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: Nature of the Problem and Relation to the Literature

Chapter Author: M. Ishaq Nadiri, Sherwin Rosen

Chapter URL: http://www.nber.org/chapters/c1248

Chapter pages in book: (p. 1 - 12)

1

NATURE OF THE PROBLEM AND RELATION TO THE LITERATURE

A. GENERAL CONSIDERATIONS

The theory of the firm shows how the demand for factors of production can be derived from knowledge of production functions and product and factor market conditions. Empirical investigations have pursued two broad lines of development. First is a series of cross-sectional studies involving direct estimation of production and cost functions on the one hand, and estimation of factor demand functions on the other (Nerlove [1967], Nadiri [1970]). Second is a vast literature using time-series data. Here attention has been divided between studies of long-term productivity (e.g., Denison [1962]) and employment of factors of production (e.g., Kuh [1965], Jorgenson [1963]). At least up until recent years, these strands have been pursued more or less independently.

The main conceptual differences between cross-sectional and time-series analyses rests on the assumption that cross-sectional observations largely reflect long-run optimizing behavior, whereas time-series observations do not. Although estimation of long-run profit-maximizing conditions may be appropriate to cross-sectional studies, no such case can be made for time series. Given the presence of large and uncertain variations in final demand and of short-run imperfections in factor and product markets, there is no reason to expect decision makers to maintain "long-run" desired input positions at every point in time. Instead, gradual adjustment to these positions is to be expected. For this reason, many economists use a partial adjustment or flexible accelerator model,

$$y_t - y_{t-1} = \beta(y_t^* - y_{t-1}), \qquad (1.1)$$

where y_t represents the level of an input at time t, y_t^* is the long-run desired level of the input and β is an adjustment coefficient bounded by 0.0 and 1.0. Functions such as (1.1) have been estimated on a wide variety of time-series data for both capital or investment demand (Eisner [1960], Koyck [1954], Hickman [1965]) and demand for labor (Brechling [1965], Ball-St. Cyr [1966], Dhrymes [1969], Ireland-Smyth [1967], Nadiri [1969]). In most of these studies, the economic meaning of the adjustment mechanisms of (1.1) is not explicitly stated, but rests on an intuitive discussion of time delays, delivery lags, installation costs, and so on. By and large, these studies do not explicitly integrate the costs of changing input levels into functions for estimating factor demand; they also treat adjustments of each input separately and independently of adjustments of other inputs.

To set out the major issues and to indicate the potential contribution of the present work, a brief discussion of econometric time-series studies of employment and investment is provided below. Only the main issues are stated; interested readers can explore the details in original sources. A brief discussion of the time-series employment function is presented first, followed by a similar discussion of empirical investment functions. Finally, the framework of this study is illustrated with a simple example of a more general disequilibrium model of factor demand. That model is fully specified in Chapter 2 and estimated in subsequent chapters of this volume.

B. TIME-SERIES EMPLOYMENT MODELS

A great deal of research in the past ten years or so has been devoted to examination of time series of production and employment, especially the behavior of these variables over the course of business cycles. The data reveal that short-term fluctuations in real output tend to be greater in amplitude than corresponding fluctuations in employment. This difference in amplitude gives rise to systematic cyclical fluctuations in measured man-hour productivity. As output falls from its peak, man-hours employed falls less rapidly, causing declines in man-hours productivity near business cycle peaks; during recovery periods, on the other hand, output grows more rapidly than man-hours, with the result that average labor productivity increases. These phenomena, along with apparently sticky money wages, which display little systematic cyclical variability, "account" for corresponding cyclical fluctuations in factor shares over the course of business cycles. Labor's share of total product tends to grow during business cycle contractions and to fall during recovery periods. Part of this is due to systematic changes in labor quality over the cycle, but observation of this behavior has also given rise to various notions of "labor hoarding." It is maintained that firms tend to smooth employment variations over the course of the cycle, to economize on transactions costs involved in the recruitment of labor and in specific investments in their employees; that is, there are costs of adjustment or costs associated with changing employment that make it economical to stabilize employment fluctuations to some extent.

