Volume Title: A Disequilibrium Model of Demand for Factors of Production

Volume Author/Editor: M. Ishaq Nadiri and Sherwin Rosen

Volume Publisher: NBER

Volume ISBN: 0-87014-261-5

Volume URL: http://www.nber.org/books/nadi73-1

Publication Date: 1973

Chapter Title: The Structure of the Model

Chapter Author: M. Ishaq Nadiri, Sherwin Rosen

Chapter URL: http://www.nber.org/chapters/c1249

Chapter pages in book: (p. 13 - 39)

2

THE STRUCTURE OF THE MODEL

THIS chapter specifies the model to be estimated. Our approach is organized as follows. Part A is a review of the theory of demand for factors of production, modified to include optimum choice of utilization rates in addition to choice of input stocks. The appropriate specification of input prices is emphasized. The data are time-series observations, and the received theory of input demand deals with static or long-run equilibria. Therefore, it is necessary to "dynamize" the model to make it refer to actual time-series observations, where the assumptions of static theory are not met. Dynamic considerations are discussed in sections B, C, and D. In section E the model is compared with others in the literature, and is seen to be a generalization of them.

A. INPUT DEMAND FUNCTIONS

Assume a continuous production function

$$Q_t \leq F(Y_1, Y_2, \ldots, Y_n, t), \tag{2.1}$$

where Q is output and the Y_i's are inputs and t is time, capturing technical progress. Assume $F_i < 0$, $F_{ii} < 0$ and $[F_{ij}]$ negative definite. It is also assumed that, for each i,

$$\lim_{Y_i \to 0} F_i = \infty \text{ and } \lim_{Y_i \to \infty} F_i = 0,$$

so that interior solutions are assured and all inputs are actually employed.

Proper specification of inputs in 2.1 must be in terms of service flows per unit of time, since output measurement is in terms of flows. Clearly, the correct flow dimension of each input is its rate of utilization per unit of stock. Hence, one possibility would be to assume that rates of utilization are constant and simply specify stocks in (2.1). Though such an assumption might be warranted in cross-sectional studies, it is evidently untenable in time series. Another possibility, and the one most commonly used, is to specify inputs in terms of the product of stock and rate of utilization, that is, in terms of total service flows. This amounts to using total "man-hours" in the case of labor inputs and total "machine hours" for capital inputs. Such an assumption would be correct only if marginal contributions to output of stock and rates of utilization were independent of each other. We see no reason to impose such severe restrictions at the outset and, therefore, allow both utilization rates and input stocks to be treated as separate objects of choice by the firm. Indeed, later evidence will show output elasticities to be considerably different for utilization and stock dimensions of inputs.

The inputs in (2.1) are defined as follows:

Y_1	stock of production labor
Y_2	utilization rate of production labor
Y_3	capital stock
Y_4	rate of capital services per unit of stock
Y_5	intermediate product
Y_6	stock of nonproduction labor
Y_7	utilization rate of nonproduction labor

Assigning a utilization rate to capital in the form of intermediate goods (Y_5) would be meaningless.

The distinction between input stocks and rates of utilization per unit of stock is made in order to capture intensity of use of inputs as well as their over-all quantity. In one sense, this is similar to a point, recognized by many nineteenth century economists in discussing agricultural production, to the effect that there exists both an "extensive" and an "intensive" margin for certain inputs. Extending that concept to manufacturing suggests that utilization rates are, in part, inversely related to the fraction of the production period during which inputs are "idle." In the case of capital stock, this concept has been studied by Marris [1964] and more formally by Lucas [1970] and is related to the amount of shift work employed over the course of a normal production day. Thus, if a normal shift is eight hours, "full" utilization of capital would consist of three shifts per day. If only two shifts were employed, capital would be idle one-third of the time, and so on. The concept of "labor hoarding," so often used in explaining short-run employment variation, can be interpreted as an attempt to apply to labor concepts similar to those used for capital stock. Alchian [1970] has extended this concept to a very general class of phenomena associated with "nonfull" employment of resources.

In the case of capital, hours per machine is an appropriate measure to use, whereas in the case of labor, hours of work per day and the amount of "overtime" are relevant. At least one other dimension to the concept of utilization should be noted, however, and that concerns intensity of use within the period of employment, or the pace of work activity within each working hour. If production is thought of as a general sequence of "processing" intermediate goods arising at some other points in the process, the flow of such goods may be accelerated or slowed down. Thus, machines may be "run faster" and employees induced to "work harder." That such variations are possible is indicated by the existence of piece-rate systems of labor payment, in which income incentives affect output per worker. Finally, all this is very much related to Stigler's important point that, in the face of output fluctuations, firms may opt for more "flexible" production arrangements as an alternative to perfect production smoothing when inventory holding costs are nonzero. Thus, inventories of inputs are substitutes for inventories of output. Indeed, they may be the only alternative when output consists of services and output storage is not possible.

In any event, variations in utilization rates serve important buffer functions when inventories of input stocks are held. The importance of this point cannot be overemphasized. We believe we have convincing empirical evidence of the crucial role of the utilization rate variations underlying time-series analysis of production and factor demand. Of course, in the empirical estimates presented below, utilization rates are approximated, since the appropriate data are unavailable. Discussion of compromises dictated by data limitations are reserved for Chapter 3. For now, we treat the problem in a more general framework and as if the ideal data were available. To make the discussion more concrete, utilization rates are treated as if they are adequately represented by some concept of hours per unit of stock. However, it should be borne in mind that these concepts are really multidimensional.

Consider the problem of minimizing costs, given some level of output constrained by (2.1). At this stage, we do not distinguish between cases where the given output constraint is in fact optimal and those where it is not. Of course, desired output depends on costs of production, or the parameters that influence choice of inputs. For the present, we choose to ignore the simultaneity and concentrate on least cost factor combinations whatever the level of output. This focuses discussion on factor substitution. Other considerations will be discussed later.

i. Specification of Input Prices

What are the costs of inputs? The first and most important costs of labor are direct rental charges or wage payments. The form of wage payment differs for different classes of labor. There are fixed price contracts, such as annual salary, which, in the first instance at least, are independent of intensity of work. There also are incentive contracts (piece rates) that are geared to some measure of productivity. Finally, there are straight- and overtime wage rates per hour of work.

First consider production labor. In this case the method of payment by annual salary can be ignored. Assume that intensity of work can be very well represented by hours of work per man and that every incentive type of contract can be converted to an equivalent hourly rate contract by measuring labor hours in equivalent "efficiency" units.[1] Let all hourly wage costs per production worker be denoted by w_p. Then if Y_2 is taken to be hours of work per production worker and Y_1 the number of production workers, total rental costs of labor services are given by $(Y_1Y_2)w_p$, where Y_1Y_2 is total man-hours. Notice that w_p conceptually includes *all* costs by the firm that are geared to man-hours, including all supplements and the dollar equivalents of fringe benefits that are awarded on that basis (Soligo [1966]). Also, most labor market contracts have provisions for differential rates of pay depending on the intensity of work or incidence of hours. Overtime and holiday work, for example, are paid at different rates from "straight time." Therefore, w_p is a function of Y_2. To facilitate discussion, the function $w_p(Y_2)$ is assumed smooth and differentiable, although typically that is not the case.

