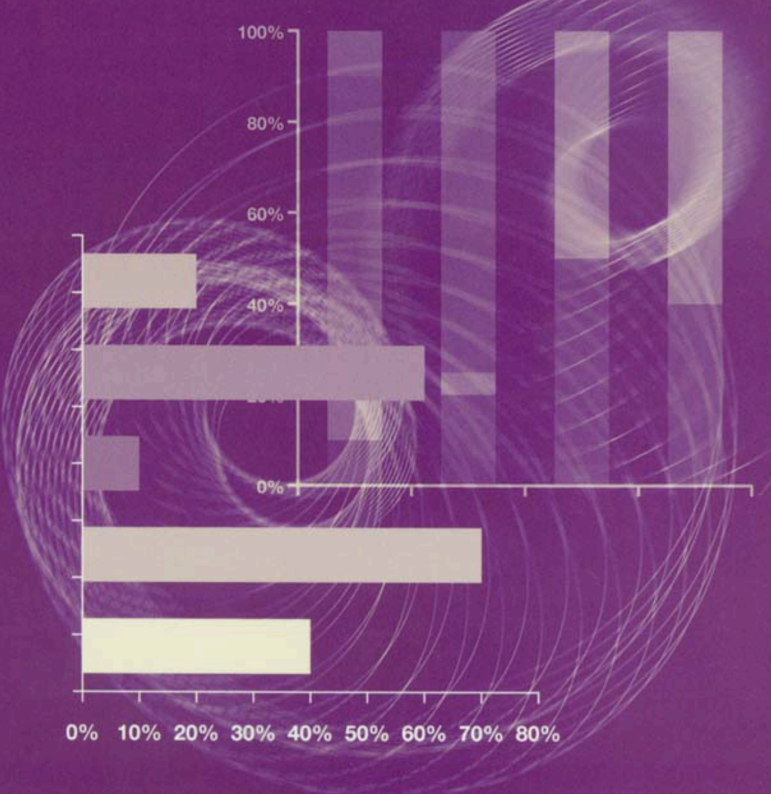
EE

Induced Investment and Business Cycles

Hyman P. Minsky

Edited and with an introduction by Dimitri B. Papadimitriou



Induced Investment and Business Cycles

Induced Investment and Business Cycles

Hyman P. Minsky

Edited and with an introduction by
Dimitri B. Papadimitriou
President, The Levy Economics Institute of Bard College, USA

Edward Elgar
Cheltenham, UK • Northampton, MA, USA

Published by
Edward Elgar Publishing Limited
Glensanda House
Montpellier Parade
Cheltenham
Glos GL50 1UA
UK

Edward Elgar Publishing, Inc.
136 West Street
Suite 202
Northampton
Massachusetts 01060
USA

A catalogue record for this book
is available from the British Library

ISBN 1 84376 216 1

Printed and bound in Great Britain by MPG Books Ltd, Bodmin, Cornwall

Contents

	List of abbreviations	vi
	Acknowledgements	vii
	Introduction: The Financial Fragility Hypothesis: the offspring of 'Induced Investment and Business Cycles' *Dimitri B. Papadimitriou*	ix
1	The analysis of business cycles: the problem and the approach	1
2	Some accelerator-multiplier models	20
3	The generation of the accelerator coefficient	64
4	The theory of the firm in relation to business cycle theory	85
5	Cost curves and investment	102
6	The survival of firms	127
7	Market constraints upon firms: vulnerability	160
8	The effect of market structure upon induced investment	177
9	Monetary behavior and induced investment	220
10	Conclusion: business cycle theory and economic policy	258
	References	261
	Index	265

List of abbreviations

AC	Average Cost
ARDD	Average Revenue Demand Curve
ATC	Average Total Cost
AVC	Average Variable Cost
DD	Demand Curve
LRAC	Long Run Average Cost
LRMC	Long Run Marginal Cost
LRTC	Long Run Total Cost
MC	Marginal Cost
MR	Marginal Revenue
MRDD	Marginal Revenue Demand Curve
SRAC	Short Run Average Cost
SRMC	Short Run Marginal Cost
SS	Supply Curve
TC	Total Cost
TVC	Total Variable Cost

Acknowledgements

I would like to thank the Board of Trustees of Bard College for sponsoring the Levy Economics Institute's ongoing research project on the *Reconstitution of the Financial Structure* conceived by my late colleague Hyman P. Minsky in 1990. As part of this project, many distinguished scholars, policymakers and business leaders visit the Institute and participate in seminars and conferences aimed at finding policy options to pressing economic problems. I want to thank most sincerely Esther Minsky, Diana Minsky and Alan Minsky for sharing many precious moments of their life with Hy. We all miss him terribly. Finally I am indebted to my assistant Deborah C. Treadway, for her skill, loyalty and devotion, without which this book would not have been possible.

To her, my warm and sincere thanks.

Dimitri B. Papadimitriou

Introduction: The Financial Fragility Hypothesis: the offspring of 'Induced Investment and Business Cycles'

Dimitri B. Papadimitriou

For more than four decades Hyman Minsky had painstakingly worked in the areas of economics and finance and in his many writings tried with vision and clarity to find a rational way to link the two. His research program has provided a definitive analysis of the linkage. As most students of his work would argue, he began with Keynes' concern with the volatility of investments, and then recognized how serious the uncertainty of cash flows from investments was since it could lead to serious repercussions on the balance sheets of firms. This, in turn, requires the government to intervene to reduce the systemic risks this process engenders by changing its fiscal and/or monetary stance to prevent a debt deflation. '[T]o Minsky a sequence of booms, government intervention to prevent debt contraction, and new booms entails a progressive build-up of new debt, eventually leaving the economy much more fragile financially' (Kaufman 1992, p. vii).

'Cash flow' to a firm, the buzzword in almost all his writings – but nowhere to be found in the neoclassical paradigm, was nevertheless crucially important in performing many functions including (1) signaling whether investments undertaken were based on sound decisions, (2) providing the funds needed by the firm to fulfill payments when due and (3) assisting in the decision-making process for future investment financial conditions (Minsky 1982, p. xvii). An analysis of cash flows documents a firm's performance by providing a connection between its ability to make payments on its debts with its cash revenues from operations. When this connection is taken in the aggregate for all business units, it determines the economic system's performance, that is 'the course of investment and thus employment, output, and profits' (Minsky 1982, p. xvii). Minsky was able to show that the observed behavior of a capitalist economy provided ample evidence that there exists an inherent instability (financial fragility) which is the result of two intervening factors: the complexity of market relations and the balance sheet adventuring of business firms. Thus investment decisions – aside from

Keynes' concern for volatility – involve complex financial relations between liability structures, committing cash flows that are (must be) generated from production and distribution of output.

It has been suggested that Minsky's development of the financial fragility hypothesis is a direct descendant of the ideas he has put forward in his doctoral thesis (Delli Gatti and Gallegati 1997). 'Induced Investment and Business Cycles' is Hyman Minsky's doctoral thesis which was supposed to be supervised by Schumpeter, but because of Schumpeter's untimely death, was completed under the supervision of Wassily Leontief. The original topic of his thesis was to explore the relations of market structure, banking, the determinants of aggregate demand, and business cycle performance (Papadimitriou 1992). The topic was set to be a macroeconomic inquiry entailing the workings of macro forces of markets and aggregate demand and how the volatility of demand leads to economic fluctuations. The dissertation that makes up this book is instead a microeconomic analysis of firm behavior encompassing the various decision-making processes regarding entry, market structure, expansion, vulnerability and survival. A firm's financing relations affecting each and every stage of the firm's development are dependent on its capacity to honor obligations in meeting commitments made today with cash flows received in the future which in their turn are dependent on the impact the business cycle has on the firm. Thus, the linkage of business investment with finance is developed in the microeconomic sphere, extending the conventional neoclassical theory of the firm. Minsky considered his approach in the dissertation to be the microfoundation for determining macro performance.

It should be understood that Harvard, where Minsky undertook graduate study, had not been as much of an influence on him as his undergraduate years at the University of Chicago. His recollections were of disappointment as topics discussed were treated in rather a mechanistic manner and also in a most unstructured fashion. '[Alvin] Hansen, the leading disciple of Keynes in America, interpreted Keynes . . . [by] virtually ignoring the significance of money and finance. Furthermore, uncertainty, which was fundamental in Keynes' understanding of the capitalist economy, was left out' (Papadimitriou 1992, p. 18). These are very important issues for Minsky and are discussed extensively in the dissertation and his subsequent research writings that span more than four decades.

Minsky describes the objective of his dissertation as the formulation of a model of particular product markets that will establish the foundation of aggregate analysis, enabling him to consider a number of public policies that are relevant to crafting an effective business cycle theory (p. 1). In so doing, he proceeds to review in the most significant detail the contributions of Paul Samuelson (1939), J.R. Hicks (1950), Wesley Mitchell (1950) and

Richard Goodwin (1951), at the time, the prevailing versions of the multiplier-accelerator model of business cycle analysis. To Hyman Minsky a satisfactory business cycle theory 'has to analyze both the interrelations among a few broad aggregates' (relating to monetary theory) and 'the behavior of individual economic units and of particular markets' (p. 1). Exogenous variables are not adequately explained in the basic structure of the linear versions (Samuelson and Hicks interpretations) of the multiplier-accelerator model and are of no real use in analysing decision-making processes that involve individual units. Minsky rejects the validity of these models and also critiques Hicks' non-linear version of 'floors and ceilings' since its parameters of 'the ceiling of full employment' and the floor at the 'level of income consistent with the maximum rate of capital consumption' (p. 8) are non-economically determined. 'What link is there between the interrelations among the variables of the model and the parameters in the Hicks floors and ceilings?' (p. 9). Minsky insists on the notion that the 'dominant factor in economic life is the interdependence of economic units and agents' (p. 12). Thus, because of its structure, the Hicksian model cannot determine the relevant factors influencing the behavior of its coefficients. The *a priori* determined coefficients α and β yield a particular level of national income and not the time path of national income the non-predetermined α, β coefficients would do (p. 17).

Furthermore, the Hicks version assumes away the effects of financial relations of non-homogeneous agents and the implications they have for investment that when taken into account would generate a model of non-linear structure closer (yet far from being adequate) to that described by Goodwin. What would be valuable, therefore, is to analyse the determinants of the value of the accelerator coefficient and thus, complete the Hicks and Goodwin models that in their present form do not explain the systematic variation of the accelerator (p. 41). These 'revised' non-linear models would posit that the value of the accelerator coefficient β be dependent on money market conditions and the balance sheets of firms, and that these factors, in their turn, be dependent on the relation between the level and the rate of change of income and the behavior of the monetary system (Minsky 1982, p. 233). Then such models could be capable of being instruments for (1) accurately describing the track of the American economy and (2) for providing alternate fiscal policy evaluations other than either 'secular stagnation' or 'continuing inflation' (p. 77). At the level of a firm, it would mean determining the 'generating relation' of its investment function that results from changes in income and 'the structure of [its] balance sheet [that] will reflect the psychological attitude toward risk taking' (p. 162). This is the Minskyan interpretation of Keynes' concept of 'induced' investment demand. In Minsky's own words, it is 'all investment which occurs because the variables which enter the investment decision function are altered by market changes associated with changes in income' (p. 135), and by extension, any changes of investment would be the cause of business cycles. Market changes include both changes in market structure and the incipient interaction of non-homogeneous firms that lead to altering investment. To Minsky, therefore, the distribution between autonomous and induced investment may be brought into question since it 'may be meaningless' and that 'all investment is induced' (pp. 135–6).

The Minskyan model of the business cycle establishes the accelerator coefficient to be pro-cyclical. It reflects 'the relation between the value of the accelerator coefficient and the structure of the markets, and the relation between the value of the accelerator coefficient and the behavior of financial markets' (p. 136). The firm's cost structure – Minsky represents it by meticulously drawn cost curves in many places in the book – is affected by both conditions (1) and (2) operating through wages as these change from fluctuations in employment and the degree of investment activity. The firm's cost curves are met by those representing demand. As market conditions vary, market structure determines individual product demand curves that shift reflecting '(1) the relation between the particular demand curve confronting the firm and the market demand curve, and (2) the way in which a firm behaves toward its particular demand curve' (p. 137). This follows more in the realm of Chamberlin's *The Theory of Monopolistic Competition* (1933) rather than Robinson's *Theory of Imperfect Competition* (1933) since Chamberlin includes the demand confronting the firm operating in non-competitive markets. Minsky's representative firm can fulfill the Schumpeterian characterization (Delli Gatti and Gallegati 1997, p. 529), in which if it possesses market power and has substantial capacity to produce with the existing plant, it will not destroy old technology unless internal and external competition encourage a 'creative destruction'. However, Minsky is careful to note that positions of previously created capital must be financed; non-depreciated capital values involve liabilities that must be absorbed and losses incurred. The willingness and ability of a firm to absorb such losses depend on current and future cash flows and balance sheets positions that are a function of past, present and future demand. As market structure changes through time, so do the financing conditions of a firm dictated by its particular balance sheet structure that determines the degree of its vulnerability and ultimately its survival. Investible funds can be obtained from three sources: retained earnings, credit and equity capital which, along with their uses, describe the structure of the firm's balance sheet (Tsiang 1951, pp. 332–3).

The firm's capacity to overcome adverse market changes derived from the stream of income flows establishes its vulnerability and survival constraints (pp. 157–8). This problem 'the usual economic theory ignores . . . and assumes a unique behavior principle for all firms [that is] profit maximization' (p. 143). However, exceeding the survival constraint will bankrupt the firm that exits the market while another firm enters, both events being sources of non-linearity affecting the value of the accelerator coefficient during the business cycle. In the case of entry of a new firm the condition that satisfies it is 'when it possesses some advantages of market position which results in the rate of return . . . greater than the rate of return generally available in the economy' (p. 254). What distinguishes Minsky's treatment of the business cycle from the traditional theory is the determination of the investment accelerator coefficient by not only the stylized facts of the income-demand interaction, but also the effects of a firm's financing conditions that inform its behavior and the behavior of money markets. This distinction is undoubtedly the origin of Minsky's *financial fragility hypothesis* for which most of his lifetime intellectual endeavors were devoted.

The world of efficient markets operates under the conditions of competition, perfect information and no financing constraints. Indeed, the well-known Modigliani-Miller theorem that reigns supreme in neoclassical economics insists that markets behave in such manner. In Minsky's world, however, starting with this book, markets are inefficient since agents are not operating in a competitive environment and most relevant information is private. Balance sheets reflect information of non-homogeneous firms and changes in them pose new constraints that take place over time and are directly related to business fluctuations.

Minsky identifies the following financial constraints to firms that engage in investment activity especially during an economic expansion: 'a) the difficulty of new firms . . . to achieve a sufficient equity base; b) the impact upon a firm of a greater tightening of financing terms, due to the imperfect elasticity of the supply of credit; c) that the rate of expansion of capital implied by the accelerator model may imply that the firm's financing condition must deteriorate' (p. 118). All three constraints listed explicitly state that balance sheet positions matter. Moreover, if firms during the course of the business cycle use their retained earnings to expand capacity (build a new plant), then this leads to a higher equity and lowered debt position, an improvement in the balance sheets positions affecting the accelerator coefficient positively. If, on the other hand, firms use their earnings to pay down their debt, this decreases their equity position and may ultimately increase debt, deteriorating the balance sheet positions and affecting the accelerator coefficient negatively (Delli Gatti and Gallegati 1997, p. 531).

Both positive and negative affects on the accelerator coefficient become pro-cyclical and may lead to the deterioration of credit terms for new loans should financial relations change thus, exacerbating the turning points of the business cycle.

As investment decisions are a function of financial conditions reflected in balance sheet structure and projections of cash flows, a firm's survival constraint becomes a serious matter. 'If we begin at any date we have that at each and every future date, in order to survive, the firm must satisfy the condition that the initial cash plus receipts minus the costs payable to that date are greater than zero' (p. 158). To avoid bankruptcy, the firm's cash flows from the sale of output must be greater than production costs and debt service commitments. 'The debts of a firm reflect the conditions that existed in the relevant financial markets at the date the debts were assumed. The survival conditions therefore are measures of the effects that financial and money market conditions have upon the behavior of firms' (p. 202). In this respect, the constraints of survival can be viewed as those of liquidity and solvency. Liquidity is the ability to meet cash commitments (it is hindered when financial conditions are altered and current or expected profit flows decline so that the liquid capital of the firm's proximate owners or producers is used up (Minsky 1982, p. 72)). Solvency is the ability to maintain some level of net worth (to avoid bankruptcy and exit the market). The firm's entry into and exit from the market are processes that are connected with liquidity and solvency and the risk of surviving is differentiated by the balance sheet position (p. 202). Firms can be classified then on the basis of vulnerability and survival into the following categories: 'wholly owned' firms and firms with 'a large volume of debt' reflecting large differences in liability structures (p. 159). The financial profiles of firms that are differentiated by their respective relations between contractual payment commitments due to their liabilities and cash flows defined as hedge, speculative and 'Ponzi', distinctions developed much later in Minsky, find their roots here.

Minsky concludes that 'the financing of a firm's expansion may result in a deterioration of its survival conditions. Therefore the liquidity crises of the downswing can be imputed to the development of the expansion. Business cycles are both monetary and real phenomena' (p. 345).

This in a nutshell is Minsky's version of a business cycle theory. When the cycle is moving toward its peak firms become more vulnerable and their balance sheet positions deteriorate, decreasing their net worth. This may result in a number of firms not surviving and thus leaving the market. Profit and income fall cumulatively until a lower point of the cycle is reached. The lower point gives the opportunity for monopoly power to rise because there are fewer firms remaining in the market with a lesser degree of vulnerability. Concurrently, individual demands shift, culminating in higher profits and improving balance sheet positions. A cumulative expansion begins engendering cyclical fluctuations.[1]

Reading Minsky's thesis in the pages that follow, a case can be made for the connections that it draws from the works of Henry Simons (1936) – the notion of the importance of the state of industrial organization; from Michal Kalecki (1937) – the linkage of profit flows to investment, in that profits become both the lure for new investment and the result of realized investment; from Schumpeter (1939) – the importance of market structure and its effects on innovative investment; and most absolutely from John Maynard Keynes.

Minsky's financial fragility hypothesis – whose relevance to the post World War II economic downturns is unquestioned – was first conceived in his dissertation and developed more fully in later years. It is a theory of business cycles connecting financial conditions of firms and markets to investment and in which coordination failures exist. Years later, Minsky (1975, 1982, 1986) articulates a theory of investment combining finance and income determination. The carefully developed 'financial fragility' inherent in advanced capitalism is based on a two price system. One is the price level of current output, and the other, the price level of financial and real assets. These two prices are determined from different relations and variables. The price of current output is the device through which production and distribution occur and costs are recovered. The price of capital and financial assets reflects uncertainty and is dependent on yields. Since yields represent streams of income over time, their current prices must reflect the current valuations of incomes that will be realized over time. An increasing divergence of the two price levels in the macroeconomy engenders instability and ultimately a business cycle ensues (Papadimitriou 1992). Market imperfections played an important role in the dissertation (chapters 4 and 7 especially), but they were not much discussed in later writings.

The Minskyan analysis of the business cycle is a new but fundamentally Keynesian species that links economics and finance. Although quite different from the standard theory, it still lies within the mainstream economic paradigm. Minsky's dissertation is an important contribution to the literature on business cycles and a reminder that efforts grown in the tradition of non-standard theory do not always remain unappreciated.

NOTE

1. This paragraph is to some extent a restatement of Delli Gatti and Gallegati (1997), 531–2.

REFERENCES

Chamberlin, Edward H. (1933), *The Theory of Monopolistic Competition*, Cambridge, MA: Harvard University Press.

Delli Gatti, Domenico and Mauro Gallegati (1997), 'At the Root of the Financial Instability Hypothesis: "Induced Investment and Business Cycles"', *Journal of Economic Issues*, **31**, 2.

Goodwin, Richard M. (1951), 'The Non-Linear Accelerator and the Persistence of Business Cycles', *Econometrica*, **29**.

Hicks, John R. (1950), *A Contribution to the Theory of the Trade Cycle*, Oxford: Clarendon.

Kalecki, M. (1937), 'The Principle of Increasing Risk', *Econometrica*, New Series 4, 440–47.

Kaufman, Henry B. (1992), 'Foreword', in Steven Fazzari and Dimitri B. Papadimitriou (eds), *Financial Conditions and Macroeconomic Performance, Essays in Honor of Hyman P. Minsky*, Armonk, NY: M.E. Sharpe.

Minsky, Hyman P. (1975), *John Maynard Keynes*, New York: Columbia University Press.

Minsky, Hyman P. (1982), *Can 'It' Happen Again?*, New York: Sharpe.

Minsky, Hyman P. (1986), *Stabilizing an Unstable Economy*, New Haven and London: Yale University Press.

Mitchell, W.C. (1950), *What Happens During Business Cycles*, New York: National Bureau of Economic Research, chapter 1, p. 3.

Papadimitriou, Dimitri B. (1992), 'Minsky on Himself', in Steven Fazzari and Dimitri B. Papadimitriou (eds), *Financial Conditions and Macroeconomic Performance*, *Essays in Honor of Hyman P. Minsky*, Armonk, NY: M.E. Sharpe.

Robinson, Joan V. (1933), *The Economics of Perfect Competition*, London: Macmillan.

Samuelson, Paul A. (1939), 'Interaction between the Multiplier Analysis and the Principle of Acceleration', *Review of Economic Studies*.

Schumpeter, J.A. (1939), *Business Cycles*, New York and London: McGraw Hill Book Company, 2 volumes.

Simons, Henry (1936), 'Rules versus Authorities in Monetary Policy', *Journal of Political Economy*, **44**, 1–39, reprinted in H. Simons, *Economic Policy for a Free Society*, Chicago: University of Chicago Press, 1948, p. 172.

Tsiang, Sho-Chieh (1951), 'Accelerator, Theory of the Firm and the Business Cycle', *Quarterly Journal of Economics*.

1. The analysis of business cycles: the problem and the approach

Paraphrasing Voltaire, we can assert that if business cycles did not exist, the economic theorist would have invented them. For if we look at the problem of business cycles, without any doctrinaire bias, it seems obvious that in this branch of economics a natural connection occurs between the often too separate compartments of economic analysis: between the 'monetary' and the so-called real phenomena. Therefore, a theory of business cycles, to be consistent with the observable material and the inherited doctrines, should be a blend of the analytical material which deals with the interrelations among a few broad aggregates – which traditionally has been the approach of monetary theory – and the analytical material which deals with the behavior of individual economic units and of particular markets – which has been the sphere of price and distribution theory. This thesis can be interpreted as an attempt to construct such an eclectic business cycle theory by utilizing a number of elements drawn from inherited economic analysis. To be complete such an attempt would have to explore such purely theoretical material as the relation between macro and micro analysis, between partial and general equilibrium analysis and between monetary and real phenomena, as all of these separate pieces of economic theory have to be used in an analysis of business cycles. That task is both too big and too general. What will be attempted here is to try to develop a technique of business cycle analysis which does draw upon the various portions of inherited doctrine.

In spite of the complexity of the phenomena which are observed during business cycles, and the seemingly obvious need for cycle theory to interpret the phenomena which occur during a business cycle, individual authors have tended to specialize in their emphasis on particular phenomena as the essential components of business cycles. To a student it is the field of business cycle theory, more than any other part of economics, which consists of a study of alternative explanations.[1]

An explanation of the continued existence of alternative theories of the business cycle is perhaps most readily made in terms of the economic policy problem which the existence of business cycles poses. The alternative theories of the business cycle are consistent with somewhat different approaches to economic policy, and business cycle theory has historically had a strong prescriptive bias. No claim is being advanced that this work differs from the others in this respect – the only claim made is that the more eclectic approach outlined in this chapter indicates that a wide variety of public policies are relevant to an effective business cycle policy.

We can distinguish two varieties of business cycle analysis, based upon the method of analysis used. One type directs its attention at a few macroeconomic variables, and studies the interrelations among these aggregates. The other variety maintains that business cycle analysis must attempt to deal with the behavior of individual economic units, and that what aggregate relations are derived are of the nature of an average. Two recent volumes in business cycle theory – that published by J.R. Hicks[2] and the posthumously published volume by Mitchell[3] are illustrative of this methodological division in business cycle analysis. The Hicks volume deals with the interrelations among a few broad aggregates – the Mitchell volume essentially denies the validity of such aggregative analysis and emphasizes the connection between the multiplicity of markets which make up economic life. In the Hicks volume, if market processes are discussed, they are taken up as asides from the main course of the volume. The core of Mitchell's theoretical apparatus is a concern with market processes. It is the contention of this volume that any reasoned perspective upon the problem of business cycles leads to the conclusion that each approach is in some measure incomplete without the other.

The theoretical framework employed by Hicks is based upon a limited number of presumably measurable economic aggregates: consumption, investment, income, and so on. The movements over time of these aggregates are interpreted as the essential characteristic of the business cycle of experience. Models are built in which interrelations among these few variables are set out. The results obtained depend upon the specification of the functional relations among the variables and upon the values assigned to the parameters of these functional relations. Such models are dynamic, connecting variables of different dates, and once such a model is constructed, its operation is 'mechanical'.

'Between the variables relevant to economic fluctuations there is, in the opinion of the theorists, a network of causal connections. Given movements in the data therefore cause movements in the endogenous variables, and it is the task of business cycle theory to show that the characteristics of observed movements in endogenous variables may be explained either by the given movement in the data or by the properties of the causal network.'[4] The essential characteristics of the Hicks model relate to the properties of the causal network – the series of lags and the nature of the functional relations among the endogenous variables. In addition various non-linear elements are introduced into the Hicks model. These are essentially exogenous, unexplained elements, which it can be claimed are included solely because they make the model work in the desired manner. The introduction of such non-linear elements may be necessary for the construction of a useful theory of the business cycle. However, no special valued element or assumption has a place in a theory of the business cycle unless the processes which generate the special element can be described, or the actual observed value of the special element is known. The elements of the Hicks model fail to meet such standards. It will be shown in Chapter 2 that economic arguments can be advanced for non-linear formulations which are alternatives to that of Professor Hicks, and the material on the theory of the firm (Chapter 4 through Chapter 8) and on the financial relations (Chapter 9) which follows constitutes an attempt to see whether the implications for investment of market phenomena generate any particular non-linear form of the accelerator coefficient.

The approach of Hicks leads to a business cycle theory which is straightforward in its exposition and which does not have so many variables that the mind is incapable of comprehending the interrelations among the variables. In its origin it is beholden to the Keynesian[5] analysis of income determination – and has many of the virtues and also the faults of its sire. The beauty of a simple set of interrelations is manifest – the emptiness which is induced by the elimination of market processes is also obvious. An attempt to repair these deficiencies is in order.

'One way out of these difficulties seems to be to construct a model consisting of an inner circle of relations between the most important macroeconomic variables and a series of supplementary relations meant to specify and analyze the inner-circle relations. The inner-circle relations might be relations using only such broad concepts as total national income, total expenditures, total imports, total exports or the general price level. . . . Corresponding to this inner-circle relation there could be supplementary relations explaining the demand for separate groups of commodities and services, for instance, for consumer goods or for investment goods.'[6] This solution to the problem of the reconciliation of the simple and the complex theories is by way of the disaggregation of the functional relations of the simple theory. 'Each inner-circle relation could in this way be illustrated and tested, and possible deviations between observed and calculated values of the macroeconomic variables "localized", i.e. it could be found out whether deviations in total imports are to be attributed to imports of raw materials or of finished products, etc.'[7]

The approach to the problem of reconciling the complexity of business cycle observations to the simplicity of macroeconomic business cycle theory suggested by Tinbergen is not the one which we are exploring in this volume. Tinbergen's suggested series of supplementary relations which aggregate to the macroeconomic relations can be interpreted as meaning that, in some sense, the complete analysis is really the analysis of the interrelations among the supplementary relations. The increase in the comprehensibility is apparent rather than real. At each step the obvious thing is to explore the interrelations in the various supplementary sets of relations. In order that these supplementary sets of relations be independent of the behavior of the aggregates, it is necessary that the aggregates themselves not be interrelated. Essentially the Tinbergen suggestion means that we go along with a complete, complex and detailed model, and we add to it a set of aggregation rules which leads to a single set of interrelated aggregates that constitutes the core model. Disaggregation involves limitations upon the nature of the functional relations which can enter both the aggregate and the particular models.[8] The idea implicit in the approach suggested by Tinbergen seems to be a fruitful suggestion to solve the dilemma of discordant theories. However, the definition of supplementary relations as formal mathematical disaggregates of the inner-circle relations is unnecessarily restrictive.

Consistent with the notion of a set of inner-circle relations, which determine overall movements of the system, and the need for supplementary relations, which lead to cyclical behavior consistent with observations, is the emphasis upon the need for special turning point analysis in business cycle theory. 'The change from prosperity to depression, from upswing to downswing, is the most crucial problem of the cycle. I still believe that we need a special theory, or rather alternative explorations of the turning points. The cumulative process is always essentially the same, but we cannot be sure that the turning point is always brought about by the same factors (even apart from possible disturbances from outside the economic system) or that the same system of difference equations will satisfactorily describe the upswing as well as the upper turning point.'[9] Both the inner-circle and supplementary relations approach suggested by Tinbergen and Professor Haberler's emphasis upon the necessity for a special theory of the turning points emphasize the separability of economic phenomena into compartments. Like the authors who divide time series into trend and cyclical components, they neglect the fact that the same economic phenomena which breed behavior of the inner-circle relations and the cumulative process also determine the behavior of the supplementary relations and the turning points. What is needed is a technique of analysis which correctly emphasizes the interrelations among economic phenomena and which nevertheless permits the separation of economic data into sectors which can be conveniently handled.

