Volume Title: Determinants of Investment Behavior

Volume Author/Editor: Robert Ferber, ed.

Volume Publisher: UMI

Volume ISBN: 0-87014-309-3

Volume URL: http://www.nber.org/books/ferb67-1

Publication Date: 1967

Chapter Title: The Theory of Investment Behavior

Chapter Author: Dale Jorgenson

Chapter URL: http://www.nber.org/chapters/c1235

Chapter pages in book: (p. 129 - 188)

The Theory of Investment Behavior

DALE W. JORGENSON

UNIVERSITY OF CALIFORNIA AT BERKELEY

1. Introduction

Business investment behavior is one of the areas of modern economic research that is being studied most intensively; empirical studies are accumulating rapidly,[1] and at the same time important developments

[1] A very detailed review of the literature through 1960 has been provided by R. Eisner and R. Strotz, "The Determinants of Business Investment," in D. B. Suits, *et al., Impacts of Monetary Policy*, Englewood Cliffs, 1963, pp. 60–338. A more concise review of developments through 1962 has been presented by E. Kuh, "Theory and Institutions in the Study of Investment Behavior," *American Economic Review,* May 1963, pp. 260–268. Empirical studies published since 1962 include: S. Almon, "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica*, January 1965, pp. 178–196; W. H. L. Anderson, *Corporate Finance and Fixed Investment,* Boston, 1964; A. Bourneuf, "Investment, Excess Capacity, and Growth," *American Economic Review,* September 1964, pp. 607–625; R. Eisner, "Investment: Fact and Fancy," *American Economic Review,* May 1963, pp. 237–246; Eisner, "Capital Expenditures, Profits, and the Acceleration Principle," *Models of Income Determination*, Studies in Income and Wealth 28, Princeton University Press for National Bureau of Economic Research, 1964, pp. 137–176; Eisner, "Realization of Investment Anticipations," in J. S. Duesenberry, E. Kuh, G. Fromm, and L. R. Klein, eds., *The Brookings Quarterly Econometric Model of the United States*, Chicago, 1965; E. Greenberg, "A Stock-Adjustment Investment Model," *Econometrica*, July 1964, pp. 339–357; B. Hickman, *Investment Demand and U.S. Economic Growth,* Washington, 1965; D. W. Jorgenson, "Capital Theory and Investment Behavior," *American Economic Review,* May 1963, pp. 247–259; Jorgenson, "Anticipations and Investment Behavior," in *Brookings Quarterly Econometric Model;* E. Kuh, *Capital Stock Growth: A Micro-Econometric Approach,* Amsterdam, 1963; J. R. Meyer and R. R. Glauber, *Investment Decisions, Economic Forecasting and Public Policy*, Boston, 1964; G. J. Stigler, *Capital and Rates of Return in Manufacturing Industries,* Princeton for NBER, 1963.

in the economic theory of investment behavior are taking place.[2] As yet, there is very little common ground between the empirical and theoretical approaches to this subject. From a certain point of view this is a desirable state of affairs.[3] Econometric studies of investment behavior date back no more than thirty years.[4] Only recently have data on investment expenditures suitable for analysis by econometric methods become available. If empirical studies are forced prematurely into a theoretical straitjacket, attention may be diverted from historical and institutional considerations that are essential to a complete understanding of investment behavior. On the other hand, if theoretical work is made to conform to "realistic" assumptions at too early a stage in the development of empirical work, the door may be closed to theoretical innovations that could lead to improvements in empirical work at a later stage.

While there is some surface plausibility in the view that empirical and theoretical research are best carried out in isolation from each other, this view is seriously incomplete. Econometric work is always based on highly simplified models. The number of possible explanations of investment behavior, which is limited only by the imagination of the investigator, is so large that, in any empirical investigation, all but a very few must be ruled out in advance. Insofar as the necessary simplifications restrict the possible explanations of investment behavior, these simplifications constitute, at least implicitly, a theory of investment behavior. Such theories can be compared with each other most expeditiously by reducing each to its basic underlying assumptions, after which empirical tests to discriminate among alternative theories can be

[2] See, for example, the following papers: K. J. Arrow, "Optimal Capital Policy, The Cost of Capital, and Myopic Decision Rules," *Annals of the Institute of Statistical Mathematics,* 1964, pp. 21–30; "Optimal Capital Adjustment," in K. J. Arrow, S. Karlin, and H. Scarf, eds., *Studies in Applied Probability and Management Science,* Stanford, 1962; K. J. Arrow, M. Beckmann, and S. Karlin, "Optimal Expansion of the Capacity of the Firm," in K. J. Arrow, S. Karlin, and H. Scarf, eds., *Studies in the Mathematical Theory of Inventory and Production,* Stanford, 1958; A. S. Manne, "Capacity Expansion and Probabilistic Growth," *Econometrica,* October 1961, pp. 632–649; E. Zabel, "Efficient Accumulation of Capital for the Firm," *Econometrica,* January-April 1963, pp. 131–150; and the following books: T. Haavelmo, *A Study in the Theory of Investment,* Chicago, 1960; F. A. Lutz and D. G. Hague, eds., *The Theory of Capital,* London, 1961; P. B. D. Massé, *Optimal Investment Decisions,* Englewood Cliffs, 1962; V. L. Smith, *Investment and Production,* Cambridge, 1961; B. Thalberg, "A Keynesian Model Extended by Explicit Demand and Supply Functions for Investment Goods," *Stockholm Economic Studies,* Pamphlet Series, No. 3, 1964.

[3] This point of view has been put forward by K. Borch, "Discussion," *American Economic Review,* May 1963, pp. 272–274.

[4] J. Tinbergen, *Statistical Testing of Business Cycle Theories,* Part I, "A Method and its Application to Investment Activity," Geneva, 1939.

designed. Far from forcing empirical studies into a theoretical straitjacket, judicious use of a theoretical framework is essential to the proper direction of empirical work.

The view that theoretical and empirical research should be carried out in isolation is incomplete in a second respect. The use of economic theory as a source of possible explanations for investment behavior frees econometric work from reliance on empirical generalizations that have not been subjected to rigorous econometric tests. There is a very real danger that econometric models of investment behavior may be made to conform prematurely to assumptions that are "realistic" by the standards of empirical work not based on econometric methods. Just as premature reliance on "realistic" assumptions may be stultifying to the development of economic theory, so reliance on historical and institutional generalizations may restrict the development of econometric models unduly. The paramount test for "realism" of an econometric model is its performance in econometric work. If a model does not perform satisfactorily by the standards of econometrics, it must be rejected, however closely it parallels historical and institutional accounts of the same economic behavior.

The point of departure for this paper is that progress in the study of investment behavior can best be made by comparing econometric models of such behavior within a theoretical framework. Ideally, each model should be derived from a common set of assumptions about the objectives of the business firm. Differences among alternative models should be accounted for by alternative assumptions about the behavior of business firms in pursuing these objectives. It will undoubtedly be surprising to some that a theoretical framework is implicit in the econometric models of investment behavior currently under study. The objective of this paper is to make this framework explicit in order to provide a basis to evaluate evidence on the determinants of investment behavior. This objective can only be attained by a thoroughgoing reconstruction of the theory of investment. Once the theory of investment is placed in a proper setting, the arguments advanced for pessimism about combining theoretical and empirical work largely evaporate.

In providing a framework for the theory of investment behavior, the first problem is to choose an appropriate basis for the theory. Two alternative possibilities may be suggested. First, the theory of investment could be based on the neoclassical theory of optimal capital accumulation. There are three basic objections to this possibility, the first of which is that a substantial body of noneconometric work on the motivation of business firms, mainly surveys of businessmen, suggests that "marginalist" considerations are largely irrelevant to the making of business decisions. This evidence has been subjected to careful scrutiny by White,[5] who concludes that the data accumulated by the surveys are so defective, even by the standards of noneconometric empirical work, that no reliance can be placed on conclusions based on them. A second objection is that previous attempts to base the study of investment on neoclassical economic theory have been unsuccessful,[6] but this argument will not withstand critical scrutiny. First, none of the tests of the neoclassical theory reported in the early literature was based on a fully rigorous statement of the theory. Secondly, the assumptions made about the lag between changes in the demand for capital services and actual investment expenditures were highly restrictive. Frequently, the lag was assumed to be concentrated at a particular point or to be distributed over time in a very simple manner. Tests of the neoclassical theory were carried out prior to the important contribution of Koyck to the analysis of distributed lags and investment behavior.[7] Despite these deficiencies, the pioneering tests of the neoclassical theory reported by Tinbergen reveal substantial effects for the price of investment goods, the change in this price, and the rate of interest.[8] Similarly, tests reported by Roos reveal substantial effects for the price of investment goods and rate of interest.[9] Klein's studies of investment in the railroad and electric power industries reveal substantial effects for the rate of interest.[10]

A third and more fundamental objection has recently been restated by Haavelmo, who argues that a demand schedule for investment goods cannot be derived from neoclassical theory:[11]

What we should reject is the naive reasoning that there is a demand schedule for investment which could be derived from a classical scheme of producers'

[5] W. H. White, "Interest Inelasticity of Investment Demand," *American Economic Review,* September 1956, pp. 565–587.

[6] J. Meyer and E. Kuh, *The Investment Decision,* Cambridge, Mass., pp. 7–14.

[7] L. M. Koyck, *Distributed Lags and Investment Analysis,* Amsterdam, 1954.

[8] Tinbergen, *Statistical Testing,* see also the discussion of Tinbergen's results by T. Haavelmo, "The Effect of the Rate of Interest on Investment: A Note," *Review of Economic Statistics,* February 1941, pp. 49–52.

[9] C. F. Roos and V. S. Von Szeliski, "The Demand for Durable Goods," *Econometrica,* April 1943, pp. 97–122; Roos, "The Demand for Investment Goods," *American Economic Review,* May 1948, pp. 311–320; Roos, "Survey of Economic Forecasting Techniques," *Econometrica,* October 1955, pp. 363–395.

[10] L. R. Klein, "Studies in Investment Behavior," in *Conference on Business Cycles,* New York, National Bureau of Economic Research, 1951.

[11] Haavelmo, *Theory of Investment,* p. 216.

behavior in maximizing profit. The demand for investment cannot simply be derived from the demand for capital. Demand for a finite addition to the stock of capital can lead to any rate of investment, from almost zero to infinity, depending on the additional hypothesis we introduce regarding the speed of reaction of capital-users. I think that the sooner this naive, and unfounded, theory of the demand-for-investment schedule is abandoned, the sooner we shall have a chance of making some real progress in constructing more powerful theories to deal with the capricious short-run variations in the rate of private investment.

We will show that it is possible to derive a demand function for investment goods based on purely neoclassical considerations. While it is true that the conventional derivation of such a demand schedule, as in Keynes' construction of the marginal efficiency of investment schedule,[12] must be dismissed as naive, there is a sense in which the demand for investment goods can be taken to depend on the cost of capital; such a theory of investment behavior can be derived from the neoclassical theory of optimal capital accumulation.

A second possible basis for the theory of investment is the assumption that business firms maximize utility defined more broadly than in the characterization of objectives of the firm in the neoclassical theory of optimal capital accumulation. This basis has been suggested by Meyer and Kuh:[13]

Partial recognition of institutional changes has led in recent years to shift the theory of the firm, and consequently of plant and equipment investment, from a profit maximization orientation to that of utility maximization. Primarily, this move represents a growing belief that profit maximization is too narrow to encompass the full scope of modern entrepreneurial motives, particularly once the previously assumed objective conditions are released from *ceteris paribus*, and the theory seeks to explain a much wider range of behavioral responses.

This position has recently been supported with much force by Simon: ". . . I should like to emphasize strongly that neither the classical theory of the firm nor any of the amendments to it or substitutes for it that have been proposed have had any substantial amount of empirical testing. If the classical theory appeals to us, it must be largely because it

[12] J. M. Keynes, *The General Theory of Employment, Interest and Money*, New York, 1936, esp. Chapter 11, pp. 135–146.

[13] Meyer and Kuh, *Investment Decision*, p. 9.

has a certain face validity . . . rather than because profit maximizing behavior has been observed."[14]

In putting forward this view, Simon ignores the entire econometric literature on cost and production functions, all of which is based on the neoclassical theory of the firm. A recent survey of this literature by Walters[15] enumerates 345 references, almost all presenting results of econometric tests of the neoclassical theory of the firm which are overwhelmingly favorable to the theory. The evidence is largely so favorable that current empirical research emphasizes such technical questions as the appropriate form for the production function and the appropriate statistical specification for econometric models of production based on this theory. We conclude that Simon's statement that the alternatives to the neoclassical theory of the firm have had no substantial amount of empirical testing is correct. However, his characterization of the empirical evidence on the neoclassical theory is completely erroneous.

One possible reaction to a proper assessment of the support for the neoclassical theory of the firm from econometric studies of cost and production functions is to reject out of hand studies of investment behavior not based explicitly on the neoclassical theory, such as the study of Meyer and Kuh. In fact, the theoretical basis for the econometric model of investment behavior proposed by Meyer and Kuh is consistent with the neoclassical theory of optimal capital accumulation. Their appeal to a less narrow view of entrepreneurial objectives is not essential to the interpretation of the empirical results they present. We conclude that the objections to the neoclassical theory of the firm as a basis for the theory of investment behavior are ill-founded. Furthermore, the appeal to a broader view of entrepreneurial objectives than that which underlies this theory is not required by evidence either from econometric studies of cost and production functions or from studies of investment behavior. The neoclassical theory of optimal accumulation of capital is a far more powerful theory than the "broader view" suggested by Simon and others in the sense that a much narrower range of conceivable behavior is consistent with it than with the amorphous utility-maximizing theory. Accordingly, we will employ a theoretical framework based on the neoclassical theory of the firm for constructing a theory of investment behavior.

[14] H. A. Simon, "New Developments in the Theory of the Firm," *American Economic Review,* May 1962, p. 8.

[15] A. A. Walters, "Production and Cost Functions: An Econometric Survey," *Econometrica,* April 1963, pp. 1–66.

The objective of explaining investment behavior on the basis of the neoclassical theory of the firm cannot be described as novel. This objective is clearly in evidence in Tinbergen's pioneering monograph, *Statistical Testing of Business Cycle Theories*. Subsequently, a similar objective was adopted by Roos and by Klein.[16] In these early studies of investment behavior, the neoclassical theory was employed to provide a list of possible explanatory variables for investment expenditures. The rate of interest, the level of stock prices, the price of investment goods, and changes in the price of investment goods were used along with other variables such as profits, output, and changes in output. Little attention was paid to the manner in which the rate of interest and the price of investment goods enter the demand for capital services or the demand for investment goods. Both variables enter only through the *user cost* of capital services.[17] There is no effect of the price of investment goods except in combination with the rate of interest and vice versa. We conclude that, although the objective of explaining investment behavior on the basis of the neoclassical theory of the firm is not new, this objective remains to be fully realized.