The second types of labor payment to be distinguished are those that depend on stocks or numbers of employees and are independent of total man-hour effort. Thus, market searching and firm-specific training of employees are not costless to firms, but involve considerable expense in

1. Of course, the form of contracts is not arbitrary and should be considered endogenous from a broader perspective. In particular, piece-rate systems serve as risk-sharing devices when information is imperfect. Our purpose in making this assumption is frankly empirical, since requisite data on various payment schemes are unavailable.

the form of personnel departments, employment of personnel to train new workers, and forgone output incurred during the search and training periods. In addition, some types of fringe benefits and indirect labor costs are not wholly geared to total man-hours, but depend on whether or not a worker appears on the payroll of the firm. Some types of unemployment compensation expenses are reckoned in this way, as are employer contributions to certain medical and other insurance schemes. Let all such costs on a per production employee basis be denoted by V_p. Then V_p must be converted from its stock concept to an annual flow. This is accomplished by defining the amortized or flow cost per employee as $s_p = V_p(r + q_p)$, where r is the rate of interest and q_p is a measure of turnover of production workers (Rosen [1968], [1969]). Thus rV_p is the annual forgone income stream derived from investing resources in workers at the time of their hire, rather than investing them in other assets (with marginal rates of return equal to r). The annual cost of replacing employees who retire or quit at rate q_p per year is $q_p V_p$ in order that the number of employees be kept constant. The rate q_p could be affected by payments that would be made rationally by the firm to reduce turnover, though such considerations are neglected here. Annual indirect payments associated with labor stocks are given by $s_p Y_1$. Therefore, total production labor payments are $(Y_1 Y_2)w_p + s_p Y_1$.

Conceptually similar issues apply to nonproduction labor. We expect user cost, s_n, for this group to exceed that of production workers because training and hiring costs are greater. Therefore, total payments to nonproduction workers are $(Y_6 Y_7)w_n + s_n Y_6$, where w_n is the wage rate and s_n is user cost for this type of labor.

Now consider the costs of physical capital services. The firm has a choice between renting its capital or purchasing it outright. However, for determining annual costs chargeable to current account or current operating expenses (the profit and loss statement), proper accounting should guarantee that expenses for renting are identical to those for owning, on the principle that true economic costs represent a claim on current-year resources. These are direct payments if capital is rented, but are opportunity costs if capital is purchased and owned by the firm. As with the case of employment costs, capital costs associated with stocks and due to ownership must be amortized over the lifetime of the capital goods (Jorgenson [1963]), that is, these costs must be converted into equivalent annual rentals. Let p_k represent the stock cost of capital goods and r the cost of capital (as a percentage). Then one portion of the annual rental is given by rp_k. Ignoring differences between borrowing and lending rates, this term may be interpreted in either of two equivalent ways. The firm may have borrowed funds at rate r to finance its purchase of equipment. Then rp_k is the annual interest charge paid on the loans used to finance the purchase.

On the other hand, it may not have been financed by debt, but by retained earnings instead. Then rp_k is the opportunity cost of tying up funds in capital goods rather than in other investments (such as bonds or other types of capital) with marginal rates of return equal to r (with the appropriate adjustment for risk). In either case, rp_k represents opportunity costs. In addition, capital deteriorates. Assume that depreciation occurs in proportion to the existing stock at geometric rate δ. Then δp_k is the per-period replacement charge, accountable to current expenses, necessary to maintain capital intact. In other words, δ ($\times$ 100) per cent of a machine dies off per period due to depreciation, and must be counted as a current operating expense. Finally, for complete consistency in following the concept of opportunity cost, some adjustment must be made in the annual rental charge for changes in valuation of capital, and is accomplished by including a term representing percentage capital gains over the period under consideration. For example, if capital goods prices rise over the period, the true cost of capital is no longer p_k $(r + \delta)$, but a number that is smaller than that by the amount of price increase per unit of capital (see Jorgenson [1963]). Thus, the net opportunity cost of capital is represented by $r+\delta$ net of capital gains. In summary, define c as the annual rental price or user cost of capital. Then

$$c = p_k\left[(r + \delta) - \frac{\dot{p}_k}{p_k}\right].$$

Hence, the stock cost of capital is the unit price times the amount employed[2]—cY_3.

Finally, there are charges associated with variations in the rate of utilization of capital services. Many of the charges that one would ordinarily include here are already captured in the labor cost accounting. For example, insofar as capital is utilized more intensely by operating

2. In the measure actually used in estimation, taxes have been taken into account (see Chapter 3, section A, below).

multiple shifts during the period, employment of labor increases and its costs rise. It is clear that the major costs of changes (other than labor costs) in the utilization of capital are not direct expenses in the same sense as hourly payments to workers, but, rather, are already included in c, once it is recognized that the level of utilization affects the rate of depreciation of capital. When the pace of production is "speeded up" and the degree of "idleness" falls, many of the capital goods components have a shorter lifetime (measured in years) and require greater maintenance expenditures to keep them in operation. Depreciation depends on the rate of use of an asset as well as time. Therefore, we write $\delta = \delta(Y_4)$ with $d\delta/dY = \delta'$ positive and increasing in Y_4. For example if $\delta(Y_4) = \delta_0 + \delta_1 Y_4$, δ_0 is the component of depreciation due to time and δ_1 is the component due to use.

Finally, the implicit rental price of inventories can be derived in a fashion similar to that of physical capital. A proper measure would take account of the composition of inventories between intermediate products and raw materials. The rental price of each component includes purchase price or opportunity (stock) cost, interest and depreciation charges, and a capital gains adjustment. Let c_I denote the user cost of total inventories. Then c_I is a weighted average of rental prices of each component, where weights are equal to value shares of each component as a fraction of total inventory valuation.

ii. The Long-Run Demand Functions

The general problem considered is to minimize costs,

$$C = w_p(Y_1 Y_2) + s_p Y_1 + w_n(Y_6 Y_7) + s_n Y_1 + cY_3 + c_I Y_5,$$

subject to the production function $Q = F(Y_1, \ldots, Y_7)$ and the definitions of the s and c terms. The solution to this problem yields "long-run" demand functions for all inputs.

For expository purposes, consider the following specific example. Assume that $F(Y)$ is Cobb-Douglas, which is a first-order logarithmic approximation to any production function.

$$Q = A \prod_{i=1}^{7} (Y_i)^{\alpha_i}; \ \alpha_i > 0;$$

where the α_i are constants and A is a function of time, capturing exogenous technical change.

The Lagrangian expression for minimizing cost, subject to the production function is

$$C = w_p(Y_1Y_2) + s_pY_1 + w_n(Y_6Y_7) + s_nY_6 + cY_3$$
$$+ c_IY_5 + \lambda(Q - AY_1^{\alpha_1}Y_2^{\alpha_2} \ldots Y_6^{\alpha_6}Y_7^{\alpha_7}).$$

Necessary conditions for minimization are

$$\frac{\partial C}{\partial Y_1} = w_pY_2 + s_p - \lambda\alpha_1\frac{Q}{Y_1} = 0\,.$$

$$\frac{\partial C}{\partial Y_2} = w_pY_1 - \lambda\alpha_2\frac{Q}{Y_2} + Y_1w_p'Y_2 = 0;\ w_p' = \frac{dw_p}{dY_2}\,.$$

$$\frac{\partial C}{\partial Y_3} = c - \lambda\alpha_3\frac{Q}{Y_3} = 0\,.$$

. . . etc.