The addition of floors and ceilings to the accelerator-multiplier type analysis by Hicks is essentially an attempt to include special turning point material in the analysis of the business cycle. The virtue that has been imputed to the accelerator-multiplier models – that they eliminated the need for specific turning point analysis – has disappeared in these later versions. However, the floors and ceilings have a non-economic cast to them. The ceiling is usually full employment, which is used by Hicks as a technological concept. The floor is that level of income consistent with the maximum rate of capital consumption possible under the existing technique of production. It also is determined from outside the economy. The use of such ceilings and floors is equivalent to the introduction of mechanical constraints from outside the system in order to have it behave properly.

'Constants' occur in theories which are successful in their application to the world. This is especially true in the physical sciences. In all such cases the 'constants' are unchanging in value only for a determinate set of problems. For other problems, the phenomena represented by these symbols have to be considered as variables, and their values at any moment of time, or under a set of conditions, has to be determined within the model. In particular, in economics, any ceiling such as the full employment ceiling, used by Hicks, has to be considered as an economic variable which is to be determined in a more general model. The assertion that 'full employment' is a non-economically determined phenomenon, and therefore not subject to an economist's analysis, is on the surface suspect; it asserts too much. Such arbitrary floors and ceilings should be replaced by endogenously determined parameters which are generated by the processes of economic life. The problem can be stated as: what link is there between the interrelations among the variables in a formal business cycle model and the parameters which in the Hicks type models are the floors and ceilings?

All of these recent developments (such as, models containing non-linear elements, containing arbitrary determinants of the turning points, containing inner-circle and supplementary relations) represent a dissatisfaction with those business cycle models which depend upon a set of linear relationships. Those models which depended upon the structure of production and sales, as expressed by a small number of functional relations and lagged variables, are in their analytical tools and intellectual derivations essentially Keynesian. 'After Keynes' General Theory was published, these limiting factors – bottlenecks, limits to bank expansion, etc. – lost much of their importance as explanations of the turning points of the cycle. As soon as the consumption function was introduced as a central feature of economic models, it was immediately recognized that a cumulative process of expansion may not be self-reinforcing but instead may inevitably lead to a crisis and a period of contraction even before the physical or financial limits to expansion have been reached.'[10] The reconsideration of such fully aggregated models may be due to the failure of the early Post World War II predictions. The dissatisfaction with too aggregative analysis can either take the form of disaggregation which leads to mathematical complexity, or of a retreat to empiricism. The simplest linear process models, which exhibited cyclical behavior, resulted in a number of possible types of behavior for the endogenous variables, and these types of behavior depended upon the values of parameters. No matter what values these parameters were assigned, the resulting model exhibited unsatisfactory behavior. This has led Mr Hicks to the reintroduction of floors and ceilings (turning point analysis) into a model which is essentially Keynesian in its derivation.

Rather than attack the problem posed by the unsatisfactory nature of the accelerator-multiplier type models by means of the mechanism which generates the floors and ceilings, we can take up the more general problem of what determines the parameters of an aggregative business cycle model. The parameters of the functions which are included in the aggregative model can be interpreted as shorthand symbols for the processes of economic life which are not included in the simple model; and they are therefore in turn determined by market processes. The variables of a macroeconomic model are such that the values which are generated by the model imply changes in the determinants of equilibrium in different particular markets. For example a change in national income affects product demand curves. This results in a change in the equilibrium conditions in particular markets which cannot be separated from the processes which determine the variables in the macroeconomic model. If we are to use such macroeconomic models, we have to integrate the relation between the particular market developments and the developments which are represented in the aggregate model. The effects of changes in the variables of the macroeconomic model upon the equilibrium conditions in particular markets are going to be interpreted as determining the parameters in the macroeconomic model. By adopting this combination, we retain the simplicity of a dated analysis in income flows and still emphasize the significance of particular market analysis. In addition this approach emphasizes that the parameters in macroeconomic models are elements whose values are to be determined by economic processes.

Adopting the language of Tinbergen, what we propose is that the inner-circle relations in the business cycle model be essentially of a Keynesian derivative type: for example a simple accelerator-multiplier type model. The supplementary relations are relations that determine the parameters in this model. In these supplementary relations the variables of the macroeconomic model appear as parameters. The changes thus indirectly induced in the parameters of the macroeconomic model by the different values of the variables of the macroeconomic model will affect the time path of the variables of the macroeconomic model. The inner-circle relation remains a straightforward few variable analysis. The supplementary relations, which determine the value of the parameters, will be complex, multivariable relations, where the entire apparatus of economic theory is brought to bear upon the analysis of economic activity. The inner-circle model retains its simplicity. As the need for floors and ceilings as such is removed, the inner-circle model can be even simpler than Hicks' model. The difficult parts – the complicated series of springs, cogs and gears that drive the clock – are removed from sight; all that remains visible is the simple two hands circling an austere dial face.[11]

The distinction that is made between the variables and the parameters of the macroeconomic model is for convenience in analysis. The dominant factor in economic life is the interdependence of elementary economic units. The aggregate variables are constants, designed for simplicity and convenience. The variables of the macroeconomic model are either sums or index numbers of measurable attributes of individual economic units. The parameters of the macroeconomic model are aggregates of individual behavior or reaction coefficients. As such they are not easily measurable and not readily aggregated. In the analysis of the particular firm's investment decision which follows, we will attempt to determine what factors are relevant in determining the behavior coefficient of individual economic units. As the behavior of the macroeconomic model depends upon the values of the parameters of the model, this will enable us to isolate those variables which lead to alternative behaviors of the economic system as a whole.

The utilization of the variables, income and change in income, to determine the value of the accelerator coefficient in the accelerator-multiplier models, leads to a non-linear theory, of the type studied by Goodwin.[12] However, rather than have the non-linearity as an economically unmotivated, or crudely motivated, relation we center our attention on the determination of the non-linearity.

In order to illustrate the approach which is adopted, it may be appropriate at this stage to indulge in a digression. Let us take the best known of the mechanical interrelation models – in its original form – and see how it can be modified in the light of the above perspective. The result is but a slight modification in the original model; but as a result the significance, the conceptual role of the model, is changed markedly. The change will make the accelerator-multiplier mechanism the core of the analysis, a framework upon which the more complete analysis can be hung, rather than an attempt to use it as it stands as a model of the business cycle.

The best known of the mechanical business cycle models is the Hansen-Samuelson model presented in its 'mathematical form' by Paul A. Samuelson.[13] The lines of the development of this model which will be undertaken here are foreshadowed in the final paragraph of this article when Samuelson asserts:

> The limitations inherent in so simplified a picture as that presented here should not be overlooked. In particular, it assumes that the marginal propensity to consume and the relation are constants; actually these will change with the level of income, so that this representation is strictly a marginal analysis to be applied to the study of small oscillations.[14]

This familiar Hansen-Samuelson model is based upon the following assumptions:[15] that National Income at any time is a sum of three components – government expenditures, consumption expenditures and private investment expenditures. Consumption expenditures are a fraction α of income at a unit of time earlier, and investment expenditures which are induced by the change in consumption are β times the change in consumption. This leads to the familiar difference equation in which income at any period of time is determined by income of two previous periods: for example

$$Y_t = 1 + \alpha[1 + \beta]\, Y_{t-1} - \alpha\beta\, Y_{t-2}$$

where Y_t, Y_{t-1}, Y_{t-2} are dated incomes, α = marginal propensity to consume and β = the accelerator relation. The 1 in the equation is due to the existence of government expenditures which, from outside the model, set the process to work. The behavior of this model, once the linear form and the lag pattern are determined, depends upon the values assigned to α and β. Four different types of behavior are possible for the variable Y_t.

1. Y_t may asymptotically approach the level of income given by $1/[1 - \alpha]$, the pure multiplier level of income (region A in Figure 1.1);
2. Y_t may take on a damped cyclical path, approaching the income level $1/[1 - \alpha]$ (region B in Figure 1.1);
3. Y_t may take on an explosive cyclical path (region C in Figure 1.1); and
4. Y_t may take on an explosive path (region D in Figure 1.1).

Under the assumptions that the relation β is a fixed technical coefficient relating output and capital stock, and that the marginal propensity to consume is a constant determined by a fundamental attribute of the society, one of the four above types of behavior is possible, aside from the

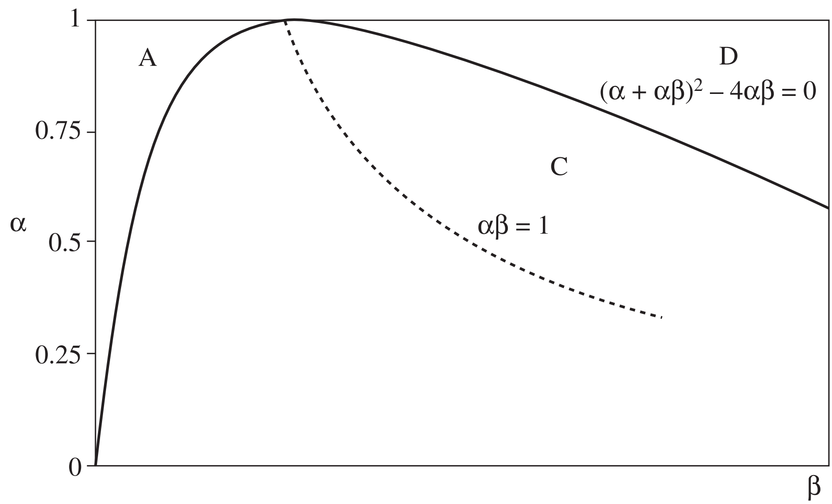


Figure 1.1 Samuelson's State Diagram

behavior which characterized the boundary conditions between two of the stages, for example if $\alpha\beta = 1$, the systems would oscillate with a constant amplitude. The assumption that $\alpha\beta = 1$ seemed to be, even to this model's strongest advocates, an unnecessarily rigid one, so the above model was recognized as a useful expository tool, but inadequate in itself to explain the cycles of experience. Both the explosive character of the development in regions C and D, and the damped character of the development in regions A and B are inconsistent with gross observations about the behavior of the economic system.

A hypothesis to be advanced here is that an accelerator-multiplier type model in which α and β are variables over the cycle can lead to a movement in time of the dependent variable of the model which is consistent with the observed values of this variable.[16] As interpreted here, the accelerator-multiplier model is meaningless without an analysis of the economic processes which generate the values of α and β. In order to be meaningful a specification of the manner in which α and β vary over the business cycle has to be advanced. The hypothesis that the business cycle of experience can be interpreted as an accelerator-multiplier model with variable coefficients, and that the coefficients of the model vary in a systematic way over the business cycle, is advanced in this chapter.

As a result of this interpretation of the accelerator-multiplier model, the business cycle analysis problem is transformed into the problem of what generates the realized values of α and β. In particular we may ask how this generation process is systematically affected by the variations in National

Income (or Employment, or generally speaking, whatever the variable whose time path is determined by the accelerator-multiplier model). The necessity for systematic variation is due to the need to have the α and β coefficients change in such a manner as to result in a sequence of values of the endogenous variables which can be considered to be consistent with observations.

This chapter focuses on theory. The analysis is directed at the construction of generating processes for these coefficients out of material drawn from the generally accepted body of economic theory. Because of these limitations, no empirical testing is undertaken here. It is obvious that the validity of the model developed depends upon its consistency with observations.

A theory which asserts that the α and β of the model vary in such a way as to generate a time path of national income which is consistent with observation is not, as stated, meaningful. For by adding α and β as 'undetermined' variables we can generate any time path of national income which can conceivably be observed. In order to make the theory meaningful it is necessary to add that the values of α and β are generated by economic processes, and that the generating process leads to such a restricted set of values of α and β that a refutable statement results.

If we ask what determines the value of the accelerator or multiplier coefficient at any time, and carry on an analysis in terms of the behavior of the households and firms, we have done more than transform a linear difference equation into a non-linear difference equation. For by turning our attention to the behavior of households and firms, we can investigate the effects upon the cyclical behavior of an economy of variation in the structure of markets and financing conditions.[17] As a result, the functional relation between the different incomes which is the core of the accelerator-multiplier model becomes dependent upon elements other than past period income and 'non-economic constants'. Once we recognize that the relations which determine the accelerator and multiplier coefficients are complex, an analysis of their determination requires more than the specification of a functional relation between these coefficients and the level of income.

The major task of this chapter, therefore, is to develop an analysis of the processes which generate the values of the coefficients in such non-economic models. Prior to undertaking this task, it seems desirable to exhibit a modified Hansen-Samuelson model which does not behave in a manner consistent with observations to show that this hypothesis is not, on the face of it, implausible.

In Table 1.1 we can see that when the marginal propensity to consume, α, is very large (0.9 or 0.95) a value of β a little greater than 1 (1.05 for explosive oscillatory, 1.6 for explosive) is sufficient to lead to an explosive development in the economy; whereas when the marginal propensity to consume is 0.7 an accelerator coefficient of the same order of magnitude

Table 1.1

Values of α	Values of β Boundary Values Between Regions		
	A & B	B & C	C & D
0.95	0.64	1.05	1.6
0.9	0.61	1.1	1.8
0.8	0.38	1.25	2.6
0.7	0.29	1.4	3.4

Table 1.2 A Hansen-Samuelson Sequence: With Variable α and β

Period	α	β	G	$C(Y_{t-1})$	$I(C_t - C_{t-1})$	Y
1	0.9	2	1	00	00	1
2	0.9	2	1	0.9	1.8	3.7
3	0.9	2	1	3.3	4.8	9.1
4	0.9	1.5	1	8.2	7.4	16.6
5	0.8	1.4	1	13.3	7.1	21.4
6	0.8	1.2	1	17.1	4.6	22.7
7	0.8	2	1	18.2	2.6	22.0
8	0.8	0	1	17.6	–	18.6
9	0.8	2	1	14.9	−5.4	10.5
10	0.9	1.0	1	9.5	−5.4	5.1
11	0.9	1.0	1	4.6	−3.9	1.7
12	0.9	1.0	1	1.5	−3.1	−0.6
13	0.9	1.0	1	−0.5	−2.0	−1.5
14	0.9	1.0	1	−1.4	−0.9	−1.5
15	0.9	1.0	1	−1.4	0	−0.4
16	0.9	0	1	−0.4	0	0.6
17	0.9	1.0	1	0.5	0.9	2.4
18	0.9	2.0	1	2.2	3.4	6.6
19	0.9	2.0	1	5.9	7.4	14.3
20	0.9	2.0	1	12.9	14.0	27.9

($\beta < 1.4$) will lead to a damped oscillatory development. It follows that an explosive development can be turned into a damped cyclical development by relatively small changes in the value of the relevant coefficients.

This can be illustrated by taking a Hansen-Samuelson sequence and exhibiting the time series it generates with the coefficients taken as cyclical variables. The example given in Table 1.2 is strictly for expository purposes: no claim is being made that the values assumed for the coefficients represent what actually happens.

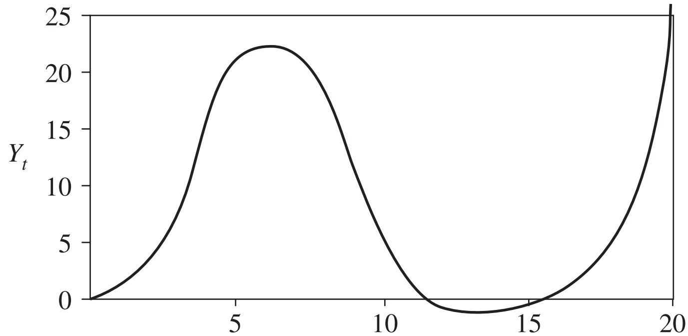


Figure 1.2 Times Series: Generated by Hansen-Samuelson Model with Variable α *and* β

The time series shown in Figure 1.2 exhibits the precipitous fall in income after the peak of the cycle. The periods at the top and at the bottom of the cycle are consistent with the behavior of firms endangered by a high degree of uncertainty as to future developments. It is obvious that by varying α and β the rise from the cyclical trough can be made as fast or as slow as desired. The model sequence exhibited in the table can be related to the Samuelson state figure (1.1). We began with values of α and β which placed the 'economy' in region D of the state diagram: the explosive state. The fall during period 4, which is taken to occur when the economy reaches a very high level of income, results in a shift to region C in the state diagram. Region C is also an explosive region; however it is explosive oscillatory. The change in α during period 5 still keeps the economy in region C, but the further fall in β during the sixth period shifts the economy to region B. Region B is the state in which the cyclical behavior is damped oscillatory. This changing value of the accelerator coefficient results in a downturn in income and the 'explosive' behavior which we assumed in the beginning has been replaced by a damped cyclical behavior.

The fall in the consumption coefficient at very high levels of income is taken to be due to the 'high propensity to save' at high incomes. The fall in β at high incomes will be 'tested' by the investment models which will be taken up later. The erratic behavior of β during periods 7, 8 and 9 is assumed because the definition of β includes both long term and short term investment. The stabilization of income at a high level during periods 5, 6, 7 and 8 is assumed to result in an increase of long term investment at the end of the boom. The zero accelerator coefficient during period 8 is taken to be composed of a tendency to continue long term investment and to decrease inventories; the value 2 accelerator coefficient during period 9 is taken to be determined by the 'inventory' clearance which accompanies the fall in income of period 8. The high propensity to consume during period 9 through 20 is taken to be determined by the desire to maintain consumption standards, whereas the low accelerator is taken to be determined by the inability to disinvest rapidly during a depression. Models of investment behavior can be developed so that the accelerator coefficient on the downswing can achieve a zero value even with falling income or with 'excess capacity'. In period 18 the accelerator coefficient is increased to allow for both short term and long term investment. The economy from period 18 on is in an explosive upward movement which, according to our hypothesis, will be choked off by means of a fall in the accelerator coefficient and a fall in the propensity to consume as the incomes reach high levels. The economy, therefore, switches from region D, the explosive state, to region C, which is explosive oscillatory, and then to region B which is damped oscillatory. The high value of the multiplier and the impossibility of disinvesting very rapidly keeps the economy in region B during the downswing. However, if the economy swings from region B to region A, which it does when the accelerator is assumed to be zero, the movement for that time is a movement toward the level of income determined by autonomous investment (in the example shown, the level of income determined by government expenditures). Rather than assume that the economy is naturally in one state, we assume that the state of the economy changes through changes in the values of the coefficients.[18]

In the model exhibited, the major characteristic to be analysed is the variability of β, especially the assumption that β turns down at very high levels of income. Unless this is to be taken as an *ad hominem* argument, a rationale from general price theory must be offered for this development. This means that a model of investment behavior must be developed which generates a relation between changes in income and induced investment, such that the relation β varies in a systematic way over the business cycle.

As used here, the β coefficient includes investment in fixed and in working capital. A fall in income can induce a large fall in working capital even though the fall in fixed capital is limited by the inability to disinvest rapidly. The result is that the β coefficient is large during the early stage of a depression, when inventories are stabilized. The inability to disinvest fixed capital rapidly does not determine a floor to the level of income, as Hicks assumes. If we work with a simple enough model, the fall in the amount of induced disinvestment when inventory disinvestment stops will lead to a rise in income. This will, under appropriate conditions, be sufficient to result in an upward movement of income.

At the same time as we varied the accelerator coefficient we varied the marginal propensity to consume. This was done in order to shift the model from one state to another without requiring large variations in the accelerator coefficient. The assumption which we made in varying the consumption coefficient is that incremental consumption falls when income is very high and that the marginal propensity to consume is low when income is decreasing. Because the propensity to consume coefficient in the Hansen-Samuelson model is related to the level of income above the equilibrium level, it is really an average coefficient over the range of incomes which the model covers. Therefore, the low marginal propensity to consume in the downturn results in a high 'average propensity to consume' out of the income variation from the base. When income turns up again we assume a high marginal propensity to consume out of incremental income until income becomes very high, when the marginal propensity to consume decreases. Such assumptions are, at this stage, just blanket assertions. To be complete, the model of household behavior which generates such a cyclical pattern of the marginal propensity to consume must be exhibited. In contrast to the investment-income relation, where there are no inherent empirical or theoretical studies of its behavior over the business cycle, the savings-income relation is supported by an extensive literature on the behavior of savings over the business cycle.[19] These results can be incorporated into our models. We can therefore expect to be on firmer ground when we vary α than when we vary β – in the sense of using inherited studies.

Because of the existence of much fine work in the cyclical behavior of savings, there is no need to concern ourselves further with this problem. Therefore, the rest of this study is directed at the generation of the accelerator coefficient, although a more complete study would also include the generation process for the marginal propensity to consume.

THE ANALYSIS OF BUSINESS CYCLES: AN ASIDE ON METHODOLOGY

If we are to analyse the processes of economic life, it is necessary to order the chaotic multiplicity of economic phenomena. This establishment of order in the observations of a particular science is a work of 'art,' of imagination or intuition; it involves the imposition of a set of constructions of the scientist upon the raw data. The Marshallian industry is such a construction. As a construction it is logically equivalent to lines sometimes drawn in the proof of a theorem in geometry: there is nothing in the hypothesis or in the conclusion that indicates such a construction is in order; yet once the construction is made, the proof is simple – without it, complicated or impossible. If, therefore, our view of the aim of economic analysis is that it is an 'engine for the discovery of concrete truth',[20] we do not look upon the analytical set-ups as pictures or images of realities. We do not even require that the processes which can be equated to the operations of the model take place: we require that the use of the tools enable us to solve problems of experience that we could not solve without the tools.

'Like other scientific concepts, ours (the business cycle) is a manmade entity, created by pulling apart items of experience that can be observed directly; then putting like parts together into a new whole that cannot be seen by the eye or touched by the fingers. Such synthetic products of the mind have often turned out to be useless or worse, in that they led to logical contradictions, conflicts with factual evidence, or futile practice. Most of the useful ones have to be reconstructed from time to time in the light of fresh discoveries that they have helped men to achieve.'[21] The interaction between the accelerator and multiplier as we use it is such a manmade entity. The standard to be applied to test its validity is not whether it presents an accurate image of what takes place in the economy, but whether this construct is a useful device by which the processes of economic life which lead to the generation of business cycles can be studied. The purpose of the core model, therefore, is to provide an orderly framework for analysis. The test of the specifications made in the accelerator-multiplier model is not the ability to measure invariant functional relations between the variables, but the ability of an analysis based upon these assumptions to lead to meaningful statements about the nature of the world. A meaningful statement about the nature of the world is here defined as a statement which is capable of being refuted by a set of observations which can be made.

The use of a formal model which simplifies and idealizes economic reality as a framework upon which an analysis of the processes of economic life can be organized is not new. The fundamental proposition in such an approach is that the values of the variables which are determined within a model, under one set of conditions, imply some changes in the markets dealing with other economic variables.[22] The satisfaction of the new equilibrium conditions in the other markets will in turn alter the set of conditions which determine the values of the variables in the original market. In our case, the variable determined within the model is the level of income. The conditions which determine the level of income are the accelerator coefficient, the marginal propensities to consume and the appropriate past incomes. The past incomes, being history, cannot be affected by the developments, excepting as what is present income necessarily becomes past income with the passage of time. The present value of income, divided, in the manner given by present values of the propensity to consume and the accelerator coefficient, among the various activities in the economic world, leads to certain conditions in the market for consumption and investment goods. These conditions, in turn, generate the value of the accelerator and multiplier. This means that the manner in which present and past incomes operate to determine future income, for example, the equilibrium value at the next date within our model, will be changed from what it would have been if present income had no effect upon the markets which generate these propensities and relations.

Much of inherited economic doctrine can be conveniently placed in such a methodological framework, and the typical method of handling these aspects of economic life is consistent with this framework. As an example, let us take a Marshallian industry consisting of N firms: the industry is competitive. The market price is determined within the model by the intersection of the demand curve with the horizontal sum of the marginal cost curve. If this price is greater than average total cost, this implies that the firms in the industry are typically earning a higher rate of return than is generally available elsewhere in the community. This is generally taken to imply that investment will take place in this industry, that is, either an increase in the capital stock of the community or a reallocation of a given stock of capital among the different industries in the community will occur. The short term equilibrium in the industry under analysis implies the existence of a disequilibrium in the capital market. The development within the capital market in response to the disequilibrium, however, implies either an increase in the number of firms or in the size of the firms within the industry. That is, investment takes place in the industry. This will shift the sum of the marginal cost curves (the competitive supply curve) to the right and result in time in the achievement of long run equilibrium in the particular market. The essential point here is not the eventual achievement of long run equilibrium, but that the analytic process is the reciprocal relations between the elements in our 'isolated' market and the equilibrium determining conditions in the capital (other) market.

Aside from its application to the Marshallian industry analysis, this perspective may be applied to some Keynesian models. Let us take a simple income equation of the form $Y = C + I + G$ where I is a function of the interest rate, C is a function of income and G is the government deficit. The government deficit is financed either by the creation of money or the sale of government bonds. In either case the effect is felt in a portfolio composition-interest rate relation: the money market phenomena encompassed under the term Liquidity Preference in the Keynesian analysis. These money market phenomena may affect the interest rate, thereby having an effect upon investment.

In addition, the investment or government deficit which is taking place does mean an increase in wealth, either in the form of government bonds or of real capital in the community. The question as to whether such an increase in wealth will affect the position of the consumption function has been raised by Pigou and others.[23] The upward shift of the consumption function as wealth is accumulated has been offered as an explanation of the constant average propensity to consume over long periods of time, even though the short run marginal propensity to consume is smaller than the long run average for some levels of income. This again is an example of the effects of the development within a model having implications for the equilibrium of another market which, in turn, as the values in this other market change, affects the equilibrium conditions within the first model.

Another example of the effects of developments within one market upon other markets was utilized by Domar[24] when he emphasized the dual role of investment, as a determinant of income in a Keynesian framework and as a change in total capital and therefore as a change in productive capacity. This duality was integrated in his theory of economic growth; however, his use of a constant average equal to marginal propensity to consume could be interpreted as an implicit acceptance of the shifting consumption function associated with the effect of investment as a change in net worth. The pattern of economic development growth analysed by Domar depends upon the duality between an equilibrium level of income and an equilibrium rate of growth. Analysis of how equilibrium in one set of markets may affect other markets is a development of the growth relations that could lead to greater relevance of the analysis.

The general perspective of this work is that we deal with a number of different sets of markets, each market developing, under given conditions, in the direction of some equilibrium value, and the values of economic variables which are generated by this market movement act upon other markets, so as to affect the equilibrium values in these other markets. We will organize our analysis of business cycle relations around the accelerator-multiplier interrelation using developments in other markets to affect the values of the parameters in the accelerator-multiplier model. This method of partial elimination of variables in order to solve a complicated moving equilibrium problem is more readily justified in business cycle theory than in price and distribution theory. The selection of the sequence in which the various markets are to be 'solved' is in reality an assumption as to the manner in which a business cycle is propagated. Such a complex propagative approach to business cycles leads naturally to the type of model which is presented below.

Methodologically this volume also emphasizes the need to relate aggregate analysis to the behavior of economic units. In particular, the relation between investment and the behavior of individual firms is investigated. This is done by using a modification of the traditional graphical analysis of the behavior of firms. As a result, the main body of this volume (Chapters 4 through 8) is characterized by an avoidance of mathematical formulations and by an abundance of graphical analysis. On the other hand, the most convenient formulation of the problems dealt with in Chapter 2 ('Some accelerator-multiplier models') and Chapter 9 ('Monetary behavior and induced investment') is symbolic. I have tried to avoid mathematical complications by using simple difference equations. To the extent that the approach oversimplifies the problem, the results are vitiated. However, the difference equation approach does fit into the schema by naturally leading to the questions of the implications of the dated changes in the aggregate model for the equilibrium conditions in the particular markets, and the reciprocal question of the implications of changes in the equilibrium in particular markets for the behavior of the aggregate model.

NOTES

1. Haberler (1941) can be considered the definitive volume on such alternative theories.
2. See Hicks (1950).
3. See Mitchell (1950).
4. See Tinbergen (1951), p. 132.
5. See Keynes (1936).
6. See Tinbergen (1951), p. 140.
7. See Tinbergen (1951), p. 140.
8. See Leontief (1947).
9. See Haberler (1941), p. 380.
10. See Metzler (1948), p. 440.
11. See Edgeworth (1925), p. 32: 'a movement along a supply and demand curve of international trade should be considered as attended with rearrangements of internal trade, as the movements of the hands of a clock correspond to considerable unseen movements of the machinery.' Similarly the cyclical movements of the variables in the core model are attended with rearrangements of particular markets: and in turn the rearrangements in the particular markets are significant elements in determining the values of the variables in the core model.
12. See Goodwin (1951).
13. See Samuelson (1944).
14. See Samuelson (1944), p. 269.
15. See Samuelson (1944), p. 215.
16. See Tsiang (1951), p. 326. 'The assumption of a constant propensity to consume in a normal and not too long period may be justified to some extent by statistical investigations of the income consumption function, but the assumption of a constant accelerator dependent solely upon exogenous factors and independent of endogenous processes of the system is without any statistical foundations.'
17. See Duesenberry (1948), p. 60. 'This suggests that in testing hypotheses we ought to operate on the following principles. First, every hypothesis ought to be stated in terms of the behavior of individual firms or households, even when we are only interested in aggregate results. This does not, of course, prevent us from considering interactions among individuals, any more than the use of the theory of the firm in the analysis of monopolistic competition prevents us from dealing with interactions among firms. Second, in so far as it is possible, we ought to test our hypotheses against data which indicate the behavior of individual households or firms.'