2. *The Neoclassical Framework*

In formulating a theory of investment behavior based on the neoclassical theory of optimal capital accumulation, a great number of alternative versions of the theory could be considered. Reduced to its barest essentials, the theory requires only that capital accumulation be based on the objective of maximizing the utility of a stream of consumption. This basic assumption may be combined with any number of technological possibilities for production and economic possibilities for transformation of the results of production into a stream of consumption. In selecting among alternative formulations, a subsidiary objective must be borne in mind. The resulting theory of capital accumulation must include the principal econometric models of investment behavior as specializations,

[16] See footnotes 9 and 10. See also L. R. Klein, *The Keynesian Revolution,* New York, 1947, esp. pp. 62–68, pp. 196–199; Klein, "Notes on the Theory of Investment," *Kyklos,* vol. 2, Fasc. 2, 1948, pp. 97–117; Klein, *Economic Fluctuations in the United States, 1921–1941,* New York, 1950, esp. pp. 14–40.

[17] A complete discussion of the concept of user cost has been given by W. A. Lewis, "Depreciation and Obsolescence as Factors in Costing," in J. L. Meij, ed., *Depreciation and Replacement Policy,* Amsterdam, 1961, pp. 15–45. See also Keynes, *General Theory,* pp. 66–73; A. P. Lerner, "User Cost and Prime User Cost," *American Economic Review,* March 1943, pp. 131–132; F. A. Lutz and V. Lutz, *The Theory of Investment of the Firm,* Princeton, 1951; A. D. Scott, "Notes on User Cost," *Economic Journal,* June 1953, pp. 364–384.

but the theory need not encompass possibilities for the explanation of investment behavior not employed in econometric work.

The essentials of a theory of optimal capital accumulation that meets this basic objective are the following: The firm maximizes the utility of a consumption stream subject to a production function relating the flow of output to flows of labor and capital services. The firm supplies capital services to itself through the acquisition of investment goods; the rate of change in the flow of capital services is proportional to the rate of acquisition of investment goods less the rate of replacement of previously acquired investment goods. The results of the productive process are transformed into a stream of consumption under a fixed set of prices for output, labor services, investment goods, and consumption goods. These prices may be considered as current or "spot" prices together with forward prices for each commodity or, alternatively, as current and future prices together with a normalization factor, which may be identified with current and future values of the rate of time discount or interest rate. Both current and forward prices are taken as fixed by the firm. Alternatively, current and future prices together with current and future values of the rate of interest are taken as fixed. Under these conditions, the problem of maximizing utility may be solved in two stages. First, a production plan may be chosen so as to maximize the present value of the productive enterprise. Secondly, consumption is allocated over time so as to maximize utility subject to the present value of the firm. In view of our concern with the theory of business investment behavior, we will consider only the first of these problems. It should be noted that, under the assumption of fixed prices, the choice of a production plan is independent of the subsequent allocation of consumption over time. Two firms with different preferences among alternative consumption streams will choose the same plan for production.

This version of the neoclassical theory of the firm is not the only one available in the literature on capital theory. From a certain point of view, the objective of maximizing the present value of the firm is only one among many possible objectives for the firm. In a recent survey paper on the theory of capital, Lutz remarks that "It is one of the surprising things about capital theory that no agreement seems to have been reached as to what the entrepreneur should maximize."[18] Alternative criteria discussed in the literature include maximization of the average internal rate of return, maximization of the rate of return on capital owned by the firm, investment in any project with an internal

[18] F. A. Lutz, "The Essentials of Capital Theory," in Lutz and Hague, *Theory of Capital,* p. 6.

rate of return greater than the ruling market rate of interest, and so on. None of these criteria can be derived from maximization of the utility of a stream of consumption under the conditions we have outlined. Maximization of the present value of the firm is the only criterion consistent with utility maximization. This approach to the theory of optimal capital accumulation was originated by Fisher and has recently been revived and extended by Bailey and by Hirshleifer.[19] The essential justification for this approach is summarized by Hirshleifer, as follows:

> Since Fisher, economists working in the theory of investment decision have tended to adopt a mechanical approach—some plumping for the use of this formula, some for that. From a Fisherian point of view, we can see that none of the formulas so far propounded is universally valid. Furthermore, even where the present-value rule, for example, is correct, few realize that its validity is conditional upon making certain associated financing decisions as the Fisherian analysis demonstrates. In short, the Fisherian approach permits us to define the range of applicability and the short-comings of all the proposed formulas—thus standing over against them as the general theoretical solution to the problem of investment decision under conditions of certainty.[20]

A second controversial aspect of the version of the neoclassical theory outlined above is the assumption that the set of technological possibilities confronted by the firm can be described by a production function, where the flow of output is a function of flows of labor and capital services and the flow of capital services is proportional to the stock of capital goods obtained by summing the stream of past net investments.[21] The concept of capital service is not essential to the neoclassical theory. A production function relating output at each point of time to inputs of labor and capital services at that point of time may be replaced by a production function relating output at every point of time to inputs of investment goods at every point of time; this description of the set of

[19] I. Fisher, *The Theory of Interest*, New York, 1930. M. J. Bailey, "Formal Criteria for Investment Decisions," *Journal of Political Economy*, October 1959, pp. 476–488. J. Hirshleifer, "On the Theory of the Optimal Investment Decision," in E. Solomon, ed., *The Management of Corporate Capital*, Glencoe, 1959, pp. 205–228.

[20] *Ibid.*, p. 228.

[21] For a discussion of this assumption and some of its implications, see J. Robinson, "The Production Function and the Theory of Capital," *Review of Economic Studies*, Vol. 21, No. 54, 1953–54, pp. 81–106; R. M. Solow, "The Production Function and the Theory of Capital," *Review of Economic Studies*, Vol. 23, No. 61, 1955–56, pp. 101–108; J. Robinson, "Reply," *Review of Economic Studies*, Vol. 23, No. 62, 1955–56, p. 247; J. Robinson, "Some Problems of Definition and Measurement of Capital," *Oxford Economic Papers*, June 1959, pp. 157–166; K. J. Arrow *et al.*, "Symposium on Production Functions and Economic Growth," *Review of Economic Studies*, June 1962.

production possibilities is employed by Fisher; moreover, it may be characterized abstractly so that even the notion of a production function may be dispensed with, as is done by Malinvaud.[22] The description of the set of technological possibilities by means of a production function as presented by Fisher is a specialization of the description given by Malinvaud. The further assumption that the relationship between inputs of investment goods and levels of output may be reduced to a relationship between output at each point of time and a corresponding flow of capital services involves a specialization of the description of technological possibilities given by Fisher.

In the neoclassical literature, two basic models of the relationship between flows of investment goods and flows of capital services have been discussed, namely, a model of inventories and a model of durable goods. At the level of abstraction of Fisher's description of the set of production possibilities, no distinction between inventories and durable goods is required. For both inventories and durable goods, the acquisition of a stock of productive goods may be represented as an input to the productive process at the time of acquisition. For inventories, the individual items "used up" at different points of time may be represented as the output of a subprocess representing the holding of stocks; these outputs may be inputs into other subprocesses. For durable goods, the outputs of the corresponding stockholding process are the services of the goods rather than the individual items of the stock; the services of the durable goods may be inputs into other parts of the productive process.

The basis for the distinction between inventories and durable goods lies in the relationship among the initial input and the various outputs from the stockholding process. For inventories, the outputs provided by the stockholding process are customarily treated as perfect substitutes. For each item held in stock, the ultimate consumption of that item can occur at one and only one point in time. By contrast, the outputs provided by durable goods are treated as if they were perfectly complementary. The output of the service of a durable good at any point of time is assumed to bear a fixed relation to the output of the same service at any other point of time. The assumptions that outputs provided by a given input of investment goods are perfectly complementary or perfectly substitutable are highly restrictive. Nevertheless, the simplification of the neoclassical theory for these limiting cases and the practical importance of these cases are very great. A far more substantial proportion of the literature on capital theory is devoted to these two limiting

[22] E. Malinvaud, "Capital Accumulation and Efficient Allocation of Resources," *Econometrica*, April 1953, pp. 233–268.

cases than to the theory of production at the level of abstraction of the descriptions of technology given by Fisher or by Malinvaud. In the following we assume that the conventional neoclassical description of a durable good is appropriate for each investment good considered.

A second assumption required for a relationship between output at each point of time and the corresponding flow of capital services is that the services of investment goods acquired at different points of time are perfect substitutes in production. Accordingly, the flow of capital services from each investment good is proportional to the stock of capital that may be obtained by simply adding together all past acquisitions less replacements. This assumption is highly restrictive; the assumption can be justified primarily by the resulting simplification of the neoclassical theory. We discuss only a single investment good. Under the assumptions outlined above, there is only a single capital service. This simplification is also completely inessential to neoclassical theory.

Finally, we assume that the flow of replacement generated by a given flow of investment goods is distributed over time in accord with an exponential distribution. This assumption implies that the flow of replacement investment at any point of time is proportional to the accumulated stock of investment goods. Again, this assumption is only one among many possibilities. Alternative assumptions employed in practice include the following: First, replacement is equal to investment goods acquired at some earlier point in time; second, replacement is equal to a weighted average of past investment flows, with weights derived from studies of the "survival curves" of individual pieces of equipment.[23] For empirical work the exponential distribution of replacements is of special interest. While empirical studies of "survival curves" for individual pieces of equipment reveal a wide variety of possible distributions, there is a deeper justification for use of the exponential distribution. This justification arises from a fundamental result of renewal theory, namely, that replacement approaches an amount proportional to the accumulated stock of capital whatever the distribution of replacements for an individual piece of equipment, provided that the size of the capital stock is constant or that the stock is growing at a constant rate (in the probabilistic sense).[24] This asymptotic result may be used as the basis for an approximation to the distribution of replace-

[23] A summary of research on the lifetimes of capital equipment as given by A. Marston, R. Winfrey, and J. C. Hempstead, *Engineering Evaluation and Depreciation*, 2nd ed., New York, 1953.

[24] For a statement of the basic theorem, see E. Parzen, *Stochastic Processes*, San Francisco, 1962, pp. 180–181.

ments; for any investment good, the stream of replacements eventually approaches a stream that would be generated by an exponential distribution of replacements. Accordingly, the exponential distribution may be used as an approximation to the distribution of replacements for the purpose of estimating the stream of replacements. A simple indirect test of the validity of this approximation has been carried out by Meyer and Kuh.[25] For any distribution of replacements except the exponential distribution, one would expect to observe an "echo effect" or bunching of replacements at lags corresponding to points of relatively high density in the conditional distributions of replacements for individual types of equipment. Meyer and Kuh report no evidence for such an effect.

To summarize, we consider a version of the neoclassical theory in which the objective of the firm is maximization of its present value. This may be derived from the objective of maximizing the utility of a consumption stream subject to a fixed set of production possibilities and to fixed current and future prices and interest rates. Since the choice of a production plan is entirely independent of the corresponding choice of a consumption stream, two individuals with different preferences among consumption streams will choose the same production plan. Secondly, we consider a description of technological possibilities in which output at each point of time depends on the flow of labor and capital services at that point of time, the flow of capital services is proportional to the stock of capital goods, and replacements are also proportional to the stock of capital goods. This description of technology is a specialization of the descriptions given by Malinvaud and by Fisher. The essential justification for this specialization is that the resulting theory of optimal capital accumulation is sufficiently broad to include the principal econometric models of investment behavior as special cases.

3. Optimal Capital Accumulation

To develop the theory of investment behavior in more detail, we must first define the present value of the firm. For simplicity, we limit the analysis to a production process with a single output, a single variable input, and a single capital input. Where Q, L, and I represent levels of output, variable input, and investment in durable goods and p, w, and q represent the corresponding prices, the flow of net receipts at time t, say $R(t)$, is given by:

$$R(t) = p(t)Q(t) - w(t)L(t) - q(t)I(t). \tag{1}$$

[25] Meyer and Kuh, *Investment Decision,* pp. 91–94.

Present value is defined as the integral of discounted net receipts; where $r(s)$ is the rate of time discount at time s, net worth (W) is given by the expression:

$$W = \int_0^\infty e^{-\int_0^t r(s)\,ds} R(t)\,dt. \tag{2}$$

For purposes of the following discussion, we may assume that the time rate of discount is a constant without loss of generality. Accordingly, the present value of the firm may be represented in the simpler form:

$$W = \int_0^\infty e^{-rt} R(t)\,dt.$$

Present value is maximized subject to two constraints. First, the rate of change of the flow of capital services is proportional to the flow of net investment. The constant of proportionality may be interpreted as the time rate of utilization of capital stock, that is, the number of units of capital service per unit of capital stock. We will assume that capital stock is fully utilized so that this constant may be taken to be unity. Net investment is equal to total investment less replacement; where replacement is proportional to capital stock, this constraint takes the form:

$$\dot{K}(t) = I(t) - \delta K(t) \tag{3}$$

where $\dot{K}(t)$ is the time rate of change of the flow of capital services at time t. This constraint holds at each point of time so that $\dot{K}$, K, and I are functions of time; to simplify notation, we will use K in place of $K(t)$, I in place of $I(t)$, and so on. Secondly, levels of output and levels of labor and capital services are constrained by a production function:

$$F(Q, L, K) = 0. \tag{4}$$

We assume that the production function is twice differentiable with positive marginal rates of substitution between inputs and positive marginal productivities of both inputs. Furthermore, we assume that the production function is strictly convex.