Eliminating λ from these expressions yields the usual optimality conditions:

$$MFC_1/MP_1 = MFC_2/MP_2 = \ldots = MFC_7/MP_7,$$

along with the production function. The marginal factor cost (*MFC*) of each input divided by its marginal product (*MP*) must be equal in all directions. For labor inputs, this means that marginal production cost at the extensive margin must equal marginal production cost at the intensive margin. At the extensive margin for production worker employment:

$$MFC_1 = w_pY_2 + s_p$$

and for nonproduction worker employment:

$$MFC_6 = w_nY_7 + s_n.$$

The cost of hiring additional employees (to work at optimum hours) is the sum of wage payments (w_pY_2 and w_nY_7) plus amortized hiring and training costs (s_p and s_n). At the intensive margin for production worker "hours":

$$MFC_2 = Y_1\left[w_p + Y_2\left(\frac{dw_p}{dY_2}\right)\right]$$

and for nonproduction worker "hours":

$$MFC_7 = Y_6\left[w_n + Y_7\left(\frac{dw_n}{dY_7}\right)\right].$$

The costs of increasing "hours per man" are the wages that must be paid existing employees to work the additional hour ($Y_1 w_p$ and $Y_6 w_n$) at the original wage per hour plus a correction for the fact that wage rates rise when hours are increased [$Y_1 Y_2(dw_p/dY_2)$ and $Y_6 Y_7(dw_n/dY_7)$]. Marginal factor cost of capital stock (Y_3) and intermediate product (Y_5) are simply c and c_I, respectively, while marginal factor cost c' at the capital-intensive margin is $p_k Y_3(d\delta/dY_4)$, reflecting the increase in depreciation charges on capital stock when utilization is increased. The solution to the necessary conditions defines input demand functions, which are log-linear under the Cobb-Douglas assumption and given by the equation on the next page, where $k_1, k_2, \ldots, k_7$ are constants, parametric on $(Y_2/w_p)w_p'$, $(Y_7/w_n)w_n'$, $\alpha_1, \alpha_2, \ldots, \alpha_7$ and A (and therefore effects of trend via technical change are incorporated in the constants and not written explicitly in this formula). Also $\gamma = \alpha_1+\alpha_3+\alpha_5+\alpha_6$. The factor demand functions may be written in more compact matrix notation as

$$Y^* = k + \xi Q + BR, \tag{2.2}$$

where Y^* is a column vector of ln Y_i^* terms, ξ is a vector of scale effects, B is a matrix of factor price effects, and R is a vector of factor prices. It is apparent from the explicit form of (2.2) that the sum of elements in each row of B is zero. Hence, B is singular, expressing the fact that factor demand functions are homogeneous of degree zero in prices and that demand functions could be expressed equally as well in terms of price ratios.

There are several interesting properties of these solutions:

a. All long-run scale effects are embedded in stock demand functions and not in service flows per unit of stock, since output enters only the demand for stocks (Y_1, Y_3, Y_5, and Y_6), not the demand for utilization rates (Y_2, Y_4, and Y_7). For example, if output doubles and there are constant returns to scale then $\gamma = \alpha_1+\alpha_3+\alpha_5+\alpha_6 = 1$, and all stock variables double, but hours per man and utilization of capital remain unchanged. This is clearly a desirable property of the solutions in view of casual observations we have made of the data. For example, hours per man and the amount of shift working have remained reasonably constant on the average during the post-World War II period of our data (see Chapter 3) in spite of massive secular changes in real output. Utilization rates display considerable variation over the sample period, but these are mainly short-run phenomena, independent of the long-run considerations under

MODEL (2.2)

$$
\begin{bmatrix} \ln Y_1^* \\ \ln Y_2^* \\ \ln Y_3^* \\ \ln Y_4^* \\ \ln Y_5^* \\ \ln Y_6^* \\ \ln Y_7^* \end{bmatrix}
=
\begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \\ k_5 \\ k_6 \\ k_7 \end{bmatrix}
+
\begin{bmatrix} \frac{1}{\gamma} \\ 0 \\ \frac{1}{\gamma} \\ 0 \\ \frac{1}{\gamma} \\ \frac{}{\gamma} \\ 0 \end{bmatrix}
[\ln Q]+
\begin{bmatrix}
\frac{\alpha_2}{\gamma} & \frac{\alpha_1-\alpha_2}{\gamma}-1 & \frac{\alpha_3-\alpha_4}{\gamma} & \frac{\alpha_4}{\gamma} & \frac{\alpha_5}{\gamma} & \frac{\alpha_7}{\gamma} & \frac{\alpha_6-\alpha_7}{\gamma} \\
-1 & +1 & 0 & 0 & 0 & 0 & 0 \\
\frac{\alpha_2}{\gamma} & \frac{\alpha_1-\alpha_2}{\gamma} & \frac{\alpha_3-\alpha_4}{\gamma}-1 & \frac{\alpha_4}{\gamma} & \frac{\alpha_5}{\gamma} & \frac{\alpha_7}{\gamma} & \frac{\alpha_6-\alpha_7}{\gamma} \\
0 & 0 & +1 & -1 & 0 & 0 & 0 \\
\frac{\alpha_2}{\gamma} & \frac{\alpha_1-\alpha_2}{\gamma} & \frac{\alpha_3-\alpha_4}{\gamma} & \frac{\alpha_4}{\gamma} & \frac{\alpha_5}{\gamma}-1 & \frac{\alpha_7}{\gamma} & \frac{\alpha_6-\alpha_7}{\gamma} \\
\frac{\alpha_2}{\gamma} & \frac{\alpha_1-\alpha_2}{\gamma} & \frac{\alpha_3-\alpha_4}{\gamma} & \frac{\alpha_4}{\gamma} & \frac{\alpha_5}{\gamma} & \frac{\alpha_7}{\gamma} & \frac{\alpha_6-\alpha_7}{\gamma}-1 \\
0 & 0 & 0 & 0 & 0 & -1 & +1
\end{bmatrix}
\begin{bmatrix} \ln w_p \\ \ln s_p \\ \ln c \\ \ln c' \\ \ln c_I \\ \ln w_n \\ \ln s_n \end{bmatrix}
$$

discussion here. There are increasing, constant, or decreasing returns to scale in the conventional sense when $\gamma = \alpha_1 + \alpha_3 + \alpha_5 + \alpha_6$ is greater than, equal to, or less than unity.