18. This analysis is carried further in Chapter 2.
19. See Duesenberry (1948, 1949) and Modigliani (1949).
20. See Marshall (1925), p. 159.
21. See Mitchell (1950), p. 29.
22. See Friedman (1949), p. 470.'In demand analysis the price of closely related commodities are the variables in Group A (variables that are expected both to be materially affected by the variable under study and in turn to affect it). They are put individually into the pound of *ceteris paribus* to pave the way for further analysis. Holding their price constant is a provisional step. They must inevitably be affected by anything that affects the commodity in question; and this indirect effect can be analysed most conveniently by first isolating the direct effect, systematically tracing the repercussions of the direct effect on each closely related commodity and then tracing the subsequent reflex influences on the commodity in question. Indeed, in many ways, the role of the demand curve itself is as much to provide an orderly means of analysing these indirect effects as to isolate the direct effect on the commodity in question.'
23. See Haberler (1941), Pigou (1943), and Patinkin (1948).
24. See Domar (1947, 1946).

2. Some accelerator-multiplier models

1. INTRODUCTION

In spite of the inability of statistical studies to establish that a measurable accelerator coefficient exists, many models of the business cycle have appeared in which the accelerator coefficient is an institutional or engineering constant. In addition to such constant valued accelerator models, models of the business cycle have appeared in which the value of the accelerator coefficient varies over the business cycle. In general the mechanism by which the change in the value of the accelerator coefficient is brought about is not explicitly stated. Assumptions as to the existence of a floor or a ceiling to investment due to the technological limitations upon disinvestment and to the existence of a full employment ceiling (total or sectoral) are made. In addition, a third variety of accelerator business cycle model may be distinguished, one in which the model is subject to random shocks. In this chapter models from each of these classes will be taken up. Variants of such models will be constructed. We will see whether or not these models are consistent with the hypothesis that the business cycle of experience can be best analysed by assuming that the value of the accelerator coefficient varies in a systematic manner over the business cycle, and that the systematic variation in the value of the accelerator coefficient can be imputed to the economic phenomena associated with the different levels and rates of changes in income.

In this chapter we are primarily interested in the formal characteristics of accelerator models, that is in the properties of the mathematical forms used. As such, we will not emphasize the content of the accelerator coefficient, leaving that to the next chapter where we will take up the process by which the accelerator coefficient is determined. Nevertheless it is necessary at this stage to recognize that three quite distinct concepts of the accelerator can be identified. The accelerator can be conceived of as:

1. a structural parameter;
2. a coefficient of induced investment;
3. a coefficient of realized investment.

The accelerator interpreted as a structural parameter depends upon the assertion that there exist fixed proportions between capital and output. Such an aggregate capital-output relation leads to a constant accelerator coefficient. If the production function for output as a whole is such that the proportions between the factors can vary, then the value of the accelerator coefficient depends upon the actual relation between capital and output in the economy. The actual value of the accelerator depends upon those prices which determine the best production technique. Therefore, an accelerator coefficient derived from such a production function is a variable which depends upon prices. As long as the variables which determine the accelerator coefficient are restricted to the determinants of the behavior of firms on the basis of a set of given prices, the accelerator is a coefficient of induced investment.

Both the accelerator as a coefficient of induced investment and as a structural parameter are statements of the demand for investment goods. If we consider the effects of the supply conditions of capital goods as a determinant of the value of the accelerator coefficient, then it becomes a coefficient of realized investment. The distinction between the accelerator as a coefficient of induced investment and the accelerator as a coefficient of realized investment is analogous to the distinction between a *ceteris paribus* and a *mutatis mutandis* demand curve.

Obviously the production function attributes of the economy underlie all three concepts of the accelerator coefficient. If we make rigid proportionality and behavior assumptions, then the structural parameter and induced investment accelerators are the same. If we assume that the supply of capital goods is infinitely elastic at given prices, then the realized investment and induced investment accelerators are the same. In this chapter we will have occasion to point out how different authors have shifted their concept of the accelerator coefficient as they went along.

The accelerator coefficient has also been used in the literature on growth, which is not our particular concern. Even though the approach used by Harrod and Domar on the equilibrium rate of growth differs, they can for our purposes be considered as equivalent in their perspective.[1] The analysis of the conditions under which steady growth is possible is more limited than accelerator and multiplier business cycle theory. The problem which is set out in the growth models is to investigate the special conditions under which Robertsonian type 'saving' is equal to 'investment', and to determine the rate of growth of income (productive capacity in Domar) which is implicit in such an equality. In terms of the models under consideration, such an analysis of growth is a special case of the more general cycle model. In the terminology to be used below, the steady growth state is an explosive state of the economy, where the rate at which income grows is a constant.

The additional factor in these models is that the explosion is taking place at a rate which is consistent with definite limitational factors, although neither Domar nor Harrod are clear or complete in their analysis of these limitational factors.[2] The floors and ceilings version of the accelerator and multiplier models (the Hicks version in particular) of the business cycle can be interpreted as models in which the short run rate of growth of income is not consistent with limitational factors.

2. LINEAR MODELS

2.1 The Hansen-Samuelson Model

Rather than begin the analysis of the accelerator-multiplier type models with the original version – which seems to be the model presented by Harrod,[3] the first accelerator model that used the Keynesian consumption function – it seems best to begin with the exposition presented by Samuelson.[4]

The definition of income in this model is Keynesian, $Y_t = C_t + I_t$, but the consumption function is a dated Robertsonian relation, $C_t = \alpha Y_{t-1}$. Investment is determined by the change in consumption and is defined as: $I_t = \beta(C_t - C_{t-1}) = \beta(\alpha Y_{t-1} - \alpha Y_{t-2})$.

The behavior of the system over time depends upon the solution of the second order difference equation: $Y_t = (\alpha + \alpha\beta) Y_{t-1} - \alpha\beta Y_{t-2})$.

For such second order difference equations four different types of behavior over time are possible,[5] and the behavior over time depends upon the values of the α and β coefficients. The four types of behavior are determined by the combination of two attributes: the time path may be either damped or explosive and either monotonic or oscillatory. Damped behavior implies that in time the value of Y_t approaches some fixed value. Explosive behavior implies that in time the value of Y_t becomes larger (smaller) than any preassigned value. Monotonic behavior means that for any Y_t, Y_{t-1} and Y_{t-2} the sign of $Y_t - Y_{t-1}$ is the same as the sign of $Y_{t-1} - Y_{t-2}$. Oscillatory behavior means that for any Y_t there exists a Y_{t-n1} and a Y_{t-n2} such that the sign of $Y_t - Y_{t-n1}$ is different from the sign of $Y_{t-n1} - Y_{t-n2}$. We therefore have that the value of Y_t can behave in four different ways:

A. damped-monotonic;
B. damped-oscillatory;
C. explosive-oscillatory;
D. explosive-monotonic.

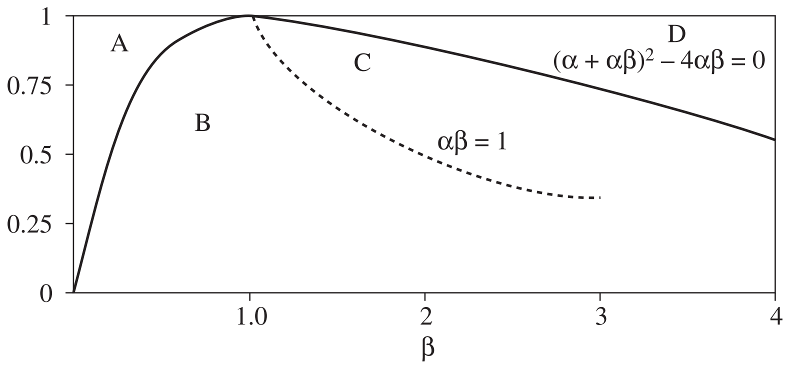


Figure 2.1 Hansen-Samuelson State Diagram

We will call these states of the economy, and the state that the economy is in will depend upon the values of α and β. For given values of α and β the economy is uniquely in one or the other state. For the Hansen-Samuelson model, the relation between the values of α and β and the above four states is shown in Figure 2.1.

The state of the economy depends upon two attributes of the second order difference equation:

1. If $\alpha\beta < 1$ the economy is 'damped', if $\alpha\beta > 1$ the economy is explosive.
2. If $(\alpha + \alpha\beta)^2 - 4\alpha\beta > 0$ the economy is monotonic, if $(\alpha + \alpha\beta)^2 - 4\alpha\beta < 0$ the economy is oscillatory.

The equation $\alpha\beta = 1$ is the border between the damped and explosive behavior. The equation $(\alpha + \alpha\beta)^2 - 4\alpha\beta = 0$ is the border between monotonic and oscillatory behavior. In Figure 2.1, for values of α and β in Region A the behavior of the system is damped monotonic, in Region B the behavior of the system is damped oscillatory, in Region C it is explosive oscillatory and in Region D it is monotonic explosive. For a given value of α, the behavior of the model depends solely upon the value of β. It is true for this form that whether or not the model is explosive or damped, for values of $\beta > 1$, depends upon the value of α.

Each of the four states defined by the values of α and β for this model is unsatisfactory from the point of view of business cycle analysis. State A and State D are not cyclical. In State A the level of income, Y_t, approaches the multiplier level of income as determined by the marginal propensity to consume. In State D the level of income explodes: either up or down depending upon the starting direction.

Regions B and C of the Hansen-Samuelson state figure (Figure 2.1) do lead to cyclical movements. However, the nature of the cycles is unsatisfactory. In Region B the cycles are damped; in Region C they are explosive. If the values of α and β were such that the economy was in State B, the cycle would, without outside influences, disappear. If the values of α and β were such that the economy was in State C, the amplitude of the cycles would become progressively greater. The boundary between C and B, which is defined by $\alpha\beta = 1$, yields a cycle of constant amplitude. This is unsatisfactory, except perhaps as a classroom expository device, because of the coincidence of values required and because the regularity of the business cycle of experience is not consistent with this result.

As all the states of the economy which the Hansen-Samuelson model yields are unsatisfactory, business cycle models based upon the interaction of the accelerator and multiplier must make some additional specifications. The specifications which have been put forward can be classified as follows:

1. specifications that there exist floors and ceilings;
2. specifications that there exists an outside energy source which maintains an otherwise damped cycle;
3. specifications that the accelerator coefficient varies in a systematic manner.

If floors and ceilings are used, the assumption in regard to β that is made is that the normal value of the accelerator coefficient is such as to place the economy in the 'explosive state' and the hitting of the ceiling or floor causes the 'turning point'. This type of business cycle theory can be considered as analogous to the Wicksellian cumulative process engendered by the disparity between the bank rate and the real rate of interest. A few models have been constructed which are based upon a systematic variation in the accelerator coefficient. These models do not explain how the systematic variation in the accelerator is generated. These models are incomplete. The aim of this volume is to analyse the determinants of the value of the accelerator, and therefore to complete such models.

Models which depend upon an outside energy source are models which normally assume values of the accelerator and multiplier that put the economy in State B. The Frisch 'Propagation Problems and Impulse Problems'[6] paper presented a model of this type, and the Schumpeterian innovation hypothesis can be considered as a case of a Region B accelerator-multiplier model with the innovations as the outside energy source. Certainly Schumpeter's outlook upon the business cycle considered the business cycle as essentially the result of secondary waves which followed upon the impulse due to the innovation.[7] Another type of accelerator and multiplier model which places the economy in Region B is one that uses a 'stochastic' energy source.

2.2 The Hicks Type of Linear Model

In the Hansen-Samuelson model investment is a function of the change in consumption: $I_t = \beta(C_t - C_{t-1})$ which becomes $I_t = \alpha\beta(Y_{t-1} - Y_{t-2})$. Hicks, in his volume on the Trade Cycle, writes the accelerator-multiplier model in the form:[8]

$$Y_t = C_t + I_t$$

$$C_t = \alpha Y_{t-1}$$

$$I_t = \beta^*(Y_{t-1} - Y_{t-2})$$

This yields the second order difference equation:

$$Y_t = \alpha Y_{t-1} + \beta^*(Y_{t-1} - Y_{t-2})$$

Again four different types of behavior of Y_t are possible depending upon the values of the α and β^* coefficients. If the accelerator coefficient $\beta^* < 1$, the solution is damped; if $\beta^* > 1$, the solution is explosive. If $(\alpha + \beta^*)^2 - 4\beta^* > 0$, the solution is monotonic; if $(\alpha + \beta^*)^2 - 4\beta^* < 0$, the solution is oscillatory. The equation $\beta^* = 1$ is the border between the damped and explosive behavior; the equation $(\alpha + \beta^*)^2 - 4\beta^* = 0$ is the border between oscillatory and monotonic behavior. The states of Regions A, B, C and D are the same as in the Hansen-Samuelson model:

Region A: monotonic-damped state
Region B: oscillatory-damped state
Region C: oscillatory-explosive state
Region D: monotonic-explosive state

This Hicks version of the Accelerator-Multiplier model has one significant difference compared with the Hansen-Samuelson version. Whether or not the economy is in an explosive or damped state depends upon the value of the marginal propensity to consume in the Hansen-Samuelson model, whereas in the Hicks version it is independent of the value of α. In the Hicks model the only influence which the value of α has is in determining whether or not the economy is oscillatory or monotonic. Inasmuch as the formal difference between the two models can be reconciled, by

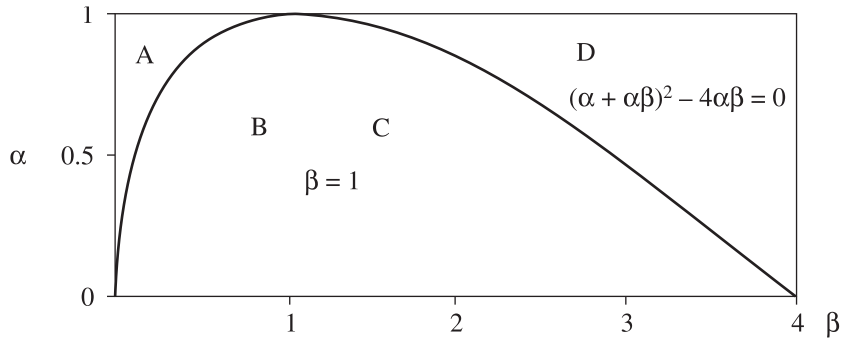


Figure 2.2 Hicks Variant of the Accelerator-Multiplier State Diagram

setting $\alpha\beta = \beta^*$, it seems more straightforward to recognize the role of the propensity to consume as a possible destabilizer or stabilizer. In addition, the Hansen-Samuelson model can be interpreted so that both the accelerator coefficient and the propensity to consume are cyclical variables. However, as the Hicks version is somewhat simpler mathematically, it will be more convenient to use the Hicks form of the linear model in most of what follows.

An analysis of the Hicks form, $Y_t = (\alpha + \beta)\, Y_{t-1} - \beta\, Y_{t-2}$, will enable us to derive the state diagrams and will throw further light on the characteristics of accelerator-multiplier models. The characteristic equation of a second order difference equation is $f(x) = x^2 - (\alpha + \beta)x + \beta$. The behavior of income over time depends upon the roots of the characteristic equation. That is, $Y_t = K_1(r_1)^t + K_2(r_2)^t$ where K_1 and K_2 depend upon the initial conditions and r_1, r_2 are the roots of the equation $f(x) = 0$.

If both α and β are greater than zero, then either both roots are real and positive or the roots are a pair of conjugate complex numbers. If a real root is greater than one, the effect of that root is explosive; if it is less than one the effect of that root is to damp the movement of income. Complex roots lead to oscillatory states. The resulting time series of income is explosive if the modulus $|r|$ is greater than one, and damped if the modulus $|r|$ is less than one. In the Hicks form the modulus is equal to β.

The derivative of the characteristic equation is: $f'(x) = 2x - (\alpha + \beta)$. If $x < [\alpha + \beta]/2$, the derivative is negative; if $x > [\alpha + \beta]/2$, the derivative is positive. Hence $f(x)$, the characteristic equation, is decreasing in the range $-\infty < x < [\alpha + \beta]/2$, and increasing in the range $[\alpha + \beta]/2 < x < \infty$. The minimum point of $f(x)$ is $([\alpha + \beta]/2)$. We also have the following tables of values for $f(x)$:

x	$f(x)$
0	β
1	$1 - \alpha$
$\alpha + \beta$	β

If $f([\alpha + \beta]/2) > 0$, then there are no real roots.[9] If $f([\alpha + \beta]/2) < 0$, then there are two distinct real roots, both of which are either in the range $0 < x < 1$ or $1 < x < \alpha + \beta$. If the roots are less than one, then the behavior is damped and if the roots are greater than one, the behavior is explosive. If $f([\alpha + \beta]/2) = 0$, then there are two identical real roots, $r_1 = r_2 = [\alpha + \beta]/2$.[10] In this case $Y_t = (K_t + K_2)([\alpha + \beta]/2)^t$, so that if $[\alpha + \beta]/2 > 1$, we have a constant rate of growth, and if $[\alpha + \beta]/2 < 1$, we have a constant rate of decline of income.

If $\beta < 1 - \alpha$ so that $\alpha + \beta < 1$, then the two roots if real are less than one. If $\alpha + \beta = 1$, the two roots if real are less than one. If $\beta = 1$ and $\alpha > 0$ then $f([\alpha + \beta]/2) = 1 - (\alpha + \beta)^2/4 > 0$; the system is oscillatory, and as β is the modulus, the result is a constant amplitude oscillation. This is the border between Regions B and C.

If $f(x)$ is such that there are two distinct real roots greater than one, then $Y_t = K_1 r_1^t + K_2 r_2^t$. Assume $r_2 > r_1$. Y_t is a weighted average of r_1^t and r_2^t. In time r_2^t will dominate the relation between Y_t and Y_{t-1}. The rate of growth of income will increase and will approach r_2 as a limit. If the solution to the characteristic equation yields two distinct positive real roots greater than one, the rate of growth of income increases in time. This is the way in which income behaves in Region D of the state diagram (Figure 2.2). The Harrod-Domar case of steady growth is where $r_1 = r_2 > 1$ so that $Y_t = (K_1 + K_2)r^t$ and $Y_t / Y_{t-1} = r$ a constant. The Harrod-Domar case is suspect because of the peculiar coincidence of values necessary to achieve its result.[11]

3. NON-LINEAR MODELS

3.1 The Goodwin Model

Both the Hansen-Samuelson version and the Hicks linear version of the accelerator and multiplier models are unsatisfactory from the point of view of business cycle theory because of the nature of the possible states in which the economy may be. 'By dropping the highly restrictive assumption of linearity we neatly escape the rather embarrassing special conclusions which follow (in linear theory). Thus, whether we are dealing with differences or differential equations, so long as they are linear, they either explode or die away with the consequent disappearance of the cycle or the society.'[12] The suggestions that non-linear relations be used as a substitute for the linear form of the simplest theory has appeared in two forms: one a 'formal exposition' by Goodwin and the second a longer informal exposition by Hicks. In what follows we shall follow Goodwin's formal exposition through the essentials of his first three models. The essential difference between the Hicks and the Goodwin non-linear models is that in Goodwin the accelerator coefficient's non-linearity is determined in the first instance by the relation between actual and desired capital whereas in Hicks the non-linearity arises due to a 'full employment' ceiling to output, which results, once full employment is reached, in making Y_t of the order of Y_{t-1}. This results in but little induced investment even though the accelerator coefficient is unchanged. Hicks' non-linearity depends upon a full employment ceiling to income whereas Goodwin's non-linearity, at least in his third model, depends upon the 'generating relation for the accelerator coefficient'. As such, nothing in Goodwin's third model is inconsistent with this thesis.[13]

'The central difficulty with the acceleration principle is that it assumes that actual realized capital stock is maintained at the desired relation with output. We know in reality this is seldom so, there being now too much and now too little capital stock.'[14] The significant relation to Goodwin is the relation between the actual capital stock k and the capital stock desired for a given output Y. Inasmuch as Y is being produced, and it is possible to produce Y without having the desired capital stock, the production function for output as a whole which Goodwin uses is one that involves alternative combinations of factors capable of producing a given output: Goodwin's production function for output as a whole involves substitution among factors. This is in and of itself an improvement upon most accelerator doctrines. The investment decision depends upon the relation between existing capital and the desired capital for a given realized output: we can assume therefore that the investment decision is based upon the difference between the 'present plant' and the 'plant which can produce today's output at the least cost'. But the 'plant which can produce today's output at the least cost' is dependent, in a production function with substitution, upon the relative prices of the factors of production. If investment costs are 'high', the amount of investment induced by a given output greater than the 'best' output for a given plant will be smaller than if investment costs are 'low'. The amount of investment necessary to bring a realized capital stock into the desired relation with output depends upon the rate of interest, among other determinants of the best way to produce a given output.

The reason given by Goodwin for the divergence between the actual quantity of capital and the desired stock of capital during the period in which there is too little capital stock is that 'the rate of investment is limited by the capacity of the investment goods industries'.[15] This is, of course, a judgement about the nature of the world as one could construct a model in which the limited capacity of the capital goods producing industries is just sufficient to satisfy the demand derived from the difference between the desired and the existing capital stock. Also, if the desired increase in capital stock is greater than the limited productive capacity of the capital goods industries, the price of capital goods can be expected to change. However, Goodwin ignores these possibilities, so we have that the maximum amount of investment possible is a constant, $\dot{k}^*$, per period, and if the desired capital is ξ, the actual capital is k, maximum investment will take place for $[\xi - k]/\dot{k}^*$ periods. An additional gratuitous observation by Goodwin that 'entrepreneurial expectations are such that, even if it were possible to expand plant in the boom, there would be great resistance to it' is made.[16] Aside from being a casual empiric assertion about entrepreneurial behavior, it would also, if taken seriously, make any investment theory of the business cycle impossible – for during the boom entrepreneurs resist investment, therefore no investment takes place at high incomes. This flies in the face of any casual observation of business cycle behavior.

The reason for the divergence between the actual quantity of capital and the desired quantity of capital during the period in which there is too much capital is that 'Machines, once made, cannot be unmade, so that negative investment is limited to attrition from wear, from time, and from innovation'.[17] This ignores, as is typical in such models, disinvestment in working capital which is not limited to wear, time or innovation, but is limited to the level of consumption and investment purchases or by the level of such stocks (for example, if output drops to zero, and all sales are out of stocks). This observation also applies to investment: the maximum amount of investment possible is not equal to the 'capacity of investment goods industries', but is equal to the production of all commodities which can be stored. Continuing to follow Goodwin, we can define a negative investment rate $\dot{k}^{**}$ per period, and if the desired capital is ξ, and the actual capital k, we have that maximum disinvestment will take place for $[k - \xi]/\dot{k}^{**}$ periods. ($k > \xi$ for disinvestment.) We can also define a situation in which $k = \xi$, which implies zero net investment. Because of these limitations we have that 'capital stock cannot be increased fast enough in the upswing, nor decreased fast enough on the downswing, so that at one time we have shortages and rationing of orders and at the other excess capacity with idle plants and machines'.[18]

The problem remains of defining the desired capital stock. Goodwin defines $\xi = KY$; ξ = desired capital stock, Y = income, K a constant. This is a linear relationship between desired capital stock and output. We therefore have a production function in which there is substitution in the short run: for Y can be produced with a capital stock not equal to ξ, but the long run production function has fixed proportions.[19] Once the substitution is admitted, there seems to be no reason why it should not also be allowed in the long run, writing $K = K(P_L, P_K)$, the accelerator coefficient K becoming a function of the relative price of labor and capital for example. We therefore would have that $\xi = K(P_L, P_K)Y$. If the price ratios of the factors change so that capital becomes relatively more expensive, the quantity of capital desired to produce a given income decreases, whereas if capital becomes relatively cheaper, the desired capital stock increases. Such a relation between the accelerator coefficient and the relative prices of the factors could be utilized to integrate money market phenomena with the accelerator-multiplier type model. For example, if the amount of financing ability in a community is inelastic, so that when induced investment is high the price of financing rises relatively faster than the price of labor, the accelerator coefficient will decrease. Conversely, if the price of financing falls more rapidly than the price of labor, the result may be a high desired relation between capital stock and output. This would tend to shorten the period $[k - \xi]/\dot{k}^{**}$ during which income is falling.[20]

Combining the above relations, Goodwin obtains his simplest model:

$$\xi = KY \quad \text{(desired capital = constant} \times \text{income)}$$
$$C = \alpha Y + \alpha_0 \quad \text{(a linear consumption function)}$$
$$Y = C + \dot{k} \quad \text{(the definition of income: } \dot{k} = dk/dt\text{)}$$

the time rate of change of capital, hence investment. We therefore have $y = \alpha_0/[1 - \alpha] + \dot{k}/[1 - \alpha]$. Goodwin assumes 'that the economy seeks the perfect adjustment of capital to output and that it does so in either of two extreme ways, capacity output of investment goods or zero gross investment'.[21] Using the relation between desired investment and income we have $\xi = (K/[1 - \alpha])\dot{k} + K\alpha_0/[1 - \alpha]$, and as $k = \dot{k}^*$, 0, or $\dot{k}^{**}$ as $\xi \lesseqgtr k$ we would have that $\xi = \xi^*$, ξ_0, or ξ^{**}. We therefore have three possible levels of income, investment and desired capital. The set of values of $y = \alpha_0/[1 - \alpha]$, $k = 0$, and $\xi = \xi_0 = k$ are equilibrium values. 'It is, however, an unstable equilibrium, since a small displacement in the phase plane leads to a large displacement from which it never returns. For example, if to ξ_0 we add $\Delta\xi$, then $\dot{k}$ changes from zero to $\dot{k}^*$ and ξ becomes ξ^*.'[22] The model operates in the following manner: for $[\xi^* - k]/\dot{k}^*$ periods the level of income is $Y = \alpha_0/[1 - \alpha] + \dot{k}^*/[1 - \alpha]$. At the end of this time, $\xi^* = k$, therefore $\dot{k} = 0$. However we have that the desired capital for $Y = \alpha_0/[1 - \alpha]$ is equal to ξ_o which is less than the desired capital ξ^* for $Y = \alpha_0/[1 - \alpha] + \dot{k}^*/[1 - \alpha]$. Therefore, investment falls to $\dot{k}^{**}$ which leads to an income $Y = \alpha_0/[1 - \alpha] + \dot{k}^{**}/[1 - \alpha]$ with a desired capital stock ξ^{**}. This income lasts for $[\xi^* - \xi^{**}]/\dot{k}^{**}$ periods ($k = \xi^*$ when the change in income occurs) at which time $k = \xi^{**}$. This leads to $Y = \alpha_0/[1 - \alpha] > Y = \alpha_0/[1 - \alpha] + \dot{k}^{**}/[1 - \alpha]$

which implies $\dot{k}^*$ investment, $Y = \alpha_0/[1-\alpha] + \dot{k}^*/[1-\alpha]$ which implies ξ^* as desired capital. This income continues for $[\xi^* - \xi^{**}]/\dot{k}^*$ periods ($k = \xi^{**}$ when the change in income occurs). Inasmuch as $\dot{k}^{**}$ is assumed to be less than $\dot{k}^*$, the time spent in the low income state is greater than the time spent in the high income state.

This crude model does illustrate the general characteristics of non-linear models:

A. The final result is independent of the initial conditions.
B. The oscillation maintains itself without any need of outside 'factors' to help in the explanation. In this sense, it is a complete self-contained theory.
C. The equilibrium is unstable and therefore the mechanism starts itself given even the smallest disturbance. Yet in spite of this instability it is a usable theory because the mechanism does not explode or break down but is kept within bounds by the non-linearity.
D. No questionable lags are introduced. The mechanism operates by its own structure.[23]

The deficiencies of this model are obvious, 'such a crude model cannot claim to be a representation of actual cycles'.[24] In particular income has only two levels, and the investment which takes place during the upswing does not increase productive capacity. As an expository device this Goodwin model may suffice, but as a framework for business cycle analysis it is even cruder than the original linear Hansen-Samuelson model.

Goodwin's second model makes allowance for technological progress. 'To make a crude allowance for technological progress, we may assume a steady growth in the desired amount of capital.'[25] We therefore write $\xi = a^t + kY$, $\dot{\xi} = a$ (the time rate of change of ξ). In this model, 'no equilibrium exists since $\dot{k} = 0$ means that k is constant and hence that ξ would become greater than k and hence $\dot{k}$ would cease to be zero'.[26]

The rate of growth of desired capital $\dot{\xi} = a$, and Goodwin again assumes that the rate of change of capital $\dot{k}$ has two values. 'If a is greater than $\dot{k}^*$, the economy can never catch up with its capital needs. Excluding this unrealistic case . . .'.[27] The unrealistic case unfortunately may not be so casually excluded during at least a portion of the business cycle. The possibility that during a strong boom the rate of desired growth of capital may outstrip the possible rate of growth of capital cannot be ignored. An inflationary period may be viewed as one in which the supply of resources for capital expansion is less than the demand of resources for capital expansion. An inflationary period may be interpreted as a time during which the financing of investment by sources outside of voluntary real savings are ineffective in yielding as high a rate of investment as entrepreneurs desire. The history of economies which have had long periods of open or suppressed inflation can be interpreted as a 'donkey-carrot' arrangement between actual and desired capital equipment. This may be particularly true in an economy that is beginning its industrial revolution and in an economy which has had a portion of its capital equipment destroyed during a war.

However, by assuming that $\dot{k}^* > a$, Goodwin achieves a cycle with growth. As $\dot{k}^* > a$ in time t, $k = \xi$, so that desired capital falls from $\xi^* = a^t + K(\beta/[1-\alpha] + \dot{k}^*/[1-\alpha])$ to $\xi^{**} = a^t + K(\beta/[1-\alpha] + \dot{k}^{**}/[1-\alpha])$. At this stage k is being decreased. However ξ^{**} increases due to 'technological change' so that in time $\xi^* = k$, which raises income to $\beta/[1-\alpha]$. This makes desired capital greater than actual capital and leads to investment at a rate $\dot{k}^*$. This model therefore operates in essentially the same manner as the simplest model, but it does succeed in reducing the relative length of the low level of income. Even though 'technological change' is taking place, the high level of income $Y = \beta/[1-\alpha] + \dot{k}^*/[1-\alpha]$ does not rise. Net investment is taking place without any change in attainable income.