As an empirical matter, most studies of short-term employment behavior use a type of flexible accelerator or stock adjustment model that was widely used in earlier studies of investment behavior. In light of the excellent survey of most of this work by Fair [1969], detailed examination of the differences among all the models is unnecessary here. However, for comparison with the present work, a brief over-all outline and summary is desirable.

The essence of these models might be captured as follows: Write the short-run production function of the form

$$Q_t = AN_t^{\gamma}h_t^{\beta},$$

where Q_t is output in period t, N_t is the number of workers employed in period t, h_t is a labor utilization rate during the period (hours per man), and A is a constant (also possibly a function of time itself, and which includes all factors that are fixed in the short run, such as physical capital). The output elasticities, γ and β, are assumed constant over the sample period. Given a value of Q and a function of h, $w(h)$, that describes how wage rates vary with hours per head, the desired levels of N and h can be solved as a standard problem in cost minimization if, in fact, employment can be changed at no cost and without delay. Let N^* and h^* denote desired quantities. For example, on the Cobb-Douglas specification above, minimization of costs implies (assuming a constant wage)

$$N_t^* = Q^{1/\gamma}w^{a_1}e^{a_2t}.$$

where the multiplicative constant term has been ignored, and a_1 and a_2 are constants. Desired labor stock is proportional to $Q^{1/\gamma}$, with adjustments for variations in the wage rate and for trend, the latter reflecting secular growth of capital and technical change. Then

$$\partial(\ln N)/\partial(\ln Q) = 1/\gamma.$$

On the basis of standard conditions on production functions, it must surely be the case that $\gamma < 1$, for otherwise the marginal product of labor would not decrease with N in the production function and the law of diminishing returns would be violated. Hence, in the short run, a change of one percentage point in output must lead to a change of more than one percentage point in labor stock employed. However, the empirical patterns described above clearly contradict this prediction: When output is rising relative to its trend, labor input does not rise by as much, and when output is falling relative to its trend, labor does not fall by as much. Therefore, at face value, such changes would imply a crude estimate of γ that is greater than unity, or estimated increasing returns to labor in the short run, in contradiction to the accepted theory of the firm.

Recognizing that costs of changing labor might be significant in layoff and hiring decisions, most investigators specify a stock adjustment hypothesis of which the following is an example:

$$N_t/N_{t-1} = (N_t^*/N_{t-1})^{\lambda},$$

with $0 \leq \lambda < 1$ representing the adjustment coefficient in proportional terms. This allows for gradual adjustment of labor stock to its long-run desired value, rather than for instantaneous adjustment. As we shall observe in section D of this chapter, such a hypothesis implies a corresponding adjustment for utilization rates in order to meet the output and production function constraints during the adjustment period.

Substituting the variables determining N^* above into the adjustment hypothesis yields a regression model in which (with all variables other than trend measured in natural logarithms) current employment is regressed on output, wage rates, trend, and lagged employment. Some writers also employ an expectational notation for output rather than output itself; this involves adding lag terms in output to the regression equation. One investigation also allows for vintage effects in the production function, specifying the augmenting effects of new investment on labor productivity by including lagged investment terms in the regression equation as well as the other variables [Dhrymes, 1969]. Thus, the empirical model is of the form

$$\ln N_t = b_0 + b_1 \ln Q_t + b_2 \ln N_{t-1} + \text{other variables},$$

where the "other variables" include all the modifications used in the various specifications. Coefficient b_1 is an estimate of short-run labor-output elasticity; the estimated value $b_1/(1-b_2)$ is identified with $1/\gamma$, or the inverse of the output elasticity of labor stock in the short-run production function. Though estimates of this parameter $(1/\gamma)$ vary from study to study, depending on the precise specification, in most of them, γ is found to be greater than unity, implying increasing returns to labor alone. No really acceptable explanation of this result has yet been provided. Moreover, in most studies, exceptionally long adjustment lags to labor alone (i.e., values of b_2 near unity) have been found, so long, in fact, as to throw doubt on the assumption of fixed capital stock for purposes of "short-run" labor decisions.