To maximize present value (2) subject to the constraints (3) and (4), we consider the Lagrangian expression:

$$\mathcal{L} = \int_0^\infty [e^{-rt}R(t) + \lambda_0(t)F(Q, L, K) + \lambda_1(t)(\dot{K} - I + \delta K)]\,dt, \tag{5}$$

$$= \int_0^\infty f(t)\,dt,$$

where:

$$f(t) = e^{-rt}R(t) + \lambda_0(t)F(Q, L, K) + \lambda_1(t)(\dot{K} - I + \delta K).$$

The Euler necessary conditions for a maximum of present value subject to the constraints (3) and (4) are:

$$\frac{\partial f}{\partial Q} = e^{-rt}p + \lambda_0(t)\frac{\partial F}{\partial Q} = 0, \tag{6}$$

$$\frac{\partial f}{\partial L} = -e^{-rt}w + \lambda_0(t)\frac{\partial F}{\partial L} = 0,$$

$$\frac{\partial f}{\partial I} = -e^{-rt}q - \lambda_1(t) = 0,$$

$$\frac{\partial f}{\partial K} - \frac{d}{dt}\frac{\partial f}{\partial \dot{K}} = \lambda_0(t)\frac{\partial F}{\partial K} + \delta\lambda_1(t) - \frac{d}{dt}\lambda_1(t) = 0,$$

and also:

$$\frac{\partial f}{\partial \lambda_0} = F(Q, L, K) = 0, \tag{7}$$

$$\frac{\partial f}{\partial \lambda_1} = \dot{K} - I + \delta K = 0.$$

Combining the necessary conditions for labor and output, we obtain the marginal productivity condition for labor services:

$$\frac{\partial Q}{\partial L} = \frac{w}{p}. \tag{8}$$

Of course, output, labor, wages, and prices are all functions of time. The difference between this marginal productivity condition and the corresponding condition of the "static" theory of the firm is that condition (8) holds at every point of time over the indefinite future whereas the marginal productivity condition of the "static" theory of the firm holds only at a single point in time. A similar marginal productivity condition for capital services may be derived. First, solving the necessary conditions (6) for $\lambda_1(t)$:

$$\lambda_1(t) = -e^{-rt}q,$$

the necessary condition for capital services may be written:

$$\lambda_0(t)\frac{\partial F}{\partial K} - \delta e^{-rt}q - re^{-rt}q + e^{-rt}\dot{q} = 0.$$

Combining this condition with the necessary condition for output, we obtain the marginal productivity condition for capital services:

$$\frac{\partial Q}{\partial K} = \frac{q(r+\delta) - \dot{q}}{p} = \frac{c}{p}, \tag{9}$$

where:

$$c = q(r+\delta) - \dot{q}. \tag{10}$$

Again, output, capital, prices, and the rate of time discount are functions of time so that these conditions hold at every point of time over the indefinite future.

Expression (10) defines the implicit rental value of capital services supplied by the firm to itself. This interpretation of the price $c(t)$ may be justified by considering the relationship between the price of capital goods and the price of capital services. First, the flow of capital services over an interval of length dt beginning at time t from a unit of investment goods acquired at time s is:

$$e^{-\delta(t-s)}\,dt.$$

If $c(t)$ is the price of capital services at time t, then the discounted price of capital services is $e^{-rt}c(t)$, so that the value of the stream of capital services on the interval dt is:

$$e^{-rt}c(t)e^{-\delta(t-s)}\,dt.$$

Similarly, if $q(s)$ is the price of capital goods at time s, then the discounted price of capital goods is $e^{-rs}q(s)$, so that the value of a unit of investment goods acquired at time s is:

$$e^{-rs}q(s).$$

But the value of investment goods acquired at time s is equal to the integral of the discounted value of all future capital services derived from these investment goods:

$$e^{-rs}q(s) = \int_s^\infty e^{-rt}c(t)e^{-\delta(t-s)}\,dt,$$

$$= e^{\delta s}\int_s^\infty e^{-(r+\delta)t}c(t)\,dt.$$

Solving for the price of capital goods, we obtain:

$$q(s) = e^{(r+\delta)s} \int_s^{\infty} e^{-(r+\delta)t} c(t)\, dt,$$

$$= \int_s^{\infty} e^{-(r+\delta)(t-s)} c(t)\, dt.$$

To obtain the price of capital services implicit in this expression, we differentiate with respect to time:

$$\dot{q}(s) = [r(s) + \delta] q(s) - c(s),$$

so that:

$$c = q(r + \delta) - \dot{q},$$

which is expression (10) given above for the implicit rental value of capital services.

The conditions describing the neoclassical model of optimal capital accumulation may also be derived by maximization of the integral of discounted profits, where profit at each point of time, say, $P(t)$, is given by:

$$P(t) = p(t)Q(t) - w(t)L(t) - c(t)K(t). \tag{11}$$

The integral of discounted profits, say, W^+, is given by the expression:

$$W^+ = \int_0^{\infty} e^{-rt} P(t)\, dt. \tag{12}$$

The side condition for investment may be disregarded, since investment does not enter into the definition of profit (11); substituting the side condition for the shadow price of capital services into the profit function, we obtain:

$$W^+ = \int_0^{\infty} e^{-rt} [p(t)Q(t) - w(t)L(t) - \{q(t)[r(t) + \delta] - \dot{q}(t)\} K(t)]\, dt.$$

To maximize this function subject to the production function, it suffices to maximize profit at each point of time subject to the production function. But this yields the marginal productivity conditions (8) and (9) and the production function (4) itself. Reintroducing the side conditions (3) and (10), we obtain the complete neoclassical model of optimal capital accumulation.

The integral of discounted profits is not the same as the integral defining present value of the firm. The difference between the two is given by:

$$
\begin{aligned}
W - W^{+} &= \int_0^\infty e^{-rt}[R(t) - P(t)]\, dt \\
&= \int_0^\infty e^{-rt}[\{q(t)[r(t) + \delta] - \dot{q}(t)\}K(t) - q(t)I(t)]\, dt \\
&= \int_0^\infty e^{-rt}[q(t)\delta K(t) + q(t)r(t)K(t) - \dot{q}(t)K(t) - q(t)\dot{K}(t) \\
&\qquad - q(t)\delta K(t)]\, dt \\
&= q(0)K(0),
\end{aligned}
$$

which is the value of capital stock on hand at the initial point of time. The present value of the firm is the sum of the integral of discounted profits and the market value of the assets of the firm. Since the market value of the assets of the firm is fixed, maximization of the integral of discounted profits results in the same path for accumulation of capital as maximization of present value of the firm. To summarize, the neoclassical model of optimal capital accumulation may be derived by maximizing present value of the firm, by maximizing the integral of discounted profits of the firm, or simply by maximizing profit at each point of time.

In taking maximization of profit as the objective of the firm, profit is defined in a special sense, namely, net receipts on current account less the implicit rental value of capital services. This concept of profit would agree with the usual accounting definition of profit only in rather unusual circumstances, for example, where the firm actually rents all the capital services it employs. The price of capital services is then a market price and the rental value of the services is an actual outlay. Where the firm supplies capital services to itself, the implicit rental value of capital services $c(t)$ is a shadow price which may be used by the firm in the computation of an optimal path for capital accumulation. For optimal capital accumulation, the firm should charge itself a price for capital services equal to the implicit rental value and should then maximize profit at each point of time in the usual way. It is very important to note that the conditions determining the values of each of the variables to be chosen by the firm—output, labor input, and investment in capital goods—depend only on prices, the rate of interest, and the rate of change of the price of capital goods for the current period. Accordingly, in the neoclassical theory of optimal capital accumulation, the firm behaves at each point of time as in the "static" theory of the firm, provided that the price of capital services is taken to be equal to the corresponding implicit rental value. Of course, in the "static" theory the marginal productivity condition (9) holds only at a single point in time.

The complete neoclassical model of optimal capital accumulation consists of the production function (4) and the two marginal productivity conditions (8) and (9):

$$F(Q, K, L) = 0, \qquad \frac{\partial Q}{\partial L} = \frac{w}{p}, \qquad \frac{\partial Q}{\partial K} = \frac{c}{p},$$

and the two side conditions (3) and (10):

$$I = \dot{K} + \delta K,$$

$$c = q(r + \delta) - \dot{q}.$$

The production function and marginal productivity conditions hold at each point of time. The side conditions are differential equations also holding at each point of time. Combined, these conditions determine the levels of output, labor input, and capital input, together with the level of investment and the shadow price for capital services.

The interpretation of condition (3) determining the level of investment is the source of some difficulty in the literature. If the level of investment is bounded, the derivative of the level of capital services must be bounded. But this implies that the level of capital services itself must be continuous. Since we have assumed that the production function is twice differentiable, a sufficient condition for continuity of the level of capital services is continuity of the prices—w, p, c.

One interpretation of condition (3) is that the initial value of the level of capital services may be chosen arbitrarily. This interpretation has been suggested by Haavelmo and by Arrow.[26] If the initial level of capital services is derived from the production function and the marginal productivity conditions and if the initial value of capital is fixed arbitrarily, optimal capital accumulation may require an unbounded initial level of investment. In management science, this interpretation of the problem may be of some interest, though even there the interpretation seems somewhat forced, as Arrow points out.[27] For empirical work this interpretation is completely artificial since firms are viewed as making new decisions to invest continuously over time. To maximize present value at each point of time, a firm following an optimal path for capital accumulation must maximize present value subject to the initial condi-

[26] Haavelmo, *Theory of Investment*, pp. 162–165. Arrow, "Optimal Capital Adjustment," in *Studies in Applied Probability*, p. 2.

[27] *Ibid.*, p. 6, fn. 1.

tion given by the optimal path up to that point. But this results in a new optimal path which is precisely the same as the old from that point forward. Accordingly, if the optimal path for capital accumulation is continuous, the initial value of the level of capital services may not be chosen arbitrarily in the maximization of the present value of the firm. At each point it is precisely that for which the initial level of investment is bounded, namely, the level of capital services derived from the production function and the marginal productivity conditions. A possible objection to this view is that firms must begin to accumulate capital at some point in time. But at such a point the initial level of capital services is not given arbitrarily; the initial level must be zero with a positive derivative.

4. The Theory of Investment Behavior

Beginning with the neoclassical model of optimal capital accumulation, we may derive differentiable demand functions for labor and capital services and a differentiable supply function for output, say:

$$\begin{aligned} L &= L(w, c, p), \\ K &= K(w, c, p), \\ Q &= Q(w, c, p). \end{aligned} \tag{13}$$

The problem of deriving the demand for investment goods as a function of the rate of interest is a subtle one. Haavelmo expresses the view that the demand for investment goods cannot be derived from the profit-maximizing theory of the firm. This is a consequence of his interpretation of the demand function for capital services and condition (3) determining the level of investment from replacement and the rate of change of demand for capital services. According to this interpretation, finite variations in the rate of interest with all other prices held constant result in finite changes in the demand for capital services. As the rate of interest varies, demand for investment goods assumes only three possible values—negatively infinite, positively infinite, or the value obtained where the initial level of capital services is precisely equal to the demand for capital services. Investment demand has a finite value for only one rate of interest. In this interpretation, the demand function for capital services is analyzed by means of comparative statics, that is, by comparing alternative production plans at a given point of time. Any attempt to derive the demand for investment goods as a function of the rate of interest by such comparisons leads to nonsensical results, as Haavelmo correctly points out.

However, an alternative interpretation of the demand function for capital services and condition (3) determining the level of investment is possible. Under the hypothesis that the firm is following an optimal path for capital accumulation and that the optimal path is continuous, the initial level of capital is always equal to the demand for capital services. By imposing this condition at the outset, the demand for investment goods as a function of the rate of interest at any point of time may be analyzed by means of comparative dynamics, that is, by comparing alternative paths of capital accumulation, each identical up to that point of time and each continuous at that point. The demand for investment goods is given by condition (3):

$$I = \dot{K} + \delta K,$$

where the level of capital services, K, is fixed; but from the demand function for capital services (13), this condition implies that for fixed values of the price of output and the price of labor services, the implicit price of capital services must remain unchanged. Holding the price of investment goods constant, the rate of change of the price of investment goods must vary as the rate of interest varies so as to leave the implicit price of capital services unchanged. Formally, the condition that variations in the rate of interest leave the implicit price of capital services unchanged may be represented as:

$$\frac{\partial c}{\partial r} = 0;$$

holding the price of investment goods constant, this condition implies that the own-rate of interest on investment goods, $r - \dot{q}/q$, must be left unchanged by variations in the rate of interest.

We assume that all changes in the rate of interest are precisely compensated by changes in the rate of change of the price of current and future investment goods so as to leave the own-rate of interest on investment goods unchanged. Under this condition the discounted value of all future capital services, which is equal to the current price of investment goods, is left unchanged by variations in the time path of the rate of interest. The condition that the time path of the own-rate of interest on investment goods is left unchanged by a change in the time path of the rate of interest implies that forward prices or discounted future prices of both investment goods and capital services are left unchanged by variations in the rate of interest. For a constant rate of interest, this condition may be represented in the form:

$$\frac{\partial^2 e^{-rt}c(t)}{\partial r \partial t} = 0.$$

Like the previous condition, this condition holds at every point of time.

To derive the demand for investment goods as a function of the rate of interest, we first differentiate the demand for capital services with respect to time, obtaining:

$$\dot{K} = \frac{\partial K}{\partial w} \cdot \frac{\partial w}{\partial t} + \frac{\partial K}{\partial c} \cdot \frac{\partial c}{\partial t} + \frac{\partial K}{\partial p} \cdot \frac{\partial p}{\partial t}.$$

For simplicity, we consider only the case in which $\frac{\partial w}{\partial t} = \frac{\partial p}{\partial t} = 0$, that is, the price of output and the price of labor services are not changed. In this case, we obtain:

$$\dot{K} = \frac{\partial K}{\partial c} \cdot \frac{\partial c}{\partial t}.$$

Differentiating the implicit price of capital services with respect to time, we have:

$$\frac{\partial c}{\partial t} = \frac{\partial q}{\partial t}(\delta + r) + q\frac{\partial r}{\partial t} - \frac{\partial^2 q}{\partial t^2}. \tag{14}$$

To derive the demand for investment goods, we combine expression (14) for the rate of change of capital services with condition (3) for the rate of investment, obtaining:

$$I = \frac{\partial K}{\partial c}\left[\frac{\partial q}{\partial t}(\delta + r) + q\frac{\partial r}{\partial t} - \frac{\partial^2 q}{\partial t^2}\right] + \delta K,$$

which depends on the rate of interest and the price of investment goods through the rate of change of capital services. Differentiating this investment demand function with respect to the rate of interest, we obtain:

$$\frac{\partial I}{\partial r} = \frac{\partial^2 K}{\partial c^2} \cdot \frac{\partial c}{\partial r} \cdot \frac{\partial c}{\partial t} + \frac{\partial K}{\partial c}\frac{\partial^2 c}{\partial t \partial r} + \delta\frac{\partial K}{\partial c} \cdot \frac{\partial c}{\partial r}.$$

But $\frac{\partial c}{\partial r} = 0$, since changes in the rate of interest are compensated by changes in the rate of change of the price of investment goods so as to leave the implicit price of capital services unchanged. This condition implies that:

$$\frac{\partial^2 q}{\partial t \partial r} = q.$$

Secondly, $\dfrac{\partial^2 e^{-rt} c(t)}{\partial r \partial t} = 0$, since changes in the time path of the rate of interest leave the time path of forward or discounted prices of capital services unchanged. This condition implies that:

$$\frac{\partial^2 c}{\partial t \partial r} = c.$$

Combining these two conditions, we obtain:

$$\frac{\partial I}{\partial r} = \frac{\partial K}{\partial c} \cdot c < 0,$$

so that the demand for investment goods is a decreasing function of the rate of interest.

