a Consistent Aggregate Index of Labor Inputs." American Economic Review 64 (1974): 391-404, a.
- _____ and _____. "The Specification of Technology in U.S. Manufacturing." Unpublished paper, University of Wisconsin, 1974, b.
- Berndt, E. and Khaled, M. "Parametric Productivity Measurement and Choice Among Flexible Functional Forms." Journal of Political Economy 87 (1979): 1220-1245.
- _____ and White, C. "Income Redistribution and Employment Effects of Rising Energy Prices." Unpublished paper, University of British Columbia, 1978.
- _____ and Wood, D. "Engineering and Econometric Interpretations of Energy-Capital Complementarity." American Economic Review 69 (1979): 342-354.
- Black, S. and Kelejian, H. "A Macro Model of the U.S. Labor Market." Econometrica 38 (1970): 712-41.
- Borjas, G. "The Demographic Determinants of the Demand for Black Labor," Unpublished Paper, University of California--Santa Barbara, 1983, a.
- _____. "The Substitutability of Black, Hispanic and White Labor." Economic Inquiry 21 (1983): 93-106, b.
- Brown, C., Gilroy, G., and Kohen, A. "The Effect of the Minimum Wage on Employment and Unemployment," Journal of Economic Literature 20 (1982):487-528.

- Brown, M. and de Cani, J. "A Measure of Technological Employment." Review of Economics and Statistics 45 (1963): 386-94.
- Chiswick, C. "The Growth of Professional Occupations in U.S. Manufacturing, 1900-73," In Research in Human Capital and Development. Ed. I. Sirageldin. Greenwich, Conn.: JAI Press, 1978.
- Chow, G. and Moore, G. "An Econometric Model of Business Cycles." In Econometric Models of Cyclical Behavior, Vol. II. Ed. B. Hickman. New York: National Bureau of Economic Research, 1972.
- Christensen, L., Jorgenson, D., and Lau, L. "Transcendental Logarithmic Production Frontiers," Review of Economics and Statistics 55 (1973): 28-45.
- Clark, K. and Freeman, R. "How Elastic is the Demand for Labor?" Review of Economics and Statistics 62 (1980): 509-520.
- _____. "Time-Series Models of the Elasticity of Demand for Labor in Manufacturing." Unpublished paper, Harvard University, 1977.
- Coen, R. and Hickman, B. "Constrained Joint Estimation of Factor Demand and Production Functions." Review of Economics and Statistics 52 (1970): 287-300.
- David, P. and van de Klundert, T. "Biased Efficiency Growth and Capital-Labor Substitution in the United States, 1899-1960." American Economic Review 55 (1965): 357-94.
- Dennis, F. and Smith, V.K. "A Neoclassical Analysis of the Demand for Real Cash Balances by Firms." Journal of Political Economy 86 (1978): 793-814.
- Denny, M. and Fuss, M. "The Use of Approximation Analysis to Test for Separability and the Existence of Consistent Aggregates." American Economic Review 67 (1977): 404-18.
- Dhrymes, P. "A Model of Short-Run Labor Adjustment." In The Brookings Model: Some Further Results. Ed. James Duesenberry and others. Chicago: Rand McNally, 1969.
- Diewert, E. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function." Journal of Political Economy 79 (1971): 481-507.
- Dixit, A. Optimization in Economic Theory. Oxford, England: Oxford, 1976.
- Dougherty, C.R.S. "Estimates of Labor Aggregation Functions." Journal of Political Economy 80 (1972): 1101-1119.
- Drazen, A., Hamermesh, D. and Obst, N. "The Variable Employment Elasticity Hypothesis: Theory and Evidence." Research in Labor Economics 6 (1984), forthcoming.