One possible explanation for these estimates relates to a very complex adjustment process and the existence of lags elsewhere in the system. If one really takes the input-output production function constraint seriously, there is a real possibility that observed long adjustment delays of labor inputs may be only "sympathetic" reflections of long lags of other inputs, such as capital. Thus, if capital stock is the ultimate source of adjustment delay, all other inputs will reflect those long lags as a matter of course, so that output and sales are maintained over the adjustment period. The small adjustment coefficients estimated in most time-series employment models (i.e., large values of b_2) may not only reflect costly labor adjustments alone, but other adjustment costs as well. Therefore, they have no ready interpretation.

In other words, adjustment lags among inputs may not be independent. Indeed, the main contribution of this study is to specify and estimate interdependent factor demand functions that show distributed lag responses to be systematically time interrelated.

Finally, in a complete dynamic model, all inputs are changing. In such a model, the conditions underlying "short-run" input demand functions can be approximated by conceptual experiments in which some factors are considered "fixed" in the short run. Results of such experiments are reported in Chapter 7 and suggest that increasing returns to labor as estimated from time-series employment studies are due to omission of certain variables such as utilization rates. Large estimates of γ in those investigations should not be considered as returns to labor alone, but are more properly interpreted as short-run returns to both employment *and* capital utilization. Thus, our specification may help resolve an important issue that has arisen in the employment function literature.

C. INVESTMENT MODELS

The theoretical and empirical research on determinants of investment in fixed capital is voluminous and controversial. This is not the place for an in-depth survey and critique of all the issues involved. Therefore, only a narrow range of issues related to the recent quarterly time-series econometric models of investment behavior in U.S. manufacturing industries is summarized below. There are useful detailed surveys of the broader issues and estimates in articles by Eisner-Strotz [1963], Jorgenson-Hunter-Nadiri [1970], and Nadiri [1970]. Nerlove [1967] has also discussed, in a different context, some theoretical aspects of investment modeling.

Perhaps the most important issues discussed in the literature on fixed investment relate to (i) output and interest elasticity of investment; (ii) specification of the correct price of capital services; and (iii) distributed lag properties of capital stock adjustment. There are, of course, many other issues that deserve consideration, but we shall confine our discussion to these topics.

i. Output and Interest Elasticities

Investment functions of an earlier vintage, or those developed by, for example, Eisner [1960], Klein [1951], Meyer and Kuh [1957], and many others, were mainly of the stock adjustment type, in which the acceleration principle was combined with some measures of profitability. Distributed lag concepts of output as a measure of expected or "permanent" sales were often used, especially by Eisner. Others, like Duesenberry [1958] and Meyer and Kuh, included financial variables as measures of risk and profitability in addition to output.

Attempts to incorporate the interest rate as a determinant of investment behavior were unsuccessful because of several statistical and conceptual problems: high multicollinearity among the variables, difficulties in specifying the lag relation between interest rates and investment decisions, the general problem of specifying expectations, and possibly improper identification of interest rates as the rental price of capital (Jorgenson [1963]). In many empirical estimates, the output variables dominated all other kinds, especially the interest rate, leading to the view that investment demand was interest-inelastic.

Meyer and Kuh argued that the investment function is essentially nonlinear; that it differs in various phases of the business cycle, with sales being the dominant exogenous factor in the expansion phase of the cycle, and liquidity factors in the contraction phase. This characterization of investment behavior was further refined by the work of Meyer and Glauber [1964], using quarterly time-series data. Approaching the problem from a different vantage point, Anderson [1964] argued that balance sheet items such as debt-asset ratios and liquidity measures representing risk and portfolio adjustment considerations affect investment behavior. A more fully specified model of interdependent decisions between financial variables and investment in fixed capital has been estimated by Dhrymes and Kurz [1964] using firm data.