Scale does not affect utilization rates in this formulation due to the nature of the assumed production function. Consider the marginal conditions for Y_1 and Y_2, $MFC_1/MFC_2 = MP_1/MP_2$. On Cobb-Douglas assumptions we have

$$\frac{w_p Y_2 + s_p}{Y_1(w_p + Y_2 w_p')} = (\alpha_1/\alpha_2)\,(Y_2/Y_1),$$

and Y_1 enters the denominator of both sides and cancels out, leaving an expression determining Y_2 independently of all other variables. If the elasticity of substitution between Y_1 and Y_2 were not unity, then it would not necessarily be true that utilization rate demand functions would be scale-free. Finally, if $F(Y_1, \ldots, Y_7)$ were not homogeneous, scale effects would not be the same in all stock equations, as they are in the Cobb-Douglas formulation.

b. Factor prices affect long-run input demand functions in various ways, which differ from the familiar solutions. Most surprisingly, a relative increase in hourly wage rates increases demand for labor stocks Y_1 or Y_6. However, it decreases demand for labor utilization, Y_2 or Y_7. The reason is that an increase in w relative to s (given output) induces substitution of stock for utilization, since hours per man become relatively more expensive than numbers, pushing out the extensive margin relative to the intensive margin. However, the former effect is not as great as the latter, since an increase in w reduces total man-hours and increases capital services, as usual. For example,

$$\partial \ln (Y_1 Y_2) \,/\, \partial \ln w = \left(\frac{\alpha_2}{\gamma} - 1\right) < 0.$$

On the other hand, an increase in user costs (s_p or s_n) reduces demand for labor stocks, since they become relatively more expensive, and induces substitution of hours per man and capital stock to maintain output. Changes in w and s only affect capital (positively), but do not influence the rate of utilization of capital. Similarly, a relative increase in some component of c induces substitution of capital utilization for capital stock and increases employment without affecting hours per man. An increase in the cost of inventories, c_I, has a negative own effect, but a positive effect on all other stock variables, which does not affect any utilization rate. Again, these conclusions might be altered slightly in a more general formulation.

c. In setting up the problem as one of cost minimization, we concentrate on factor substitution and ignore price-induced scale effects. All relative factor price changes increase marginal cost of output and eventually lead to decreases in desired output, working toward reductions in all inputs. Thus, if the latter effect were to be included, an increase in w_p would increase Y_1 (labor stock) due to substitution, but decrease it due to scale. The net change depends on the magnitudes of both effects.

d. Long-run utilization rates are independent of cross-price effects. Thus, c and c_I do not enter demand functions for Y_2 and Y_7 (labor utilization) nor do w_p, w_n, s_p, and s_n affect the demand for Y_4 (capital utilization). Such independence evidently is due to the assumption of Cobb-Douglas production functions and is not a general consequence of the theory for any production function. However, it is not obvious a priori how the theory predicts signs of cross effects on the usual general assumptions of production functions.

B. SHORT-RUN ADJUSTMENTS

As noted previously, there is no reason to expect firms to be in long-run equilibrium at every point in time. Therefore, time-series data reflect temporary and short-run influences that are not fully captured by any long-run model. Our brief review of the literature suggests that the most satisfactory empirical time-series specification postulates lagged adjustment to some "desired" targets, and we adopt a modified version of that hypothesis.

Specify a log-linear adjustment hypothesis (with all variables measured in natural logarithms):

$$Y_{it} - Y_{it-1} = \sum_{j=1}^{7} \beta_{ij}(Y^*_{jt} - Y_{jt-1}) + \varepsilon_{it};\ i = 1, \ldots, 7; \qquad (2.3)$$

where Y^*_{jt} is the desired or target level of input ($\ln Y_j$) in period t, defined by (2.2), ε_{it} is a random variable, and the β_{ij} are fixed adjustment coefficients. On specification (2.3), $Y^*_{jt} - Y_{jt-1}$ is the proportional divergence between actual and desired input levels at the start of the period under consideration, or relative "excess demand" (or "excess supply" if negative) for factor Y_j. The systematic portion of (2.3)—that is, excluding ε_{it}—asserts that the relative change in each input is proportional to the divergence between desired and actual levels of *all other* inputs as well as to its own level of excess demand. This specification introduces *feedbacks and interrelated adjustments among all factors of production.* For example, suppose that there exists excess demand for production workers (Y_1). If product demand has increased due to a business cycle expansion, disequilibrium in the stock of production workers is likely to call forth extraordinary adjustments in factors such as utilization rates that are more easily altered in the short run. Firms may well find it desirable to increase both hours of work (overtime) and capital utilization, both of which can be altered at less cost in the short run, in order to take up the slack of less than optimal labor stock. The equation system (2.3) is designed to capture all such effects on a symmetrical and internally consistent basis. To elaborate on these complex issues we need to consider the conceptual basis of adjustment processes, some qualifications to these arguments, and the nature of disturbances in the system. These issues are considered in what follows.

i. Costs of Adjustment: Theoretical Considerations

Note that system (2.3) can be derived for a firm from principles of wealth maximization across an infinite horizon, when changes in factors generate "costs of adjustment" (see Lucas [1967]) as mentioned in Chapter 1. Indeed, (2.3) is simply a generalization of the well-known flexible accelerator, or partial adjustment model, and it is appropriate at this point to review the principles underlying its derivation.

Consider a competitive firm selling a single good x, producing with a vector of n inputs y. Let μ be a vector of fixed "depreciation" rates, some of which may be zero; p is product price, and v is a vector of fixed input prices. Let $g(\dot{y}+\mu' y)$ be a function denoting costs of gross changes in inputs, or costs of adjustment, with $dy/dt = \dot{y}$. The function $g(\ )$ represents market search and training costs for such inputs as labor, and installation and "gestation" costs for inputs such as capital. Assume that $g(0) = 0$, and g_i, $g_{ii} > 0$ for $|\dot{y}+\mu y| \neq 0$, and that the derivatives do not change sign at the origin. Further, assume

$$\lim_{z_i \to 0} g_i(z) = 0 \quad \text{and} \quad \lim_{z_i \to \infty} g_i(z) = \infty.$$

Then the present value of the firm is

$$W = \int_0^\infty [px - vy - g(\dot{y} + \mu' y)]e^{-rt}\,dt, \tag{2.4}$$

where r is the rate of interest. Maximization of W, subject to the initial conditions, y_0, and the production function (2.1), $x = F(y)$, requires choosing a set of functions $y_i(t)$, describing optimum input levels at each point in time. Note that its solution requires knowledge of $p(t)$ and $v(t)$, the future course of prices over the horizon. Conditions for maximization of (2.4) are

$$pF_i - v_i = g_i[r + \mu_i - (\dot{g}_i/g_i)]; \; i = 1, 2, \ldots, n;$$
$$\lim_{t \to \infty} g_i e^{-rt} = 0; \tag{2.5}$$

where F_i and g_i are derivatives of F and g with respect to their ith arguments. The first equation of (2.5) must hold at every point in the planning horizon and characterizes the optimum functions $y_i(t)$. In this formulation, marginal value products are not set equal to factor prices for maximization. Instead, marginal value products exceed factor prices by an amount that reflects costs of changing inputs along the optimal path. The term on the right-hand side of (2.5) is simply the marginal stock cost of adjusting inputs, amortized to a periodic flow. Note that this term is not constant along the optimum path, but varies, depending on the actual changes in y. The second equation in (2.5) is a condition guaranteeing that (2.4) has a finite maximum. Given the course of $p(t)$ and $v(t)$ over the horizon and the initial conditions, (2.5) is a simultaneous set of nonlinear differential equations in y. These may be solved for optimal paths of each input, $y_i(t)$, over time. The complete solution to (2.5) is in general very difficult to obtain, and most analyses proceed by linearizing the nonlinear terms F_i and g_i, which amounts to taking second-order or quadratic approximations to the production and cost functions (Eisner and Strotz [1963], Lucas [1967], and Schramm [1970]). Under certain conditions, the solution to the linearized form of (2.5) can be expressed in the form of the systematic portion of (2.3). A crucial condition for derivation of generalized flexible accelerators (2.3) is the constancy of p and w over the horizon (Gould [1968]), an assumption known as "static expectations." In such a case, the process (2.5) converges to the equilibrium

$$pF_i = v_i + \bar{g}_i(r + \mu_i); \; i = 1, 2, \ldots, n; \tag{2.6}$$

where $\bar{g}_i$ is g_i evaluated at $\mu'\bar{y}$, and $\bar{y}$ is the equilibrium value of y.[3]