- Ferguson, C. The Neoclassical Theory of Production and Distribution.
Cambridge, England: Cambridge, 1969.
- Field, B. and Grebenstein, C. "Capital-Energy Substitution in U.S.
Manufacturing." Review of Economics and Statistics 62 (1980): 207-212.
- Freeman, R. "Demand for Labor in a Nonprofit Market: University Faculty."
In Labor in the Public and Nonprofit Sectors. Ed. D. Hamermesh.
Princeton, N.J.: Princeton University Press, 1975.
- _____. "The Effect of Demographic Factors on Age-Earnings Profiles."
Journal of Human Resources 14 (Summer 1979): 289-318.
- _____ and Medoff, J. "Substitution Between Production Labor and Other
Inputs in Unionized and Nonunionized Manufacturing." Review of
Economics and Statistics 64 (1982): 220-233.
- Fuss, M. and McFadden, D. Production Economics: A Dual Approach to Theory
and Applications. Amsterdam: North-Holland, 1978.
- Government of Australia, Bureau of Labor Market Research. Youth Wages,
Employment and the Labour Force. Canberra: Australia Government
Publishing Service, 1983.
- Grant, J. "Labor Substitution in U.S. Manufacturing." Ph.D. dissertation,
Michigan State University, 1979.
- _____ and Hamermesh, D. "Labor-Market Competition Among Youths, White
Women and Others." Review of Economics and Statistics 63 (1981):
354-360.
- Griliches, Z. "Capital-Skill Complementarity." Review of Economics and
Statistics 51 (1969): 465-68.
- Grossman, J.B. "The Substitutability of Natives and Immigrants in
Production." Review of Economics and Statistics 64 (1982): 596-603.
- Hamermesh, D. "Econometric Studies of Labor Demand and Their Application to
Policy Analysis." Journal of Human Resources 11 (1976): 507-525.
- _____. "Factor Market Dynamics and the Incidence of Taxes and
Subsidies." Quarterly Journal of Economics 95 (1980): 751-764.
- _____. "Minimum Wages and the Demand for Labor." Economic Inquiry 20
(1982): 365-380.
- _____. "New Measures of Labor Cost: Implications for Demand
Elasticities and Nominal Wage Growth." In The Measurement of Labor
Cost. Ed. J. Triplett. Chicago: University of Chicago Press, 1983.
- _____ and Grant, J. "Econometric Studies of Labor-Labor Substitution and
Their Implications for Policy." Journal of Human Resources 14
(1979): 518-542.

- Hanson, R., Klotz, B. and Madoo, R. "The Structure of Demand for Low-Wage Labor in U.S. Manufacturing, 1967." Unpublished paper, Urban Institute, 1975.
- Hicks, J.R. The Theory of Wages. London: Macmillan, 1964.
- Hopcroft, M.P. and Symons, J. "The Demand for Labour Schedule in the Road Haulage Industry." Discussion Paper No. 169, Centre for Labour Economics, London School of Economics, 1983.
- Johnson, G. "The Demand for Labor by Educational Category." Southern Economic Journal 37 (1970): 190-204.
- _____ and Blakemore, A. "The Potential Impact of Employment Policy on the Unemployment Rate Consistent with Non-Accelerating Inflation." American Economic Review 69 (1979): 119-23.
- Kesselman, J., Williamson, S. and Berndt, E. "Tax Credits for Employment Rather Than Investment," American Economic Review 67 (1977): 339-49.
- King, W. "A Multiple Output Translog Cost-Function Estimation of Academic Labor Services." Unpublished Ph.D., Michigan State University, 1980.
- Layard, R. "Youth Unemployment in Britain and the United States Compared." In The Youth Labor Market Problem: Its Nature, Causes and Consequences. Eds. R. Freeman and D. Wise. Chicago: University of Chicago Press, 1982.
- Liu, T.C. and Hwa, E.C. "A Monthly Econometric Model of the U.S. Economy." International Economic Review 15 (1974): 328-65.
- Lovell, C.A.K., "CES and VES Production Functions in a Cross-Section Context," Journal of Political Economy 81 (1973):705-720.
- Lucas, R.E. and Rapping, L. "Real Wages, Employment and Inflation." In Microeconomic Foundations of Employment and Inflation Theory. Ed. Edmund Phelps. New York: W.W. Norton & Co., 1970.
- Magnus, J. "Substitution Between Energy and Non-Energy Inputs in the Netherlands, 1950-1976." International Economic Review 20 (1979): 465-484.
- McKinnon, R. "Wages, Capital Costs and Employment in Manufacturing." Econometrica 30 (1963): 501-21.
- Merrilees, W. "Labor Market Segmentation in Canada: An Econometric Approach." Canadian Journal of Economics 15 (1982): 458-473.
- Morrison, C. and Berndt, E. "Short-run Labor Productivity in a Dynamic Model." Journal of Econometrics 16 (1981): 339-365.
- Nadiri, M. "The Effect of Relative Prices and Capacity on the Demand for Labor in the U.S. Manufacturing Sector." Review of Economic Studies 35 (1968): 273-88.