We conclude that it is possible to derive the demand for investment goods as a function of the rate of interest on the basis of purely neoclassical considerations. However, the demand for investment goods depends on the rate of interest through a comparison of alternative paths of capital accumulation, each continuous and each depending on a time path of the rate of interest. Although this conclusion appears to be the reverse of that reached by Haavelmo, his approach to the demand for investment goods is through comparative statics, that is, through comparison of alternative production plans at a given point of time. The demand function for investment goods cannot be derived by means of such comparisons. As a proposition in comparative statics, any relation between variations in the rate of investment and changes in the rate of interest is nonsensical.

To summarize, the complete neoclassical model of optimal capital accumulation consists of the production function (4), the two marginal productivity conditions (8) and (9), and the side condition (10). An alternative form of this model consists of the demand functions for capital and labor services, the supply function for output:

$$L = L(w, c, p),$$

$$K = K(w, c, p),$$

$$Q = Q(w, c, p);$$

and the demand function for investment goods:

$$I = \frac{\partial K}{\partial c}\frac{\partial c}{\partial t} + \delta K,$$
$$= I\left(w, c, p, \frac{\partial c}{\partial t}\right).$$

The demand for investment goods depends on the change in the demand for capital with respect to a change in the implicit price of capital services, the time rate of change in the price of capital services, and the level of replacement demand. Where the time rates of change of the price of labor services and the price of output are not zero, the demand function for investment goods may be rewritten:

$$I = \frac{\partial K}{\partial w}\frac{\partial w}{\partial t} + \frac{\partial K}{\partial c}\frac{\partial c}{\partial t} + \frac{\partial K}{\partial p}\cdot\frac{\partial p}{\partial t} + \delta K,$$
$$= I\left(w, c, p, \frac{\partial w}{\partial t}, \frac{\partial c}{\partial t}, \frac{\partial p}{\partial t}\right).$$

5. *Alternative Theories of Investment Behavior*

The neoclassical theory of demand for investment goods just outlined may be contrasted with the theory current in the literature. Most recent accounts of the theory of demand for investment are based on Keynes' *General Theory,* in which the criterion for optimal investment behavior is that any project with an internal rate of return greater than the ruling rate of interest is undertaken.[28] An investment demand schedule is constructed by varying the rate of interest and plotting the quantities of investment undertaken for each value of the rate of interest. The criterion for optimal investment behavior used by Keynes is inconsistent with maximization of the present value of the firm, as Alchian and Hirshleifer have pointed out.[29] Nevertheless, a substantial portion of the current literature on the investment demand function is based on a straightforward reproduction of Keynes' derivation. Alchian lists a number of examples from the literature prior to 1955; examples from the

[28] Chapter 11, "The Marginal Efficiency of Capital," especially p. 136.

[29] A. A. Alchian, "The Rate of Interest, Fisher's Rate of Return over Costs and Keynes' Internal Rate of Return," in *Management of Corporate Capital,* p. 70; and J. Hirshleifer, in *ibid.,* pp. 222–227. This conclusion of Alchian and Hirshleifer contradicts the position taken by Klein in *The Keynesian Revolution.*

current literature are provided by the recent work of Duesenberry and Tarshis.[30] Keynes' construction of the demand function for investment must be dismissed as inconsistent with the neoclassical theory of optimal capital accumulation.

An alternative construction of the demand function for investment goods has been suggested by Fisher.[31] In Fisher's theory any project with positive present value is undertaken. Keynes appears to have identified his construction of the marginal efficiency of capital schedule with that of Fisher, as Alchian points out.[32] There are two difficulties with Fisher's construction. First, the construction is carried out by means of comparative statics so that the resulting schedule may be interpreted as a theory of demand for capital services for which no demand function for investment goods exists. Second, the construction is not internally consistent in a second sense pointed out by Alchian, since ". . . we cannot in full logical consistency draw up a demand curve for investment by varying only the rate of interest (holding all other prices in the impound of *ceteris paribus*)."[33] The relevant prices are forward prices of all commodities; but altering the rate of interest amounts to altering certain forward prices. It is inconsistent to vary the rate of interest while holding such prices fixed. This inconsistency may be eliminated by stipulating that variations in the rate of interest must be precisely compensated by changes in the time rate of change of the price of investment goods. The price of investment goods at a given point of time is held fixed; the rate of change of the price of investment goods varies with the rate of interest. The construction of the demand function for investment goods involves a comparison among alternative paths of optimal capital accumulation; all paths are identical up to the point of time for which the investment function is constructed. Such a theory of investment behavior is internally consistent and may be derived by means of comparative dynamics.

Klein has attempted to derive a demand function for investment goods on the basis of profit maximization. His treatment, though suggestive, is

[30] J. S. Duesenberry, *Business Cycles and Economic Growth*, New York, 1958, pp. 49–85. Duesenberry asserts that Keynes' derivation is based on "profit maximization" (p. 85). L. Tarshis, "The Marginal Efficiency Function," *American Economic Review*, December 1961, pp. 958–985. Tarshis asserts that the Keynesian theory is based on that of the "profit-maximizing firm" (pp. 958–959).

[31] Fisher, *Theory of Interest*, pp. 159–176.

[32] Alchian, in *Management of Corporate Capital*, p. 67. Klein (*Keynesian Revolution*, p. 62) follows Keynes in identifying these two distinct approaches to the construction of the marginal efficiency schedule.

[33] Alchian, *Management of Corporate Capital*, p. 71.

marred by a number of inconsistencies. In his first attempt, the stock of investment goods is defined as the integral of past flows of investment, but the flow of investment is employed as a stock in the production function and in the definition of "discounted profit."[34] A second attempt involves the identification of the flow of capital services with the flow of depreciation.[35] In both attempts, quantities measured as rates of capital service per unit of time are added to quantities measured as rates of investment per unit of time, which is self-contradictory. This inconsistency carries over to the empirical implementation of the resulting investment function, where the price of investment goods is identified with the price of capital services.[36] An internally consistent treatment of the theory of investment along the lines suggested by Klein leads to a comparative statics theory of demand for capital services in which no demand function for investment goods exists.

Another branch of the current literature is based on the view that no demand function for investment goods exists. We have already cited Haavelmo's support of this position. A similar view may be found in Lerner's *Economics of Control.* Lerner argues that, under diminishing returns, the firm has a downward sloping demand curve for capital services but that, except where there is no net investment, the rate of investment is unbounded:[37]

> . . . there is no limit to the rate per unit of time at which [the individual] can acquire assets by buying them, borrowing money for the purpose if he has not enough of his own. This indefinitely great rate of "investment" means that he can move at once to the position . . . which makes the (private) marginal productivity of capital equal to the rate of interest. Once he gets there, there is no tendency for further expansion. . . .

This view is the same as that expressed by Haavelmo. A recent restatement of this position has been given by Witte, who concludes, with Lerner and Haavelmo, that ". . . the continuous function relating the rate of investment to the rate of interest at the micro level has no foundation in the ordinary theory of the firm."[38] We have demon-

[34] Klein, *Keynesian Revolution,* esp. pp. 196–199.

[35] Klein, in *Kyklos,* Vol. 2, fasc. 2, 1948, pp. 97–117; and his *Economic Fluctuations.*

[36] *Ibid.* The price of investment goods (p. 21 and p. 85) is identified with the price of capital services (p. 15).

[37] A. P. Lerner, *The Economics of Control,* esp. pp. 330–338.

[38] James G. Witte, Jr., "The Microfoundations of the Social Investment Function," *Journal of Political Economy,* October 1963, pp. 441–456.

strated that it is possible to derive the demand for investment goods from the comparative dynamics applied to the ordinary neoclassical theory of the firm. The conclusion reached by Haavelmo, Lerner, and Witte concerning a demand function for investment goods derived on the basis of comparative statics is, of course, correct.

An attempt has been made by proponents of the view that the demand function for investment goods does not exist to rehabilitate the Keynesian marginal efficiency of investment schedule. Alternative versions of this rehabilitation are presented by Haavelmo, Lerner, and Witte.[39] The essentials of the argument are that, at a given rate of interest, a certain price for investment goods is required to equate the marginal productivity of capital with the implicit price of capital services; but the higher this price the lower the rate of interest, so that a rising supply curve for investment goods implies that the amount of investment goods produced will increase as the rate of interest falls. A fundamental difficulty with this view is that it fails to account for the purchase of new investment goods by the users of capital equipment.[40] Witte summarizes this consequence of the view as follows: ". . . the rate-of-investment decision is the rate-of-output decision of supplying enterprises and not the rate-of-input decision of capital-using firms."[41] In the same vein Haavelmo writes, ". . . it is, actually, not the users of capital who demand investment, it is the producers of capital goods who determine how much they want to produce at the current price of capital."[42] A further attempt along these lines of the rehabilitation of the Keynesian marginal efficiency of investment schedule has been presented by Clower.[43] His argument follows that of Haavelmo, Lerner, and Witte in assuming that demand for capital services is equal to supply. However, Clower intro-

[39] Haavelmo, *Theory of Investment,* pp. 194–197. See also: B. Thalberg, "An Analysis of a Market for Investment Goods," in Lutz and Hague, *Theory of Capital,* pp. 161–176, and "A Keynesian Model Extended by Explicit Demand and Supply Functions for Investment Goods," in *Stockholm Economic Studies,* Pamphlet Series, No. 3, 1964. Lerner, *Economics of Control,* pp. 333–334. Witte, in *Journal of Political Economy,* October 1963, pp. 445–447.

[40] A second difficulty with this view is that an increase in the price of investment goods may result in a rise or a fall in the supply of investment goods, depending on the relative capital intensity of the investment goods and consumption goods industries. Lerner, for example, assumes implicitly that investment goods are produced with no capital services. This difficulty was pointed out to me by James Tobin.

[41] *Ibid.,* p. 448.

[42] Haavelmo, *Theory of Investment,* p. 196.

[43] R. W. Clower, "An Investigation into the Dynamics of Investment," *American Economic Review,* March 1954, pp. 64–81.

duces a demand for investment goods which is not necessarily equal to the supply of investment goods. The excess or deficiency of demand over supply is net accumulation of capital. This view also fails to account for the purchases of new investment goods by the users of capital equipment.

For internal consistency, the rehabilitation of the Keynesian marginal efficiency of investment schedule requires either a changing rate of interest, as suggested by Haavelmo, or a changing price of capital goods, as suggested by Lerner.[44] For if the rate of interest and the price of investment goods are fixed over time and the marginal productivity of capital is equal to the implicit price of capital services, the firm's demand for investment is determinate; this demand is precisely equal to replacement demand so that net investment is zero. Under these circumstances, the rate of investment demand by users of capital equipment is independent of the rate of interest so that the price of investment goods must be that at which this rate of investment will be supplied by investment goods producers. But then if the marginal productivity of capital is to be equal to the implicit price for capital services, the rate of interest is uniquely determined, which is inconsistent with variations in the rate of interest from whatever source.

To complete the rehabilitation of the Keynesian marginal efficiency of investment schedule, interpreted as the level of investment resulting from a market equilibrium in investment goods corresponding to a given rate of interest, market equilibrium must be studied in a fully dynamic setting. The demand for investment goods must be derived from a comparison among alternative paths of optimal capital accumulation. It remains to be seen whether such a rehabilitation can be carried out in an internally consistent way.

[44] Haavelmo, *Theory of Investment,* p. 196. Lerner, *Economics of Control,* diagram, p. 336.

COMMENT

On Crockett-Friend and Jorgenson

BY JAMES TOBIN, YALE UNIVERSITY

I agree with Jorgenson's general defense of the neoclassical theory of the firm. As he says, its usefulness is by no means confined to static conditions. As long as expectations are assumed certain, maximization of the present value of the firm is as powerful a principle for dynamic theory as profit maximization has been for static theory. A dynamic theory based on this principle has much more to say, and can handle many more complexities, than is often appreciated.

Jorgenson's specific example, however, is only barely dynamic. His firm can maximize present value simply by maximizing profits at every point in time. The firm confronts no intertemporal trade-offs, in which profits now must be weighed against profits later. It purchases capital *services* at a market rental, just as it purchases labor at a market wage. There is a perfect market in capital goods; capital is homogeneous in quality regardless of its vintage; and capital evaporates exponentially, so that future depreciation is also independent of vintage. Thus, any surviving capital can always be sold at the prevailing price of new capital goods. Therefore if, as Jorgenson assumes, the rental of capital services correctly reflects interest, depreciation, and the change in price of capital goods, the firm will be indifferent in choosing between renting and owning. The present value of such future rentals just equals the current price of capital goods.

I would like to make a parenthetical semantic remark: Jorgenson calls the rental just discussed, specifically $q(r + \delta - \dot{q}/q)$, user cost. To anyone who learned about user cost from the appendix to Chapter 6 of Keynes' *General Theory,* this terminology seems surprising. Keynes assumed that the decline in the value of a stock of goods during a period depends on the intensity of use, not just on the passage of time, hence the term *user* cost. Keynes' assumption is notably absent from most modern capital theory, including Jorgenson's. I find it confusing to see a rental which is just a time or ownership cost called user cost.

By assuming diminishing returns to scale, Jorgenson makes the size of his firm determinate within the framework of pure competition and certain expectations. However, the sale of the services of owned capital is an activity with constant returns to scale, and in Jorgenson's world of perfect competition and perfect knowledge, the scale of ownership by any one individual is indeterminate.

Given the time path of the price of the product p, the wage rate w, and the rental on capital c, Jorgenson's firm decides upon the paths of output Q, employment L, and use of capital services K. Indeed, these paths will simply maximize profits $pQ - wL - cK$ at each point in time, subject to the production function. If the time paths of p, w, and c are continuous, then so are the paths of Q, L, and K.

However, as Jorgenson points out, there is no reason to assume that markets will never present an individual firm with jumps in p, w, and c. If they do so, the firm's profit-maximizing response involves jumps in Q, L, and K. In Jorgenson's firm there are no frictions or speed-of-adjustment costs to make profitable any delay at all in responding to new conditions.