Goodwin expands the above models in three directions. He introduces a dynamical multiplier, an investment lag, and he generalizes his non-linear accelerator. Only the third need concerns us – his generalization of the non-linear accelerator. 'The investment, $\dot{k}$, consists of an autonomous part, $\ell(t)$, and an induced part φ. About induced investment we may make the less crude (than the previous one) assumption that the acceleration principle $\xi = KY$ holds over some middle range but passes to complete inflexibility at either end as is shown in (Goodwin's) Figure 4. The upper limit is the $\dot{k}^*$ of the previous models and the lower limit the $\dot{k}^{**}[d\varphi(\dot{y})/d(\dot{y})]$ is

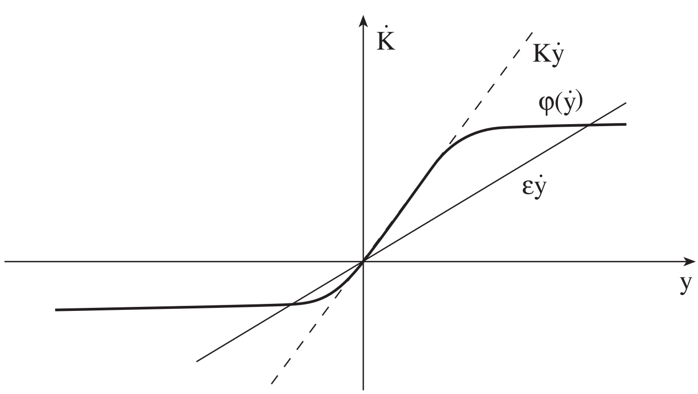


Figure 2.3 Goodwin's Figure 4

equal to the acceleration coefficient, κ, in its middle range and zero (or some quite small value) at either extremity.'[28] The function $\varphi(\dot{y})$ interests us; the function $\varepsilon(\dot{y})$ (his lagged consumption as a function of the change in income) need not concern us.

Goodwin's $\dot{k} = \varphi(\dot{y})$ is a function relating the change in capital to the change in income. The largest quantity of investment possible is $\dot{k}^*$, and when the rise in income is large enough so that $\kappa\dot{y} \geq \dot{k}^*$ no further increase of investment can take place. Therefore, the 'effective' accelerator coefficient falls to $\dot{k}/\dot{y}$. The argument for the effects of a fall in income is the same. No argument is advanced for the assumption that there exists an invariant fixed ceiling to the amount of investment that can take place, either on the upswing or on the downswing. As far as the technical mathematical exposition we can accept all his results. The interesting economics, however, is involved in the relation between investment and the change in income.

In order to compare Goodwin's non-linear accelerator with our previous material, let us first change his model into one in terms of difference rather than differential equations, so that investment depends upon the differences between today's income and yesterday's income and secondly, let us define a realized accelerator coefficient β as the relation between the change in income and investment. If we graph the realized accelerator coefficient β against the change in income, we derive from Goodwin's assumptions that the graph of the value of the accelerator coefficient can be broken into three segments.

1. The central part, where $[\dot{k}^{**}/\kappa] < Y_t - Y_{t-1} < [\dot{k}^*/\kappa]$, has $\beta = \kappa$ and κ is the technological constant relating desired capital to output. In this range the accelerator β is a structural parameter.
2. The left-hand portion, where $[\dot{k}^{**}/\kappa] < Y_t - Y_{t-1}$, β is a variable, determined by the equation $\beta(Y_t - Y_{t-1}) = \dot{k}^{**} < 0$. $\dot{k}^{**}$ is a technological

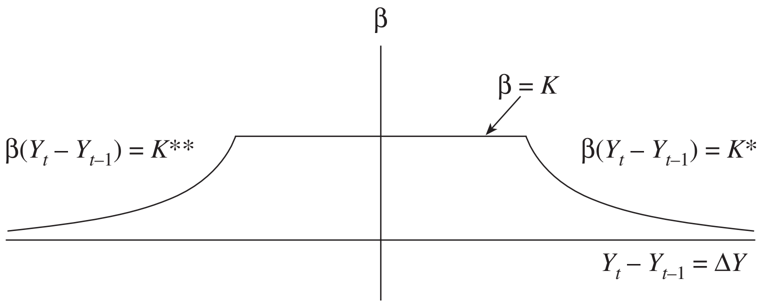


Figure 2.4 Relation between β *and* Δ*Y*

constant, the ceiling rate of disinvestment. β in this range is a coefficient of realized investment, where the ceiling on disinvestment determines the effect of the stimuli given by the change in income.

3. The right-hand portion, where $\dot{k}^*/\kappa > Y_t - Y_{t-1}$, β is again a variable, determined by the equation $\beta(Y_t - Y_{t-1}) = \dot{k}^*$. $\dot{k}^*$ is a technological constant, a ceiling rate of investment. β again is a coefficient of realized investment.

Note that in these definitions of the realized accelerator coefficient, the desired amount of capital ξ which played such a large role in the first two models of Goodwin's has disappeared. Instead of having the mechanism of the accelerator depend upon the relation between desired and actual capital, and having the limited investment and disinvestment rates determine the time spent in bringing the actual capital into equality with the desired capital, the non-linearity of the β coefficient depends upon the ceilings to investment and disinvestment. Rather than being purely a technical concept, β becomes a technical concept and a coefficient of realized investment.

If we combine the relation between the change in income and the value of the realized accelerator coefficient with the state diagrams of the Hansen-Samuelson model, we get that, for a fixed value of α, as the value of β changes, the economy moves from one state to another. If we assume that the value of κ is such as to place the economy in the monotonic explosive state, State D, then as $Y_t - Y_{t-1}$ increases the value of β falls sufficiently to place the economy into States A, B or C. However, in these states the amount of induced investment which takes place is, according to Goodwin's Hypothesis, a constant equal to $\dot{k}^*$. Income therefore is $Y_t = \alpha_0/[1 - \alpha] - \dot{k}^*/[1 - \alpha]$, and as when $\beta = \kappa$ total induced investment is less than $\dot{k}^*$, Y_{t+1} will be greater than Y_t. If $Y_{t+1} - Y_t$ is sufficiently great

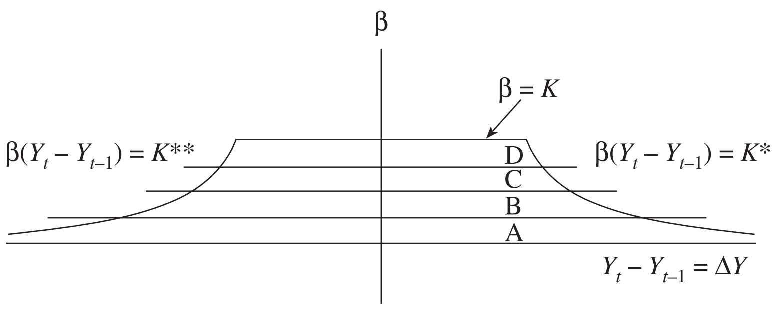


Figure 2.5 Hansen-Samuelson States and the Relation between β *and* Δ*Y*

to leave the economy in the region determined by $Y_{t+1} - Y_t > \dot{k}^*/\kappa$, then investment remains at $\dot{k}^*$, so that $Y_{t-2} = Y_{t-1}$. This results in a fall of $Y_{t-2} - Y_{t-1}$ to zero; therefore, zero investment is induced which leads to a fall in income. If the behavior of the economy when income falls is symmetric with its behavior when income rises, then we have a cycle which is essentially like the cycle sketched in the first chapter which was based upon a cyclically variable accelerator.

If $\beta = \kappa$ is sufficiently large to lead to a monotonic explosion, the economy will in time reach the state where $K(Y_t = Y_{t-1}) > \dot{k}^*$, so that investment will become equal to $\dot{k}^*$. Nevertheless $Y_{t+1} > Y_t$. If $Y_{t-1} - Y_t$ is small enough so that β again equals K. We know that Y_{t-2} will be smaller than Y_{t-1} for

$$Y_{t-1} = \frac{\alpha_0}{1-\alpha} - \frac{\dot{k}^*}{1-\alpha}$$

$$Y_{t-2} = \frac{\alpha_0}{1-\alpha} - \frac{K(Y_{t-1} - Y_t)}{1-\alpha}$$

and

$$K(Y_{t-1} - Y_t) < \dot{k}^*$$

This leads to a downward movement of income which continues until the disinvestment ceiling $\dot{k}^{**}$ is reached. That the economy enters such a cycle independently of the initial conditions is obvious. This mechanical model with β variable over the cycle can be modified in a number of directions. It is obvious that the downturn in income that takes place in the model is independent of whether or not the actual capital stock becomes equal to the desired capital stock.

Goodwin's model can be modified in a number of ways, not all of which lead to satisfactory models of the business cycle. In particular the assumption that K, the 'technological accelerator coefficient', is large enough so that the economy is in the explosive (monotonic or cyclical) state when $\beta = K$ has to be made.[29] If, when $\beta = K$, the economy is in either of the damped states (States A or B), the model will be equivalent to the linear accelerator-multiplier models with the need for an external energy source if the cycle is not to die out.

If we modify the model by allowing $\dot{k}^*$ to increase with time, this will mean that the peak of income will rise. However, as the peak capital stock also increases, the rate of depreciation of capital $\dot{k}^{**}$, which can be taken as a function of the volume of capital, will also increase. This nonlinear business cycle model will therefore have a result that the amplitude of the cycle (which is $[\dot{k}^* - \dot{k}^{**}]/[1 - \alpha]$) will increase in time. The Goodwin type model, with this modification, will behave in a similar way to a linear accelerator model if it is in State C. This implication of the model is not satisfactory, and the assumption which is most vulnerable, as far as the empirical evidence is concerned, is whether $\dot{k}^{**}$ remains a constant proportion of peak capital equipment. If business cycles are a part of business experience, one would expect arbitrage behavior to result in a fall of $\dot{k}^{**}$ during later stages of the cycle: that the maximum fall in capital will decrease, for the expectation of a rise in income will make it profitable, at some stage, to maintain capital or even begin to expand capital in spite of the present downward movement of income. The entire assumption that capital decreases, that there is net disinvestment during a cyclical downswing, is of course suspect. What data there are on national income indicate that net investment in fixed capital is rarely, if ever, less than zero.

Goodwin's assumption, that the realized value of the accelerator coefficient β is at a maximum when $Y_t - Y_{t-1}$ is small, is also suspect. That small changes in income, upward or downward, would induce large amounts of investment, whereas large changes in income will not, seems contrary to both theory and evidence. Small changes in income can be considered as ineffective in inducing investment, whereas large changes may be efficient. If we wished to incorporate this into Goodwin's model of the business cycle, which is based upon the existence of an investment ceiling, we have to consider that the accelerator coefficient is an increasing function of the rate of change of income (both for positive and negative changes in income) until it reaches the rectangular hyperbola segments of the relation between a change in income and the accelerator coefficient. Figure 2.6 indicates how β varies with changes in income in this case. Such a variable accelerator coefficient business cycle model, unless the 'peaks' in State D

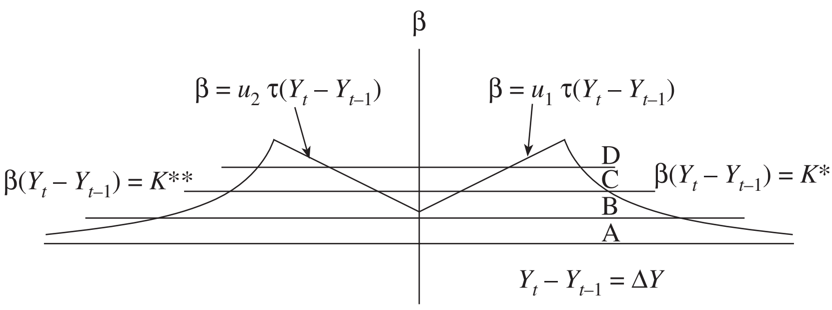


Figure 2.6 Hansen-Samuelson States and the Relation between β and ΔY

are large, would have a strong tendency to settle down into the damped states. Such a model is taken up in the next section.

The above comments on the non-linear accelerator models of Goodwin are sufficient to show the following:

A. The assumption of a non-linear accelerator coefficient is insufficient by itself to yield a satisfactory cycle model.
B. Once we make the assumption that the accelerator coefficient is not a constant over the cycle, the possibilities of the nature of the non-linearity which can occur are legion.
C. The assumptions as to the foundations of the non-linearity of the accelerator coefficient which Goodwin made are not the only assumptions which can be made.
D. Goodwin changes his assumptions as to the basis of the non-linearity of the accelerator coefficient during the course of his argument. At one stage his non-linearity is based upon a difference between the desired and the existing capital stock; at another it is based upon a ceiling rate of change of capital stock.
E. If non-linear accelerator models are to serve any useful purpose aside from exhibiting the mathematical virtuosity of their authors, it is necessary to base the non-linearity of the accelerator coefficient upon a foundation of household and firm behavior.

Therefore, as non-linear accelerator-multiplier models can, but do not necessarily, yield an appropriate cyclical model, it is necessary if we are to use such models in business cycle theory, to have a theory of the determination of the value of the accelerator coefficient. This theory should be designed to enable us to select, from the multiplicity of the possible accelerator generating functions, a particular set of accelerator determining relations as being consistent with the known behavior of elementary economic units, in particular, the behavior of business firms.

3.2 Alternative Non-Linear Models

An interesting formulation of the accelerator principle, alternative to Goodwin's, which yields a non-linear accelerator, and which is based upon an analysis of the behavior of particular firms, can be derived from a suggestion by Tinbergen:

> During a period of increasing production, not all firms and not all branches attain at the same moment the point of full capacity. Suppose that for individual firms the principle acts only, and then of course in its rigorous form, when

> full capacity is reached; then to a given increase in the total production for all firms a smaller percentage increase in the total stock of capital goods may correspond . . .[30]

The proportion of firms that attain full capacity production during any period of time can be related to the level of income in comparison with the 'previous peak' level of income which is used as a measure of the economy's productive capacity, and to the size of the change in the level of income. The reason for using the above is that the value of the accelerator coefficient can be considered as a function of the proportion of firms that attain full production during any period and therefore as a function of the level of income, and previous peak income. The size of the change in income will be a measure of the impact of rising incomes upon firms, and therefore can be expected to affect the inducement to investment.

The relation between the value of the accelerator coefficient and the level of income can be made more precise if we examine the relation between present income Y_t and the highest previous income that the economy has achieved $\overset{*}{Y}$. If Y_t is taken during a period of rising income $Y_t > Y_{t-1}$ then $Y_t \gtreqless \overset{*}{Y}$. If $Y_t < \overset{*}{Y}$, then historically the economy has 'excess capacity'; although if we allow for 'capital consumption' during the time for which Y_t has been smaller than $\overset{*}{Y}$, the productive capacity at time t may not be greater than that necessary to produce Y_t. If $Y_t = \overset{*}{Y}$ then the economy is operating at its 'historical' full capacity. If $Y_t > \overset{*}{Y}$ then either $Y_{t-1} = \overset{*}{Y}$ or $Y_{t-1} < \overset{*}{Y}$. If $Y_{t-1} = \overset{*}{Y}$, then we would expect a 'larger' proportion of the firms to have reached the point of full capacity than if $Y_t < \overset{*}{Y}$ or if $Y_t > \overset{*}{Y}$ but $Y_{t-1} < \overset{*}{Y}$. We would then have a lower accelerator coefficient if present income is not the 'peak income' than we would have if present income were the peak income; and if present income is the peak income, we would expect a lower accelerator coefficient if the previous period's income were not the peak income to date than if the previous period's income were Y. We could write that if

$$
\begin{aligned}
Y_t - Y_{t-1} > 0 \text{ and if } Y_t < \overset{*}{Y} &\quad \text{then } \beta = \beta_0 \\
Y_t = \overset{*}{Y} &\quad \text{then } \beta = \beta_1 \\
Y_t > \overset{*}{Y} \text{ and } Y_{t-1} < \overset{*}{Y} &\quad \text{then } \beta = \beta_2 \\
Y_t > \overset{*}{Y} \text{ and } Y_{t-1} = \overset{*}{Y} &\quad \text{then } \beta = \beta_3
\end{aligned}
$$

and have $\beta_0 < \beta_1 < \beta_2 < \beta_3$. This relation between the value of the accelerator coefficient and the relation between the present income, the previous period's income and the highest income to date is similar to the relation between the value of the marginal propensity to consume and income in previous periods that has been used to explain the secular upward drift of the consumption function.[31]

The second factor that can influence the value of the accelerator coefficient is the size of the change in income. If $Y_t - Y_{t-1}$ is 'large' a greater number of firms will approach full capacity than if small, and therefore the accelerator coefficient will vary directly with $Y_t - Y_{t-1}$.

As a first step in constructing a relation between the accelerator coefficient, the level of income and the change in income, we shall focus our attention upon the relation between the accelerator coefficient and the change in income, $Y_{t-1} - Y_{t-2}$. Let us assume that the aggregate accelerator coefficient β_t for a time period t, a period of increasing production, is the 'sum' of the acceleration coefficients of particular firms. The accelerator coefficient for a particular firm is a function of the shift in the demand curve confronting the industry. If we define $Q_\lambda(t-2)$ as the amount taken of a product at the market price in period $t-2$, and $Q_\lambda(t-1)$ as the amount that would be taken during period $t-1$ at the price that ruled in the market during period $t-2$, $Q_\lambda(t-1) - Q_\lambda(t-2)$ can be taken to measure the shift in the demand curve of the product. If we define $\gamma_{\lambda\rho}$ as the amount of investment that the ρth firm in the λ industry will make as a result of a shift in the industry demand curve $Q_\lambda(t-2) - Q_\lambda(t-1)$ we have that aggregate induced investment I_t will be

$$I_t = \sum_\lambda \sum_\rho \gamma_{\lambda\rho}[Q_\lambda(t-1) - Q_\lambda(t-2)]$$

and as the accelerator relation is of the form

$$\beta_t(Y_{t-1} - Y_{t-2}) = I_t$$

we have that

$$\beta_t = \frac{\sum_\lambda \sum_\rho \gamma_{\lambda\rho}[Q_\lambda(t-1) - Q_\lambda(t-2)]}{Y_{t-1} - Y_{t-2}}$$

Now if for each firm in each industry $\gamma_{\lambda\rho} = 0$ when

$$Q_\lambda(t-1) - Q_\lambda(t-2) < \xi_\lambda$$

and if the size of the shifts in demand curves is a function of the size of the change in income $|Y_{t-1} - Y_{t-2}|$, we have as $|Y_{t-1} - Y_{t-2}|$ increases, the number of industries for which $Q_\lambda(t-1) - Q_\lambda(t-2) \geq \xi_\lambda$ increases. Therefore the value of the accelerator coefficient is of the form $\beta_t = \varphi(Y_{t-1} - Y_{t-2}, \xi)$ and if $Y_{t-1} - Y_{t-2} > \xi_0$, $d\beta_t/[d(Y_t - Y_{t-1})] > 0$, β_t increases when the change in income increases.

This may also be formulated as a 'sampling problem', a formulation which will be discussed in greater detail in the following section. Let us consider the set of all firms $\lambda\rho$. The value of the accelerator coefficient is determined by the number of firms 'activated' or 'drawn' by means of a change in income $Y_{t-1} - Y_{t-2}$. When production is increasing and $Y_{t-1} - Y_{t-2} > \xi_1$, ($Y_{t-1} - Y_{t-2}$ is large) the entire population of firms is drawn in the sample and the value of β_t is large. When $Y_{t-1} - Y_{t-2} < \xi_0$, ($Y_{t-1} - Y_{t-2}$ is small) no firm's investment is activated by the change in income, the sample size is zero, and the value of β_t is zero. That is, the size of the sample is zero when $Y_t - Y_{t-2} < \xi_0$, the sample is equal to the population when $Y_{t-1} - Y_{t-2} \geq \xi_1$, and if $\xi_0 < Y_{t-1} - Y_{t-2} < \xi_1$, the size of the sample of firms affected by the change in income is an increasing function of the change in income. Therefore, the value of the aggregate accelerator is zero when income is changing slowly, and large when income is changing rapidly and, as the relation between the change in income and the accelerator is assumed to be continuous, there will exist a change in income $Y_{t-1} - Y_{t-2} = \xi_2$ $[\xi_0 < \xi_2 < \xi_1]$ which will yield an aggregate accelerator coefficient β_t equal to any number between $\beta_t = 0$ and β_t large. The large β_t is assumed to place the economy in State D, monotonic explosive.

These relations between the accelerator coefficient's value and the change in income can be graphed. Figure 2.7 assumes that there exists a meaningful accelerator coefficient in the downswing. It will be shown that if $Y_{t-1} - Y_{t-2}$ is in the region $\xi_0 - \xi_1$, then the β coefficient will change in

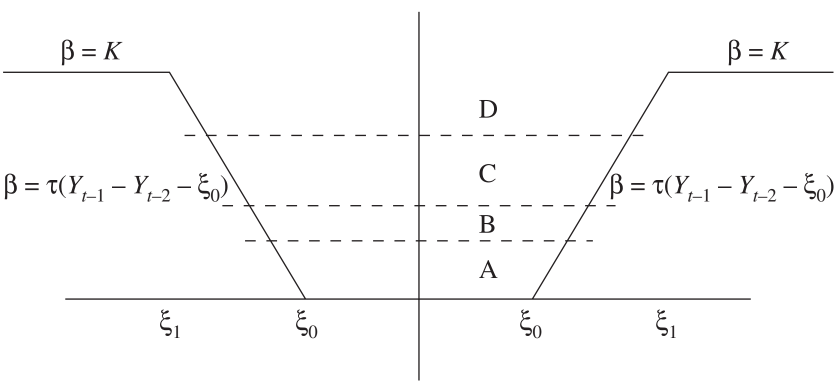


Figure 2.7

value so that β_t will reach its stable states A or D, implying a stable income movement.

Let us consider an accelerator multiplier model of the form $Y_t = \alpha Y_{t-1} + \beta_t(Y_{t-1} - Y_{t-2})$. Let us assume α is a constant. We have that if $Y_t - Y_{t-1} > \xi_1$, so that β_t is large, the economy will be in a monotonic explosive state. Therefore, $Y_t - Y_{t-1} > Y_{t-1} - Y_{t-2} > \xi_1$, and β_{t+1} will also be large. By recursion we can see that the economy will remain in the monotonic explosive state.

If $Y_{t-1} - Y_{t-2} < \xi_0$, so that no – or only a few – firms will be induced to invest, then ($\beta_t = 0$) the economy is in a monotonic damped state. Therefore $Y_t = \alpha Y_{t-1}$ which means that $Y_t < Y_{t-1}$ so that $|Y_t - Y_{t-1}| < |Y_{t-1} - Y_{t-2}| < \xi_0$. Therefore $\beta_{t+1} = 0$. By recursion, we have that the economy will remain in the monotonic damped state.

Therefore we have that once the value of β_t is such that the economy is in the monotonic-explosive or in the monotonic-damped state, the economy will, unless 'disturbed' from outside, tend to remain in that state. We can call States A and D of the Samuelson State Diagram stable.

If, for a given value of α, the value of β is such that the economy would be in States B or C, the economy will not remain in that state. States B and C are not stable. In order to show this we assume that β_t is a monotonic function of $Y_{t-1} - Y_{t-2} - \xi_0$. For example β_t is a function of the difference between the change in income and the accelerator threshold change in income. We therefore have

$$\beta_t = \tau(Y_{t-1} - Y_{t-2} - \xi_0)$$

Substituting in the difference equation

$$Y_t = \alpha Y_{t-1} + \beta_t(Y_{t-1} - Y_{t-2})$$

we get

$$Y_t = \alpha Y_{t-1} + \tau(Y_{t-1} - Y_{t-2} - \xi_0)(Y_{t-1} - Y_{t-2})$$

If Y_t is to equal Y_{t-1} so that $I_t = S_{t-1}$ then

$$\tau(Y_{t-1} - Y_{t-2} - \xi_0)(Y_{t-1} - Y_{t-2}) = (1 - \alpha) Y_{t-1}$$

If $Y_t = Y_{t-1}$ then

$$\beta_{t+1} = \tau(Y_{t-1} - Y_{t-1} - \xi_0) = 0$$

So we have that

$$\tau(Y_{t-1} - Y_{t-2} - \xi_0)(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$$

is a necessary condition for $\beta_{t+1} > 0$, for then $Y_t > Y_{t-1}$. However as $\beta_{t+1} = \tau(Y_t - Y_{t-1} - \xi_0)$ so that if $Y_t - Y_{t-1} < Y_{t-1} - Y_{t-2}$ then $\beta_{t+1} < \beta_t$ and $Y_{t+1} - Y_t < Y_t - Y_{t-1}$ which through recursion will result in $(Y_{t+n} - Y_{t+n-1} - \xi_0) \leq 0$ so that $\beta_n = 0$.

If $\tau(Y_{t-1} - Y_{t-2} - \xi_0)(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$ so that $Y_t > Y_{t-1}$ but $Y_t - Y_{t-1} = Y_{t-1} - Y_{t-2}$ then $\beta_{t+1} = \tau(Y_t - Y_{t-1} - \xi_0)$ will be equal to β_t. Then

$$Y_{t+1} = \alpha Y_t + \beta_{t+1}(Y_t - Y_{t-1})$$
$$Y_t = \alpha Y_{t-1} + \beta_t(Y_{t-1} - Y_{t-2})$$

and

$$Y_{t+1} - Y_t = \alpha(Y_t - Y_{t-1}) < Y_t - Y_{t-1}.$$

Therefore $\beta_{t+2} < \beta_{t+1}$ and through recursion will result in $\beta_n = 0$. If $\tau(Y_{t-1} - Y_{t-2} - \xi_0)(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$ so that $Y_t > Y_{t-1}$ and if it is sufficiently greater for $Y_t - Y_{t-1} > Y_{t-1} - Y_{t-2}$ then $\beta_{t+1} = \tau(Y_t - Y_{t-1} - \xi_0)$ will be greater than β_t. Then $Y_{t+1} - Y_t > Y_t - Y_{t-1}$ which by recursion will continue until $Y_{t+n} - Y_{t+n-1} > \xi_1$ so that β_{t+n+1} is large, with the resultant stable monotonic explosive state.

The operations of this model can be illustrated by numerical examples assuming that $\tau(Y_{t-1} - Y_{t-2} - \xi_0)$ is a linear function.

Table 2.1 Initial Conditions: $Y_{t-2} = 5$; $Y_{t-1} = 10$; $\tau = 1/2$; $\xi_0 = 2$; $\alpha = 0.8$

Time	Income	Consumption	Accelerator Coefficient $1/2(Y_{t-1} - Y_{t-2} - 2)$	Investment
$t-2$	5	–	–	–
$t-1$	10	–	–	–
t	15.5	8	1.5	7.5
$t+1$	22.0	12.4	1.75	9.6
$t+2$	32.2	17.6	2.25	14.6
$t+3$	66.6	25.8	4.1	40.8

Note: An accelerator of 4 is assumed to be large.

Table 2.2 Initial Conditions: $Y_{t-2} = 5.2$; $Y_{t-1} = 10$; $\tau = 1/2$; $\xi_0 = 2$; $\alpha = 0.8$

Time	Income	Consumption	Accelerator Coefficient $1/2(Y_{t-1} - Y_{t-2} - 2)$	Investment
$t-2$	5.2	–	–	–
$t-1$	10	–	–	–
t	14.7	8.0	1.4	6.7
$t+1$	18.1	11.8	1.35	6.3
$t+2$	17.9	14.5	0.7	2.4
$t+3$	14.3	14.3	−0.0	0

Numerical example 1 begins with initial conditions $\alpha = 0.8$, $\beta = 1.5$ which puts it in Region C – explosive cyclical. By the third time period it had moved into Region D where, by the nature of the accelerator generating relation, it will remain. Numerical example 2 begins with initial conditions $\alpha = 0.8$, $\beta = 1.4$ which also puts it in Region C – explosive cyclical. However, this does not generate a large enough change in income (in numerical example 1 $(Y_t - Y_{t-1} = 5.5) > (Y_{t-1} - Y_{t-2} = 5.0)$, whereas in numerical example 2 $(Y_t - Y_{t-1} = 4.7) < (Y_{t-1} - Y_{t-2} = 4.8)$) so that the movement of the economy is toward the stable State A – monotonic-damped.

This model is in some ways the inverse of the models which Goodwin constructed. He constructed non-linear models in which the stable state was a steady cycle. That non-linear models can result in stable states which are non-cyclical, damped or explosive has been shown. The economic interpretation of the model is also, I believe, meaningful.

Consider an economy in which the rate of increase of income is small relative to its existing level of income. In such an economy firms can, without much straining of their productive capacity, satisfy market demand. Whatever increases in capacity occur will take place as a result of the improvement in capital equipment that takes place in a technically progressive economy whenever replacement takes place. Net induced investment is small or zero. If the economy is displaced from this state, by, for example, government deficit expenditure, and the displacement is not large, induced investment will be small, the accelerator will operate only fitfully and uncertainly and the economy will 'stagnate'.