3. A more general formulation of total costs is $g(y, \dot{y} + \mu' y)$. In (24.), $g_{12} = 0$ since terms in y and $\dot{y} + \mu' y$ are independent. If $g_{12} \neq 0$, and adjustment costs are nonseparable,

Conditions (2.6) are in fact the usual ones for maximum profit under static conditions, namely, that marginal value product equals marginal factor cost. Indeed, the simultaneous solution of (2.6) yields y_i as functions of p and v_i, and these functions define the desired *fixed* targets y_i^* in (2.3) and, in fact, are similar to our equation (2.2). At that point, the firm must be minimizing costs, and an equivalent set of targets can be found by specifying the ordinary factor demand functions [such as (2.2) in section A above] at the stationary value of output that is consistent with the solution to (2.6). Finally, the adjustment coefficients β_{ij} are clearly related to the properties of $g(\dot{y}+\mu y)$. In cases where p and v are not constant, the particular solutions to (2.5) depend on the explicit evolution of prices, and flexible accelerator formulations do not necessarily apply. The reason is apparent from the specification of fixed targets in (2.3). Evidently, the fixed targets arise because p and w are fixed. If p and w are not constant, the targets at which the firm aims undergo change, and (2.3) cannot hold.

ii. Some Qualifications

It is important to point out that the data to be analyzed are generated by *markets* and are aggregate in character. Therefore, micromodels of the firm are not testable in the absence of consideration of market repercussions on firm behavior. In particular, the assumption of static expectations is not tenable when one considers market feedbacks. Consider a unit once-and-for-all shock in the market demand function for output in a competitive industry. Market price, p, immediately rises, since short-run supply is inelastic. If we now attempt to reason along the lines of the model above for a "representative" firm, it will turn out that static price expectations are unrealized at every point in time. Actual price will always turn out to be lower than expected price as industry output rises and price falls along the new demand schedule. Ex post, firm output decisions during the adjustment period will have been too high, and in that sense non-optimal. In the face of such losses, it is probable that learning will occur, and firms will begin to anticipate market reactions more or less correctly. If so, such anticipations by all firms will, in themselves, have repercussions on market prices as all firms react and short-run supply is affected.

---

some of the conclusions in the text are slightly altered. For a discussion of this issue see Treadway [1966] and Nerlove [1971].

Perhaps a concept such as "rational expectations" (Muth [1960]) is appropriate here. But in any event, p cannot be considered as constant over the adjustment period. In addition, it may even be true that attempts by all firms to expand will have repercussions on the price of inputs. Even if wages remain fixed, capital goods prices must certainly change. Firms have access to capital goods either by purchasing other firms or by ordering new equipment from capital goods producers and these supplies are not perfectly elastic. Therefore, the fixed targets hypothesis of cost-adjustment models is untenable in a market setting.

In view of these limitations of cost-adjustment models of the firm for market behavior, some modifications are necessary. Our modifications take the form of retaining the highly suggestive lag specification of (2.3). However, the target and disturbance terms are altered in a manner that captures some of the *industry* variation in factor demand. In particular, we relax the specification of *fixed* targets toward which the system moves.

Our basic view is that there exists *short-run* monopoly power in product and factor markets; that is, during the adjustment process, firms do not regard product and factor prices as fixed parameters, at which they can sell or buy all conceivable quantities of output and inputs, but in fact, exercise some control over their prices. It is only in the long run, under conditions of reasonable stability, that the forces of potential and actual entry and exit of firms force the stringent conditions of competition on markets (Arrow [1959]). The empirical evidence for this view is well known and can be briefly stated. Prices vary, but simply do not fall by as much as might be expected during recessions and do not rise by as much as might be expected during recovery periods (Stigler and Kindahl [1970]). Indeed, money wage rates have hardly fallen at all during the entire postwar period (see Chart 3.8). Since resources cannot move freely in and out of industries in the face of "adjustment costs" and uncertainty about the future, firms must take account of the future reactions of the market and, thereby, of other firms in making their current decisions. This in itself is sufficient to produce a kind of short-run imperfection in the market.[4]

The upshot of this hypothesis is that desired or target factor demands are not wholly a function of current product and factor prices. Instead, the representative firm exercises some of its short-run control, depending

4. Formal models along these lines are given in Phelps [1970].

on the type and extent of disturbances in the market, by stabilizing prices and thereby fixing output, input, and inventory targets.

iii. Characterization of Disturbances

More specifically, (2.3) represents dynamic responses of inputs whose precise movements depend on "shocks" to the system. As noted above, these shocks can best be analyzed in terms of final demand or *sales*. So far, it has been assumed that additions to finished goods inventories are part of current output. Sales, output, and additions to inventories of finished goods are related by the identity $S_t + \Delta I_t = Q_t$, where S is the level of sales and ΔI_t is output inventory investment. If it is true that disturbances arise from fluctuations in the demand for final goods and that marginal production cost is increasing, firms will meet part of their final demand from stocks of finished inventory. This suggests a sales "production function," $S_t = f(Q_t, I_t)$, that allows for the holding of buffer stocks and production smoothing (Lovell [1969]; Holt et al. [1960]). Therefore we may write as an approximation:

$$S_t = H(Y_{it});\ i = 1, \ldots, 7;$$

where all Y_i's are defined as in section A, above, except for Y_5, which now refers to total inventories of both finished and unfinished goods. Thus, total inventory is a decision variable in the model.[5]

It is useful to make a distinction between what are regarded as "permanent" changes and those that are regarded as "temporary" shocks. Such analysis has proven useful in a wide variety of time-series models and may have some payoff for the problem at hand (Eisner [1967] and Friedman [1957]). Accordingly, let $Y_j^* = Y_{jt}^P + Y_{jt}^T$, where Y_{jt}^P is the permanent component, and Y_{jt}^T is the transitory component of the target level of input j. Y_{jt}^P is meant to capture all secular long-run forces that drive the system and that are clearly foreseen by firms. These forces result from growth in the economy in general; secularly increasing demand for industry output, such as population growth; technical change

5. Strictly speaking, inventory decision models are most meaningful in the presence of uncertainty, where prices and sales are random variables (Arrow, Karlin, and Scarf [1958]; Zabel [1967]). We have adopted the present method in order to introduce finished goods inventory into an essentially deterministic system. However, stochastic elements are introduced in the manner discussed below. Marketing and advertising costs might be included, too, but we have not done so.

in the economy; and general capital accumulation. For the most part, Y_{jt}^{p} is to be identified with trend. However, it may also include nontrend extrapolations based on some kind of long-term smoothed average of past experience. This hypothesis suggests that Q_t in expression (2.2) above should be regarded as S_t^p, an unobserved target level of sales that would be maintained under competitive market conditions in a long period of relative stability (Nerlove [1967a]). That is, we envision a relatively fixed long-run supply function in the industry (possibly increasing with respect to price, possibly not) and a demand function that is rising at a rather steady rate, with superimposed disturbances. The long-run rate of growth of demand depends largely on economic growth in general and on the income elasticity of demand for the product in particular.