Many economists—Jorgenson cites Haavelmo and Lerner—have concluded that such an individual firm has no demand schedule for *net investment* $\dot{K}$ but only a demand schedule for capital K. These theorists think that if conditions change the optimal rate of use of capital services, the firm will immediately shift to the new optimum—by renting more or less capital or by buying or selling capital goods. This is not a surprising conclusion. It is the use of capital services, proportional to the stock not to the flow, which is related to the determining prices. Similarly, the firm has a demand schedule for labor services, not for their rate of change. No one is dismayed that a frictionless firm is expected to shift in no time from one employment level to another.

The investment demand schedule which these economists have sought and not found is a relationship at a given point in time t_0 between investment $\dot{K}(t_0)$ (or $\dot{K}(t_0) + \delta K(t_0)$) at that time and the rate of interest $r(t_0)$, *holding constant all other current and expected prices.* This is the marginal efficiency schedule which Keynes purported to draw at the aggregative level, which Lerner and Haavelmo doubted existed for an individual firm, which Lerner tried to justify on macroeconomic grounds. Now varying $r(t_0)$ to the individual firm, holding all other relevant prices constant, is bound to cause the break-even rental $c(t_0)$ to vary also. Indeed all hypothetical values of $r(t_0)$ except one involve a jump at t_0 in the optimal $K(t)$. Moreover, one cannot escape the conclusion that, except for the one value of $r(t_0)$ which keeps $c(t_0)$ adjusted to the existing capital stock, $\dot{K}(t_0)$ must be either $+\infty$ or $-\infty$.

Jorgenson does not escape this conclusion either, but by asking a different question he arrives at what he identifies as an investment demand schedule. He does not hold all other prices, present and future, constant while he varies $r(t_0)$. Instead he compensates the variation of $r(t_0)$ by changes in present and future $q(t)$ so that $c(t_0)$ remains the same.[1] Thus $K(t_0)$ is independent of these compensated variations in r, but the subsequent path $K(t)$ is not. And in particular $\dot{K}(t_0)$ will depend on r!

Maybe there is some question to which this is the answer, but it is not the question to which Jorgenson finds previous answers so unsatisfactory. There is no reason to assume that expected prices of capital goods accommodate themselves so obligingly to interest rate variations. Unless they do so, Jorgenson's investment demand schedule cannot serve the analytical purposes for which such a schedule is desired, and one must look elsewhere for a determinate theory of investment. At the level of a single firm, this may be derived from frictional or adjustment costs; at the level of the whole economy, it may be derived from capacity limitations on production of investment goods (although here Lerner's famous solution is, as Jorgenson points out, far from foolproof).

It would be desirable to have a neoclassical theory of consumer investment to place alongside the theory of business investment. In such a theory it would be necessary to state payoffs in utility rather than in money, to recognize imperfections in rental and second-hand markets, and to allow for true Keynesian user cost. A model of this kind would, I think, suggest some differences between real and financial investments by households which do not appear in the Crockett-Friend paper.

Their model is considerably less theoretical. In their view, each household has a desired total and composition of net worth, depending on its normal income and its tastes, and on the yields and risks of various assets and debts. Crockett and Friend explain flows of household investment and saving as a process of stock adjustment, without worrying with Jorgenson why adjustments should not be instantaneous.

While I am sympathetic to this approach to empirical data, I think the authors' formulation is too static. Not all flows should be interpreted simply as efforts to eliminate discrepancies between actual and desired stocks. Desired stocks change, and there would be nonzero flows even if the household were continuously in adjustment. Even for the same normal income, for example, a household's desired wealth will change in total and in composition with time and age. I suspect that reformulation along these lines might improve the authors' empirical estimates of adjustment speeds, which are so far rather unsatisfactory.

The main purposes of the Crockett-Friend project, of which this con-

[1] For example, if $\dot{q}(t)/q(t)$ is increased for all $t \geq t_0$ by the same amount as a once-and-for-all rise in r at t_0, then $c(t_0)$ is unchanged. Future $c(t)$ are increased, but since they are discounted more heavily their present value is still $q(t_0)$.

ference paper is a progress report, are to estimate normal income elasticities of demand for wealth and its components, and to estimate speeds of stock adjustment. The data are cross sections, and the authors rely especially on cross sections containing observations in the same households for more than one year.

A principal finding is that the ratio of net worth to normal income increases with wealth and income. Crockett and Friend suggest that this finding is inconsistent with saving theories which contend that "permanent" saving is a constant fraction of permanent income. However, their finding is relevant to this suggestion only when age is controlled. When net worth and normal income are compared across age groups, wealth will appear to have an income elasticity above one, even if saving does not. Crockett and Friend do try to control for age, but their age brackets are so broad as to leave the issue in doubt.

A potential test of great interest concerns households just retired or about to retire. If those which had enjoyed larger earned incomes had by this age accumulated relatively larger net worth, simple permanent income models that assume all saving is for retirement would be called into question. The Crockett-Friend findings for retired households do appear to be inconsistent with those models and to suggest an estate motive for saving. But these findings must be interpreted with caution because of the vagaries of "income" reporting for persons already retired.

Other problems in interpreting the apparent high income elasticity of demand for wealth arise from the possibilities that the net worth of high-income households may be disproportionately swollen by inheritances and unrealized capital gains.

In line with much recent work, Crockett and Friend devote considerable attention to the measurement of normal income free from transient components. They use two devices—averaging of several annual incomes reported for the household and averaging of incomes of members of an occupational group. Neither device adds appreciably to the explanatory power of two-year disposable income. However, calculations based on groupings which allow for age are yet to be reported. When this is done, it may be possible to use the age profile of income for people with a given occupation and education in computing their permanent incomes. In principle, normal income should be forward looking not backward looking.

So far the authors' calculations of speeds of adjustment are not very encouraging. It is scarcely surprising that total wealth at the end of 1961 is related to wealth two years earlier. It is disconcerting that wealth at the end of 1959 is not much help in explaining 1960-61 flows. With respect to individual assets, few stocks were available for use in the flow regressions, and these were used only in their "own" regressions.

Crockett and Friend are properly concerned with eliminating spurious relationships due to persistent differences among households in "tastes" —both general thriftiness and preferences for the services of particular consumer assets. As is well known, these differences can obscure stock-flow relationships in cross sections. The authors' device of classifying households by saving behavior in a year not used in the regression does not, on the whole, produce significant results. It would be better, as far as possible, to exploit the panel nature of the data to examine changes in the behavior of identical households.

The paper is a progress report on a large-scale empirical research project, and the main thing a discussant can do is to cheer the authors on their promising line of inquiry. Perhaps it is not too irreverent for this discussant, who has in the past labored in the same field, to remind the authors of the challenge to all of us presented by the near constancy of the aggregate ratio of household saving to disposable income in the U.S. since the Korean War. Have our detailed researches yet provided the forecaster and policy-maker with a better guide than the rule of thumb that 5 per cent of disposable income is saved? Will they give advance warning if and when this rule of thumb is breached? We should gear our research to these questions and not be satisfied with statistical explanations of household differences for their own sake.

On Crockett-Friend and Jorgenson

BY ZVI GRILICHES, UNIVERSITY OF CHICAGO

In the conventional approach, "theory" gives one the demand for capital as a level and comparative statistics tells how much capital stock will be demanded at different relative prices, but from neither can one derive a unique optimal adjustment path from one equilibrium position to another. There are two aspects to this position:

First, defining equilibrium as the stationary solution ($dk/dt = 0$) concedes the possibility that markets are out of equilibrium during the investment process. Given full adjustment to the previous situation, there is no positive net investment unless something (e.g., prices) *changes* and disrupts the previous equilibrium. In this sense, net investment is viewed as a disequilibrium phenomenon.

Second, without adding additional constraints on the possible range of adjustment or a concept of "cost of change," the instantaneous *rate* of investment could be infinite in response to a once-and-for-all shift in the exogenous variables.

Jorgenson's contribution, and it is an important one, is to show that under certain conditions, when prices are and have been changing smoothly, both problems need not arise and it is possible to derive a unique relationship between the rate of investment and the variables influencing it.

If things are continuously and smoothly changing, one may assume that the firm is always in equilibrium—that all marginal conditions are satisfied everywhere along the accumulation path. This allows one to define different paths of accumulation and associate comparative dynamic statements saying, e.g., that accumulations paths differing only in the level of the ruling rate of interest can be characterized by larger or smaller investment rates.

It should be pointed out, though, that the solution to these problems is achieved through a very severe restriction on the scope of the original question. In the Jorgenson model, one cannot answer the question of what happens to the rate of investment if the rate of interest or other prices *shift* to a new permanent level in one move or if a change occurs in depreciation rules. A discontinuous jump to a new accumulation path is not admissible. Since these are the types of questions that Haavelmo and others wanted to answer, solving a more restricted problem, while very useful, does not necessarily imply that they were wrong or that *their* problem has been solved.

The conventional position, having got as far as theory would carry it —to the demand for capital but not for investment, proceeded to "solve" the problem by introducing ad-hoc "partial adjustment," "cost of change," or "liquidity constraint" theories, which explained why and how a particular desired change in capital levels is spread out over a substantial period of time. The theoretical underpinnings of these additions were very weak, but they did force one immediately into a consideration of lags and of a larger list of possible variables, making the theory empirically much more promising and effective.

By limiting himself only to continuous changes, Jorgenson shows that this type of ad-hockery is redundant as far as the original problem is concerned. It can be solved in a smooth world within the original theoretical model without invoking various dubious lag hypotheses. But this may be an illusory gain. To be effective econometrically, the Jorgenson theory will also have to be broadened to include some lag or "cost of change" hypotheses. As of now, it implies that dk/dt (net investment) $= 0$, whenever wages, prices, or interest rates are constant, irrespective of their previous paths. Adding some lag hypothesis did solve the infinite derivative problem in the original model. Since some lag hypothesis will also be necessary in this model, it is not all that clear what will be the final contribution of solving the infinite derivative problem separately. It is clear, though, that the comparative dynamics apparatus developed by Jorgenson will prove very useful in future elaborations of this and similar models. A very important problem still remains unsolved, however: the form and determinants of the optimal adjustment path from one equilibrium position to another. We hope to be able one day to derive it as an implication of our theoretical model, instead of just tacking on something "reasonable" at the end. Showing that these lag hypotheses are not necessary to solve one problem (the derivation of an internally consistent investment function) does not make them any less important.

I have only a brief comment on the Crockett-Friend paper. Their theory should allow for a replacement component of saving, since their saving is gross saving (at least in some of its components). Thus, the coefficient of assets is equal to the difference between the rate of depreciation (replacement) and the rate of adjustment. This may explain why they get, on the face of it, such unreasonably low estimated rates of adjustment. One should add to these the appropriate average maintenance and replacement coefficient associated with the given level of assets.

ON JORGENSON

BY ROGER F. MILLER, UNIVERSITY OF WISCONSIN

Jorgenson's paper does a great deal to expose the misunderstandings at the heart of the controversy on whether or not an investment demand function is derivable from the neoclassical theory of the firm. In brief, the neoclassical theory contains a demand-for-capital-services function; to get capital services, the firm acquires capital assets (or another firm acquires them and rents them to the producing firm); and acquisition of additional capital assets is *defined* to be gross investment. The demand for investment is derived from the demand for capital assets, which in turn is derived from the demand for capital services. There are, thus, three demand functions involved, all intimately related, and either all exist or none exists. The existence of any one is unaffected by the fact that it may be a simple transformation of another in a simple model. Nor is it affected by the fact that it is a derived demand. Most demands *are* "derived"! It may be that there is little point to introducing the *concept* of investment in such a model, but this is a distinct objection unrelated to the controversy.

More consequential is the problem of the continuity and continuous differentiability of the demand-for-capital-assets function. At any point where this function is not continuously differentiable, the investment function becomes discontinuous. If the demand-for-capital-assets function is also discontinuous, the fact that the neoclassical model allows instantaneous adjustments has been interpreted as implying that the amount of investment at such a point is unbounded when expressed as a rate per instant of time. Jorgenson's paper adds nothing to the solution of this problem because he merely finesses the problem completely.

Following the apparent intent of the neoclassicists, Jorgenson makes adjustments instantaneous, and he also *imposes* continuity on the variables he discusses. In particular, his introduction of $\dot{K}(t)$ in (3) and its treatment in the present value maximizing exercise which follows is tantamount to assuming that $\dot{K}(t)$ is continuous and differentiable from the beginning. His later interpretation of condition (3) is less than helpful because it seems to imply that the assumed continuity is a result of the analysis. In condition (3) Jorgenson defines investment as "$I(t) = \dot{K}(t) + \delta K(t)$, where $\dot{K}(t)$ is the time rate of change of the flow of capital services at time t." If $\dot{K}(t)$ is not differentiable or is discontinuous at t, this is inappropriate because $\dot{K}(t)$ is undefined. Jorgenson has simply assumed that such occasions do not arise, and thus sheds no light on this aspect of the controversy.

Jorgenson's contribution is interesting and valuable in spite of his having finessed the unboundedness issue, to which I will return below. It is, however, unfortunate that Jorgenson muddied the waters by discussing, however briefly, the arguments and evidence for aggregate investment functions, which might be very closely approximated by continuous functions even if firm or plant investment functions are not, but which are at best very tenuously related to the microfunctions mentioned in the first paragraph above. Apart from this, I believe the Jorgenson paper is a worthwhile opposite extreme to the case of once-and-for-all adjustment where the capital stock for a given "firm" is a fixed amount.[1] In the latter case, it is clear that in determining the initial (and permanent) capital stock of a given plant, the amount of capital (and thus the amount of investment in *that* enterprise) is nega-

[1] See Vernon L. Smith, "The Theory of Investment and Production," *Quarterly Journal of Economics,* February 1959, and Roger F. Miller, "A Note on the Theory of Investment and Production," *Quarterly Journal of Economics,* November 1959.

tively related to the rate of interest and to the cost of capital assets.[2] Because this conclusion carries through in Jorgenson's instantaneous adjustment model, there is a strong presumption in favor of this relation holding for intermediate lagged adjustment cases.

While I welcome Jorgenson's explication of the neoclassical framework, I feel that one of his contributions is exposing the rather severe restrictions one must impose upon the model in order to deal with some of the important questions of concern to economists. I strongly doubt that the prominent neoclassicists, were they alive and well read today, would find much interest in a model which assumes away uncertainty with regard to the future, lags in adjustment, difficulties of aggregation and composition, discontinuities, etc. Jorgenson's analysis should help to bury this Caesar, as well as praise him.