If this same economy is displaced from a low income state again by deficit financing, but the displacement is large, induced investment will be large and the accelerator will operate strongly and the economy will begin to explode. If we remove all 'financial' constraints and consider our income to be money income, so that the change in income can be maintained by a change in the price level, then the 'explosive' state can continue indefinitely as a 'full-employment-inflation economy'. If the full-employment-inflation state is broken, then the economy sinks into the low accelerator stagnant state. Such an economy will therefore be mainly in these two states: either in a state of secular stagnation or a full-employment-inflation state. Full employment at a constant price level is unstable if the rise in real income that occurs is such as to lead to a fall in the accelerator coefficient which will return the economy to its stagnant state.

This model can be used to interpret cyclical behavior in the United States since the end of World War I. The 1920s can be characterized by a 'high accelerator-high employment' explosive stable state. The financial crisis can be interpreted as the 'shock' operating through the financial positions of investors which displaced the economy toward its 'stagnant state'. The 1930s, the period of secular stagnation, can be characterized as a low accelerator-low employment monotonic damped state. World War II can be interpreted as the 'shock' which displaced the economy to its 'explosive' state. The last six years can be characterized as a high accelerator-high employment-explosive state. In contrasting the 1920s with the post World War II period, we would have to say that the 'explosion' in the 1920s was at a rate consistent with a fairly stable price level whereas the explosion of the past six years has been at a rate inconsistent with a stable price level.

This non-cyclical business cycle model is in some ways similar to the Hicks' Trade Cycle Model. However, it is different to the extent that there is no endogenous element which leads to a break in the cumulative process or in the stagnant state. Also it differs from Hicks' model (and the Goodwin article in part) in that the function which generates the non-linearity is an accelerator coefficient generating function which can be explicitly related to the behavior of business firms.

The formal model which we considered was composed of two parts:

1. the accelerator generating relation: $\beta = \tau(Y_{t-1} - Y_{t-2} - \xi_0)$
2. the linear accelerator-multiplier model: $Y_t = \alpha Y_{t-1} + \beta_t(Y_{t-1} - Y_{t-2})$.

We now wish to consider the effect upon the value of the accelerator coefficient β_t of the relation between Y_{t-1} and $\overset{*}{Y}$, the previous peak income, which we discussed earlier. We had written that if $Y_{t-1} < \overset{*}{Y}$ the accelerator coefficient would be lower than if $Y_{t-1} = \overset{*}{Y}$, and the accelerator coefficient would be at its maximum if $Y_{t-1} > \overset{*}{Y}$ and $Y_{t-2} = \overset{*}{Y}$. We could, of course, construct a model based solely upon this relation. If the high accelerator coefficient (when $Y_{t-1} > \overset{*}{Y}$ and $Y_{t-2} = \overset{*}{Y}$) is sufficiently great to place the economy in Region D, and if the low accelerator when ($Y_{t-1} < \overset{*}{Y}$) places the economy in Region A or even B, the economy would either stagnate or explode. (Region C is unstable for, in the upward part of the explosive cycle, either $Y_{t+1} > \overset{*}{Y}$ and $Y_{t+2} > \overset{*}{Y}$ resulting in a value of the accelerator coefficient sufficiently large to continue the upward explosion or we get that the explosive cycle turns down when $Y_{t-2} < \overset{*}{Y}$ making the accelerator, as a result of the downturn, sufficiently low to place the economy in State A or B.) This model, which is similar in its behavior to the previous model, is essentially a trivial model, although it is consistent with the existence of 'strong booms' requiring breaking.

However we can combine the two factors which we have distinguished as determinants of β and write the β generating function

$$\beta_t = \varphi(Y_{t-1} - Y_{t-2}, \xi_0, \overset{*}{Y}).$$

We have distinguished four relations between Y_{t-1}, Y_{t-2} and $\overset{*}{Y}$, which we used to determine our β coefficients. We can define μ as 'the coefficient of activation of the accelerator coefficient' which is dependent upon the relation between Y_{t-1}, Y_{t-2} and $\overset{*}{Y}$, specifically:

$$\begin{aligned}
&\text{If } Y_{t-1} - Y_{t-2} > 0 \text{ and } Y_{t-1} < \overset{*}{Y} \quad \text{then } \mu = \mu_0\\
&\text{If } Y_{t-1} - Y_{t-2} > 0 \text{ and } Y_{t-1} = \overset{*}{Y} \quad \text{then } \mu = \mu_1\\
&\text{If } Y_{t-1} - Y_{t-2} > 0 \text{ and } Y_{t-1} > \overset{*}{Y} \text{ but } Y_{t-2} < \overset{*}{Y} \quad \text{then } \mu = \mu_2\\
&\text{If } Y_{t-1} - Y_{t-2} > 0 \text{ and } Y_{t-1} > \overset{*}{Y} \text{ but } Y_{t-2} = \overset{*}{Y} \quad \text{then } \mu = \mu_3
\end{aligned}$$

and $\mu_0 < \mu_1 < \mu_2 < \mu_3$. Then for a given change in income $Y_t - Y_{t-1}$, the value of the accelerator coefficient generated depends upon the coefficient of activation which is defined by the relation between Y_t, Y_{t-1} and $\overset{*}{Y}$. For our equation for β, $\beta_t = \tau(Y_{t-1} - Y_{t-2} - \xi_0)$, we would write $\beta_t = \mu\tau(Y_{t-1} - Y_{t-2} - \xi_0)$ where μ takes on one of the four values given by the set of equations.

In such a model, the efficiency of a large change in income in activating a large accelerator coefficient depends upon whether or not the large change in income resulted in an income equal to or greater than the previous peak income. A large change in income may be sufficient, even if it does not result in a level of income greater than, or equal to, the previous peak income, to generate an accelerator coefficient large enough to place the economy in an explosive state. However, this change in income would have to be larger than if the productive capacity of the economy were smaller. The existence of a previous peak income larger than present income, that is the existence of idle capacity, acts as an inhibiter or damper to the accelerator generating process and therefore would tend to increase the size of the shock necessary to shift the economy out of the stagnant state. This effect would tend to increase the run of low income.

For example, the numerical example 1, utilized earlier, entered the explosive state at $t+3$. The development of that model would have to be revised depending upon the relation between the income at $t-1$ and $t-2$ with previous peak income. If the income at $t-2$ had been the previous peak income then the model would have developed as it did, $\mu_3=1$. If the previous peak income had been greater than income at $t-2$, then the accelerator coefficient would have been lower than the 1.5 which is given in the table. If the previous peak income had been greater than Y_{t-1}, then the accelerator coefficient would have been still lower. This effect of the previous peak income upon the time path of the level of income therefore is to make the accelerator coefficient, for a given change of income, higher or lower depending upon the relation of present income to peak income. When the economy is in the 'stagnant state', State A, with a previous income Y larger than the present income, the change in income which is necessary to induce a large amount of investment is larger than when the economy is in a state where present income is the peak income.

When the present income is the peak income, all or almost all firms are operating at capacity. In such an economy, any increase in income will result in all or almost all firms believing that an expanded plant will be profitable. On the other hand if present income is below the peak income, all or almost all firms will have excess capacity. Any increase in income will result in only a 'fraction' of the firms believing that an 'expanded plant' will be profitable. A given arithmetical change in income will induce less investment with excess capacity than with full employment.

The effect of the existence of such a threshold to the operations of the accelerator is to introduce a strong discontinuity in the behavior of the economy. If the economy is in the stagnant State A, a rise in income which does not raise income above the previous peak income will not induce much investment, particularly if it takes place over a number of years so that the rate of change of income is small. However a 'rapid' rate of change of income, which carries present income above the peak past income and keeps it increasing for a number of time periods, can be efficient in inducing private investment. This may serve as an 'abstract' rationalization of the inefficiency of deficit financing during the 1930s and the efficiency of the World War II deficits in stimulating an explosion.

If these models can be identified with the American economy, the alternatives available to policy makers are 'secular stagnation' or 'continuing inflation'. The only hope of something like full employment with a stable price level is a slow enough rise in income so that the accelerator never becomes large. This implies that there exists a sufficient volume of non-accelerator investment either autonomous (due to changes in production functions) or honorary (due to a government deficit) to maintain such a rate of growth of income. If the rate of growth of income becomes larger than some critical value, an inflationary explosion takes place. If government deficits that have been maintaining stable growth stop, secular stagnation results.

4. STOCHASTIC MODELS

4.1 The Error Approach

So far we have taken up two varieties of accelerator-multiplier models, the linear and the non-linear variety. We have examined each type, and have shown that they each cannot stand alone as business cycle models. We have constructed a 'new' model which exhibits secular stagnation and explosion as alternative 'stable' states. We have isolated as the significant problem for study the way in which the behavior of firms generates the accelerator coefficient. We still have to take up the stochastic variety of accelerator and multiplier models.

Let us consider a Hicks type (induced investment is a function of the change in income) accelerator-multiplier model in which the accelerator coefficient $\beta < 1$. The time path of income will be damped. If we assume that the value of α is sufficiently small for the given β coefficient, the economy will be in State B: damped oscillatory. This state implies that the cycle will die out and therefore unless it is modified it is unsatisfactory for business cycle analysis. It can be rendered satisfactory by imposing upon the damped cycle an outside energy source. The Schumpeterian hypothesis can be considered as such an outside energy source,[32] as can the suggestion by Frisch[33] in regard to the use of erratic shocks. A model of the Frisch type was taken up recently by G.H. Fisher,[34] and we can begin by following his exposition.

Consider a consumption function and an investment function subject to random errors; that is $C_t = \alpha Y_{t-1} + \mu_t$ and $I_t = \beta(Y_{t-1} - Y_{t-2}) + \upsilon_t$. The μ_t and υ_t are random errors. We therefore have

$$Y_t = (\alpha + \beta)\, Y_{t-1} + \beta\, Y_{t-1} + \omega_t$$

where $\omega_t = \mu_t + \upsilon_t$. Fisher assumes that ω_t is normally distributed with a mean of zero and a variance σ^2_ω. Fisher also assumed that $\alpha = 0.7$ and $\beta = 0.5$ which places the economy in State B, the damped oscillatory state. In order to indicate how a model subject to random shocks would behave over time, Fisher computed Y_t for 100 periods. ω_t was estimated by using random selections from a simulated normal population of the ω_t. These ω_t

were the 'energy source' that maintained the cycle. The time series which resulted did not exhibit the highly damped characteristics that the non-stochastic model exhibits with the values of α and β which were assumed. The random shocks served as an energy source which counteracted the damping influence of the assumed values of α and β.

Such a result is inconsistent with Hicks' conclusion that 'the theory of damped fluctuations and erratic shocks prove unacceptable'.[35] Hicks' rejection of the erratic shock-damped oscillation hypothesis is based upon the conclusion that 'the correlation between corresponding terms of successive cycles is quite small'.[36] It is true that the correlation between these corresponding terms of successive cycles as determined by the accelerator-multiplier mechanism in the erratic shock cycle is small. In order to use this as the basis of the rejection of the erratic shocks model it is necessary to assume that the period of the cycle which results when such random shocks are added to the damped oscillations of the accelerator model is fundamentally the period of the cycle generated by the accelerator-multiplier mechanism.

The random shocks can be considered as elements of a sample drawn from a given universe. In such drawings of a sample you expect, with a probability distribution which is dependent upon the nature of the universe, to have runs of similar valued shocks of various lengths. Such runs tend to 'build up' the amplitude of a deviation from the equilibrium level of income. Such large deviations lead to the persistence of the oscillatory movement. The resulting time series would tend to have a small correlation between corresponding terms of successive cycles as determined by the accelerator-multiplier mechanism. Nevertheless the overall time series would exhibit a cyclical movement with varying amplitudes to the individual cycles. In many ways the hypothesis that there exist random stochastic elements in the mechanism which generates the business cycle seems worthy of serious consideration. The 'error' or shock approach adopted by Fisher seems to be an unmotivated theory of the cycle, for if the damping is large, the systematic mechanism does not explain enough. 'A quite moderate reduction in the investment coefficient leaves us with fluctuations which are mainly random – with fluctuations, that is, which remain unexplained.'[37] The value of the accelerator coefficient which Fisher uses ($\beta = 0.5$) does lead to a highly damped cycle, so that the cycle he exhibits can be said to be due, for the main part, to the random shocks.[38]

The approach to stochastic processes in economic life which Fisher used, and, as far as I could recognize, Hicks criticized, is the approach which assumes that a random shock is attached or added on to a systematic generating function. This approach in its modern dress can be imputed to Haavelmo.[39] The admirable work which has followed upon the publication of Haavelmo's essay is not in question; the validity of his approach to the problem he attacked is not being doubted. The question raised is whether, in effect, a giant has been harnessed and used in an ineffective manner. The ideology of Haavelmo's approach is best expressed by the following quotations:

> What we want are theories that, without involving us in direct logical contradiction, state that the observations will as a rule cluster in a limited subset of the set of all conceivable observations, while it is still consistent with the theory that an observation falls outside this subset now and then . . . [40],

and

> The question is not whether probabilities exist or not but whether – if we proceed as if they existed we are able to make statements about real phenomena that are correct for practical purposes.[41]

The approach embodied in these two quotations can be derived from two sources: (1) the residual variations in correlation analysis after the systematic effect of the 'variables' has been eliminated, and (2) errors of observation where the fallibility of humans and of the measuring instruments combine to yield observations which do not, in detail, conform to the real world values. The Haavelmo approach leads to the formulation of economic problems in the light of statistical testing techniques. This is an appropriate transformation of economic models where the problem is to apply such tests to economic data. However, it is not the appropriate approach to the construction of a 'stochastic model'.[42]

4.2 Stochastic Variables

As an alternative to the Haavelmo errors of observation and unexplained residuals approach to the use of stochastic variables in economics, we can contrast a truly stochastic formulation of an accelerator-multiplier model. Such a model postulates that the economy and its processes contain elements which are in their very nature random variables. This postulate will be embodied in statements which assert that the values of certain attributes of the elementary economic units, firms or households, after allowing for the constraints of market conditions, technological production or utility relations, and specified behavior principles, may still take on any of a set of values. These attributes will be characterized by a probability distribution. Therefore, in any model in which such an attribute enters as a parameter, the values of the elements determined by the model are not strictly determinate. What may be true is that the time process involved in the generation of these attributes may yield a strictly determinate asymptotic distribution of the values of the parameters and therefore of the variables. 'To characterize the economic process with the aid of a random process implies that certain parameters, for example, the output of a firm, its profitability during a given period, its investment decisions are regarded as variables that with given probabilities assume given values; i.e. they are considered random variables. The probability distribution of a random variable or of a combination of such random variables at a certain moment is determined by the past of the economic process.'[43]

In attempting to set up models for investment behavior economists have to rely upon expected values to achieve a meaningful statement. Expectation relations are inherently of the nature where for the different economic units different expectations can coexist; and the 'distribution of expectations' becomes an element in the aggregate investment relation. The investment relation – the amount of investment forthcoming during any period of time – is one that is not strictly determined by the observable and measurable variables of the economic system.

If we attempt to apply the random process approach as defined by Lundberg to an accelerator-multiplier model of income determination, we have, naturally, to assume that the 'αs' and the 'βs' are the random variables. For the observable and measurable determinants of income in these models are the previous period's income. Present income is not strictly determined by these historic variables, assuming that the α and β parameters of the income generating model are random variables.

It is easy to set up a model of the economy which yields an accelerator coefficient that naturally is of the nature of a probability distribution. Let us assume that each firm is an element in a Marshallian industry, that it is a unit in a set of firms producing a homogeneous product. The firms in the industry are different in a manner which is consistent with the doctrine of the representative firm: differences in their cost structure, production function, and in the nature (perhaps spread) of their reaction to changes. The economy consists of many such industries, and in each industry we assume that the behavior of the firms is determined by the industry parameters and not by the situation in other industries.

A change in income implies that the set of demand curves for the products of the particular industries shifts. However, firms are the investing units. What is needed for each industry is a transformation of the shift in the industry demand curve into a change in a parameter upon which the firms in an industry base their investment decisions.[44] The impact upon a firm of a shift in the industry demand curve depends upon the market structure of the industry. In a competitive industry a shift in demand affects firms by means of a change in the market price of the product. This change in price implies that at the old price a quantity different from the quantity actually taken would now be taken. Let us assume that the investment decision of firms is based upon the firms' estimate of the change in the quantity that the market will take at the price that ruled prior to the shift in demand. Each firm will estimate the quantity of the product which it would be profitable for it to produce by allowing its plant size to vary. The investment by a particular firm which is induced by a change in income will be the change in fixed capital necessary to alter its plant size plus whatever change in working capital is needed to produce the new optimum output. Such induced investment in a competitive industry may take place by means of a change in the number of firms in the industry rather than by means of an alteration in the size of the plants of existing firms.

We therefore have a particular firm investment relation of the form:

$$i_{\lambda\rho}(t) = \gamma(Q_{\lambda\rho}(t-1) - Q_{\lambda}(t-2))$$

where i = investment by a particular firm; λ = industry index; ρ = firm index; γ = coefficient of induced investment for a particular firm; $Q_{\lambda\rho}(t-1)$: estimate by the ρ firm of the quantity demanded at the price of $t-2$ during the period $t-1$; $Q_{\lambda}(t-2)$: quantity actually taken at the price of $t-2$ during the period $t-2$; $Q_{\lambda p}(t-1) - Q_{\lambda}(t-2)$: the firm's estimate of the industry demand curve shift.

The amount of investment that takes place in an industry will be given by:

$$i_{\lambda}(t) = \sum_{\rho} i_{\lambda\rho}(t) = \sum_{\rho} \gamma_{\lambda\rho}(Q_{\lambda\rho}(t-1) - Q_{\lambda}(t-2)).$$

Heroically assuming that estimates of $Q_{\lambda\rho}(t-1)$ by all firms in the λ industry are the same[45] we have

$$i_{\lambda}(t) = (Q_{\lambda}(t-1) - Q_{\lambda}(t-2))\sum_{\rho} \gamma_{\lambda\rho}.$$

The amount of investment induced in the economy is the sum of the investment of the different industries:

$$I_t = \sum_{\lambda} i_{\lambda}(t) = \sum_{\lambda}\left[(Q_{\lambda}(t-1) - Q_{\lambda}(t-2))\sum_{\rho} \gamma_{\lambda\rho}\right].$$

However, by the aggregate accelerator relation we also have that $I_t = \beta_t(Y_{t-1} - Y_{t-2})$. Therefore we have that

$$\beta_t(Y_{t-1} - Y_{t-2}) = \sum_{\lambda}\left[(Q_\lambda(t-1) - Q_\lambda(t-2))\sum_{\rho}\gamma_{\lambda\rho}\right]$$

$$\beta_t = \frac{\sum_{\lambda}\left[(Q_\lambda(t-1) - Q_\lambda(t-2))\sum_{\rho}\gamma_{\lambda\rho}\right]}{Y_{t-1} - Y_{t-2}}$$

If the set of shifts in industry demand curves which is implied by a change in income is determinate, and if the impact of these shifts in industry demand curves upon particular firms investment is determinate, then β_t, the coefficient of induced investment, will be determinate. For the aggregate coefficient of induced investment to be a constant we have to assume that each $\gamma_{\lambda\rho}$ is independent of the size and direction of the shift in its particular industry demand curve and that the shift in each demand curve $Q_\lambda(t-1) - Q_\lambda(t-2)$ is a fixed ratio to $Y_{t-1} - Y_{t-2}$. Then we would have that

$$\beta_t = \frac{\sum_{\lambda}\left[(Q_\lambda(t-1) - Q_\lambda(t-2)\sum_{\rho}\gamma_{\lambda\rho}\right]}{Y_{t-1} - Y_{t-2}}$$

where $\Sigma_\rho\gamma_{\lambda\rho} = \bar{\gamma}_\lambda$ a constant; $\bar{\gamma}_\lambda$ is the 'industry' coefficient of induced investment; $Q_\lambda(t-1) - Q_\lambda(t-2)/[Y_{t-1} - Y_{t-2}] = \xi_\lambda$ = constant: as $\rho_\lambda(t-2)$ is used to estimate $Q_\lambda(t-1)$; ξ_λ is equivalent to the marginal propensity to consume a particular good; $\beta_t = \Sigma_\lambda\xi_\lambda\bar{\gamma}_\lambda$ = constant.

If the set of shifts in industry demand curves which is associated with a change in income is determined by a process which can be considered as analogous to sampling, then at any time the shift in a particular industry's demand curve, which is the immediate cause of inducing investment, can be considered as a sample drawn from a universe. Alternatively, the amount of investment which a given shift in an industry demand curve will induce can be interpreted as depending upon the reactions of the affected firms, and the firms' reactions to particular stimuli may have a probability distribution.[46] In both circumstances the β_t coefficient for the economy is a random variable.

Combining the two, we have that the probability distribution of β_t depends upon: (a) the probability distribution of particular shifts in industry demand curves given a particular change in income $Y_{t-1} - Y_{t-2}$; (b) the probability that a particular set of firms ω from the set of all firms Ω being affected by a particular change in income $Y_{t-1} - Y_{t-2}$. If the above probability relations apply in the determination of β_t we would no longer expect a fixed relation to exist between a change in income and a change in investment. The main body of this thesis emphasizes the systematic relations between the value of the accelerator coefficient and the changes in income. In the main body we are interested in a 'generating relation' of particular firm investment as the result of change in income. We therefore derive, for particular industry structures, conditions under which the investment is not strictly determinate. As a result, the main body of this thesis is consistent with a formulation of the accelerator coefficient as a 'random variable'. However, the characteristics (mean, variance, and so on) of the distribution of random variables are determined by the systematic relation.

In order to contrast a true stochastic process with the error process which is Haavelmo's approach, we assume that β is a random variable whose value at any moment of time is drawn from a probability distribution which depends upon (a) the structure of demand curve shifts which result from a given change in income; (b) the set of firms for which the resultant demand curve shifts imply investment; (c) the relation between output and capital stock for each firm. As β is a function of a subset of firms drawn from the set of all firms, it is a true random variable.

If we assume that the structure of demand curve shifts which result from a given change in income is independent of the level of income or of the change in income (that the marginal propensity to consume particular goods is constant), and if we assume that the magnitude of the individual firm's accelerator coefficient is independent of the magnitude of the shift in the industry demand curves, then we have that the aggregate accelerator probability distribution is independent of the level or change in income. The probability distribution of β will be independent of the time path of income. Such a stochastic model of the accelerator process states that, given the value of α, the realized value of β will be in the interval which results in the economy being in States A, B, C or D a certain percentage of the time. Such a proposition is truly stochastic, as it is based upon a frequency distribution of the βs from which the observed values are drawn. For example, let us assume an accelerator-multiplier model of the type $Y_t = (\alpha + \beta) Y_{t-1} - \beta(Y_{t-2})$. In this model, Table 2.1 gives the range of values of β which, for given αs, place the economy in each state.

The probability of the economy being in any state depends upon the probability of β having the value appropriate to that state. For example, with $\alpha = 0.9$ the probability of the economy being in explosive cyclical depends upon the probability of β having a value of between 1 and 1.73. As we are using a Hicks type model, the probability of the economy being damped (States A or B) or being explosive (States B or C) depends solely upon the value of the β coefficient.

Table 2.3

Values of α	Values of β			
	States of the economy			
	A	B	C	D
0.9	0–0.47	0.47–1	1–1.73	1.73–
0.8	0–0.30	0.30–1	1–2.10	2.10–
0.7	0–0.20	0.20–1	1–2.40	2.40–
0.6	0–0.14	0.14–1	1–2.66	2.66–
0.5	0–0.08	0.08–1	1–2.92	2.92–

In an attempt to illustrate how such a purely random β would affect the operations of the accelerator and multiplier model, two test runs were made using the values of the constant λ, of the marginal propensity to consume α, and of Y_{t-1} and Y_{t-2} that Fisher used in his 'random variable' model. In the first run β was assumed to have a rectangular distribution, with the values of $\beta = 0$, $\beta = 0.25$, $\beta = 0.50$, $\beta = 0.75$, $\beta = 1.0$, $\beta = 1.5$, $\beta = 2.0$, $\beta = 3.0$, $\beta = 4.0$, all being equally probable. The resultant series exhibited a great amplitude of fluctuation in the first half of the series. Then because of a run of values of β coefficients which lead to a highly damped movement of income, the series exhibited a very damped cycle. Of course in such a series if $Y_{t-1} = Y_{t-2} = \lambda/[1 - \alpha]$, the cycle would die out. The damping of the series was so great that in the latter part of the sample the cycle wellnigh disappeared.

A second test of β as a random variable was made using a triangular frequency distribution of β which deviated from the rectangular distribution in that the extreme values of β had a lower probability of occurring than values of β in the neighborhood of 1. The frequency distribution from which the sample of β values were drawn was:

β	Relative Frequency
0	1
25	2
50	3
75	4
1.00	4
1.50	4
2.00	3
3.00	2
4.00	1

The time series which resulted does not exhibit the extreme fluctuations that the time series derived from a rectangular distribution of β exhibited. The 50 period time series also did not show the 'damping' of the cycle that the rectangular distribution exhibited. The reason is obvious: with the probability distribution in the second case, the changes of a 'run' of values of the β coefficient which leads to a highly explosive or a highly damped movement is much lower than in the rectangular distribution. As a result the extreme fluctuations and the extreme damping associated with the rectangular distribution do not occur.

We could continue to analyse the implications of β being a probability distribution independent of the level of income, or the path of income, by constructing additional frequency distributions of β, taking samples with replacements from these frequency distributions and observing, for specified values of α, the resultant time series. However, the assumptions that were made – that the structure of demand curve shifts which result from a given change in income is independent of the level of income or of the change in income and that the magnitude of the individual firm's accelerator coefficient is independent of the magnitude of the shift in the industry demand curve (in order to derive the probability distribution of β independently of the level or the change in income) – are strong. Let us weaken our assumptions by assuming that the expected value (mean) of the frequency distribution of β depends upon the change in income and the difference between last period's income and the previous peak income:

$$\overline{\beta}_t = \varphi(Y_{t-1}, Y_{t-2}, Y_{t-1} - \overset{*}{Y})$$

where $\overline{\beta}_t$ is the mean of the frequency distribution of β. If we adopt the conventions we used earlier, that $Y_{t-1} - Y$ determines a coefficient of activation of the accelerator coefficient whereas $Y_{t-1} - Y_{t-2}$ enters into the generating function of $\overline{\beta}_t = \tau(Y_{t-1} - Y_{t-2} - \xi_0)$, we derive a relation that determines the mean value of the accelerator coefficient in the same manner in which the value of the accelerator coefficient was determined earlier. Let us assume that the value of the variance of the frequency distribution of β is independent of the value of β_t, and that the mean and the variance are the only relevant moments of the frequency distribution of β_t. We can now write the income generating function as:

$$\begin{aligned} Y_t &= \lambda + \alpha(Y_{t-1}) + (\overline{\beta}_t + \mu\sigma_{\beta_t})(Y_{t-1} - Y_{t-2}) \\ &= \lambda + \alpha(Y_{t-1}) + \overline{\beta}_t(Y_{t-1} - Y_{t-2}) + \mu\sigma_{\beta_t}(Y_{t-1} - Y_{t-2}) \end{aligned}$$

where the actual value of the accelerator coefficient at time t is $\beta_t = \overline{\beta}_t + \mu\sigma_{\beta_t}$. Depending upon the relation between σ_{β_t} and $\overline{\beta}_t$, and upon the frequency distribution of $\overline{\beta}_t$, we have that we can assign a probability to $\overline{\beta}_t$ falling within any range, for example P[$\beta_t > \beta_o$ and $\beta_t \leq \beta_1$]. This can be written as:

$$P_\beta = \int_{\beta_0}^{\beta_1} \varphi_1(\overline{\beta}_t, \sigma_{\beta_t}) d\beta_t.$$

For each value of α we can define the values of β_0 and β_1 which put the economy in any of the four states, and the given values of Y_{t-1} and Y_{t-2} determine the mean value of β_t. Therefore we can determine the probability that the economy will be in State A, B, C or D. For example, the probability of State A, for $\alpha = 0.8$ is

$$P_1 = \int_0^3 \varphi_1(\overline{\beta}_t, \sigma_{\beta_t}) d\beta_t$$

where

$$\overline{\beta}_t = \varphi(Y_{t-1} - Y_{t-2}, Y_{t-1} - \overset{*}{Y}).$$

As the probability distribution of β_t is a function of the time path of income, we have that the probability of β_t being such as to place the economy in each of its four stages is a function of the path of income. If the variance of the probability distribution is small, we have that the probability of the economy remaining in State D, where the value of $\overline{\beta}_t$ is high, is greater than the probability of the economy remaining in State A, where the value of $\overline{\beta}_t$ is small.

If the variance of the probability distribution of β is large with respect to the mean, and the mean is small, and the distribution is symmetric, we have to interpret the meaning of a negative β coefficient. A negative β coefficient means that a rise in income results in disinvestment and a fall in income results in investment. Although inconsistent with the specialized accelerator coefficient of the mathematical models, in our accelerator generation relation a small expected value of the accelerator coefficient is generated by a small change in income. If for example the economy had been exploding upward, and a small value of the coefficient β_t is generated, then the value of the realized accelerator $Y_t - Y_{t-1}$ is small. This generates a small expected value of the accelerator. It is not inconsistent with what happens during business cycles for decreases in the rate of expansion to lead to disinvestment. For the disinvestment to be consistent with the use of an accelerator-multiplier model, the value of the accelerator coefficient will have to be negative. This can occur with a probability given by the frequency distribution of β_t under our stochastic assumption.

If the economy is in State A and is approaching the 'equilibrium' level of income from above [$Y_t > Y_{t-1}$] then a series of decreasing arithmetic falls in the level of income occurs. It is not inconsistent with what happens during business cycles for positive investment to occur during period $t + 1$ even though $Y_t < Y_{t-1}$. For such a 'reversal' to be consistent with the use of an accelerator-multiplier model, the value of the accelerator coefficient will have to be negative.