To see what this entails, suppose there were no random shocks in the economy. Assuming system (2.3) to be stable, it will eventually "damp" down to a long-run equilibrium trend. If the initial conditions are not along the trend line, all inputs converge to their long-run rates of growth (possibly zero, as with Y_2), which in general equals the rate of growth of sales, corrected for returns to scale and technological changes in production of the given industry. Long-run factor proportions will depend on prices, as usual. In such a case, S_t^p will eventually become the observed S_t, rather than an unobserved component, if the data are available for a sufficiently long period of time.

Evidently, long-run trends are not the only forces moving the system. There are transitory shocks as well. In this regard, it is useful to distinguish two types of disturbances. The first are purely random shocks that are strictly uncorrelated with each other over time. Since this is the usual econometric specification, no further elaboration is necessary.

The second, and for present purposes more interesting, type of disturbances that shock the system is identified with certain types of business cycle activity around the long-run trend. It is here that our distinction between short-run monopoly power and long-run competitiveness comes into play. It is clear that since business cycle activity displays substantial serial correlation, some of its components are regular enough to be in part predictable on the basis of past observations. Firms attempt to maintain short-run sales targets on the basis of such predictions. Thus, we specify short-run deviations from the long-run targets based on short-run monopoly power in the product market, as part of the desired target, Y_t^*. As an empirical proposition, and in the face of sticky product prices, these are taken to be related to predicted sales during transitional periods. As will be discussed in detail below, short-run sales targets are specified functions of past observations, new orders, back orders, and so on. Thus, in the relevant "long run" (never observed), the expected values of the short-run deviations from long-run trend targets are zero.

In sum, our specification is: (a) equation (2.3), allowing interactions and feedbacks between factor demand functions; (b) specification (2.2) for the moving long-run target variables in Y_t^*, with S^P replacing output and representing an unobserved "permanent" component of sales closely associated with trend; and (c) an additional specification for short-term deviations in these targets, depending on predicted sales in the immediate future. The model is (in matrix notation):

$$\left.\begin{aligned} Y_t &= \beta Y_t^* + (I - \beta)\, Y_{t-1} + \varepsilon_{1t}; && \text{(a)}\\ Y_t^* &= Y_t^P + Y_t^T; && \text{(b)}\\ Y_t^P &= k + \rho S_t^P + BR_t + \varepsilon_{2t}; && \text{(c)}\\ Y_t^T &= \varphi(Z_t - S_t^P) + \varepsilon_{3t}. && \text{(d)} \end{aligned}\right\} \quad (2.7)$$

All variables are in logarithms: β is the matrix of adjustment coefficients, $\{\beta_{ij}\}$, and I is the identity matrix. The vector of inputs at time t is Y_t, and Y_t^* is the vector of desired target levels of inputs at time t. The second relation partitions targets into permanent and transitory components, as above. In the third relation, Y^P is specified to be a vector of log-linear functions, as in equation (2.2), that depend on the permanent component of sales, S_t^P, and a vector of factor prices R_t. In the empirical work, the singularity restriction on the price response matrix B is imposed by using price ratios in R_t, rather than their absolute level. ρ is a vector of long-run sales elasticities of each input. The fourth relationship specifies the cyclical and temporary but systematic shocks that drive the system, Y_t^T, as linear functions of the difference between predicted sales during transitional periods, Z_t, and the permanent components. φ is a vector of constants. The precise content of Z_t is a matter of empirical judgment, and discussion of it is deferred to a later chapter. Finally, the ε terms are vectors of unsystematic, serially uncorrelated random disturbances, with zero means and finite variances. The hypothesis of time independence is maintained for these variables. Of course, contemporaneous disturbances may, in fact, be correlated across equations, that is, ε_{ijt} and ε_{klt} do not necessarily exhibit zero covariance.

C. RESTRICTIONS ON ADJUSTMENT COEFFICIENTS

On our interpretation of short-run adjustment mechanisms, firms maintain their position along the production surface at every point in time; that is, firms are not "off" their production functions during the adjustment process. Note that this does not mean that such phenomena as labor "hoarding" or temporary excess capacity are not possible, for utilization rates can vary in an opposite way. Moreover, the essence of the adjustment mechanism in (2.7) is that excess demands or supplies of factors "held" by firms exist during the adjustment period. Imposition of production function constraints means that if some excess demands exist, some excess supplies must exist as well, in order to maintain output.

Some examples will clarify the point. Suppose a recession occurs. If changes are so rapid that they are not perfectly foreseen by all firms, it is reasonable to suppose that holdings of capital stock at current rates of output and sales will be greater than would be desirable under stationary conditions at recession rates of output. In this sense there would be excess holdings of capital stock. But the productive capacity of capital depends not only on stock magnitudes, but also on rates of stock utilization. Thus, in the present case, capital utilization declines, producing the recession-induced lower rates of output (this discussion ignores production for inventory). In a similar vein, suppose the firm finds itself in the position of carrying "excess" workers on its payroll, or of hoarding employees. Then, if it were possible properly to measure utilization or intensity of work of these employees, use of a measure of real input would result in a measure of actual current output.

On the basis of these examples, certain relationships between the adjustment coefficients β_{ij} in (2.3) or (2.7) are implied by the constraint. This was also illustrated in the example of Chapter 1, and can be analyzed by considering equation (2.7a) in detail. Repeating it for convenience,

$$Y_t = \beta Y_t^* + (I - \beta)\, Y_{t-1} + \varepsilon_t. \tag{2.7a}$$

In addition, assume the sales production function can be approximated by[6] $S_t = \alpha' Y_t$, where α' is a vector of constants ("sales elasticities" for each component of Y). Substituting the adjustment hypothesis (2.7a) into

6. The need for approximation here arises because sales, output, and inventory changes are linearly related, whereas, on Cobb-Douglas assumptions, output and inputs are related in a log-linear fashion.

the sales production function yields

$$S_t = \alpha'\beta Y_t^* + \alpha'(I - \beta)\,Y_{t-1}.$$

Now, consider the condition

$$\alpha'(I - \beta) = 0, \tag{2.8}$$

or, equivalently, $\alpha' I = \alpha'\beta$, implying certain restrictions on possible values of β_{ij}. Equation (2.8) is in fact a set of n equations relating α_j and β_{ij}. By the equality in the sentence above, it is seen to amount to the condition

$$\sum_j \alpha_j \beta_{ij} = \alpha_i;\ i = 1, \ldots, n;$$

that is, a weighted column sum (over j) of β_{ij}, with weights equal to the sales elasticities, must sum to the appropriate sales elasticity itself. Evidently, since the terms in α_j are all nonzero (otherwise, input j would not be a proper input), equation (2.8) implies that matrix $(I - \beta)$ is singular, or that $|I - \beta| = 0$. It also implies that all elements in any row of $(I - \beta)$ cannot be of the same sign. That is, inputs must react positively to excess demands for some Y_i's and negatively to excess demands for others. In principle, $|I - \hat{\beta}| = 0$, where $\hat{\beta}$ is a matrix of estimated values, provides a test of the production function restriction. We emphasize "in principle" because sampling distributions of the roots of $|I - \hat{\beta}|$ are not readily available. However, these roots are computed as a matter of course in what follows, and the smallest root should give some indication of the restriction. Alternatively, the restrictions on β could be imposed on the estimation procedure at the outset. This alternative is considered in a later section. Notice that restriction (2.8) has an additional implication regarding S_t and Y_t^*. In particular, if (2.8) holds,

$$S_t = \alpha'\beta Y_t^* + \alpha'\varepsilon_t = \alpha' Y_t^* + \alpha'\varepsilon_t. \tag{2.9}$$

The expected value of a weighted sum of the target values of Y^* must equal sales. The meaning of this should be apparent from the discussion of Y^* itself. Recall that Y^* consists of two components: a long-run trend or permanent component Y^P and a short-run cyclical component, Y^T. The latter term reflects transitory deviations from long-run sales; the input targets resulting from cyclical fluctuations are due to short-run monopoly power exercised by firms. Condition (2.9) simply means that the target sales are in fact produced on the average, or that sales forecasts are realized on average. Thus, a component of Y^T changes in the short run to guarantee this to be so. If exogenous disturbances make it desirable for the firm to produce something other than S^P, then Y^T is nonzero to reflect that decision.