I believe the day has passed when our analyses have surpassed our observational and computational capabilities. Jorgenson's introductory remarks are to the point. It is just because of this that I think it is unfortunate that Jorgenson chose to sidestep the unboundedness problem. I wish to make it clear that I am not concerned with the "realism" of the model, but with the domain of its application. For most purposes, it may be perfectly satisfactory to regard a fully continuous model as a sufficient approximation to our essentially discrete activities. I strongly suspect that not all purposes are served equally by this approximation, and that for investment timing for a particular firm or individual discontinuities may be of the essence. If this is so, the relevant discontinuities should be recognized and the model constructed so as to allow for them.[3] In the conference discussion it was pointed out that adoption of "period analysis" using discrete time intervals, or of a lagged adjustment function, represent two ways of avoiding the discontinuity problem. Neither of them is fully consistent with the instantaneous-adjustment full-equilibrium framework of the neoclassicists, however, and both merely sidestep the controversy in another dimension. Because both Jorgenson and the discussants thus leave the controversy in an unsatisfactory state, I should like to put forward a few comments and suggestions on dealing with nondifferentiability and discontinuity problems which I hope may resolve the present controversy and have much wider applicability as well.

As a preface to my suggestions, I feel it is necessary to point out that the concept of a function is independent of the concepts of differen-

[2] The result comes from finding $(\vartheta X_2/\vartheta r)$ and $(\vartheta X_2/\vartheta W_2)$ from (68) in Miller (*ibid.*, p. 678).

[3] This is, of course, one of the principal motivations of the model presented in the Miller-Watts paper included in this volume.

tiability, continuity, or boundedness, although the latter properties make functions more tractable to traditional mathematical manipulations. Thus, to say that the investment function is unbounded at any time for which the demand-for-capital-assets function is discontinuous does not imply that the investment function does not exist. However, it does raise questions as to the economic sense of the function as it is defined. We do a disservice to the science of economics (and to the discipline of mathematics as well) if we bind ourselves too rigidly to conventional and convenient mathematical formulations and definitions. I believe this is precisely the heart of the problem in this controversy: it is much ado about nothing, where the nothing in question is the time between t and t (i.e., $dt = 0$). In particular, there is a confusion between the instantaneous time rate of change of capital assets and the quantity of investment which takes place at any instant. The latter is clearly what we are interested in; the former is useful only if it leads to the latter.

Without loss of generality, the ensuing discussion is simplified and clarified by assuming that we have the following function for the quantity of capital assets demanded as a function of the continuous variable t over the interval from $t = 0$ to $t = 4$:

$$\left.\begin{array}{lll} \text{(a)} & K(t) = 2.0t & \text{for } 0 \leq t < 1. \\ \text{(b)} & K(t) = 2.0 - 0.5(t-1) & \text{for } 1 \leq t < 2. \\ \text{(c)} & K(t) = 4.0 + 0.2(t-2) + 0.1(t-2)^2 & \text{for } 2 \leq t \leq 4. \end{array}\right\} \quad \text{(I)}$$

The units in which $K(t)$ is measured are whatever is appropriate for the way K is defined, say, tons of machinery. This yields the following diagram:

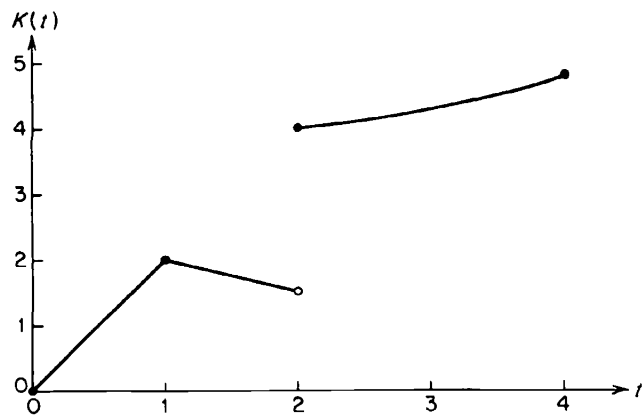

FIGURE 1

Clearly, the amount of net investment that has taken place over any finite interval of t between 0 and 4 is a finite and determinable quantity. This is true despite the fact that the instantaneous time rate of change of the stock of capital assets is unbounded at $t = 2$. For example, if $0 \leq \epsilon \leq 1$, then the cumulative amount of net investment that takes place over the interval from $(2 - \epsilon)$ to $(2 + \epsilon)$ is equal to $2.5 - 0.3\epsilon + 0.1\epsilon^2$ tons of machinery. This is simply derived by subtracting $K(2 - \epsilon)$ found in (b) from $K(2 + \epsilon)$ found in (c). As $\epsilon \to 0$, this converges on $I(2) = 2.5$, however, which understates the actual amount of investment taking place at $t = 2$.

To develop the correct formulation of the investment function which can be applied to demand-for-capital-assets functions of this type, it is convenient to start with Jorgenson's definition of gross investment at t:

$$I(t) = \delta K(t) + \dot{K}(t), \tag{3}$$

where δ is a positive fraction representing the quantity of capital assets which disappear through depreciation. This investment function serves perfectly well for any instant except $t = 1$ or $t = 2$ in our example above.

(a) $\dot{K}(t)$ is not defined at either of these critical values. The economic sense of this term, however, is the amount of additional K demanded to provide for immediate future production, so that we are only interested in the right-hand derivatives of $K(t)$ with respect to time. We may define such a right-hand derivative as $\lim_{\epsilon \to 0} [\dot{K}(t + \epsilon)]$ and substitute this for $\dot{K}(t)$ in the expression for $I(t)$, removing this difficulty.

(b) At $t = 2$ we face another difficulty with respect to depreciation. At any t, depreciation applies to the pre-existing stock of capital assets, not to the amount being newly acquired (otherwise $\delta K(t)$ would have to be included in (3) above as a third term). To capture this feature, consider $K(t - \epsilon)$ as a sequence and find $\lim_{\epsilon \to 0} [\delta K(t - \epsilon)]$ as a replacement for the first term in (3) above. At $t = 2$ this limit is 1.5δ, not 4.0δ.

(c) Finally, at $t = 2$ there is nothing in (3) to capture the instantaneous jump from $K = 1.5$ to $K = 4.0$. This can be remedied in the same manner as the depreciation technique by including in $I(t)$ the $\lim_{\epsilon \to 0} [K(t + \epsilon) - K(t - \epsilon)]$, which in our example is $4.0 - 1.5 = 2.5$.

The modifications of the preceding paragraph, plus the recognition that $\dot{K}(t) \geq 0$, yield the following gross investment function:

$$\left.\begin{aligned} &\text{(a)}\ I(t) = \lim_{\epsilon \to 0} \delta K(t - \epsilon) + \lim_{\epsilon \to 0} \dot{K}(t + \epsilon) + \lim_{\epsilon \to 0} [K(t + \epsilon) - K(t - \epsilon)] \\ &\text{(b)}\ I(t) \geq \lim_{\epsilon \to 0} (\delta - 1)K(t - \epsilon), \end{aligned}\right\} \text{(II)}$$

where (IIb) overrides (IIa) in case of a conflict, and merely says that it is impossible to disinvest more capital than is available. Applied to our example, the gross investment function is, with arguments ordered as in (IIa) above:

$$\left.\begin{array}{lll}
\text{(a)} & I(t) = 2.0\delta t + 2.0 + 0.0 & \text{for } 0 \leq t < 1 \\
\text{(b)} & I(t) = \delta(2.5 - 0.5t) + (-0.5) + 0.0 & \text{for } 1 \leq t < 2 \\
\text{(c)} & I(t) = 1.5\delta + 0.2 + 2.5 & \text{for } t = 2 \\
\text{(d)} & I(t) = \delta(4.0 - 0.2t + 0.1t^2) + (0.2t - 0.2) + 0.0 & \text{for } 2 < t \leq 4
\end{array}\right\} \tag{III}$$

Notice that the third term is always zero except where $K(t)$ is discontinuous. The resulting diagram for *net* investment ($I(t)$ less depreciation) is:

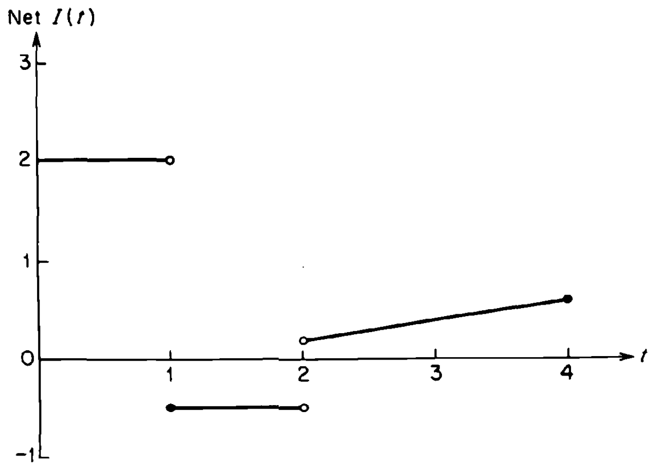

FIGURE 2

The "limiting" processes I have introduced above are simply rules for finding which numbers are the appropriate ones to enter into the function at a given t. As such, they are matters of definition and should not be confused with the distinct limiting process which is involved in defining a derivative.[4]

[4] Furthermore, the investment function defined in (II) is Stieltjes-integrable back to the demand-for-capital-assets function (given the appropriate constants of integration) if we assume that in the neighborhood of any point of nondifferentiability ($t = 1$) or discontinuity ($t = 2$) of the demand-for-capital-assets function there exists some interval including that point over which the function is continuous and differentiable. The relative unfamiliarity of Stieltjes-integration (as opposed to the more common Reimann-integral) is a mathematical, and not an economic, consideration.

The question remains in what units $I(t)$ is expressed. This is not a trivial question since it is not obvious that the third term has the same time dimensionality as the first two. The appearance of the terms in an equation can be deceptive, however, since any term can have a coefficient (necessarily equal to one and therefore not apparent) which is expressed in units appropriate to make the term have the desired units. In Jorgenson's formulation, since $\dot{K}(t)$ is a time derivative, it has (by itself) units of capital assets flowing per instant of time. If no other coefficient is added, $I(t)$ and all other terms must have the same units. This requires, for example, that δ be defined as the fraction of existing assets that disappear (flow away) per instant of time due to depreciation. Both terms, and the corresponding terms in my (IIa) and (IIb) above, represents an amount of capital assets per instant (e.g., tons of machinery per instant) such that, *if* continued at a constant level over the interval from t to $t + 1$, the total change in the stock of capital assets *would* exactly equal the sum of the terms in the equation. The third term in my formulation has exactly the same interpretation: it is the change in the stock of capital assets that takes the form of a discrete jump *at the instant t,* and is thus an instantaneous rate in the same sense as the other terms. In the example above, notice that the rate of net investment at $t = 2$ is 2.7. If that rate of net investment were to continue constant at that level over the interval from $t = 2$ to $t = 3$, the stock of capital assets would increase by precisely 2.7 tons of machinery (from 4.0 to 6.7), and the demand-for-capital-assets function (*if* that rate of net investment were maintained) would have to be modified accordingly to be:

$$K(t) = 4.0 + 2.7(t - 2) \text{ for } 2 \leq t \leq 3. \tag{Ic$'$}$$

In this case, of course, we would also have

$$\lim_{\epsilon \to 0} \dot{K}(t + \epsilon) = 2.7 \text{ for } 2 \leq t < 3.$$

I can see no mathematical or economic objections to the manner in which I have redefined the investment function. I would not have pursued it to this extent if I did not feel that the technique employed was sufficiently useful and unknown to make its exposition a useful contribution per se. In addition, it should lay to rest the unfortunate controversy over whether or not a sensible investment function is derivable from the neoclassical model of the firm. My investment function may not be so easy to manipulate as a continuous and differentiable one, but that is a small matter of mathematics and not a fundamental matter of economics. As redefined in this comment, the investment function frees us from the necessity of assuming continuity of prices or of capital services while allowing the retention of the assumption of instantaneous adjustment to new optimal levels of capital services input. This may be a small gain attained at a high price. If so, it is only because we are slavishly pursuing the letter rather than the spirit of neoclassical economics.

Reply to Tobin and Griliches

JEAN CROCKETT AND IRWIN FRIEND

To begin with Tobin's last question as to whether our detailed microeconomic studies provide better forecasting devices for aggregate personal saving than the "7 per cent rule," we have several rather obvious answers to make. We agree that sophisticated models now in existence probably could not have given more accurate predictions of saving over the last twelve years than that saving would be 7 per cent of disposable income. However, we hope that our models will eventually be able to improve on this rule, since the saving-income ratio has departed substantially from 7 per cent within the memory of man and is quite likely to do so again.

The interesting stability of the ratio in recent years may be the product of offsets among the effects of a number of changing variables. For example, the increasing proportion of retired with their relatively low savings ratios may offset the increasing proportion of homeowners with their relatively high savings ratios; or the increased economic confidence which has made households willing to assume continually increasing amounts of indebtedness relative to disposable income—a process which can hardly go on indefinitely—may offset a natural tendency for the savings ratio to rise with income. Even if such offsets are not the explanation of the recent stability, there are still many savings-income functions (including our own) which may give approximate constancy of the savings ratio over a particular income range but which would have quite different implications for higher incomes. If the normal income elasticity of assets and savings is significantly above one, as our analysis strongly implies, the constant savings ratio cannot be expected to persist except through other influences offsetting the income effects. Our analysis, if it is correct, gives insights into the implications of alternative economic policies which cannot be obtained from observation of the approximate constancy of the savings ratio.

It is our belief that the best path to an adequate understanding of aggregate saving behavior involves two steps: (1) the development and estimation of a satisfactory microeconomic model, toward which we believe that we have made some progress in this paper, and (2) the development of aggregate forecasting procedures based on the microeconomic parameters. The second is a far from trivial problem to which we have hardly addressed ourselves here, except insofar as we have tried to free our estimated income effects from biases due to the correlation of other cross-sectional variables with income. In addition to this, it is necessary to solve the aggregation problem and to allow for the influence of factors which are variable over time but whose effects cannot be determined in the cross section.

Quite apart from the question of forecasting aggregate savings, the present kind of investigation of the size and composition of household portfolios has implications for the capital markets, since consumers are very important elements in the supply of and demand for various types of funds. We find it rather amusing that Tobin is concerned with the implications of a "constant" ratio of aggregate personal saving to disposable income in recent years for our "detailed researches" in this area, without experiencing or at least expressing a similar concern about the corresponding implications of a "constant" investment-income ratio.