Differences in measurement of the variables and absence of comparability of basic data used in these studies make it very difficult to summarize all the empirical results. The general impression one gets from reading them is that the output variable is probably the most significant determinant of investment behavior, while financial variables may also influence investment behavior, albeit with small effect, except possibly in recession periods.

ii. The Neoclassical Model of Investment

Most of the recent discussion in the literature on investment behavior has been stimulated by the pioneering work of Jorgenson and associates (Jorgenson [1963], Jorgenson and Siebert [1968], Jorgenson and Stephenson [1967]). Jorgenson's argument is that substitution parameters have been improperly neglected or ignored in most work on investment behavior. He accepts the widely held specification that demand for capital is a function of output produced, but argues that it is also a function of the relative price of output and capital. Investment itself, then, consists of replacement of depreciating capital and distributed lag adjustment of capital to its equilibrium. Though Jorgenson suggests that, in quarterly data at least, a particular generalization of the techniques used by Chenery and Koyck for estimating distributed lag relations is essential, that question remains open ended.

In this work, the importance of factor prices, especially the cost of capital, is emphasized. Jorgenson points out that the correct measure of cost of capital services is an implicit rental or flow price, taking account of taxes, interest, depreciation, and capital gains or losses over the adjustment period. The empirical formulation of his models contains such measures (excluding capital gains, which may involve specification error in an inflationary economy), and adds distributed lag adjustments to an essentially long-run equilibrium model. Lags are rationalized in terms of institutionally determined states of investment-planning completion (Jorgenson [1963]). The empirical results reported by Jorgenson and associates based on quarterly postwar time-series data for manufacturing industries suggest the following:[1]

a. Investment demand is highly responsive to changes in relative prices, which include policy variables such as the interest rate and taxes.

b. The distributed lag response of investment to changes in its determinants is fairly long, about eight to nine quarters on the average, and there is no response in the first few quarters.

c. The distributed lag structure of investment behavior in each industry has a bell-like shape, implying that gross investment increases at an increasing rate in the short run and then increases at decreasing rates as long-run equilibrium is approached.

These conclusions, however, have recently been questioned. The controversy centers on the technological assumption of a Cobb-Douglas production function and the distributed lag specification of the model. The main issue concerning technology is whether or not the elasticity of investment with respect to relative prices equals unity. If the Cobb-Douglas production function is assumed, then the hypothesis of unitary elasticity of investment follows as a necessary consequence. Eisner and Nadiri [1968], on the assumption of a CES (constant elasticity of substitution) production function and data used by Jorgenson, show that the output elasticity of investment seems to be high—close to unity—and that its price elasticity is very low. This result was recently confirmed by Mayor [1971].

Jorgenson develops a sophisticated rational distributed lag mechanism, which is then added to "an essentially static theory of demand for capital services" (Nerlove [1971]). The lag distribution is interpreted as an unforeseen delivery lag, that is, firms can adjust their capital stock instantaneously but are prevented from realizing the optimal stock because of suppliers' delivery lags. Thus, it can be argued that, like most other investment functions, Jorgenson's model does not explicitly in-

1. The basic Jorgenson model has also been applied to time-series firm data (Jorgenson and Siebert [1968]) and public utility data (Jorgenson and Handel [1971]).

corporate the adjustment hypothesis as an integral part of the dynamic theory of capital accumulation (Nerlove [1971]).

What emerges from these discussions is that the role of price elasticity of investment and its distributed lag properties are unsettled questions. The magnitude of the price response depends on how the rental price is measured, what type of data are used, what sample period is used, what seasonal adjustment procedures are applied, what splicing techniques are used, how aggregated the data are, etc. The nature of the lag structure of investment behavior is an active area of theoretical research in the context of the models that explicitly account for adjustment costs.

With the possible exception of some of Jorgenson's writings, there is very little treatment of macromarket models in the literature. Most writers dealing with adjustment cost models concentrate on micro behavior and ignore the repercussions of these decisions on the capital goods market and the valuation of capital. It is implicitly assumed that the unit price of capital goods is fixed. Such an assumption would be tenable only in the unlikely event that the supply of investment goods (regardless of length of "run") were perfectly elastic. In all other cases, exogenous disturbances that affect the profitability of firms must generate changes in capital values, at least over some short intervals of time, and these are bound to influence actual behavior. Though Jorgenson does stress the inclusion of capital gains and losses in his (ideal) measure of implicit prices of capital services, the term is ignored in the empirical implementation of his model.