D. DYNAMIC PROPERTIES OF THE SYSTEM

Equations (2.7) may be called the structural specification of the model. It is useful to examine the implications of the reduced form to check on the consistency of the structure. In particular, all dynamic models can always be cast in the form of weighted sums of previous values and initial conditions by iteration.

To this end, consider the systematic portion of (2.7a) and assume some vector of initial inputs Y_0. Then by recursion, it follows that

$$Y_t = \beta Y_t^* + (I - \beta)\, \beta Y_{t-1}^* + \ldots + (I - \beta)^{t-1} Y_1^* + (I - \beta)^t Y_0, \qquad (2.10)$$

and Y_t is a weighted sum of all past desired values, Y_t^*, and of the initial condition, Y_0. Analysis of (2.10) consists of a set of conceptual experiments designed to determine dynamic responses to various shocks.

Consider equilibrium properties first. For this purpose assume all values of Y_t^* are constant and equal over time: $Y_1^* = Y_2^* = Y_3^* = \ldots = Y^*$. Then equilibrium requires $Y_t = Y^*$. If the system runs for a sufficiently long period of time, there must come a point where actual values of Y_t settle arbitrarily near Y^* and remain there. Denote the equilibrium value of Y_t by $\bar{Y}$. Then, at equilibrium, $Y_t = Y_{t-1} = \bar{Y}$. Substituting into (2.7a), we have

$$\bar{Y} = \beta Y^* + (I - \beta)\, \bar{Y}$$

or

$$[I - (I - \beta)]\, \bar{Y} = \beta \bar{Y} = \beta Y^*$$

as required.

Next, consider the question of stability. Given equilibrium values of Y^* in our conceptual experiment, will the actual values of Y_t eventually converge to Y^* from any initial condition Y_0 if left to run for a sufficiently long period of time? The answer to this question can be obtained from (2.10). Setting $\{Y_t^*\} = \{Y^*\}$,

$$Y_t = [I + (I - \beta) + (I - \beta)^2 + \ldots + (I - \beta)^{t-1}]\beta Y^* + (I - \beta)^t Y_0. \qquad (2.11)$$

Clearly, the convergence of this series depends on the properties of $(I - \beta)$. As is well known, stability requires $(I - \beta)^t$ to converge to the zero matrix as t approaches infinity. If that is true, then the matrix sum $[I + (I - \beta) + (I - \beta)^2 + \ldots]$ converges to β^{-1}, from which it follows [from (2.11)], that, in the limit, Y_t approaches its equilibrium value, Y^*.

$(I - \beta)$ is similar to a diagonal matrix, $(I - \beta) = P\Lambda P^{-1}$, where Λ is a diagonal matrix of characteristic roots of $(I - \beta)$ and P is a square matrix.[7] Also,

$$(I - \beta)^m = (P\Lambda P^{-1})\,(P\Lambda P^{-1}) \ldots (P\Lambda P^{-1}) = P\Lambda^m P^{-1}$$

Since Λ is diagonal, Λ^m can be expressed as the diagonal elements all raised to the power m. Therefore, as m increases $(I - \beta)^m$ approaches zero if each element of Λ approaches zero, requiring all characteristic roots to lie within the unit circle.

We now know the properties guaranteeing convergence to "equilibrium." An equally important question concerns the speed of response, or properties, of the approach to equilibrium. For this purpose, it is necessary to investigate transient responses of the system, defined as the response to a one-time unit impulse. Consider an initial equilibrium at which $Y_t = \bar{Y}$ and $Y_t^* = Y^*$ and $\bar{Y} = Y^*$. Denote

$$\tilde{Y}_t = Y_t - \bar{Y};$$
$$\tilde{Y}_t^* = Y_t^* - Y^*.$$

Then (2.7a) may be written

$$\tilde{Y}_t = \beta \tilde{Y}_t^* + (I - \beta)\tilde{Y}_{t-1}. \tag{2.12}$$

The transient response of the system is obtained by ana lyzing the conditions $\tilde{Y}_0 = 0$, $\tilde{Y}_1^* = 1$, $\tilde{Y}_2^* = \tilde{Y}_3^* = \ldots = \tilde{Y}_m^* = \ldots = 0$. Iterating (2.12) with these assum ptions yields

$$\tilde{Y}_1 = \beta Y_1^* = \beta;$$
$$\tilde{Y}_2 = \beta(0) + (I - \beta)\,\tilde{Y}_1 = (I - \beta)\,\beta\tilde{Y}_1^* = (I - \beta)\,\beta; \tag{2.13}$$
$$\cdots\cdots\cdots\cdots$$
$$\tilde{Y}_t = (I - \beta)^{t-1}\beta.$$

7. The number of nonzero characteristic roots for Λ equals the rank of $(I - \beta)$. Thus, if the restrictions hold, we know that $(I - \beta)$ is singular and has rank $n - 1$. If they do not hold, it has rank n.

In this experiment, Y_t^* has been increased by unity in the first period and then reduced to its initial value thereafter. Conditions (2.13) show that the effects of this unit impulse are not confined to the first-period response, but are *distributed* over time. Thus, $\tilde{Y}_1$ is the first-period or impact response, $\tilde{Y}_2$ is the second-period response, and so on. From the general form of $\tilde{Y}_t$ in (2.13) stability properties of $(I - \beta)$ guarantee that the effects of the impulse gradually converge to zero. The (normalized) patterns of $\{\tilde{Y}_t\}$ in (2.13) are in fact equivalent to distributed lag patterns found in all lag models, as will be shown in a moment.

It is also interesting to investigate the response to a unit step-function impulse, or once-and-for-all-time shock to the system, rather than to a one-time unit shock. That is, let $Y_1^* = Y_2^* = \ldots = 1$. By the same reasoning as above, responses in each period are simply the sum of the distributed lag effects in (2.13):

$$\tilde{Y}_1 = \beta;$$

$$\tilde{Y}_2 = \beta + (I - \beta)\,\beta;$$

$$\cdots\cdots\cdots\cdots$$

$$\tilde{Y}_t = \beta + (I - \beta)\,\beta + (I - \beta)^2\beta + \ldots + (I - \beta)^{t-1}\beta;$$

which converges either to the step value, 1, itself, or to the new equilibrium level.