As to the more specific criticisms which Tobin makes of our paper, he first argues that our model is too static since we do not allow for changes in desired asset stocks over time. We have specifically allowed for changes in desired stocks when normal income varies, as it must if it is based on anything less than expected lifetime income, and even then if expectations are revised as additional information is accumulated. In addition, we entirely agree with Tobin that desired asset stocks also change with age. This is implicit in the balancing of the utility of an extra dollar of consumption against the utility of the present and discounted future services of an extra dollar of assets, particularly for assets whose major services occur in the future, since the discounted value of such services rises over time. While we did incorporate age as an explanatory variable for desired assets in the preliminary version of our paper to which Tobin's comments refer, we have made much greater use of age in the present version than we were able to do earlier. The various techniques we have used for holding age constant do not improve our empirical estimates of adjustment speeds in linear regressions, and these are in any case quite reasonable for the logarithmic regressions.

Second, Tobin suggests that we have not adequately controlled for age in arriving at the conclusion that the income elasticity of net worth is greater than one. In the present version of our paper, age is controlled by (a) including age as a continuous variable in net worth regressions, (b) fitting separate regressions within four age groups, and (c) including age as a continuous variable in the regressions within age groups to take care of the possibility of strong but nonlinear age effects. Our results have not been altered in any significant way by this extension of our earlier analysis.

Third, Tobin mentions the possible problems introduced by disproportionately high inheritances and unrealized capital gains for the upper-income groups in interpreting the high income elasticity of demand for wealth. Disregarding the effect of capital gains, we do not see how inheritances per se could result in an upward bias in the estimated income elasticity of wealth if there is a unitary income elasticity of demand for saving. However, capital gains do pose a problem which we considered in the original version of our paper. In addition to the evidence presented there that this problem does not seriously affect our conclusion on the income elasticities for wealth and saving, we have introduced a crude proxy for capital gains in the present version of our net worth regressions, and while this reduces income elasticities slightly, they remain well above unity. The capital gains proxy also improves somewhat the estimated adjustment speeds in the linear regressions.

Fourth, Tobin criticizes our saving tastes variables and suggests that it would be better to hold tastes constant by considering changes in the behavior of individual households over time. Here we agree entirely with the desirability of such an approach and had pointed this out in our paper. We were greatly disappointed that the body of data which we analyzed did not permit the use of this approach. Data for two distinct time periods were available only for three items, and even here the periods were too close together to produce much change in normal income and thus permit accurate estimation of a normal income elasticity. We hope to utilize the 1950-60 BLS consumer expenditures data to study changes in the saving behavior of socioeconomic or other groups over a ten-year period, somewhat in the manner of Duesenberry and Kistin. One of the authors has already used this approach in a forthcoming analysis of the aggregate postwar data for different countries, the other in an analysis of Greek household expenditure data.

Finally, Tobin notes that neither our separation of income into normal and transitory components nor our introduction of initial asset levels adds much to our correlations. This is not true for total assets, where the introduction of transitory income (as well as initial assets) improves the correlations and has a significantly different impact from normal income. Unfortunately, Tobin's caveat is true for total saving, although it should be emphasized that the primary reason for both of these devices was to produce (we hoped) a relatively unbiased estimate of the income elasticity rather than to raise correlations. Thus, turning to the major components of saving, we find that the effect of normal income on contractual saving is significantly higher than that of transitory income in the linear regressions for employees, even though the separation of the two effects does not raise the correlation, while in the quadratic regressions for liquid saving (which provide much the best fit) both transitory income and the second-degree term in normal income are highly significant for employees and the self-employed. Furthermore, for both groups, lagged assets are highly significant and raise the correlations for liquid saving in both the linear and quadratic models, though the implied adjustment speeds are rather low for employees.

For mortgage debt also, a very important savings component for homeowners, the introduction of initial debt levels raises the correlations; and since there was some tendency to increase mortgages, even though no purchases of new homes were involved, this is not quite so mechanical as it may seem. Incidentally, the comment that assets stocks were used only in their "own" savings regressions is not quite correct. Total net worth was frequently included, in addition to specific asset stocks, to represent all other assets, but did not prove significant or add to the correlation.

As to Griliches' comment that our estimated speed of adjustment for net worth may be understated because of our failure to allow for depreciation in housing, he is quite correct if we wish to consider our regressions as referring to total saving rather than merely to saving in the form of financial assets and if we consider only the saving but not the assets regression; but the adjustment is not quantitatively important even for the total saving regressions. With a depreciation rate of .035 per year for housing, which seems high but is used by Muth in the study discussed in our paper, and with a value of house estimated to account for about one-third of total net worth, depreciation should amount to perhaps 1 per cent of net worth. Thus .01 should be added to the estimated speed of adjustment in the total saving regressions. However, in the assets or net worth regressions, it is not necessary to make any adjustment of this type since assets are measured at a market value rather than on an undepreciated cost basis.

Reply to Tobin, Griliches, and Miller

by Jorgenson

The comments by Griliches, Miller, and Tobin should convince even the most blasé observer that the theory of investment behavior is a difficult and far from settled branch of economic theory. Even within the extremely simple framework I have used, elementary confusions arise, ambiguities persist, issues remain unresolved.

Tobin is correct in pointing out that there is no reason to assume that markets will never present an individual firm with jumps in prices. But it would be equally correct to say that there is no reason *not* to assume that firms will never be presented with jumps. The selection of an appropriate assumption is entirely a matter of analytical convenience. If jumps have interesting consequences, these consequences should be studied and tested against data. If continuity of prices has interesting consequences, these consequences are equally deserving of study.

In the theory of investment behavior, the assumption of jumps in price levels rules out any consequences at all. On the other hand, the assumption of continuous price levels has interesting and unsuspected consequences, namely, a rigorous theory of investment behavior based on the neoclassical theory of optimal capital accumulation. Keynesians receive the additional benefit of a "correct" sign for the change in investment with respect to variations in the rate of interest. In view of these consequences, it is difficult to interpret Tobin's remark to the effect that the resulting investment demand schedule "cannot serve the analytical purposes for which such a schedule is desired" as anything but a simple misunderstanding.

To sum up, the answer to the question whether demand for investment goods is a function of the rate of interest is that it all depends on what you hold constant. If Tobin insists on holding constant all present and *future* prices of investment goods (while varying the rate of interest), investment is unbounded except for a single value of the price of capital services. On the other hand, if present and *forward* prices of investment goods are held constant, there exists a perfectly well-defined investment demand function that depends on the rate of interest.[1] Tobin follows Haavelmo and Lerner in identifying two separate questions: (1) Is demand for investment goods a function of the rate of interest? (2) What happens to investment when the rate of interest varies with all present and *future* (not *forward*) prices held constant? Only when it is realized that there is no necessary connection between the two questions can a complete and unambiguous answer to the first question be given.

I intended the theory of investment behavior developed in my paper and econometric work on investment to be less directly related than Griliches supposes. Two different theoretical positions are commonly employed to rationalize empirical work. One is based on the Keynesian marginal efficiency of investment schedule, and the other on a theory of demand for capital services.

In view of the previous literature on the theory of investment, it may be surprising that both these positions can be developed within the same theoretical framework. Now that this fact has been demonstrated, tests to discriminate between the two approaches can be undertaken. As Griliches suggests, in empirical applications both positions are associated with substantial ad-hockery. Before the two positions can be tested against each other in any definitive way, it will be necessary to reduce the ad-hockery in each.

Miller's suggested modification of my theoretical framework is based on an unfortunate slip. The problem is one of appropriate dimensions. Using discrete time, we often write something like:

$$K_{t+1} = I_t + (1 - \delta)K_t.$$

A relationship like this can also be written using continuous time:

$$\frac{K(t + \epsilon) - K(t)}{\epsilon} = I(t) - \delta K(t),$$

where $\epsilon = 1$. When we employ such a relationship only at discrete points of time—t, $t + 1$, and so on—the time interval, $\epsilon = 1$, may be suppressed. However, where we pass to continuous time, letting $\epsilon \rightarrow 0$, it is important

[1] Tobin asserts that "there is no reason to assume that expected prices of capital goods accommodate themselves so obligingly to interest rate variations." In the conventional approach, one might argue similarly that there is no reason to assume that present and future prices obligingly hold themselves constant. Both of these arguments are beside the point. The investment demand schedule, like most economic relationships, is based on conjectural variation. Real income does not obligingly stay constant while we study changes in the demand for a commodity resulting from changes in its price. We hold it constant *by assumption*. Similarly, in studying investment demand, we hold whatever is held constant to be constant *by assumption*. Needless to say, changing the assumption usually changes the results.

to make the time interval explicit. The dimensions of the left-hand side variable are units of investment goods *per period of time;* these units correspond to those of $I(t)$ and $\delta K(t)$, both of which are measured as investment goods *per period of time*. Now taking the limit:

$$\lim_{\epsilon \to 0} \frac{K(t + \epsilon) - K(t)}{\epsilon} = \dot{K}(t) = I(t) - \delta K(t),$$

we obtain quantities which are still measured as investment goods *per period of time*.

The difficulty with Miller's expression II (a) is that the quantity $K(t + \epsilon) - K(t - \epsilon)$ is measured in investment goods, not investment goods per period of time. The appropriate expression is $[K(t + \epsilon) - K(t - \epsilon)]/2\epsilon$, since 2ϵ, the time interval, is measured in units of time and the ratio is measured as investment goods *per period of time*. Thus, Miller adds investment goods, a stock, to investment goods per period of time, a flow, which is self-contradictory. This is an elementary point, but it is essential to a correct formulation of the continuous time version of the basic relationship between gross and net investment. Miller's results are vitiated by this error.

PART II

Financial Aspects

Estimates of the Cost of Capital Relevant for Investment Decisions Under Uncertainty

MERTON H. MILLER
UNIVERSITY OF CHICAGO
AND
FRANCO MODIGLIANI
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

For discussion at the Conference, we distributed a lengthy and detailed paper in which we attempted to develop methods for estimating the cost of capital relevant for investment decisions under uncertainty and to apply these methods to a cross-sectional sample of large electric utilities for 1954, 1956, and 1957. Much of this material had a direct and important bearing on the subject of the conference, particularly the latter sections of the paper which contrasted our estimates with several of the alternative measures of the cost of capital currently used in empirical studies of investment behavior.

Other parts of the paper, however—especially the review of the underlying theory of valuation under uncertainty, the discussion of the various theoretical and practical problems involved in the estimation, and the fairly extensive testing of the basic specification—were clearly of less direct concern to the Conference's central theme, except as supporting material for a critical evaluation by the discussants and those with a direct interest in the area of finance.

In revising our paper for inclusion in this volume, therefore, we decided to confine ourselves primarily to those portions of the original document most closely related to the central purpose of the Conference. Thus we have focused on our specific estimates of the cost of capital, on comparisons with alternative measures, and on the problems inherent in trying to develop continuous historical series. The remaining portions have been summarized and cut down to the bare minimum necessary for explaining and interpreting the results. Readers interested in a fuller development are referred to the *American Economic Review* for June 1966, which contains an unabridged, though slightly revised, version of the original paper (referred to hereafter as the "unabridged version").

I. Introduction

In its simplest form, the central normative proposition of the micro-theory of capital is that the firm should adjust its capital stock until the marginal rate of return on further investment (or disinvestment) is equal to the cost of capital. Under conditions of perfect certainty—which is the assumption on which most of classical theory has been developed—the concept of the cost of capital presents no particular difficulty; it is simply the market rate of interest. Since all securities must have the same yield in equilibrium under certainty, there is only one such rate per period and it is, in principle, a directly observable magnitude. Under real world conditions, however, we are confronted not with one, but with a bewildering variety of securities, with very different kinds and priorities of claims to portions of the (uncertain) future earnings of the firm. Since these securities will also, in general, have different anticipated yields, it is by no means clear which yield or combination of yields is the relevant cost of capital for rational investment planning. Nor, because it is based on anticipations, is the cost of capital any longer a directly observable magnitude. It must, rather, somehow be inferred from what *is* observable, namely, the market prices of the various kinds of claims represented by the different securities.

Although most (but not all) recent studies of investment behavior have shown some awareness of these difficulties, a common approach in empirical work has been simply to ignore the problem and to use, without comment or explicit justification, some standard index of current, nominal yields on high-grade corporate bonds (or even government bonds) as a measure of the cost of capital. Other writers use both a series on current bond yields to represent the cost of debt capital and a current profit series to measure the "availability" and hence, presumably, also the "cost" of equity capital. Still others have tried indexes of share prices, current dividend yields, or current earnings yields alone or in various averages with bond yields along the lines suggested in the standard texts on corporation finance. How much error is involved in the use of such measures is still unknown, though even a cursory survey of the underlying theory suggests many grounds for apprehension on this score; but we cannot be sure. Too little work has yet been done to permit even a rough calibration of these series as proxies for the cost of capital, let alone to provide acceptable alternative series.

The results that follow should be thought of as first steps toward closing this gap in our understanding and measurement of the cost of funds relevant to investment decisions. They are only first steps partly because of the very limited coverage of the estimates (three years for one industry)[1] and partly because the underlying model from which the estimates are derived is a special and still incompletely tested one. In particular, the interpretation of our estimates as the "cost of capital" rests on the assumptions of perfect capital markets and rational behavior by investors and by the corporate managers responsible for the actual investment decisions. Neither of these assumptions, needless to say, is likely to be enthusiastically accepted by those working in this field. In their defense, however, the following points are perhaps worth noting. (1) Our concern in this paper (and in the series of earlier papers on which it builds) is almost entirely with large, well-established firms. Though relatively few in number, these firms account for a disproportionately large share of total investment and in some major industries (such as our utilities or the steel industry) for virtually all of it. For such firms, we feel that the assumption of perfect capital markets—which implies, among other things, that, over the relevant range of funds requirements and except for very short intervals of time, there is no constraint on the total of funds from all sources that a firm can obtain at the going "cost of capital" to finance its investment outlays—cannot be ruled out as implausible, at least to a first approximation. For smaller firms, on the other hand, which are known to face severe limitations in their ability to expand their equity capital, the assumption may be largely inappropriate, and we would regard other kinds of models (stressing "availability" considerations) as more promising. (2) Insofar as rational behavior is concerned, the great virtue in that assumption is that it leads to a direct and simple connection between

[1] The sample consists of sixty-three separate firms representing all of the (consolidated) systems classified as of 1950 as Class A by the Federal Power Commission; plus those of the smaller Class B systems, whose assets devoted to electricity generation were at least $15 million in 1950. The sample years are 1954, 1956, and 1957.

the cost of capital and market valuation. To some, of course, such a defense will smack of looking for a missing wallet under the lamppost because the light is better there. But we do not yet know that the wallet is not there! There will be time enough to try working with more complicated behavioral assumptions when the evidence shows that the assumption of rationality fails (and precisely how it fails). As it turns out, some of the implications of the assumption of rational investor behavior stand up quite well when confronted with the data (see the unabridged version). The issue of the rationality of investment policy by corporate management is a more delicate one, and the final answer will have to wait until estimates of the cost of capital of the kind developed here have been tried out in studies of investment behavior.