The study of investment behavior has been revolutionized in recent years by the direct integration of dynamic adjustment costs into the theory of the firm, as opposed to the simple grafting of a dynamic adjustment hypothesis such as (1.1) onto an essentially static theory. In a remarkable paper, Eisner and Strotz [1963] demonstrated that the flexible accelerator model actually could be derived directly from first principles of optimum accumulation. In their model, investment costs were specified as a quadratic function of the rate of investment. Thus, firms would not find it profitable to adjust instantaneously to long-run equilibrium, because of the increasing marginal costs of doing so. Instead, they would find it optimal to adjust slowly and distribute the adjustment costs over time, as shown by (1.1). This model was generalized, by Lucas [1967], Gould [1968], Treadway [1966], and Chetty and Sankar [1967], to include several factors of production. By examining approximations to first-order conditions along the paths of optimum accumulation, they derived generalized flexible accelerator models. Little empirical analysis along these lines has been reported, except by Schramm [1970], who assumed quadratic profit and cost functions.

Though the question of price expectations was seldom considered very crucial in the original versions of the accelerator model, the introduction of adjustment costs into the firm's decisions at the outset makes it imperative to treat both the expectations and the optimizing problem. In deriving flexible accelerators for the firm, most authors assume static price expectations. Gould [1968] has shown that under the more reasonable assumption of nonstatic price expectations current decisions depend on the entire future course of prices, and that characterization of optimum paths by flexible accelerator approximation is not generally possible.

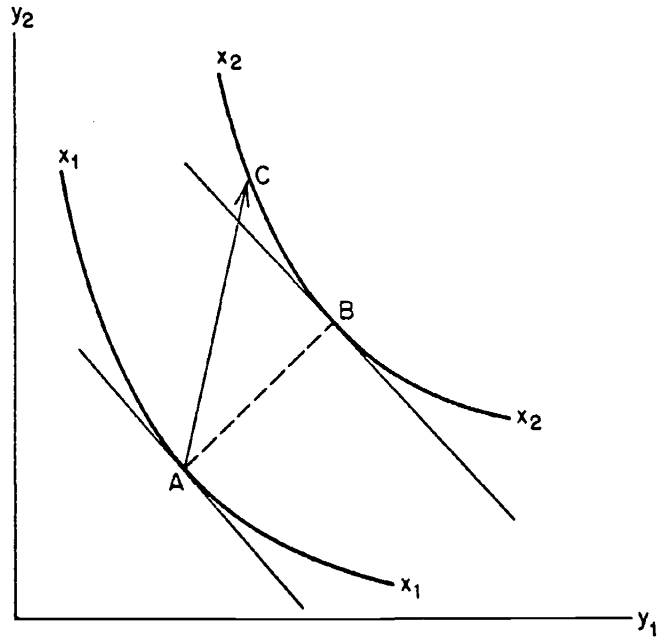
In this work, we attempt to take account explicitly of adjustment costs of several inputs together, and jointly estimate an entire set of input demand functions that are mutually consistent and generated by a unified underlying structure. Before turning to a detailed examination of the model, we will illustrate with a simple example the nature of dynamic interactions among time paths of inputs and set up the more general discussion that will follow. We hope to demonstrate that the model provides a rationale for the high estimated output elasticity of labor input. It also provides new evidence on the price and output elasticity of investment and on their distributed lag properties.

D. AN EXAMPLE

Suppose the production function is $x = f(y_1, y_2)$, where x is output and the y_i $(i = 1, 2)$ are inputs, with f displaying the usual continuity properties. Two isoquants are illustrated in Figure 1.1. The points A and B, derived in the usual way, represent efficient input combinations at which total costs are minimized. Though this may be an adequate description of long-run behavior, there is plenty of evidence to suggest that firms do not remain at points such as A and B at every moment of time. Since it is assumed that the changing of input levels is costly, some kind of partial adjustment model is called for. The conventional way of incorporating such lags is the partial adjustment model (1.1).