On the other hand, it is inconsistent with what happens during business cycles for a period in which there is a large change in income, either positive or negative, to be followed by reversal in aggregate investment: for example, if $Y_t > Y_{t-1}$ and $Y_t - Y_{t-1}$ is large, $I_{t+1} < 0$ is not observed. Therefore, for changes in income which generate large values of the accelerator coefficient, we would expect $P(\beta < 0)$ to be zero; that is $\int_{-\infty}^{0} \varphi(\overline{\beta}_t, \sigma_{\beta_t}) d\beta_t = 0$ if $\overline{\beta}_t$ is large; $\int_{-\infty}^{0} \varphi_1(\overline{\beta}_t, \sigma_{\beta_t}) d\beta_t > 0$ if $\overline{\beta}_t$ is small.

We have assumed that the effect of the rate of change of income is to shift the mean value of the frequency distribution of the accelerator coefficient, leaving the variance of the frequency distribution of the accelerator coefficient unchanged. We have also assumed that this variance of the frequency distribution of the accelerator coefficient is small in respect to the 'explosive' values of the accelerator coefficient and relatively large with respect to the damped values of the accelerator coefficient. We therefore have a model in which the probability that random variation will lead to a change in the direction of the movement of income is high when income is changing slowly, but the random process has a small probability of affecting the value of the accelerator coefficient sufficiently to change the state of the model when income is changing rapidly.

A succession of high values of the accelerator coefficient in relation to the expected value of the accelerator coefficient may, if the economy is in a damped state, lead to an explosive movement. A succession of small values of the accelerator coefficient in relation to the expected value of the accelerator coefficient may, by decreasing the rate of growth of income, lower the expected value of the accelerator coefficient through a number of time periods so that if the economy has been in an explosive state, it enters a damped state. This stochastic formulation of the accelerator generation process can be combined with the model of the accelerator generating relation which leads to either explosive or stagnant states as stable states. This inflation-stagnation model did not contain a satisfactory mechanism which would result in a change of the economy from State A to State D and vice versa. A combination of the random element and the systematic element makes it unnecessary to posit 'shocks' or 'crises' of the magnitude of the stock market crash of 1929 or of World War II in order to have the economy shift from one of its stable states to another of its stable states. Although not a determinate relation, the allowance for the variance of the accelerator coefficient is an endogenous economic phenomenon, for it is simply a statement to the effect that the investment reaction of a particular economic unit to a given economic change (a change in income) is, to some extent, indeterminate. As a result, overall economic behavior which is due to the coefficients of macroeconomic models, such as the accelerator coefficient, is, to some extent, indeterminate.

In some ways the stochastic element in macroeconomic relations can be likened to the inability of economists to predict 'fashions', and also the inability to predict the time rate at which an innovation will catch on. For example, introduction and dispersion of television in the United States during the post-war period can be imputed to the high level of consumer income. But the rate, the dating, and even its wide acceptance can be considered as 'random' processes. As an example of the problem of forecasting the acceptance of innovations, we can take the case of air conditioning. Room air conditioning units first won wide acceptance during the summer of 1952. Is it sufficient to assume high incomes during the period 1953–56, to be able to assert that by 1956 room air conditioning units will be as widespread as television was by 1952? The acceptance of an innovation by households is inherently a random process. The reaction of firms to the impact that a change in income has upon them also has an inherent random element. Therefore, the aggregate accelerator coefficient is not strictly determinate. However, the change in income, the level of income, and the other elements which systematically determine the value of the accelerator coefficient dominate the behavior of the economy.

Such a combination of systematic determining relations for the accelerator coefficient, and inherent random elements which imply that the systematic determining relation is not complete, seems to be a meaningful hypothesis to use in business cycle analysis. It is essentially a combination of non-linear and stochastic process assumptions. It has previously been assumed that the stochastic process was essentially an energy source which maintained an otherwise damped cycle. This model can assume that the stable states of the economy are two such distinct states as monotonic damped and monotonic explosive. The stochastic process serves to 'destabilize' these otherwise stable states. We also have that if the indeterminacy of the accelerator generating process is large, such destabilizing would occur relatively often as compared to a world in which the indeterminacy is relatively small.

Another conclusion which follows from the stochastic assumption, when combined with a non-linear theory, is that the non-linearity no longer has to be such that it will generate the business cycle. In Goodwin's non-linear model, his assumption of a ceiling rate of disinvestment and investment was shown to be equivalent to an assumption that the value of the accelerator coefficient falls when the change of income is great. This yields the type of cyclical behavior desired, but as is obvious unless the special assumptions as to the nature of the accelerator are rationalized by observations or by economic theory, this is a *post hoc ergo propter hoc* argument. That is, the model generates a time series which superficially resembles the time series of business cycle observations; therefore the model is a business cycle theory. We have shown that a wide variety of accelerator models can be constructed which yield time series consistent with the nature of the data. It follows that such superficial agreement with gross observations cannot be considered sufficient ground for acceptance of a model. A useful model of business cycle has to result in more than a time series which is similar to the time series of business cycles. A business cycle model to be useful must yield an apparatus that enables prediction and perhaps control of the business cycles.

A number of desirable results follow from the combination of the stochastic and non-linear assumptions. One is that the inability to measure the value of the accelerator coefficient is no longer an argument against using this relation in cycle theory. This is true of the non-linear assumption as Goodwin pointed out.

> Another advantage (of non-linear theory) lies in the possible treatment of the acceleration principle. Because statistical studies (for example, Tinbergen's 'Statistical Evidence on the Acceleration Principle') have shown that it does not correspond to the facts, many economists favor dropping it entirely. Yet this would be mistaken since it is merely the statement of a simple consequence of the one omnipresent, inconvertible dynamic fact in economics – the necessity to have both stocks and flows of goods. In any case, it is worthwhile to try assumptions which take account of this fact but do not require any rigid proportionality. In doing this we may avoid another shortcoming of linear theory – the requirement that the upswing be essentially the same thing as the downswing . . .[47]

The 'nature' of the upswing and the 'nature' of the downswing may differ more markedly if stochastic processes are allowed as elements in the cyclical process than if the non-linearity is the sole cycle generator.

Another result of the combination of stochastic and non-linear assumptions as to the nature of the accelerator is that the smoothness of the time series which is generated by the model is not greater than the smoothness of the time series of experience; and also instead of having one cycle essentially the same as another cycle, the impact of the random element may lead to successive cycles which differ in their duration and in their amplitude. Instead of the smooth periodic movements of income, the result is a movement of income in which the amplitude and duration of the cycles varies.

The deficiency of such cycle theory is that it can be made so complicated, with so many alternatives being consistent with the theory, that disproof is impossible. The key parameter, the value of the accelerator coefficient, can no longer be measured from time series data, for it is not a constant. The problem therefore is to restrict the assumptions so that the theory can be used for prediction purposes. This means that we need a specification of the model so that not anything that could conceivably occur is consistent with the theory. In order to do this the nature of the accelerator generating relation has to be estimated. What I intend to do is to examine (a) whether or not the theory of the firm throws some light upon the nature of the accelerator generating relation; and (b) what the effect of alternative assumptions about the behavior of financial institutions throws upon the nature of the accelerator generating process. Therefore, the essential problem of this thesis is to investigate the systematic portion of the accelerator generating process by using inherited material from economic theory.

5. CONCLUSION

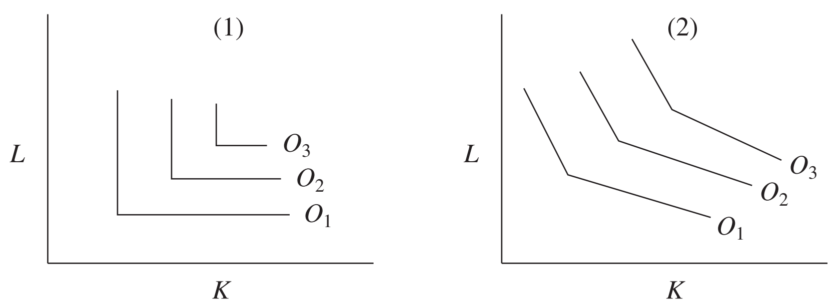
In this chapter a number of different accelerator and multiplier models of the business cycle were taken up. They were two varieties of the original linear model, the Goodwin non-linear model (which in many ways is equivalent to Hicks' model), an alternative non-linear model, a stochastic error model and a stochastic coefficient model. The conclusion, if any, is that a stochastic coefficient–non-linear accelerator model is consistent with the observed irregularity and non-symmetry of business cycle experience. Therefore, it seems to be a worthy hypothesis for further examination. But the assumption of non-linearity requires specifications, and an effort must be made to see whether the theory of the firm and a theory of financial or monetary behavior will lead to a selection of the nature of the non-linearity of the accelerator generating relation.

NOTES

1. See Harrod (1939, 1948); Domar (1947, 1946) and Pilvin (1952). The Harrod version of the growth model is phrased in terms of the familiar accelerator-multiplier interaction. The Domar version substitutes a capital coefficient (a ratio between capital stock and output, a fixed capital coefficient production function for output as a whole) for the accelerator coefficient. This makes the Domar accelerator a structural parameter

whereas the Harrod accelerator could be a coefficient of induced or realized investment as well as a structural parameter.

2. In this atomic age, the Harrod-Domar steady growth models may be likened to the atomic pile which would be used in an atomic power plant, whereas the explosive movement in the Hicks' business cycle model may be likened to an atomic bomb. The limitational factor which is ignored in both the Harrod and the Domar models is that if the rate of growth is to be steady, and if their production functions are of the fixed proportions type, a rate of growth of the labor force consistent with the rate of growth of capital equipment must take place. If the growth is to be the result of deepening rather than the widening of capital, their models need revision.
3. See Harrod (1936).
4. See Samuelson (1944). I am simplifying the model.
5. The border between the states is ignored for now.
6. See Frisch (1933).
7. See Goodwin (1946).
8. See Hicks (1950), appendix to chapter 6. I am changing symbols and ignoring his 'lagged' relations. For the formal Hicks' 'Accelerator-Multiplier' model, see pp. 184–93.
9. $f([\alpha + \beta]/2) = ([\alpha + \beta]/2)^2 - (\alpha + \beta)([\alpha + \beta]/2) + \beta > 0$; $(\alpha + \beta)^2 - 4\beta < 0$. This is Regions B and C in the state diagram. Reversing the inequality yields Regions A and D.
10. $f([\alpha + \beta]/2) = ([\alpha + \beta]/2)^2 - ([\alpha + \beta]/2)(\alpha + \beta) + \beta = 0$; $(\alpha + \beta)^2 - 4\beta = 0$. This is the equation of the line which separates Regions A and D from Regions B and C.
11. See Alexander (1949). Alexander discusses the Harrod-Domar case as a case of two different positive roots greater than one. His error lies in not recognizing that the border between the explosive oscillatory state and the monotonic state is a state of steady growth.
12. See Goodwin (1951).
13. As will be shown later, even though Goodwin's argument as to the accelerator's mechanism takes the form of a relation between actual and desired capital, in his third model he assumes that there exists a ceiling to the productive capacity of the investment goods industries, and the non-linearity occurs when this ceiling in investment goods productive capacity is reached, independently of whether or not 'actual' capital is equal to 'desired' capital. Therefore, even though the formal Goodwin relations are consistent with the hypothesis of this thesis, the interpretation of the relations has to be somewhat different.
14. See Goodwin (1951), p. 4.
15. See ibid., p. 4.
16. See ibid., p. 4.
17. See ibid., p. 4.
18. See ibid., p. 4.
19. If we draw the isoquants for a fixed proportion production function (graph 1), there is no possibility of producing O_2 by adding short run factors L to a fixed amount of capital K. However, a production function of the sort given in graph 2 would allow for both phenomena: $\xi = KY$ as the long run production function which uses fixed proportions



of the factors in spite of wide variations in their relative prices. Nevertheless, short run substitution in production is consistent with the production function. A specification of the production function as in graph 2 is necessary for Goodwin's model.

20. Goodwin allows a level of income to be produced with a capital equipment less than the amount appropriate or desired for that level of income. This must mean that this level of income is being produced by the use of more labor than will be used to produce the same level of income when capital has increased to the desired level. Therefore unemployment increases during the period in which investment is high. As the relative factor prices can be assumed constant in his model, it also follows that all investment is designed to reduce costs, not to expand productive capacity. That is, investment is desired to lower the cost at which a fixed output is being produced.
21. See Goodwin (1951), p. 5.
22. See ibid., p. 5.
23. See ibid., p. 6.
24. See ibid., p. 6.
25. See ibid., p. 7.
26. See ibid., p. 7.
27. See ibid., p. 7.
28. See ibid., p. 9.
29. The operation of the model when $K = \beta$ places the economy in State C (cyclical explosive) is essentially the same as when $K = \beta$ places the economy in State D. The difference is that the economy may go through some 'free' cycles as Hicks calls them before having a change in income $Y_t - Y_{t-1}$ so great that the investment ceiling is reached.
30. See Tinbergen (1938), p. 166.
31. See Duesenberry (1949) and Modigliani (1949).
32. See Goodwin (1946). The Schumpeterian shock is systematic, and as such can be considered as a particular non-linearity: for example, that $\beta_t = \tau(Y_{t-1} - Y_{t-2}, \xi.) + V$ where V is large. Thus 'innovational investment' will push the economy rapidly forward after the lower turning point.
33. See Frisch (1933).
34. See Fisher (1952).
35. See Hicks (1950), p. 91.
36. See ibid., p. 195.
37. See ibid., p. 91.
38. See Fisher (1952). Fisher assumed that the variance of the random shock ($\sigma^2_{\omega_t}$) is 5 billion dollars. His stochastic income generating equation is $Y_t = 17 + 1.2Y_{t-1} - 0.5Y_{t-2} + \omega_t$. The equilibrium level of income is 57 billion dollars. A positive shock equal to or greater than one standard deviation will occur 16% of the time. In Fisher's simulated normal distribution they occur 19.5% of the time. If we assume that $Y_t = Y_{t-1} = Y_{t-2} = 57$ billion and that two successive shocks of $+5$ will occur we get $Y_{t+1} = 62$ billion, $Y_{t+2} = 68$ billion. In Fisher's simulated normal population the mean value of the shocks ≥ 5 is 7.7. If we use this mean value we get: $Y_{t+1} = 65$, $Y_{t+2} = 74$. If we look at the time series of Fisher (1952, pp. 532–33) and estimate by observations the value of income, we get that an income ≥ 68 billion occurred six times; ≥ 74 billion occurred three times. Using the values in Fisher's simulated normal distribution we would get an income ≥ 68 billion by means of two successive positive shocks of 5 billion or more four times out of 100. If we take into account the possibility of achieving an income of 68 or more by means of a succession of smaller positive shocks (if $Y_t = 57$, $Y_{t+1} = 65$, for $Y_{t-2} \geq 68$ it is sufficient that $\omega_{t+2} = 1.5$ which in Fisher's simulated population will occur .38% of the time), we conclude that the series which Fisher derived is essentially the result of the 'random shocks'. The effect of the accelerator-multiplier mechanism is to make each period's income a weighted average of the two previous incomes plus or minus the shock. We could derive essentially the same type of time series by writing $Y_t = 1/2\ Y_{t-1} + 1/2\ Y_{t-2} + \omega_t$. This Fisher model therefore is vulnerable to the contention of Hicks that it leads to a cycle which is essentially due to the random shocks.

39. See Haavelmo (1944).
40. See ibid., p. 40.
41. See ibid.
42. The tendency in such 'testing' stochastic models to assume that the random term is distributed normally, with a zero mean, is carrying the assumption made in errors of measurement analysis into their theory, and this is consistent with the origin of this approach. Certainly in economic analysis we should expect to find that the probability distribution of the random variables which are generated by economic processes may deviate from the 'normal' probability distribution. We should expect the processes of economic life would often lead to 'equally likely' alternatives (a rectangular distribution) and to distributions in which 'almost all' events would have the same value: 'Poisson' type distributions.
43. Paraphrase of a statement of Lundberg, (1940, p. 3).
44. This is essentially the problem taken up in Chapters 4 through 8.
45. That is, every firm in the λ industry has the same estimate of the elasticity of demand for the product.
46. If an industry consists of a large number of firms, and if the probability distribution of reactions by firms is the same, then aggregate investment will be a summation of the reactions of a large number of firms. By 'laws of large numbers' the summation will tend to be stable. If an industry consists of a small number of firms, the summation of random variables which is the aggregate investment relation will tend to be unstable. This can be interpreted as implying that for competitive industries the stochastic element in the determination of β is relatively unimportant, whereas for oligopolistic industries the stochastic element will tend to be more significant.
47. See Goodwin (1951), pp. 3, 4.

3. The generation of the accelerator coefficient

Even though the accelerator coefficient has been used in a considerable number of business cycle models, there has been relatively little in the literature on the mechanism by which its value is determined. In this chapter I wish to do the following: (a) examine statements by a number of authors as to the mechanism by which the value of the accelerator is determined, and (b) set out the relevant factors which enter into the determination of the value of the accelerator coefficient. These relevant factors will form the basis for the theoretical analysis of the behavior of business firms which follows. The examination of the statements of the mechanism which determines the value of the accelerator will not be comprehensive; rather it will be limited to the remarks of a few authors.

As was mentioned earlier, at least three different concepts as to the nature of the accelerator coefficient can be identified. These are:

1. The accelerator as a structural parameter.
2. The accelerator as a coefficient of induced investment.
3. The accelerator as a coefficient of realized investment.

Prior to taking up the rationalizations for the use of the accelerator which have been put forward, it is necessary to clarify these differing concepts.

If the aggregate accelerator coefficient is interpreted to be a structural parameter, then it is derived from the production function. Typically in such models the production function states that output is a constant times the capital stock, for example $Y = \upsilon K$, where υ is the capital coefficient. Hence $dy/dt = \upsilon\,[dk/dt]$; $I = dk/dt = 1/\upsilon\,[dy/dt]$ and therefore, in this case, β (the accelerator coefficient), is the reciprocal of υ, the capital coefficient. For this form of the accelerator coefficient to be accepted there can be no possibility of substitution among factors in production. In addition either the production function for every sector of the economy must have the same capital coefficient or the distribution of a change in income among the various sectors (the income elasticities of demand for the various outputs) must always be the same. If the distribution of a change in income among different outputs is always the same, then the aggregate accelerator coefficient is a weighted sum of the particular sectoral accelerator coefficients with constant weights.

> Perhaps we should do well to get into the habit of thinking not of one accelerator coefficient, relating total output to total investment, but rather – as a more tenable approximation – of a series of such coefficients operating in each of the major industries. The task of summation will then present itself more clearly to our minds as a further complication of some importance which must be dealt with before the behavior of total investment and total output can be adequately explained. Or to make the point in a different way, the aggregate accelerator is the weighted average of the accelerator coefficients in each sector of the economy: not only are these component coefficients liable to vary themselves, but their respective weights in the national average will not be even approximately constant over the cycle.[1]

This implies that even if sectoral production functions can be written in a form which yields the accelerator as a structural parameter, the aggregate inducement to invest can vary.

> 'Suppose that within each firm the accelerator principle is working smoothly: $I = \upsilon[do/dt]$. Even in this highly simplified case care must be exercised in moving from the firms investment to aggregate business investment, there are likely to be wide variations in the capital coefficients between different branches of production, and if the income elasticities of demand are also markedly different with the result that output expands more in some lines than in others, the aggregate capital coefficients which relate total output to total investment will be altered accordingly. That is to say, constancy in the accelerator coefficient within each firm does not necessarily imply constancy in the aggregate accelerator coefficient.'[2]

Wilson's comments show the difficulties involved in taking the aggregate accelerator coefficient to be a constant even if the assumption of constant structural coefficients in the various sectors is made. If we assume that the production function of the particular firms is such that their investment coefficient can vary, then the derivation of the accelerator coefficient becomes even more complicated. In order to derive the sectoral accelerator coefficients it is necessary to assume that investment is taking place on the basis of 'given' conditions. These conditions include the relative prices of the different factors of production. Assume that firm plans are made on the basis of such a set of given conditions. Then the effective capital coefficient in each sector is determined. This, together with the weights of each sector, derived from the income elasticities of demand, is sufficient to yield a quantitative inducement to invest. But the sectoral coefficients' values depend upon relative prices – that is, upon the existing equilibria in the relevant markets. All the inducement to invest represents the quantity of investment goods demanded under given conditions. This is the second concept of the accelerator coefficient: that of a demand for investment goods.

> To sharpen the issue let us suppose that for technical reasons υ is not merely the ideal relationship between capital and output, but the only possible one, so that no change in capital intensity can be made by an individual concern. Sales are rising and are expected to rise for some time, but the funds needed for expansion are everywhere becoming more difficult to obtain. The expansion will then be retarded, but it will be retarded to a greater extent in the more capitalistic than in the less capitalistic lines of production. Thus the composition of output will change in such a way that the ratio of total capital to total output tends to decline: the aggregate ratio may then fall as Professor Hayek said it would, and its fall may weaken the stimulus for continued widening.[3]

The terms under which the funds needed for expansion can be obtained are part of the supply conditions of capital. We could also include the repercussions of a limited productive capacity of capital goods industries upon the price of capital goods as a determinant of the effect of a given change in income upon realized investment. This clearly leads to a third concept of the relation between a change in income and investment, one in which the effects of the inducement to invest upon the other markets is taken into account. In this concept the resultant investment depends upon the effects of the induced investment upon the equilibria in the relevant markets. An accelerator coefficient defined in such a way is a coefficient of realized investment.

It is obvious that the time path of income depends upon realized investment: not in the sense that realized investment determines productive capacity, but in the sense that realized investment is a portion of income. Accelerator models in which the accelerator is a coefficient of realized investment are not vulnerable to the criticism that they are solely a demand theory of the business cycle. However, with such an accelerator the simplicity of the accelerator model vanishes. The most that the formal accelerator model can be in this case is a framework for analysis: the content of the model for purposes of prediction and control depends upon the determinants of the coefficient of realized investment.

Accelerator models based upon these three alternative formulations investigate different problems. In the first type, where the accelerator is a structural parameter, the implications of assuming a constant ratio between capital and output are explored. In the second type the parameters other than income and the change in income which affect investment by a particular firm are assumed to be constant. Induced investment (aggregate) can vary due to the effects of changes in the structure of demand and in the level and rate of change of income. Such a model investigates the effects upon national income of the allowed changes, and investigates the effects of allowing the accelerator to vary. Both the first and the second types of model assume that income depends upon aggregate demand. Slight modifications of both the first and the second types of model which depend upon exogenously determined ceilings and floors have been produced. The time series that such a modified accelerator generates is altered, even though the content of the accelerator has not been changed.

In the third version of the accelerator, the supply conditions of capital are taken into account. The accelerator coefficient represents the relation between a change in output and investment, allowing these parameters of the individual firm's investment functions which are affected by the repercussions of the inducement to invest to vary. The third type is the broadest accelerator concept and such an accelerator does not result in the simple models to which the accelerator concepts lead.

Typically, the accelerator is introduced into a model by means of a simple assertion that the accelerator exists. 'There is a well-established relation, vouched for by experience and the laws of arithmetic, between the demand for consumable goods and the demand for durable goods, the essence of which is that the absolute amount of the latter depends primarily on the rate of increase of the former.'[4] The experience to which Harrod referred boils down to the well-known greater relative amplitude of fluctuations in investment as compared with the fluctuations of consumption in the trade cycle. The laws of arithmetic are examples of stock and flow relations such as the textbook example of shoe and shoe machinery demand. The statistical foundations for the use of a fixed value accelerator coefficient are flimsy. Harrod's volume on the trade cycle contains no statement on the necessary conditions to be imposed upon the behavior of business firms for the relation between the change in income and investment which the accelerator implies to be valid.

An analysis of the conditions necessary for the accelerator to operate was set out by J. Tinbergen in 1938:

> In this simplest form the (acceleration) principle states that percentage changes in the production of consumers goods are equal to percentage changes in the stock of capital goods. As the latter is usually considerably larger than the annual production of capital goods, the corresponding percentage changes in the latter are much larger than the percentage change in the production of consumer goods. The principle has two aspects between which it is useful to distinguish:
> a) the correlation aspect: there must be correlation between new investment in durable capital goods and the rate of increase in consumer goods production;
> b) the regression aspect: the percentage fluctuations in consumer goods production are equal to the percentage fluctuation in the stock of capital goods.[5]

Tinbergen asserted that:

> In its more rigorous form, the acceleration principle can only be true if the following conditions are fulfilled:
> a) Very strong decreases in consumer goods production must not occur. If the principle were right, they would lead to a corresponding disinvestment, and this can only take place to the extent of replacement.
> b) There should be no abrupt change in technique leading to a sudden increase in the amount of capital goods necessary to the production of one unit of consumer goods.[6]

The limitations to the value of induced disinvestment pointed out by Tinbergen has been incorporated in most of the later models. The acceptance of such a ceiling to disinvestment tends to 'draw out' the length of the depression phase of the resulting business cycle which leads to difficulties for cycle analysis. The assumption of 'strict complementarily' among the factors of production has usually been made, so that the capital coefficient is constant which in turn implies a constant accelerator, for example the accelerator is a structural parameter. This implies an assumption that whatever technological changes occur are, on the average, neither capital saving nor capital consuming. 'In its more rigorous form, the principle is equivalent to saying that a constant part of productive capacity is idle and that enterprises never increase production of consumer goods before having increased correspondingly their capacity. In the case of the constant part being zero – i.e. full occupation of capacity at any moment – that is, at least for increases a necessity, in all other cases this policy would have to be followed deliberately, and there are hardly enough reasons to suppose this occurs in reality.'[7] The possibility of output increasing without inducing any investment has been accepted by some authors who use such accelerator models, and Chenery in *Econometrica* has analysed the operation of the accelerator coefficient under conditions where 'excess capacity' exists.[8]

Tinbergen modified the accelerator coefficient to allow for the above shortcomings, and in so doing he indicated a mechanism of firm behavior which had to operate.

> The acceleration principle may, however, be given a less rigorous form. Instead of equality of percentage changes in consumers goods production and capital goods stock there may be assumed to be only proportionality or even only a linear relationship . . . . Two reasons exist for giving the principle its less rigorous form:
> a) During a period of increasing production, not all firms and not all branches attain at the same moment, the point of full capacity. Suppose that for the individual firms the principle acts only – and of course in its rigorous form – when full capacity is reached, then to a given increase in total production for all firms a smaller percentage increase in total stock of capital goods may correspond.

> b) A second reason for the less rigorous form of the acceleration principle might be that even with idle capacity a firm would expand its plant proportionately to the rate of increase in consumer goods production, but not by an equal percentage. This means that there would not be an immediate necessity for investment but that the willingness to invest would depend chiefly on the rate of increase in consumer goods production.[9]

These two factors which Tinbergen identified: that not all firms would reach 'full capacity' at the same moment of time and that a firm may, with idle capacity, 'anticipate' full capacity production and therefore 'expand its plant proportionately to the rate of increase in consumer goods production' lead naturally to a variable accelerator coefficient. The accelerator coefficient's value would be a function of the level of income and the rate of change of the level of income. The recognition that with less than full capacity production induced investment depends upon 'anticipations' would lead to a stochastic formulation of the accelerator coefficient. This concept of the accelerator is as a coefficient of induced investment.

Tinbergen offers as a substitute for the accelerator investment relation that 'an explanation of investment fluctuations by profit fluctuations is more natural'.[10] The use of profits in the investment function, rather than the change in income, is not, in and of itself, inconsistent with an accelerator formulation. If we recognize that profits are a function of income, and perhaps the change in income, and have an investment-profits relation we get:

$$I = f(p)$$
$$P = \varphi(Y, \Delta Y)$$
$$I = f(\varphi(Y, \Delta Y)) = \psi(Y, \Delta Y)$$

which is the accelerator formulation of induced investment. However, profit is a variable that directly impinges upon a firm's behavior; as is the relation between its capacity output and its present output. These factors therefore can be considered as the immediate determinants of investment behavior. If the change in profits, the level of profits, and the relation between output and capacity for individual firms are functions of income then, within our framework, the income investment relation belongs in the core of the model and the relation between profits and capacity and firms' investment decisions belongs in the supplementary relations. As such, the profit-income and capacity-income relations are appropriate places in which the mechanism which generates the accelerator coefficient allows for a variable accelerator coefficient.

An article by Sho-Chieh Tsiang contains a critique of the use of a constant accelerator coefficient over the business cycle. In his paper Tsiang recognizes the need to relate the value of the accelerator coefficient to the behavior of particular firms and realizes that the behavior of firms depends upon the structure of the market within which they are working.[11]

Tsiang emphasizes the inability to measure the accelerator coefficient, the dependence of the value of the accelerator coefficient upon industrial structure and the relation between the value of the accelerator coefficient and the supply of 'finance' to a particular firm. He uses these elements to advocate the use of a profit-income relation in the analysis of investment. The elements which he presents, and their similarity to the doctrine to be put forward here, is recognized. However, the use to which they are put seems unduly restricted. At this point I wish to take up his analysis of two phenomena: the relation between the value of the accelerator coefficient and industrial structure and the relation between the value of accelerator coefficient and the supply conditions of finance.

Tsiang's analysis of the manner in which investment is induced in a competitive industry by means of a change in income is, in its essentials, the same as the one to be presented later in this volume.