With tests of this sort kept in mind as the ultimate goal, we turn now to the immediate task at hand. We shall begin in section II by providing an operational definition of the cost of capital and developing therefrom a link between the cost of capital and market valuation. The perfect market and rational behavior model of valuation will then be sketched out and the basic estimates of market-required rates of return and their relation to the cost of capital will be presented along with a very brief account of the estimating methods employed. Section III compares the estimates of the cost of capital with various popular alternative measures, with particular emphasis on average yield measures of the kind emphasized in the conventional literature on corporation finance. We conclude in section IV with some tentative suggestions on the problem of developing time series estimates of the cost of capital.

II. Valuation and the Cost of Capital

As used throughout, the term cost of capital (C) will be taken to mean the minimum prospective rate of yield that a proposed investment in real assets must offer to be worthwhile undertaking from the standpoint of the current owners of the firm. Under conditions of perfect capital markets, there is a one-for-one correspondence between "worthwhileness" in the above sense and the current market value of the owners' interest. If the management of the firm takes as its working criterion for investment (and other) decisions, "maximize the market value of the shares held by the current owners of the firm," then it can be shown (see, e.g., Hirshleifer [6])[2] that this policy is also equivalent to maximizing the economic welfare or utility of the owners. Thus, under the assumptions, valuation and the cost of capital are intimately related.

[2] Numbers in brackets refer to Bibliography at end of paper.

1. THE SIMPLE CERTAINTY MODEL

The precise relation between them is most easily seen in the context of a simple certainty model, in which all real assets are assumed to yield uniform, sure income streams in perpetuity and in which the market rate of interest (r) is given and constant over time. If, in addition, we assume perfect capital markets, rational investor behavior, and no tax differentials on different sources of income, then it can readily be shown that the equilibrium current market value (V) of any firm (i.e., the sum of the market values of all securities or other claims to its future earnings) is given by

$$V = \frac{1}{r}X \tag{1}$$

where X is the (uniform) income per period generated in perpetuity by the assets presently held. The term $1/r$ in (1)—the reciprocal of the interest rate—is commonly referred to as the market "capitalization rate" for sure streams since it represents the factor the market applies to a unit income flow to convert it to a capital stock.

For an expansion of real assets to be worth undertaking from the standpoint of the current owners of such a firm, the investment must lead to an increase in the market value of their holdings. If we let dA equal the purchase cost of the assets acquired, dS^o the change in value of the holdings of the original owners, and dS^n the market value of the additional securities issued to finance the investment, then differentiating (1) with respect to A yields:

$$\frac{dV}{dA} = \frac{dS^o}{dA} + \frac{dS^n}{dA} = \frac{dS^o}{dA} + 1 = \frac{dX}{dA}\frac{1}{r}. \tag{2}$$

It follows that the cost of capital (C) must be the reciprocal of the market capitalization rate for earnings since from (2), $dS^o/dA \geq 0$ if and only if $dX/dA \geq r$, i.e., if and only if the rate of return on the new investment is equal to or greater than the market rate of interest.[3]

2. EXTENSION TO THE CASE OF UNCERTAINTY

When we turn from a world of certainty to one of uncertainty, the problem of relating the cost of capital to market valuation becomes a

[3] We have here stated (and shall continue to state) the conditions for optimality of investment decisions in terms of the rate of return or internal yield on investment. Although it is well known that there may be cases in which such a rate of return cannot be adequately or unambiguously defined (see, e.g., Hirshleifer [6]), such cases are largely ruled out by our additional simplifying assumption.

much more formidable one for which no completely general solution is yet available. We have at least been able to show, however (see [10]), that if we retain the assumptions of perpetual streams, rational investor behavior, perfect markets, no taxes (and no "growth" in a sense to be more precisely defined later), then an analog of the certainty valuation formula (1) does carry through to the case of uncertainty. In particular, if we restrict attention to what we have called a "risk equivalent class" of firms, then the equilibrium market valuation of any firm in such a class can be expressed as

$$V = \frac{1}{\rho_k}\bar{X} \tag{3}$$

for all firms in class k, where V is the sum of the market values of all the firm's securities, $\bar{X}$ is the *expected* level of average annual earnings generated by the assets it currently holds, and $1/\rho_k$ can be interpreted as the market's capitalization rate for the expected value of uncertain, pure equity earnings streams of the type characteristic of class k. Hence, by a straightforward extension of the reasoning in the previous section, the cost of capital for a proposed expansion in scale by any firm in the class is simply ρ_k. The precise value of ρ_k will, of course, be different from class to class, presumably increasing with the market's uncertainty about the level of future long-run earnings in the class (and reflecting also the nature and extent of the covariation with the returns in other classes). Though the various ρ's themselves are not directly observable, they can, in principle, be inferred from the market valuations, e.g., by regressing the observed V on estimates of $\bar{X}$ over a cross section of firms within a class.

An important implication of (3) is that the market value of a firm depends only on its real earning power and on the market capitalization rate for pure equity streams of its class and not at all upon the particular mix of security types that characterize its financial structure. This independence of value and financial structure is basically a reflection of the assumption of perfect capital markets—an assumption implying, among other things, that for comparable collateral, the supply curve of borrowed funds for individuals is the same as that for corporations. Hence, if corporations making heavy use of borrowed funds should sell, say, at a premium relative to unlevered corporations in the same class, rational investors could always obtain a more efficient portfolio by selling the "overvalued" levered shares, purchasing the "undervalued" unlevered shares, and restoring the previous degree of leverage by borrowing against the shares on personal account. And the converse is true if levered shares should sell at a discount, in which case the "arbitrage" operation involves selling the unlevered shares, buying the levered shares, and unlevering them by also buying a pro rata share of the firm's debts.[4]

With reference to the cost of capital, the independence of market value and financial policy implies, of course, that the cost of capital relevant for investment decisions is also independent of how the investment is to be financed, even though the particular securities considered may, and in general will, have very different expected yields. This seeming paradox disappears as soon as it is recognized that the independence property also requires that the common shares in levered corporations have higher expected yields than those of less levered corporations in the same class—a differential which can be thought of as compensation for the greater "riskiness" attaching to levered shares. Thus, the apparent gain in terms of the cost of capital coming from the ability of a firm to finance an investment with "cheap" debt capital is offset (and with rational behavior in a perfect market exactly offset) by the correspondingly higher cost of equity capital.

3. THE EFFECT OF CORPORATE INCOME TAXES

When we extend the analysis to allow for the existence of corporate income taxes and the deductibility of interest payments, the picture changes in a number of respects, the most important being that market value and financial structure are no longer completely independent. To see what is involved, let us again denote by $1/\rho_k$ the capitalization rate in a given class for pure equity streams available to investors (i.e., streams of expected net profits after taxes in unlevered firms), by $\bar{X}$ a firm's expected total earnings (now to be taken as earnings before taxes as well as interest), and by τ the (constant) marginal and average rate of corporate income taxation. Then the market value of an unlevered firm can be expressed as:

$$V_u = S_u = \frac{\bar{X}(1 - \tau)}{\rho_k}, \tag{4}$$

where $\bar{X}(1 - \tau)$ is the unlevered firm's earnings after taxes. The value of a levered firm with D of debt or other securities whose payments are tax

[4] A fuller account of the arbitrage mechanism and proof of the independence proposition is given in our [10]. It is perhaps worth noting that the independence proposition can be proved under assumptions much weaker than those necessary to develop equation (3). In particular, neither the perpetuity assumption nor the concept of a risk equivalent class is essential (see, e.g., the discussion in our [9], pp. 429–430, and also Hirshleifer [5]).

deductible and P of preferred stock or other nondeductible senior securities can be shown to be (see our [12]):

$$V \equiv S + D + P = \frac{\bar{X}(1-\tau)}{\rho_k} + \tau D. \tag{5}$$

Note that in (5), the expression $\bar{X}(1-\tau)$ no longer represents the firm's earnings after taxes or any other standard accounting concept when D is not zero and hence when $\bar{X}$ includes some tax deductible interest. To avoid confusion, therefore, we shall hereafter refer to $\bar{X}(1-\tau)$ as "tax adjusted" earnings using the symbol $\bar{X}^\tau$ for earnings after taxes in the ordinary accounting sense (i.e., for the sum of expected net profits after taxes, preferred dividends, and interest payments as they actually come on to the market for sale to the various security purchasers).[5]

As to the meaning of (5), it says, in effect, that the government pays a subsidy to firms using certain sources of capital which under current law would include bonds, notes, and other firm contractual obligations of indebtedness but not preferred stocks (with some minor exceptions) or common stocks. The addition to the present worth of the firm occasioned by these tax savings is the corporate tax rate times the market value of the debt—the latter being, of course, the present worth, as judged by the market, of the future stream of tax deductible payments.

Since the deductibility of interest payments thus makes the value of the firm a function of its financial policy, it must also make the required yield or cost of capital a function of financial policy. To see the precise nature of this dependence, let dS^o, as before, stand for the change in the market value of the shares held by the current owners of the firm, dS^n for the value of any new common shares issued, dP for the value of any new preferred stock issued, and dD for the value of any new tax-deductible debt issued, with $dS^n + dP + dD = dA$. Then from (5) we have:

$$\frac{dV}{dA} = \frac{dS^o}{dA} + \frac{dS^n}{dA} + \frac{dP}{dA} + \frac{dD}{dA} = \frac{dS^o}{dA} + 1 = \frac{d\bar{X}(1-\tau)}{dA}\frac{1}{\rho_k} + \tau\frac{dD}{dA}, \tag{6}$$

from which it follows that the cost of capital or required yield on a tax-adjusted basis is

$$C = \rho_k\left(1 - \tau\frac{dD}{dA}\right) \tag{7}$$

[5] The relation between the various concepts can easily be established by observing that expected taxes paid will be $\tau(\bar{X} - \bar{R})$ so that $\bar{X} = \bar{X}^\tau + \tau(\bar{X} - \bar{R})$ and hence $\bar{X}(1-\tau) = \bar{X}^\tau - \tau\bar{R}$. Further discussion of these concepts along with the basic proofs for the tax case will be found in our [12].

since $(dS^o/dA) \geq 0$ if and only if $(d\bar{X}(1-\tau)/dA)$ is equal to or greater than the right-hand side of (7).

In connection with (7), the two extreme cases of financing methods are of particular interest. For an investment financed entirely by equity capital—and in this context equity capital includes nondeductible preferred stock—dD/dA will equal zero. Hence the required tax-adjusted yield or "marginal cost of equity capital" is ρ_k. For an investment financed entirely by debt or other sources of capital whose payments are tax deductible, dD/dA is unity, implying that the "marginal cost of debt capital" is $\rho_k(1-\tau)$.

The term marginal cost has been put in quotation marks to emphasize that, while these extreme cases serve to illuminate the meaning of (7), neither is directly relevant for actual decision-making at the level of the firm. For companies with reasonable access to the capital markets, as would certainly be true of those in our sample, investment and financing decisions (including decisions to retire outstanding securities) are made continually and largely independently. Since particular investment projects thus are not, and in general cannot be, linked to particular sources of financing, the relevant cost of capital to the firm must be thought of as essentially an average of the above costs of debt and equity capital, with weights determined by the long-run average proportions of each in the firm's program of future financing. If we denote this "target" proportion of debt as L, then the weighted average cost of capital can be expressed as $C = C(L) = \rho_k(1-\tau)L + (1-L)\rho_k$, or, more compactly, as $C = C(L) = \rho_k(1-\tau L)$, where the notation $C(L)$ will be used when we want to emphasize that the cost of capital is a *function* of the target debt ratio L.

Notice, finally, that while the definition of the cost of capital has become a good deal more complex as a result of the introduction of income taxes, the problem of estimation remains essentially the same. It still involves only the estimation of a single capitalization rate—in this case, $1/\rho_k$, the capitalization rate for unlevered, pure equity streams in the class. The difference between the cost of equity and debt capital introduces no new difficulties because the cost of debt capital does not depend on the market rate of interest on bonds, but only on the above capitalization rate and the tax rate. Hence $1/\rho_k$ remains a sufficient parameter both for economists seeking to explain rational investment behavior and for firms planning their investment programs on the basis of given financial policies.[6]

4. GROWTH AND VALUATION[7]

Up to this point, we have focused attention entirely on the role of current earning power and financial policy as determinants of the value of the firm. There are, of course, very many other factors that influence real world valuations and some that may well be large enough and systematic enough to warrant incorporating them directly into the model rather than impounding them in the general disturbance term. Of these, one of the most important is "growth potential," in the sense of opportunities the firm may have to invest in real assets in the future at rates of return greater than "normal" (i.e., greater than the cost of capital).

Clearly, translating such a concept into operational terms is a task of formidable proportions and one for which no universally applicable solution can be expected. For industries such as the electric utilities, however—where the growth in earnings has been (and will presumably continue to be) reasonably steady—rough, but tolerable, approximations to growth potential can probably be obtained by exploiting the so-called constant-growth model. In particular, suppose that a firm has the opportunity to invest annually an amount equal to 100 k per cent of its tax-adjusted earnings ($k \lesssim 1$), on which investments it will earn a tax-adjusted rate of return of ρ^*, greater than $C = C(L)$, its average cost of capital. (These assumptions imply, among other things, that earnings will grow at the constant rate of $k\rho^*$ per year.) And suppose further that these especially profitable opportunities are expected to persist over the next T years, after which only normally profitable opportunities will be available. Then, by analogy to the solution we have derived for the

[6] The independence of the cost of equity capital (and hence also of debt capital) from the interest rate is, of course, an independence only within a partial equilibrium framework. In a general equilibrium setting, there is necessarily a very direct connection between the interest rate (which may be regarded, to a first approximation, as the yield on assets generating sure streams) and the various ρ_k (which are the expected yields on assets generating streams of various degrees of uncertainty). But while the connection is direct (since they are mutually determined in the process of market clearing and jointly reflect such underlying factors as the level of wealth, the composition of the stock of real assets, and attitudes toward risk), there is no reason to believe that they will always move closely together over time.

[7] Since all the main earnings and cost of capital concepts have now been introduced, we shall hereafter, in the interests of simplicity, drop all subscripts and superscripts on the variables where there is no danger of ambiguity.