E. COMPARISON WITH OTHER MODELS: THEORETICAL CONSIDERATIONS

It is useful to rewrite the system in another way, to facilitate comparison with other models in the literature. "Partial" reduced form expressions may be obtained in which each dependent variable is expressed in terms of lagged desired targets, Y^*, and lagged own values, Y_{it-1}. To simplify the algebra, let L be the lag operator: $LY_t = Y_{t-1}$, $L^2Y_{t-1} = Y_{t-2}$, etc. In this notation, (2.7a) may be rewritten (ignoring stochastic terms) as $[I - (I - \beta)L]Y_t = \beta Y_t^*$, and an equivalent reduced form is

$$Y_t = [I - (I - \beta)\,L]^{-1}\beta Y_t^*. \tag{2.14}$$

Each element of $[I - (I - \beta)L]^{-1}$ is a rational polynomial function of L. In fact, each element is the ratio of two polynomial functions of the lag operator; the parameters depend on the particular values of β_{ij}. The denominator of each of these functions is the determinant $|I - (I - \beta)L|$, or a polynomial function of L, $\Theta(L)$,

$$\Theta(L) = b_0(1 - b_1L - b_2L^2 - \ldots - b_mL^m),$$

where the coefficients b_0, b_1, . . . are functions of β_{ij}. In fact, m, the order of $\Theta(L)$, is $n - 1$ if the production function restrictions apply [$(I - \beta)$ is singular]. Otherwise, it is of order n. Similarly, the numerator of each term in the inverse matrix is another polynomial in L, $\Theta_{ij}(L)$, with

$$\Theta_{ij}(L) = a_{0j} + a_{1j}L + \ldots + a_{n-1j}L^{n-1};\ i = 1, \ldots, n - 1;$$

where a_{0j}, a_{1j}, . . . are also functions of β_{ij}. The order of $\Theta_{ij}(L)$ is $n - 1$ irrespective of the production function restrictions. We have

$$[I - (I - \beta)L]^{-1} = \{\Theta_{ij}(L)\} / \Theta(L). \qquad (2.15)$$

Carrying out the multiplication in (2.14) after substituting (2.15) yields

$$Y_{it} = [\sum_j \Theta_{ij}(L)\beta_{.j}Y^*_{jt}] / \Theta(L);\ i = 1, \ldots, n. \qquad (2.16)$$

Finally, multiplying both sides of (2.16) by $\Theta(L)$ yields an equivalent distributed lag formulation,

$$Y_{it} = \frac{1}{b_0}[\sum_j \Theta_{ij}(L)\beta_{ij}Y^*_{t}] + b_1Y_{it-1} + b_2Y_{it-2} + \ldots + b_mY_{it-m};\ i = 1, \ldots, n. \qquad (2.17)$$

Examination of (2.17) indicates that the structure (2.7), including feedbacks and interactions in the time demand for factors of production, can be reduced to n separate distributed lag functions, in which *only own past values* of the demand for each factor appear as arguments, without feedback effects apparently present. In addition, the arguments of (2.17) other than Y_{it} are current and lagged values of all variables included in the specification of each Y^*_{it}. Thus, equation (2.17) should be familiar as a general version of the commonly assumed distributed lag structure. All equations in (2.17) could be further reduced to infinite distributed lags of all current and past values of the variables included in Y^*. This amounts to what has been set forth above [see equation (2.10) and related discussion], and there is no need to repeat it.

Evidently, distributed lag models of the form (2.17) characterize most of the literature on time-series input demand. For example, the index of i relating to capital stock in (2.17) yields a formulation that is identical in form to neoclassical investment functions (Jorgenson [1963], Eisner and Nadiri [1968], Bischoff [1971]). Existing short-run employment demand function studies (Brechling [1965], Dhrymes [1969]) can also be considered as special cases of (2.17). Thus, the present model integrates these two apparently unrelated branches of the literature, and alternative estimates of such functions are possible. Note, however, that we have taken into account the *cross* and own adjustments in each equation of (2.17). Thus, a substantially different interpretation of the adjustment process is suggested in our model, compared to what exists in the literature.

As one application of the model, we examine the many distributed lag investment and employment functions that display lag distributions with complex roots and implied oscillatory patterns (Griliches and Wallace [1965], Griliches [1967], and Nadiri [1968]). Such results are questionable on economic grounds, if one considers the source of adjustment lags strictly in terms of own lags with no interrelations present. Under the usual interpretation, it is difficult indeed to account for nonmonotonic convergence to new equilibria on the basis of a single equation model. However, (2.17) shows that the adjustment hypothesis embedded in equations (2.7) has a definable interpretation: Each term of $\{\beta_{ij}\}$ is real; and some inputs show a positive reaction to excess demands of factors by the firm, while others show a negative reaction. However, if $(I - \hat{\beta})$ has complex roots, the distributed lag models of (2.17) must display values of b_i in $\Theta(L)$ that also imply complex roots, generating distributed lag patterns that have cyclical components.

One other implication of (2.17) is worth mentioning. Notice that the own-lag terms in each equation *all* have the *same* set of coefficients b_k. This is a well-known property of the reduced form of a system of difference equations. Previous studies of investment behavior have indicated that adjustment lags for demand for capital are of very long duration (Mayer [1960], Jorgenson [1963]). Such findings have been rationalized in terms of very long gestation periods and large costs of adjustment necessary to change productive capital stock. On the other hand, independent investigations of the demand for employment and hours have also found long adjustment lags (Dhrymes [1969], Nadiri [1968]). A priori, logic suggests that the lags in production worker employment should be substantially shorter than for capital, since adjustment costs to the firm are probably smaller. Thus, the long lags estimated for capital and employment have been something of an empirical puzzle. However, if one accepts the basis of the current model, the puzzle disappears. The terms in (2.17) are identical across equations, so that anything producing long lags in the system as a whole tends to produce a long lag for each and every input. Thus, the adjustment process for employment and hours might display long lags, simply because the adjustment for capital—probably the ultimate source of lags—displays long lags. *If the firm is not in long-run equilibrium with respect to capital stock for very long periods, a complementary disequilibrium must appear elsewhere in the system*, if factor demand functions are time-interrelated and firms operate near their production possibility frontiers.

Another major difference between (2.17) and distributed lag formulations of factor demand that exists in the literature relates to the disturbance terms. It is clear that the omitted disturbance terms in (2.17) are complicated weighted averages of all contemporaneous and several lagged values of the disturbances in the original structural equation (2.7). Hence, if the original disturbances in (2.7) are serially independent, those in (2.17) cannot be independent, and techniques of estimation must take that into account. If, on the other hand, the disturbances in (2.17) are merely "tacked on" and taken to be serially independent, the disturbances in (2.7) must be serially correlated. Though the latter formulation is the one implicit in the reduced form estimates of the literature, in this study we adhere to the former view. We do so for two reasons: First, there has been extensive experimentation with single-equation models such as (2.17), and estimated parameters of such models have some undesirable features. Therefore, perhaps some contribution can be made by estimating (2.7) directly, rather than in reduced form. In this way, we hope to obtain additional information and further insights into the adjustment process than are available in reduced form estimates. Secondly, if (2.17) is estimated under the hypotheses of our model, it is necessary to estimate all equations simultaneously, and to impose the restriction that regression estimates of lagged endogenous variables in each equation are identical. This is a difficult and expensive procedure, and is not necessary if (2.7) is estimated directly. Of course, the real question concerns the true properties of disturbance terms in (2.7) and (2.17). Notice, however, that in both forms, lagged values of endogenous variables appear as regressors. It is well known that tests for serial correlation are biased under such circumstances. In any event, estimation of (2.7) seems promising. However, it is true that ordinary least squares estimators are inconsistent in the presence of serially correlated residuals. Hence the estimation procedure must take that possibility into account.