> For the increase in aggregate effective demand that is registered in the minds of the individual entrepreneurs may add up to a sum quite different from the real increase in aggregate output. This is particularly obvious for a competitive industry. There an increase in aggregate demand would generally lead first to a rise in the price of the product, although the tendency to rise in price might be held in check somewhat by speculators. It is the rise in the market price of this product that would be registered in the mind of an individual entrepreneur; but he would have no idea of the magnitude of the increase in aggregate demand, let alone his share of the increase. If he assumes that the increase in demand, that is to say the price of his product, is permanent, the increase in output and the requisite optimum investment to produce it, which he would plan in response, would be that which would make his own long-run marginal cost curve equal to the new price. It is evident that there is no guarantee that the planned increase in the output of all the entrepreneurs concerned would add up exactly to the original increase in demand, had there been no increase in price. It is therefore quite unlikely that the induced investment would bear any rigid relation to the original increase in demand.[12]

The importance of the impact effect upon firms of a rise in aggregate demand, the rise in the market price of a competitively produced product, and the significance of the reaction of the entrepreneurs to this impact (including the importance of the entrepreneur's price expectations) is correctly emphasized. That, in time, the long run equilibrium of the industry is consistent with a long run normal profit situation for a representative firm in the industry is overlooked by Tsiang when he questions whether or not the change in output will 'add up exactly to the original increase in demand'. As is well known, the long run supply curve of the industry will result in an equilibrium output equal to, less than or greater than the output 'had there been no increase in price', depending upon the existence of external economies or diseconomies. If such long run equilibrium considerations are to be relevant to business cycle analysis, the problem of the relevance of long run adjustments to cyclical phenomena has to be analysed. The assumption that *ceteris paribus*, the development within the industry would be in the direction consistent with long run equilibrium must be modified in business cycle analysis on two grounds.

(a) Inherent in the nature of the business cycle is the statement that the relevant parameters for the firm's behavior do not remain the same over the cycle.
(b) The relation between the time periods of the business cycle and the time necessary for long run adjustments to work themselves out must be considered.

We therefore are left with the problem of how to actually use the theory of the firm in order to generate the accelerator coefficient, which Tsiang does not adequately handle.

A problem recognized by Tsiang is that the short run change in output is not the same as the long run change in output for a given shift in the demand curve for a product. Therefore the accelerator coefficient relating investment to a short run change in output would be different from the accelerator coefficient relating investment to a long run change in output. However, the short run change in output and the long run change in output can be the same if 'the demand curve is kinked because of the firm's asymmetric expectation with respect to other competitors' reaction to its own price changes, or if a firm simply adheres to a conventional price'.[13] Figure 3.1 represents the discontinuous oligopoly demand curve analysis as given by Tsiang. If the demand curve as visualized by the firm shifts to D_1 from D_0, the marginal revenue curve shifts from MR_0 to MR_1. The short run marginal cost curve (SRMC) and the long run marginal cost curve (LRMC) both pass through the kink in the marginal revenue curve. Output therefore increases from O_0 to O_1, and the induced investment which takes place is designed to produce a fixed output O_1 at a lower cost. Induced investment, therefore, is designed to reduce the cost for a particular output, rather than to expand output. The Tsiang critique of the use of output in the accelerator relation indicates that it may be preferable to use a change in expenditures upon the product, that is to use money income rather than real income in accelerator business cycle analysis. In fact such a use of money income rather than real income in the accelerator relation can make the cycle models which use the accelerator relation consistent with the

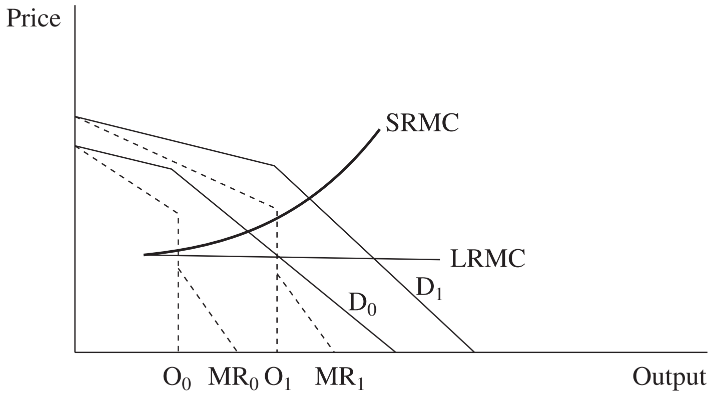


Figure 3.1 Discontinuous Oligopoly Demand Curve

changes in the price level and the changes in the money and financial relations that occur in the business cycle of experience.

On the relation between the accelerator coefficient and the financing of investment Tsiang again identifies some of the relevant variables. The supply of capital (finance) to a firm is limited because 'lenders will rely primarily upon the net value of the business of a going concern for their security, and will be unwilling to lend to any firm more than a given proportion of its equity capital'.[14] Therefore, 'as the firm's demand for capital approaches the limit set by the security provided by the firm's own capital and its net value is a going concern, the curves must rise steeply into a vertical wall, as Hawtrey puts it, regardless of the fact that the rates of interest prevailing in the market may be constant all the time'.[15] The conclusion that follows is that 'if a firm starts from a position where it has already reached the inelastic section of its supply curve of capital, which is not an unreasonable assumption to make for the upswing, further supply of investible capital funds will become available only through retained profits and savings of the investing class, plus the additional credit accommodation which it may secure at reasonable subjective and objective costs in its expanded equity basis; that is through the shift to the right of its previous supply curve of capital. This fact would certainly impose a great limitation upon the acceleration principle in the mechanical sense, which takes the constant accelerator and the changes in output as the sole determinants of the actual rate of induced investment. For if the demand of a firm for new investment generated by the increase in effective demand for its product in accordance with the accelerator, supposed to be constant is in excess of the supply of investible funds available to the firm, all the demand simply cannot become effective.'[16]

It is true that for a particular firm, the rate at which it can invest, during any period, is limited by its ability to finance investment. It is true that the ability to finance net investment for a firm is determined by retained earnings, increase of equity through security sales, the ability to borrow and whatever 'excess' liquidity it possesses. The assumption that at the upswing a firm has already reached the inelastic section of its supply of investible capital funds and that therefore all firms have done so is unwarranted. In an expanding industry, it is conceivable that the expansion of particular firms is discontinuous even though the expansion of the industry is continuous. Firms which expand to the limit of their financial resources at the beginning of an upswing will have repaid at least a part of their loans, and increased their equity base during the upswing. Firms which did not expand at the beginning of the upswing will earn high profits and will generally have a favorable balance sheet position during the upswing. Both of these classes of firms will have an adequate capital base to finance expansion toward the latter part of the upswing. The profit record that firms would make in an industry which has not expanded capacity during an upswing is such that these firms would have little difficulty in financing investment at 'market' rather than 'infinite' rates. The financing problems of firms during an expansion arises from three factors: (a) the difficulty of new firms, without an adequate profit record, to achieve a sufficient equity base; (b) the impact upon a firm of a general tightening of financing terms, due to the imperfect elasticity of the supply of credit; (c) that the rate of expansion of capital implied by the accelerator model may imply that firm's financing conditions must deteriorate.

The retention of profits by a firm during an upswing depends upon the use that the firm can make of such retained funds. The impact effect upon an industry of the accelerator mechanism works through the profitability of the firms in the industry. This means that in industries where there are investment opportunities in the accelerator sense, the profits of the firms will be good. The rate of growth of the firms which can be financed without any deterioration of credit conditions is equal to such retained earnings times a borrowing factor. If such retained earnings times a borrowing factor are insufficient to finance the requisite expansion, then two factors would still operate to enable the industry to expand at the accelerator rate: (a) the flow of new equity funds to the industry: the organization of new firms being one aspect of this; (b) the increase in the willingness to lend and the willingness to borrow due to the improvement in the profit position of the firms in the industry.

If we look at the economy as a whole, and recall the Robertsonian definition of savings which is the basis of the difference equations of the accelerator and multiplier model, we recognize that the volume of savings during an upswing is not sufficient to finance all of the induced investment. The amount of savings is equal to the difference between last periods's income and consumption, which is last period's investment. As income is assumed to be rising due to the accelerator effect, the present period's investment can be greater than the previous period's investment. Symbolically:

$$I_{t-1} = S_{t-1} = Y_{t-1} - C_{t-1} = Y_{t-1} - \alpha Y_{t-2}$$
$$I_t = \beta(Y_{t-1} - Y_{t-2})$$

We can assume that the present period's investment is to be financed by S_{t-1}. As $I_t > I_{t-1}$, by hypothesis, the financing of a quantity of investment equal to $I_t - S_{t-1}$, it requires a 'monetary' expansion. The rate of growth of real income has been of the order of magnitude of 3 per cent to 4 per cent a year. If the accelerator coefficient is a constant, we would have that I_t is approximately 3 per cent to 4 per cent greater than S_t. The problem as to whether the financing of an expansion can be accomplished without a deterioration of financing terms depends upon whether (a) the money supply is infinitely elastic and (b) the proportion of S_t that is available for equity financing is consistent with the existing balance sheet structures. Of course, to the extent that the price level changes, the amount to be financed by monetary expansion is changed.

For example, if balance sheet structures, on the average, involve an equal amount of equity and debt financing, and if I_{t+1} is 4 per cent greater than I_t, then 52 per cent of S_t must be available for equity financing either through retained earnings or through net equity investments on the part of households.[17] The danger involved in firms' retained earnings arises from the second purpose for which firms may retain earnings. Firms may retain earnings either to expand capacity or to improve their balance sheets by means of repaying debt. If at one stage of the business cycle a larger percentage of firms retains earnings in order to repay debt (such savings are assumed not to be available for equity financing), whereas if at another stage of the business cycle a large percentage of firms retains earnings in order to expand capacity, then during the stage of the cycle in which firms are repaying debt, the proportion of total savings available for increases of equity is low. This implies that a larger percentage of the accelerator induced investment (if income is rising) has to be financed by debt instruments. This increase in the proportion of the expansion of capital equipment that is financed by debt involves a deterioration of the credit terms available to the expanding firms. This may result in a decrease in the realized accelerator coefficient. The problem of the relation between the realized accelerator coefficient and the terms upon which investment can be financed, broached by Tsiang, will be analysed further later in this volume.

As has been emphasized, much of contemporary business cycle analysis takes the value of the accelerator coefficient as a given: a non-economic 'constant' which relates the volume of capital in a community to the volume of output.[18] The demand for capital, for example investment, is a function of the change in output in such models, and the cycle is generated by assuming that this quantity of capital goods demanded is forthcoming without any analysis of intervening market processes. In our analysis a change in total output (real or monetary) will affect the demand conditions for investment, but that the quantity of investment forthcoming, as the result of a particular change in output, depends upon the behavior of the pertinent intervening markets.[19] In order to be able to analyse the effects of changes in real or money income upon the level of investment, it is necessary to see how such changes in income may generate demand for investment goods; and how the effect of a change in income upon the volume of investment which takes place may vary with the nature of the affected intervening markets. For our hypothesis to be meaningful, this relation between the quantity of realized induced investment and the change in income must vary in a systematic manner with the level of income and the rate of change in income. Random variation in the accelerator coefficient due to the nature of the market processes by which changes in income affect the volume of investment may be an element in an economic model of the business cycle: but the analysis of such an inherently random process is a peripheral element in our central model. It is sufficient at this time to say that if the processes which generate the accelerator coefficient are inherently random with a sufficiently large variance, an interesting alternative model of the generation of business cycles of experience results. Models based upon such random processes have been considered earlier. We are primarily interested in the market processes which generate systematic variation in the accelerator coefficient. Nevertheless to the extent that the market phenomena do not completely determine the value of the accelerator coefficient, it is of the nature of a random variable. The argument of this work is that the systematic variation in the accelerator coefficient is sufficient to explain most of the significant elements of the business cycles.

In contrast to the mechanical pendulum business cycle models (including the non-linear theories) a theory of the business cycle which considers the generation of the values of the coefficients in such accelerator models as the meaningful economic problem must investigate the relevant market processes. 'We cannot expect any activity to respond regularly to business cycles unless it is subject to man's control within the periods occupied by cyclical phases, and unless this control is waged, consciously or not, by short period economic considerations.'[20] In order to transform a change in income into a change in investment it is necessary to investigate the effect of a change in income upon the variables that enter into a firm's investment decision function. Realized investment not only depends upon demand conditions, but also depends upon the supply conditions for the investment good: so the accelerator determination problem does evolve into an analysis of supply and demand conditions in a set of markets.

A number of steps are necessary for the accelerator to operate. The first step in the detailed analysis of the process by which the accelerator is generated is that the rise in income implies an upward shift in the demand curves for particular products.[21] This is inherent in the value of the marginal propensity to consume being greater than zero. We can use a concept of the marginal propensity to consume particular goods. The marginal propensities to consume particular goods add up to the marginal propensity to consume, and it states the change in expenditure upon the particular goods that will occur when income rises.[22] There is no need for the marginal propensity to consume a particular good being between zero and one in value. The marginal propensity to consume a particular good may be negative or it may be greater than one. The only constraints upon the marginal propensity to consume particular goods is that their sum must be equal to the aggregate marginal propensity to consume. As each rise in income must be transformed into a rise in particular product demand curves, the constant valued accelerator models have to assume that the effect upon demand curves of equal changes in income is independent of the level of income, or that the relation between change in consumer expenditures and the demand for investment goods is independent of the goods demanded for consumption. The first assumption is that the marginal propensities to consume particular goods are independent of the level of income. There is no inherent reason to believe that the marginal propensity to consume particular goods is stable; rather the experience of the post World War II period would indicate that the aggregate marginal propensity to consume is more stable than the particular goods propensity to consume.[23] A theory would have to cover the case where increments in consumption are distributed differently among the varying products in a market. In particular, for consumer durable goods, we can expect that with stable income the percentage of consumer income being spent on each such good will vary due to the stock and flow of services relations.

To a certain extent, the secular rise in real income in different economies may enable us to predict the distribution of increments of real income among different commodities for all but the 'most advanced' economy, by neglecting cultural differences: a truly heroic abstraction. However, if we assume that an economy has sufficient cultural independence that such arguments by analogy with other economies are invalid, we have no reason to expect that successive increments of real incomes will result in equivalent shifts of demand curves.

Any economy may be divided into a group of non-competing consumption classes. Duesenberry has shown that to posit the existence of such consumption classes is a meaningful hypothesis.[24] Among these non-competing consumption classes he has shown that the existence of a hierarchical pattern of consumption relations is a meaningful hypothesis. A step – logically prior to the distribution of income increments among products – is the distribution of the incremental income among these consumption classes. It is obvious that a rise in real income may lead to different distributions of increments of real income to the varying Duesenberry consumption classes and that the distributions of these increments of income within the consumption class may occur.

A rise in income implies a rise in consumption. This means that the upward shift in particular commodity demand curves outweighs the downward shift in the demand curves for 'inferior products'. The upward shift of particular demand curves depends upon the distribution of the income increment within the population and the manner in which each consumption class distributes its increment of income among the particular products. How an increment of consumption will be divided among the different commodities is to a great extent a random phenomenon. In particular there is no reason to assume that equal increments of consumption expenditures will lead to the same distribution of shifts among the different demand curves. Therefore at the stage in the accelerator generating process where incremental income is distributed as expenditures upon the different commodities, elements of variability and of randomness enter into the accelerator coefficient.

Another step in the accelerator generating process is the transformation of the rise in the demand curves for products into a change in the demand curves which confront particular firms. At this stage the distribution of industries among the different varieties of industrial organization becomes a significant element in the accelerator generating process. For the sake of conceptual simplicity we can differentiate two extreme[25] cases of industrial organization: competition and monopoly, and recognize that most firms are neither elements in a competitive industry nor monopolists. The analysis of the region between competition and monopoly, however labeled, is an unsatisfactory portion of economic theory. We will have to handle the effect of income changes upon the individual firm demand curve, and upon the individual firm's behavior in this region between competition and monopoly. A hypothesis at which our analysis will be directed is that the weight, in an economy, of these different market structures affects the aggregate accelerator coefficient. Hence, the business cycle experienced by an economy depends upon the structure of markets in the economy.

Two patterns of transformation of market demand curves into the demand curves confronting firms are well worked out in economic theory: the pattern of competition and that of monopoly. The competitive transformation is that the demand curve for the particular firm is a horizontal line at a price given by the intersection of the product demand curve and the sum of the marginal cost curves. The monopoly transformation is, of course, that the market demand curve is identical with the firm demand curve.

The region between competition and monopoly, whether it be labeled monopolistic or imperfect competition, or oligopoly, does not lead, by the definition of the market situation, to any such clear cut transformation of the product demand curve into the demand curves for the particular firms. The economics of this type of market structure is uncertain, for the inherent instability of such not large groups leads inevitably to the formation of quasi-cartels. In such a situation, it is the rule by which the demand curve for the product is transformed into the demand curves for the particular firm which is determinate rather than the particular demand curves. These rules require a lengthy analysis. In this introduction to the problem we can select a few examples to indicate the nature of the problem that arises: leaving a more complete analysis for later.

The quasi-cartel, into which an industry is organized, may have as its rule a principle by which the market is shared. Price competition is ruled out unless the market shares are infringed upon: at such times the rule by necessity breaks down. This market share rule is not inconsistent with price leadership, although price leadership alone may be the rule of a quasi-cartel. These industries may be characterized by large or small numbers. As the group becomes larger, the solution approximates the large group solution of Chamberlin. However, the market share-quasi-cartel approach may also be utilized for the analysis of small groups if we assume that a market sharing rule is adopted by the firms in the industry.

The rule in a quasi-cartel may be cost plus pricing. In this case a rule relating unit cost of production of the product to its price is established, and the division of the market demand curve into firm demand curves depends upon non-price competition. All of these market organizations between competition and monopoly leave a certain vagueness to the individual demand curves. This vagueness is not the same as the indefiniteness of a particular firm's position within a competitive Marshallian industry. The existence of firms of different size and profitability in a Marshallian industry is a function of the efficiency of those factors of production which are uniquely associated with the particular firm. The range and distribution of size of firms in the Marshallian industry is associated with variations in the equity and entrepreneurial skill of firms which allows variation within the industry to exist. In quasi-cartels the vagueness of the particular demand curve is due to the inability of the market constraints to completely determine the demand curves confronting firms. The demand curves for the output of the firms in imperfectly competitive industries are derived in different ways, depending upon the industry rule.[26] These alternative ways of transforming an industry demand curve into a firm demand curve will lead to different changes in the demand curve confronting a firm for a given shift in the product demand curves.

So far in this chapter we have identified three steps which have to take place for a rise in income to induce investment. These steps are:

1. The rise in income occurs as a rise in the income among the various consumption classes. At this step, those factors which relate National Income to Disposable Income are of importance.
2. Each such consumption class in turn distributes the rise in its disposable income among the various 'products' in the economy, and between savings and consumption. As each product is identified with a demand curve, step 2 implies that in general, as a result of the rise in income, the product demand curves have shifted to the right.[27]
3. The rise in industry demand curves results in shifting the demand curves confronting particular firms upward.

The analysis has taken us as far as the impact of the rise in income upon the particular firms. We have to carry it through to the reaction of the firms to this impact, that is, we have to analyse what takes place within the firm. For the firm, we must have that its reaction to the change in its demand curve, which is the third step stated above, must be such that investment is induced, if the accelerator is to be a meaningful concept.[28] That is, the rise in the demand curve for the output of a firm must affect the firms so that they are willing to expand the capital equipment which they use; or the conditions of each firm within an industry must now be such that new firms – which implies a net increase in capital in the industry – will enter.

If we retreat, for the time being, from the recognition that all is not competition in the real world, and take the competitive case as our model, an upward shift in the demand curve for the product of a competitive industry results in the short run in an upward movement of the infinitely elastic demand curve which confronts each firm in the industry. If we assume that the cost of the factors which enter into variable costs are not affected by the upward shift in the demand curve, this upward shift increases the profitability of the industry. This profitability increase is presumably the trigger for investment to increase, at least in the pure model of a competitive industry. For the increase in profitability to imply investment we need the following: (1) that the price of capital goods and/or the price of non-capital factors of production do not rise sufficiently to absorb the increase in profitability; (2) that financing is available for an expansion of capital in the industry. The assumption of financing being available means that the money market supply conditions are not changed so as to erase the increase in the profitability of the industry.[29]

This last step, that the rise in income results in an increase in the profitability of the firms in an industry and that this rise in profitability (at least in a competitive industry) is the relevant change in the economic situation of the firms in an industry so that investment is induced, leads us to an investment decision relation derived from the accelerator analysis that formally is similar to the relation in a typical analysis of an investment decision independent of the accelerator relation. The typical analysis of an investment decision in a competitive situation, under the assumptions of the absence of risk and uncertainty, involves a comparison of the marginal internal rate of return on investment and a market rate of interest. The marginal internal rate of return on investment is that rate which equates a future stream of returns with the present cost of the investment good. The investment decision is positive when this marginal rate passes some threshold which is determined by a given to the firm, usually taken to be some function of the market rate of interest. If the rise in profitability of a firm, which is a measure of the marginal internal rate of return being earned by an existing firm in the industry, is to imply investment, it means that the investment decision is based upon some extrapolation of the present situation, for example, that in a market situation, where risk and uncertainty exists, the estimates of future returns are based upon present and immediate past experience.[30]

In the literature which has grown up around the general theory of Keynes, the major attention has been centered around the money-interest rate relation and the consumption-income relation – while the relation between the interest rate and investment has been little discussed. The acceptance by Keynes of the prevailing inherited notions in relation to investment, in particular the acceptance of the work of Irving Fisher, made this seem natural.[31] However, for the use of a Keynesian type consumption relation and money-interest rate relation in an analysis of the business cycle, we have to see to what extent the interest-investment relation is to be accepted. It is obvious that if the Keynesian interest-investment relation is to be accepted as the determinant of investment, the role of the accelerator is to 'shift' this function. The aggregate interest-investment relation is then the horizontal summation of the interest-investment relations, which are sensitive to change in the level of income, and of interest-investment relations, which are insensitive to changes in the level of income.

Investment which occurs independently of changes in income is called autonomous investment by Hicks in his volume on the trade cycle. An autonomous investment decision must differ from the induced investment decision by what changes in order to make the investment take place. However, if both the autonomous investment decision and the induced investment decision are based upon a comparison of the relation between a schedule of internal rate of interest and an external rate of interest, they would both as a quantity be sensitive to changes in the external rate of interest. Therefore, they must differ in the element that affects the internal rate of return.

In the induced investment case the change in the values which enter into the investment decision function is due to the rise in demand for the product. In the autonomous investment case it must be due to the changes in the conditions under which the output may be produced. Let us take as the difference between autonomous and induced investment whether the rise in the internal rate of return over cost is due to a rise in the demand situation or to a change in the supply of the product situation. This, however, can lead us into difficulties if the industry is one which normally operates under short run increasing costs (consistent with our, for the time being, competitive assumption). The immediate effect of the rise in demand for a product is that the costs of production (marginal) have increased. This means that the rise in demand has affected the supply price as well as the demand conditions. Any investment which results from such a situation must be considered as induced.

There is one further situation which must be considered. Assuming that production functions are such that substitution among factors is possible, changes in the relative cost of capital goods and of non-capital factors may induce investment. Situations in which capital goods are relatively cheaper than labor would be favorable for investment. If such situations are associated with changes in income, should such investment be included in the category of autonomous or induced? In such a condition we may even have to allow for negative accelerator coefficients. If the world is such that a fall in income results in a sharp fall in the price of capital goods and no fall in the price of labor, then an increase in the capital used per unit of output may be desirable for a sufficiently large segment of the economy so that the fall in income induces positive investment. This means that for such industries the demand curve for investment goods shifts upward when income falls – a negative accelerator coefficient. The possibility of such a phenomenon cannot be ruled out on any *a priori* grounds. It may be found that such relative price sensitive stabilizing accelerator behavior exists under some set of conditions. If so, of course, the immediate policy prescription follows – establish the conditions so that for a significant portion of the economy the accelerator is negative,[32] and the reaction of this segment of the economy will be stabilizing. In fact a negative accelerator or a zero accelerator coefficient could also be a stabilizing element at the peak of a cyclical movement.

The similarity between induced and autonomous investment is that they are both based upon an internal-external rate of return calculation. A definition of induced investment which we can use, is that all investment which occurs because the variables which enter the investment decision function are altered by market changes which are associated with the change in income is called induced. This means that not only the investment due to the change in profitability which arises from the upward shifts of particular demand curves is induced but all investment which is the result of changes in relative prices of factors of production – if the change in the relative prices of the factors of production is due to income changes – is induced. Autonomous investment is thereby reduced to innovational investment – and if the investment decision function of innovators is affected by phenomena associated with changes in the level of income, the autonomy of innovators' investment may be brought into question.

We are left with the result that the distribution between induced and autonomous investment may be meaningless. We reduced induced investment to a rate of return over cost schedule – which made it equivalent to autonomous investment. And then we showed how the rate of return over cost schedule for all investment may be affected by income changes – so we made all investment induced. At this stage we are ready to turn to a more intensive analysis of the problems raised by this chapter. In particular we will focus our attention upon two points: the relation between the value of the accelerator coefficient and the structure of markets, and the relation between the value of the accelerator coefficient and the behavior of financial markets. We will see whether these two relations generate a value of the accelerator coefficient which varies in a systematic manner over the business cycle.

NOTES

1. See Wilson (1953), p. 83.
2. See Wilson (1953). Wilson's symbols.
3. See ibid., p. 78.
4. See Harrod (1936).
5. See Tinbergen (1938), p. 165.

6. See Tinbergen (1938).
7. See ibid., p. 166.
8. See Chenery (1952).
9. See Tinbergen (1938), p. 167.
10. See Tinbergen (1938).
11. See Tsiang (1951).
12. See ibid., p. 328. Note that Tsiang identifies the increase in demand with a rise in price.
13. See ibid., p. 330.
14. See ibid., p. 332.
15. See ibid., p. 333.
16. See ibid., p. 334. The independence of Hicks' autonomous investment from Hicks' induced investment is also questioned by Mr Tsiang. 'Furthermore, once it is realized that the rate of supply of finance may be a limiting factor of the rate of investment, in so far as it is carried out by private firms, autonomous investment cannot be taking place at a constant rate of growth independent of induced investment. For these two types of investment would compete with each other for a limited supply of capital funds . . .'.
17. This assumes that all monetary expansion takes place by means of lending by banking institutions.
18. See Hicks (1950), pp. 39–40. 'Now, suppose that there is an increase in output (we need not inquire for what reason) and that the increase is expected to be permanent – the new level is to be maintained indefinitely. For the production of the new output the existing capital stock will no longer be appropriate. In the *short period*, the enlarged output can be produced, but only by using the capital equipment at more than its optimum intensity, and therefore (in all probability) increasing its effective rate of depreciation, this provides a motive for increasing the stock of equipment. Thus the increase in output induces investment; investment occurs as part of the process of moving from one equilibrium to another.'

    And see Tsiang (1951), p. 326. 'Most of these models (accelerator and multiplier business cycle models) have one characteristic in common in assuming that there is a constant accelerator, given exogenously by the prevailing technique of production, which determines the rate of investment in proportion to the rate of change in income or output, except when full capacity of production is reached, when redundant capacity prevails and when the change in income is negative.'
19. See Tsiang (1951), p. 327. 'Just as the old fashioned quantity theory of money is vitiated by the discovery that the velocity of circulation is not exogenous but is a pliable endogenous variable of the economic system, so may the theory of the business cycle, which is based on the mechanical interaction between a constant accelerator and a constant multiplier, be vitiated by the fact that the accelerator is probably not an exogenous constant but is rather a pliable endogenous variable.'
20. See Mitchell (1950), p. 95.
21. See Tinbergen (1938), p. 168. 'Dr. Staehle, in a private discussion, suggested that the correlation would exist between new investment activity and the *shift in the demand curve* for consumer goods instead of in the increase in the actual quantity demanded or produced. This seems indeed more natural especially in the following situation: suppose that with full capacity used, demand increases, but as productive capacity cannot be expanded immediately, the quantities produced cannot rise correspondingly. Prices will rise instead. Nevertheless it is natural to assume that there will be new investment. Dr. Staehle's hypothesis can be given an especially convenient form if the elasticity of demand is unity (which approximately, will be the case for "all consumption"): then the shift in the demand curve is simply equal to the change in the money value of consumption or consumption outlay. Therefore, his device would simply be equivalent to saying that new investment is correlated with the rate of increase in consumption outlay.' Note that it is 'natural' to assume that a shift in the demand curve will lead to new investment. Also to say that new investment is correlated with the shifting consumption outlays implies that it is irrelevant to the investment process which commodity

demand curves are shifted; that is the investment-output relation is independent of the product demanded.