Suppose that for some reason the firm desires to increase output in Figure 1.1 from x_1 to x_2, given initial condition A. Then if factor prices are defined correctly, the long-run target or stationary values of the inputs (y_1^*, y_2^*), are given by point B. Consider the partial adjustment

FIGURE 1.1



mechanism for input y_1:

$$y_{1t} - y_{1t-1} = \beta(y_1^* - y_{t-1}). \tag{1.1'}$$

Equation (1.1′) implies an immediate move from A to (say) C, with convergence along isoquant x_2 to the new stationary point B. Therefore, given the production function and hypothesis (1.1′), an adjustment path for y_2 is *automatically implied.*

To illustrate, suppose f is Cobb-Douglas, i.e., $x = Ay_1^a\, y_2^b$. Taking logs and rearranging,

$$\ln y_{2t} = -\frac{1}{b}\ln A + \frac{1}{b}\ln x_t - \frac{a}{b}\ln y_{1t}.$$

Using a log-linear form of (1.1′), $\ln (y_{1t}/y_{1t-1}) = \beta \ln (y_1^*/y_{1t-1})$ and substituting for y_{1t} in the expression above:

$$\ln y_{2t} = \frac{1}{b}\ln x_t - \frac{a}{b}\left[\beta \ln y_1^* + (1-\beta)\ln y_{1t-1}\right] - \frac{1}{b}\ln A$$

$$= \frac{1}{b}\ln x_t - \frac{a}{b}\beta \ln y_1^* - a\frac{(1-\beta)}{b}\ln y_{1t-1} - \frac{1}{b}\ln A.$$

In general, two *independent* hypotheses such as (1.1′) for both y_1 and y_2 imply an additional hypothesis concerning the role of the production function during the adjustment period. There are many possibilities: If the production function always holds as an equality, independent adjustments imply an output decision rule. But in this case, there is no reason to expect that to be optimum. Moreover, in estimating (1.1′) and its counterpart for y_2, the production function parameters are overidentified.

If output is taken to be exogenous, two independent adjustments imply that firms are "off" their production functions and capable of producing more than they actually do during the adjustment period. In such a case, some dimensions in the measurements of y_1 and y_2, such as utilization rates, must be unmeasured and less than potential.

It is apparent then that a very general specification, allowing firms to remain on their production functions at every point in time given very different adjustment costs for inputs, is the natural generalization of (1.1):

$$\begin{bmatrix} y_{1t} - y_{1t-1} \\ \\ y_{2t} - y_{2t-1} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_1^* - y_{1t-1} \\ \\ y_2^* - y_{2t-1} \end{bmatrix} \tag{1.2}$$

In (1.2), the production function constraint can be satisfied, since input adjustments are *not* independent and include cross adjustments or feedback effects β_{12} and β_{21}. Reconsider Figure 1.1. If the true path is described by ACB, β_{21} and β_{22} must be sufficiently greater than zero to push y_2 initially above its ultimate value, y_2^*. From period 1 onward, there is "excess supply" of y_2, $(y_2^* - y_{2t-1}) < 0$, and "excess demand" for y_1, $(y_1^* - y_{1t-1}) > 0$. Necessarily, y_2 overshoots its target to maintain the output level if y_1 is slowly being adjusted and then y_2 slowly adjusts down to y_2^* as y_1 is increased to y_1^*. Obviously, there must be restrictions on β_{ij} to insure that the firm remains along the isoquant.[2]

This example makes clear the importance of cross-adjustment mechanisms among inputs. The omission of these effects in previous studies has been a serious defect, and we hope to remedy it in our work.

2. A related point has been made by Brainard and Tobin [1968] in the context of portfolio adjustments among assets, though the constraints in their case are simpler. The sum of all asset values must add up to total wealth at every point in time.