- _____ and Rosen, S. A Disequilibrium Model of Production. New York: National Bureau of Economic Research, 1974.
- Nickell, S. "An Investigation of the Determinants of Manufacturing Employment in the United Kingdom." Discussion Paper No. 105, Centre for Labour Economics, London School of Economics, 1981.
- O'Connell, J. "The Labor Market for Engineers: An Alternative Methodology," Journal of Human Resources, 7 (1972):71-86.
- Pindyck, R. "Interfuel Substitution and the Industrial Demand for Energy: An International Comparison." Review of Economics and Statistics 61 (1979): 169-179.
- Rosen, H. and Quandt, R. "Estimation of a Disequilibrium Aggregate Labor Market." Review of Economics and Statistics 60 (1978): 371-379.
- Rosen, S. "A Note on Aggregation of Skills and Labor Quality." Journal of Human Resources 18 (1983): 425-431.
- _____. "Short-Run Employment Variation on Class-I Railroads in the U.S., 1947-1963," Econometrica 36 (1968):511-529.
- Sato, R. "Homothetic and Nonhomothetic CES Production Functions." American Economic Review 67 (1977): 559-569.
- _____ and Koizumi, T. "On the Elasticities of Substitution and Complementarity." Oxford Economic Papers 25 (1973): 44-56.
- Schott, K. "The Relation Between Industrial Research and Development and Factor Demands." Economic Journal 88 (1978): 85-106.
- Sosin, K. and Fairchild, L. "Nonhomotheticity and Technological Bias in Production." Review of Economics and Statistics 66 (1984): 44-50.
- Stapleton, D. and Young, D. "The Effects of Demographic Change on the Distribution of Wages, 1967-1990." Unpublished Paper, Dartmouth College, 1983.
- Symons, J. and Layard, R. "Neo-classical Demand for Labour Functions for Six Major Economies." Discussion Paper No. 166, Centre for Labour Economics, London School of Economics, 1983.
- Waud, R. "Man-Hour Behavior in U.S. Manufacturing: A Neoclassical Interpretation." Journal of Political Economy 76 (1968):407-27.
- Tinsley, P. "A Variable Adjustment Model of Labor Demand." International Economic Review 12 (1971): 482-510.
- Varian, H. Microeconomic Analysis. New York: Norton, 1978.

Waud, R. "Man-Hour Behavior in U.S. Manufacturing: A Neoclassical Interpretation." Journal of Political Economy 76 (1968): 511-529.

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

_____. "Education in Production." Journal of Political Economy 78 (1970): 764-71.

_____ and Cunningham, J. "Effects of Minimum Wages on the Level and Age Composition of Youth Employment." Review of Economics and Statistics 60 (1978): 140-145.

Woodbury, S. "Is the Elasticity of Demand for Labor Subject to Cyclical Fluctuation?" Unpublished paper, University of Wisconsin, 1978.

FOOTNOTES

¹The little trick to derive (17a) and (17b) is to remember that, after having done the grubby arithmetic, the numerator is just Y raised to the power $1-\rho$.

²The arithmetic that demonstrates this is in Varian [1978, p. 18].

³One might wonder how, if $\eta_L = -[1-s_L]\sigma$ in the two-factor case,

$\eta_{LL} = s_L\sigma_{LL}$ in the multi-factor case when we assume $N = 2$. Remembering that $s_L\sigma_{LL} + s_K\sigma_{KL} = 0$, $\eta_{LL} = -s_K\sigma_{KL}$. Since $s_K = 1 - s_L$, and σ_{KL} is just alternative notation for σ , the two representations are identical.

⁴A good mnemonic for these distinctions is that the q and p refer to the exogenous quantities and prices whose variation is assumed to produce changes in endogenous input prices and quantities respectively.

⁵To derive σ_{ij} , perform the required differentiation and remember that $g_i = X_i$.

⁶One should note that the slope parameter on $\ln w_L$ in (48) is not the usual constant-output labor-demand elasticity, and that the latter needs to be calculated from the estimate using (4a). It is also worth noting that (48) is a transformation of the equation used by Arrow et al (1962) to estimate the elasticity of substitution in the CES function they had proposed: Under constant returns to scale (48) can be written as:

$$\ln Y/L = -a_0 - \sigma w_L,$$

the form originally used to estimate σ .

⁷The issues from 1975-1982 of a large number of journals were searched. For years before 1975 the references are taken from Hamermesh (1976). While we make no claim that our survey is exhaustive, it should give a fair representation of work on this subject.

⁸Remember that this is an economic issue, not a problem of inferring the partial elasticities of substitution or complementarity. In the translog case, for example, those can always be inferred, either easily or by inverting a matrix involving all the coefficients estimated.

⁹Indeed, one should be able to demonstrate that workers can be aggregated linearly, not merely that those within a subgroup are separable from those in other subgroups.

¹⁰Stapleton and Young (1983) have attempted to apply this view to the United States for 1967-1977. The results support many of the findings summarized in the next section, though they are not uniformly consistent with the theory of production.

¹¹The issues from 1979-1982 of a large number of journals were searched. For years before 1979 the references are taken from Hamermesh-Grant (1979).

¹²This finding implies nothing about how quickly an economy can adjust to such a change. Even though the required change in relative wages may be slight, adjustment costs may be sufficiently large to lead to long periods of disequilibrium in the markets for some of the groups of labor.

¹³By itself, though, this shows very little, since small changes in the estimated parameters in a translog or generalized Leontief system often lead to large changes in the estimates of the underlying production or cost parameters, as the discussion in Section III indicates.