22. If α = the marginal propensity to consume and there are η goods in the community we define α_i as the ratio of the increment of expenditures upon good i to the change in income. We therefore have that $\Sigma\alpha_i = \alpha$. If we write $I_t = \beta(C_t - C_{t-1}) = \beta(\Sigma\alpha_i Y_{t-1} - \Sigma\alpha_i Y_{t-2})$ for the induced investment relation, we can also break down investment into the η industries: $I_i(t) = \beta_i(\alpha_i Y_{t-1} - \alpha_i Y_{t-2})$ and have that $I_t = \Sigma I_i(t)$. This is essentially the process we are engaged in: the focusing of attention upon the investment induced in a particular industry.
23. See Klein (1946), p. 298. 'It is one thing to say that there is a stable peacetime relationship between total consumption and income, and something else again to say that there is a stable relationship between each of the several categories of expenditure and income.'
24. See Duesenberry (1949).
25. Competition and monopoly are extremes in terms of the number of firms that share a market. In terms of the stability of the firms' demand curve with respect to the behavior of other firms, and in terms of the applicability of unconstrained profit maximization by a firm with respect to its demand curve, competition and monopoly are more alike than either one of them is with those markets characterized by oligopoly or conditional monopoly.
26. The particular rule of behavior adopted by the firms in an oligopolistic industry depends upon the history of the industry. Given a generally similar market situation – competition among a few firms – it is necessary to have an industry rule in order to have stable (or predictable) demand curves confronting each firm. The various rules adopted by such industries are economically equivalent, but the particular pattern adopted by an industry depends upon the accidents of history. That is, given an oligopolistic market situation we can predict that there will exist a rule and we can indicate what the rule accomplishes, but we cannot from the known economic data predict the particular content of the rule.
27. No analysis of what happens to the elasticity of demand for particular products as the income shift in demand curve occurs is offered. The only element in step 2 that is relevant is the 'in general' upward shift of demand curves confronting particular firms.
28. See Tsiang (1951), p. 327. 'In a capitalistic economy, the increase in aggregate effective demand which is supposed to bring the accelerator into operation must first be registered in the minds of individual entrepreneurs and then affect their investment decisions through their expectations.'
29. See Tsiang (1951), p. 331. 'The greatest weakness of the acceleration principle, however, is its complete disregard of the supply of capital to individual firms as a determinant of the rate of their investment activities.'
30. The relation between financing costs and investment decisions under alternative behavior principles is taken up in Chapter 8.
31. The difference between Keynes' flow concept of investment, which is natural for income determination problems, and the classical stock concept of investment, which is natural for equilibrium analysis, is not significant for our purposes.
32. The automatic compensatory deficit financing suggestions are equivalent to saying that the accelerator coefficient for government fiscal policy is negative.

4. The theory of the firm in relation to business cycle theory

If non-linear accelerator type models are to be used in business cycle theory, it is necessary to understand the process by which a change in income induces investment and whether or not the effect of a change in income upon investment varies systematically over the business cycle. We will examine the hypothesis that the effect upon investment of a change in income depends upon the relation between investment decisions of individual firms and (1) the structure of the product markets in which the firm is operating and (2) the financing conditions which confront the firm. This leads us to a study of the investment behavior of business firms.

Market structures determine two relations for a firm: (1) the relation between the particular demand curve confronting the firm and the market demand curve and (2) the way in which a firm behaves toward its particular demand curve. Market structure is the manner in which the market for a product is organized. Organization is really an improper description of the concept as no formal organization of producers or consumers need exist. Aside from the generalization of cost curves to allow for financing conditions, we do not need anything more than the traditional analysis of competitive and monopolistic markets.[1] Oligopoly, the region between competition and monopoly, does cause us concern, and we will have to set up a working model of such markets.

Financing conditions include the effects of different balance sheet structures upon the behavior of firms and the behavior of the money market. The traditional theory of the firm does not explicitly deal with such problems. The terms under which a firm can obtain financing is a price. Cyclical variations in this price affect all firms in their decision as to the best scale of operation. Therefore, financing conditions are particularly relevant to any theory of firm investment behavior designed for business cycle analysis.

For the accelerator to operate, a change in national income must affect at least some of the business firms that make up the economy in such a manner that a different quantity of investment takes place than would have occurred without the change in national income. Therefore the change in national income must affect some of the variables which enter into the investment decision relations of firms.[2] A change in national income shifts the demand curve for the product of an industry. This is transformed into a change in the demand conditions confronting each firm. The market structure determines the manner in which a change in the product demand curve is transformed into a change in the demand curve confronting a firm, and, within limits, the firm's reaction to its demand curve. The relation between induced investment and market structure therefore depends upon the effect that market structure has upon the determination of the firm's demand curve and upon the firm's behavior principle.

That the difference in firm behavior due to variations in industrial structure is relevant to the problem of business cycle analysis has been conjectured by many. A variety of business cycle theory that emphasizes industrial structure is associated with the assertion that business cycle phenomena can be imputed to the perverse behavior of particular product markets, which, in turn, can be imputed to the existence of 'monopoly.'

The industrial structure variant of business cycle theory asserts that – 'there is a predisposing element in our present-day economic organization, – which is a necessary factor in contribution to fluctuations and that their structural element consists of price rigidities; that price rigidities themselves are not an active agent, but that their existence enables other factors of initiation or disturbance to produce fluctuations in the ratio of employed to employable resources, and that, moreover, in the absence of price rigidities substantial fluctuations in this ratio are inconceivable'.[3] It follows that – 'if the prices of commodities and the prices of factors were perfectly flexible, then whatever happened to the volume, or velocity of money, you might get cycles in price levels, you might get cycles in income measured in money, but you would not get cycles in the ratio of employed to unemployed resources'.[4] These price rigidities are at different stages in economic development, due to different causes. 'But the cause of price rigidity has undoubtedly undergone a change in the past century. The older price rigidities were probably attributable in large part to the force of custom and to sluggish market mechanism, factors which have since shrunk greatly in importance, whereas the modern price rigidities are obviously due in the main to other factors.'[5] The factors which made for price rigidity, in the present (1936) were, 'the prominent item in the moral code of business which looks upon price stability as an economic virtue and price cutting as the most heinous of economic sins';[6] and 'the general accounting practice of computing costs for all purposes on the average rather than the marginal or differential basis'.[7]

Probably more important 'is monopolistic control of prices',[8] including 'the tendency of competition to become imperfect as the size of the concerns in a particular industry grows and the number of concerns decreases, and of price rigidity to result even if the complete monopoly control of the industry fails to develop'.[9] 'Another factor of increasing importance contributing to price rigidities is to be found in the expanding areas of governmentally regulated prices . . .'.[10] and 'finally, in the important range of what might be called government prices – taxes, fees and other changes – there is, at the best, rigidity of prices and at the not infrequent worst, marked perverse flexibility of prices'.[11]

All of the causes of price rigidity listed above, except for 'government prices' are examples of the effect of market structure upon the behavior of firms. The price rigidities operate so as to transform any decrease in total (or particular product) demand into a decrease in the quantity taken of the product. With price flexibility the contention is that a decrease in total demand will result in movements down in the incremental cost curve. If factor prices are flexible, the downward shift of the incremental cost curve due to the fall in factor prices will tend to leave the quantity produced, and therefore employment, unchanged. This business cycle theory does not take into account the reciprocal effect of flexible prices of factors and commodities upon the volume of total demand. This phenomenon, due to the assumption of 'induced' expenditures (the consumption function), is the key to Keynesian and later business cycle analysis.

In contrast with the Keynesian and later business cycle theory, the price flexibility cycle theory of Viner did not specifically distinguish between consumption demand and investment demand. The lack of a separate analysis of the factors relevant to the determination of these broad categories of aggregate demand had led to its general abandonment. In addition, this theory relies upon autonomous changes in the quantity of money or in the velocity of money to initiate the cycle. The artificial separation between monetary and real phenomena is inconsistent with the hypothesis that the analysis of the determinants of investments is necessary for business cycle theory. As investment behavior is related to prices (the interest rates) which are determined, at least in part, in the financial or money markets, no theory of the business cycle which does not consider the interrelations between financial and real factors is consistent with the key hypothesis as to the significance of investment behavior. As a result of the emphasis upon investment, the later analysis of business cycles has not investigated the relation between market organization and business cycles. The particular virtue in this older price flexibility business cycle theory is its emphasis upon the relevance of market structure to the behavior of firms, and its assertion that the behavior of firms determines business cycle processes. We will emphasize the relation between market structure and the determinants of the investment behavior of firms rather than the part played by market structure as a determinant of the degree of price flexibility. The result is a synthesis of the emphasis of this older business cycle theory with the income-investment business cycle theory.

Evidence for the relevance of industrial structure to the problem of business cycles can be found in the work of the National Bureau of Economic Research. 'When producers lack effective short-run control over output, prices conform better than production to business cycles, and have higher reference cycle amplitudes. When producers possess such control, production conforms better than prices, and has higher reference cycle amplitudes.'[12] And again Mitchell sums up the evidence by asserting:

> This, the empirical generalization that prices rise when goods are turned out in greater abundance, and vice versa, should be confined to industries in which producers have effective short period control over prices . . . . Another qualification suggested by our evidence, though less emphatically, is that the rule applies better to competition than to administered prices. The prices of plate-glass, asphalt, passenger automobiles and iron ore do not rise on the average in the expansion covered by Mills' analysis.[13]

Before we go further, it is necessary to clarify the relation between a plant and a business firm. A plant can be identified with a collection of fixed productive factors. A firm can be identified as an entrepreneurial or decision making unit. The plant is a set of productive factors which cannot be depreciated or used up rapidly. The size of plant is determined by a firm on the basis of the following information (some of which may be conjectural):

(a) the production function;
(b) the relative prices of plant and short run factors;
(c) the demand curve for the output of the plant as conceived by the firm;
(d) the rule of behavior which the firm uses in determining the optimum plant size.

The elementary analysis which is based upon the production function and the prices of the factors of production is concerned with the determination of plant size.

The firm is an economic unit which makes the following decisions:

(a) the product to be produced;
(b) the rule of behavior to be used in determining the optimum size plant;
(c) the rule of behavior of a given plant with respect to a given demand curve;
(d) the financing technique to be used.

The attribute of a firm that makes it necessary to have both a firm and a plant in economic analysis is the problem posed by financing. The usual economic theory ignores financing problems and assumes a unique behavior principle for all firms (profit maximization), leaving only the trivial problem of the choice of the product to be produced to the firm.

We will treat the problem of the financing technique to be used by a firm as the problem of balance sheet structure. At any given time a firm can be identified with a fixed equity. As a result of the fixed equity of a firm, the relative prices of the factors of production to a firm are not constant but vary as the ratio of equity to debt financing changes.[14] The scale of plant for different firms in the same industry depends upon the production function and upon differences in the relative prices of the factors of production to firms due to the difference in their equity base.

The production function may be such that one and only one scale of plant is efficient. A scale of plant is efficient if the price compatible with normal or acceptable profits for this plant is not significantly higher than for any other size plant.[15] If one and only one scale of plant is efficient, the long run cost curve that results is the typical U-shaped cost curve. Firms which are too small, in terms of their financing ability, to acquire plants of the efficient size will in time be eliminated from the industry. Their productive equipment will either be modified and utilized in plants of efficient size or their productive equipment will be redundant. Firms which have large equities in relation to the cost of acquiring this technically optimum scale of plant will in time be multiple plant firms.[16]

On the other hand, the production function may be such that a wide range of scale of plant is efficient. In this case the long run average cost curve for the relevant set of factor prices will be flat bottomed. In such cases the scale of plant which a particular firm operates will depend upon the way in which the relative prices of the factors to a firm change as a result of the financing problem. In industries where the production function leads to such flat bottomed long run cost curves, equilibrium will be compatible with the existence of plants of varying sizes.

As a result of the firm being identified with a fixed equity, the profit maximizing principle can be equated with the maximization of the rate of return on the fixed equity. Only by means of some such approach can the equality of some measure of internal returns and a market rate of interest be related to investment behavior.

The definition of the production function is to a great extent arbitrary. It could be defined so that the entrepreneur and the equity base are contained in the production function. Each such production function is unique, as the entrepreneur and the equity base are fixed. In such a production function each economic unit is operating in the short run with a set of fixed factors, and the expansion of output by adding other factors is equivalent to a 'rent' problem. On the other hand, the production function can be defined independently of the entrepreneur and the equity base, as an engineering input-output relation. It states how outputs would be related to inputs, and the relation holds for any firm that would produce the output. This objective production function is, I believe, more useful. It naturally leads to the plant-firm dichotomy used here. A firm, which is an entrepreneur and an equity base, operates upon a technologically given production function.

An analysis, such as this, which depends upon the relations between a market demand curve, the demand curve confronting individual firms, and the supply curve of a product, is basically Marshallian, and the industries which are analysed are Marshallian industries.[17] As such the industry is not the industry to be found in statistical compilations, which are often based upon production function similarities. The industry is the supply correlative of a demand curve. The demand curve for a product is made up of two parts, the first a transformation of household preferences, the second the demand for a product derived from its use as a factor of production. This conceptually objective demand curve is the beginning of our analysis, states the quantities of a commodity which will be taken at a given set of prices, under the assumption that the prices of closely related commodities are given.

As the supply correlative of this objective demand curve, the industry is the set of firms which produce the output represented by the demand curve. A firm may belong to several industries at the same time. The attribute of the firm which is associated with the given demand curve is, under competition, a schedule which asserts how much of the product will be forthcoming at each price. In general, for a profit maximizing firm, the relevant schedule of the firm states the amount of the product forthcoming at each marginal revenue. The constraint upon the firms in the industry imposed by the demand curve is that the quantity which they produce must be sold at a price consistent with the given demand curve for the product.

In order to derive this price-quantity value for each firm we need a demand curve for the firm, and a relation which indicates how much the firm will produce at each market price. To get this we need to know not only the demand conditions confronting the firm, but how the firm will behave in relation to these demand conditions; we need a behavior principle. A unique behavior principle can be derived only in the case of competition.

Much of present day price theory tends to identify price theory with marginal analysis. Marginal analysis is far from being a primitive concept; it is derived in the analysis of perfect competition. Marginal analysis is derived from profit maximization, and in the analysis of non-competitive markets a postulate is explicitly made that the entrepreneur 'is assumed always to choose the output which will maximize his net receipts'.[18] Beginning with Mrs Robinson's work the carrying over of profit maximization behavior to non-competitive markets has been the typical approach as far as price theory is concerned. Economists can be accused of being parrots who say that marginal *x* equals marginal *y*, and the position so determined is where the firm will operate.

Observed behavior of firms,[19] any casual observation of the behavior of certain non-competitive firms during an inflationary period, can be interpreted to indicate that firms either lack the knowledge or the desire to behave as the marginal analysis indicates a firm should behave. The use of unmodified profit maximization as the sole basis for the analysis of firm behavior is a carry-over by the economist from the analysis of competitive markets. A firm confronted with a horizontal demand curve and selling in a market where there is free entry and exit is effectively constrained to maximize profits. If the firm in this competitive case did not maximize profits it would earn a lower rate of return on owned factors of production, capital and the services of individuals uniquely linked to the firm, than that freely available to the owners of these factors (the free entry assumption).

Implicit, therefore, in the assumption that a firm must maximize profits are assumptions about the nature of the capital market. Each owner of capital who invests in a firm has available alternatives which will earn as much (or an equivalent amount if risk factors are taken into account) upon his capital as he can earn upon it by his investment in a particular firm given that the firm is maximizing profit. Each owner of wealth has freely available alternatives, to either invest in equities or to invest in debt, which, taking into consideration risk differentials, will yield him the same return as he can earn in any particular profit maximizing firm. This means that, for firms operating in a competitive product market and in an economy in which the capital market is perfect, there is one and only one cost item for the financing of control over the amount of capital equipment the firm requires. This is the market value of these items times the going rate of return. The cost to the firm of the plant factors is given; the fixed costs are determined. Hence, for each plant there is a unique average total cost curve associated with each marginal cost curve. The unique average cost curve for each plant, and for variations in plant, as drawn in traditional price theory, depends upon the existence of competitive product markets and a free movement of capital.[20]

The traditional theory of imperfect or monopolistic competition, derived from Joan Robinson and Professor Chamberlin,[21] eliminated either one or both of the conditions which made profit maximization the only possible behavior rule for a firm, that is, it eliminated the horizontal demand curve confronting the firm and the free and costless entry or exit of firms to the industry. Nevertheless, they still kept the formal rule of profit maximization as the behavior principle of the firm. If the necessity for profit maximization follows from the structure of the perfectly competitive market, then the analysis of the behavior of a firm selling in a market which is not perfectly competitive cannot be based upon the assumption that a firm necessarily maximizes profit. Profit maximization, rather than being the only behavior policy that a firm can possibly follow, becomes, for market conditions which are not competitive, but one of a set of alternative policies.

In order to determine the behavior rule for a firm in a non-competitive market, it is necessary to inquire into the conditions under which the negatively sloped demand curve confronting the firm can be changed into a less advantageous (shifted to the left or more elastic) demand curve. The factors which constrain profit maximization by a non-competitive firm are the cost of entry into the industry by new firms, the relation between the production technique the firm is using and alternative production techniques, the imperfections, aside from the entry costs, of capital markets, and the effects of legal and other institutional pressures.

The demand curve confronting the firm in a market which is neither competitive nor monopolistic is a hazy concept at best. The firms in this oligopoly market share a market demand curve for the product. Their individual demand curve depends upon the prices of the products of the other firms in the industry. This interdependence leads to an indeterminant demand curve for each firm.

Aside from the imperfections of the capital market, the constraints upon profit maximization enumerated above can be treated as limitational prices, quantities or profits for the firm. For a given set of the prices of the products of the other firms in the oligopolistic industry, a negatively sloped demand curve for the product of a particular firm can be derived. The constraints upon the behavior of the firm will modify this narrowly defined demand curve. These constraints operate upon all the firms in the oligopoly industry, so that they all have modified demand curves with common limitational factors. This makes more tenable the assumption of a given set of prices of the products of the other firms in determining the narrow oligopoly demand curve. For a particular firm the effective demand curve is determined by these limitational factors and the negatively sloped demand curve narrowly defined. As an operating hypothesis we will assume that the firm profit maximizes with respect to this effective demand curve, although we recognize that even such constrained profit maximization is not a necessary behavior rule.

The limitations upon the firm's demand curve which are derived from the market structure define a best attainable position for a firm. This best attainable position may be inferior to the unconstrained profit maximization position with respect to 'profits'. A lower limit to the firm's profits is given by the survival conditions of the firm. As a result of transforming the demand curve confronting the firm by these market structure constraints, we decrease the range in which the firm can deviate from maximizing behavior. Therefore the possible error introduced by assuming profit maximization is reduced by introducing the market structure constraints into firms' demand curves.

The advantages to a firm of operating in a non-competitive market center around the return that is being earned by the firm on a given amount of owners' equity. The advantages to owners of a firm which sells in a non-competitive market is that the returns upon their investment are higher than can be earned in the available alternatives. In an economy in which different degrees of monopoly and competition exist, we should expect to find that firms earn different rates of return upon investment. We should also not expect to find that single rate of return upon investment being earned which follows from the free entry and exit assumptions of competitive theory. We should also expect to find that there is no real tendency toward an equalization of the rate of return earned upon owners' investment among industries and among firms in the same industry as long as the 'industrial structure' remains stable.

The assets owned by a firm yield a revenue to the firm. The net revenue which such assets yield to the firm, divided by the current replacement costs of these assets, is the average internal rate of return. If we subtract from the numerator the debt charges and from the denominator the value of debt outstanding, the result is the average rate of return on owners' equity. If assets are freely transferable from one industry to another and from one firm to another, then the effect of owners of wealth seeking the best return among the available alternatives will be that all such collections of assets will tend to earn the same rate of return. Allowing for risk, this uniform rate of return earned upon assets will hold for all firms and for all industries.

If assets are not freely transferable from one industry or firm to another industry or firm, different going rates of return can be earned upon assets valued at their replacement costs.[22] Such assets, which can be assumed to have a market price that is independent of the particular firm which purchases them (absence of monopsony), will be more valuable to those firms which earn a higher rate of return than to those firms which earn lower rates of return. For a firm whose internal rate is higher than the market rate of return, the capitalized value of its net revenue will be greater than the market price of its assets. This difference is 'the value of the organization', which includes the 'value of market position'.[23]

To illustrate the above argument, let us make the heroic assumption that we can rank the various industries in an economy according to the 'freedom of entry'. No owner of wealth need earn a net return lower than that which is available in the 'freest entry industry',[24] taking into account the cost of entry (installation charges) into this freest entry industry in determining the denominator for the measurement of the going rate of return. If there are industries such that the terms of entry are not the same for all owners of wealth, so that those who can enter on favorable terms (cheaply) will earn a rate of return greater than that available in the freest entry industry, whereas those who cannot enter cheaply would, if they paid the entry price, earn a rate of return lower than that available in the freest entry industry, then the favored entrants will, in purchasing assets at the market price, earn a capital gain. This capital gain is due to the valuation of the assets owned by the firm on the basis of 'capitalization of earnings' rather than on the basis of the market price of the asset; and the rate of return at which earnings are capitalized is the going rate of return in the freest entry industry. This going rate of return in the freest entry industry could be labeled the competitive or market rate of return.

An equivalent statement is that in an economy in which the various industries have different terms upon which they can be entered, and if the owner's investment in the firm is measured by the replacement or market cost of its capital goods minus the firm's debts, firms will earn different rates of return upon their owner's investment. If firms can be ranked according to their rates of returns earned in such a ranking, the rate of return earned by competitive (the most easily entered) industries will be, under appropriate static assumptions as to entry, the lowest rate of return earned. The value of a non-competitive position therefore can be measured by the differential rate of return over the minimum rate of return earned by the competitive industries.[25]

In order to analyse firms operating in non-competitive markets we will begin with the foundations, and some of the consequences, of the non-competitive position of the firm. The foundations of the non-competitive position of the firm will result in some constraints being imposed upon the firm, these constraints being independent of the market demand curves for the product and the cost curve of the firm. To the extent that these constraints can be interpreted as implying conditions upon the profitability or price of the product of the firm, they can be represented as constraints upon the profit maximizing behavior of the firm. Rather than seek a substitute rule of behavior for the firm, we will posit that firms attempt to profit maximize, but that the constraints upon their profit maximization behavior are broader than those usually considered in price theory.

For each firm which is selling in a non-competitive market we posit:

1. The firm is not earning a lower rate of return upon owner's investment than could be earned in the freest alternatives available to the owners of the firm.
2. This 'monopoly' position which results in the higher rate of return is vulnerable. The vulnerability of a firm that possesses the advantage of operating in an imperfectly competitive market is a key to the understanding of the apparent deviations from simple maximizing behavior. It is sufficient to say that all other things being equal a firm would maximize profits, and to remark that the 'other things which may not be equal' are the conditions necessary for survival, and the limitations upon entrepreneurial behavior, real or imagined, so as to assure the maintenance of an advantageous market position. These advantages are measured by the excess of the return earned upon investment above the competitive rate of return.[26]

A firm that earns a rate of return greater than the competitive rate is either operating so as to maximize profits with respect to a particular demand curve or there exists some effective constraint (objective or subjective) that prevents the firm from simply maximizing profits. If the firm is maximizing profits with respect to a given demand curve the analysis is straightforward, and is contained in all textbooks. A firm will not maximize profits if the consequences of an attempt to maximize profits is to undermine the foundation of its rate of return being greater than the competitive rate. Such an undermining of the favored market position of a firm depends upon the reaction of other economic units to the result of profit maximizing behavior. The deviation from simple short-run profit maximizing behavior by a firm is based upon a recognition that it must protect its favored market position.

What we will call the vulnerability of firms takes into account the effect upon market structure of the price, and quantity produced, of the firm's product and of the return which the firm is earning upon its investment. If certain values of these variables result in reactions by other economic units which affect the given firm's market position, the firm is vulnerable. Hence, vulnerability determines bounds to these variables, and the particular effective demand curve has to take these bounds into account. If these bounds, which are determined by market structure, are effective they lower, in at least one dimension, the values of the variables which result from profit maximization by the firm with respect to the unmodified demand curve.

A constraint which operates upon firms and which is independent of the market structure as defined above, is the limitations upon firm behavior introduced by financing considerations.[27] Firm behavior can be considered as ruled not only by the desire for profits but also by the desire on the part of the firm's management to avoid failure or bankruptcy: for in failure or bankruptcy the value of the organization disappears. These monetary or financial constraints upon the firm I have labeled the survival constraints. These conditions[28] can be stated in terms of the excess (marginal equality) of some measure of total revenue over some measure of total cost. Each business enterprise not only is constrained by the technological conditions of production, and the relative prices of the factors of production, but it also has to meet contractual obligations payable in money, which are the result of the production process. These obligations are normally met by the proceeds of the sale of the product of the firm[29] – that is, the monetary obligations are a measure of total cost, the receipts from the sale of the product are a measure of total revenue. The payments that any firm has to make are dated, some are immediate results of the process of production, other factors receive payments after intervals of time, some of the obligations of a firm are essentially contingent. If we begin at any date we have that at each and every future date, in order to survive, the firm must satisfy the condition that the initial cash plus the receipts minus the costs payable to that date are greater than zero. That is, in order to survive a firm must be able to pay debts when due. As the elements of cost that have to be paid and the going rate increase with time, the liquidity condition states that the survival conditions are more stringent the longer the time interval being considered. An important concept for the analysis of survival is this time rate of change of the cost items to be paid, a time marginal concept of cost in contrast to the usual output marginal concept. A sufficient condition for the survival of a firm is that the sum of revenues be equal to or greater than the sum of costs over any time period.

As the survival conditions are limited to the balance sheet structure of the firm, it is necessary to distinguish among firms on the basis of how they are financed. Some firms will have balance sheets in which the only liability is owners' equity; such firms will be called wholly owned. Other firms will have a large volume of debt. Ignoring both the possible use of a firm as an 'investment trust' by the owners and the existence of various types of rental contracts, firms using the same production process and operating at the same scale and rate must have the same dollar value of assets distributed in the same manner among the different factors of production. These firms may have widely different liability structures. The liability structure is a relevant variable in the determination of the conditions under which a firm can survive. If a firm is to survive at the original scale, it is sufficient for the wholly owned firm that it earn a zero rate of return on the total assets, whereas a firm which finances part of its operations by debt instruments must earn a sufficient amount on its total assets to pay its debt charges.[30] That is, the minimum survival conditions for a wholly owned firm involve a zero internal rate of return on total investment, whereas for a firm partially financed by debt instruments the minimum survival condition implies that the firm earns that positive internal rate of return on total investment which is equivalent to a zero rate of return on owners' equity. Because of these factors, we will use the zero rate of return on owners' equity as our effective survival condition.

If there exist some industries with free entry, then there is a positive rate of return on owners' equity at which the owners of a firm will prefer to liquidate rather than continue operations. This voluntary exit rate of return is greater than the zero rate of return on owners' equity. We can rephrase the above by noting that a negative rate of return on owners' equity forces either a reduction in the scale of the plant, an increase in debts or the introduction of new equity. A rate of return equal to or greater than zero but smaller than the positive minimum for continued existence does not force either the cut in scale, an increase of debt, or the introduction of new equity. Such a rate of return can induce the firm to exit (with the owners' investment unimpaired if sufficient time is allowed to elapse prior to exit). It is necessary that the firm earn a positive return on owners' equity before the preferences of the firm's owners can result in any deviation from profit maximization. We conclude that a zero rate of return on owners' equity is the true minimum survival condition for a firm and that rates of return between the zero rate and the freely available rate can be considered as voluntary exit rates.[31]

The balance sheet constraints upon a firm are that the firm maintain a positive net worth (be solvent) and that the firm be able to pay debts when due (be liquid). The solvency aspect of a balance sheet structure depends upon the certainty of the future value of assets. A firm may become insolvent due to two causes: (1) the firm may make losses, or (2) the firm's assets may lose their value independently of the firm's profit or loss position. A firm may have a positive net revenue from its production process and may lose in its net worth due to a decrease in the value of assets which it owns. A firm, for example, may become insolvent if a bank failure occurs and its cash balance is wiped out or it may speculate on inventories and as a result make large losses, or a fire not covered by insurance may destroy its assets. Aside from such contingencies, the possibility that a firm's assets will lose their market value depends upon the net revenues that the firm earns; for example the capitalized value of returns earned by its plant and equipment. In this case, the possibility of an unsatisfactory balance sheet will only occur simultaneously with the occurrence of unsatisfactory profits. Aside, therefore, from the best possible profit not being good enough, and the possibility that the firm loses on 'outside' investments, the insolvency attribute of survival adds nothing which is not included in profit maximization unless the balance sheet variables are introduced as a technique for the analysis of a firm's behavior under conditions of risk.

A firm may be confronted with investment alternatives, in which the greater profit possibility also carries the greater chance of insolvency. A firm has to weigh the relation between the profit possibilities and the chance of success in the undertaking. That which leads to the larger profit possibility may be so uncertain that an entrepreneur rejects it. This evaluation of profits in relation to risk premium is nothing new in economic analysis and does not call for any essential modification of profit maximizing assumptions. To a certain extent, business decisions involving risk reflect a significant variation among entrepreneurs in their enjoyment of risk. Psychological theories of the business cycle emphasize the waves of optimism and pessimism which, it is claimed, run through the business community. The structure of the balance sheet of business firms will reflect the psychological attitude toward risk taking. Balance sheets which contain large cash margins and a large volume of government bonds and other fixed value assets which are essentially unrelated to the firm's production process indicate a cautious attitude toward risk bearing whereas the ownership of no assets not immediately relevant to the business of the firm indicates a more adventuresome attitude toward risk. We will allow for the risk associated with different balance sheet structures when we modify firm's cost curves to take financing considerations into account.

The relation between liquidity and survival is a reflection of imperfections in the capital market. Consider a plant the construction of which is desirable from the point of view of profit maximization. The firm is both solvent and liquid. If the firm constructs the plant using its own funds, the firm will become illiquid. In a perfectly competitive market for financing there would be no need for the firm to decrease its liquidity in order to construct the plant. In such a financing market all of the financing alternatives (such as rental, borrowing, or use of own funds) would be equivalent to the firm.

In an imperfect capital market the above are not all equally possible alternatives. The price (both money costs and the increased chances of illiquidity) which the firm may have to pay for financing the project without using internal funds may transform a potentially profitable undertaking into an unprofitable prospect. We will allow for both the objective costs and risks associated with different financing techniques when we modify cost curves to take the liquidity considerations into account.

The addition of balance sheet structure as a determinant of the behavior of firms essentially adds a dimension to the firms' cost curves. The use of a survival constraint upon profit maximization directly emphasizes the effect of varying balance sheet structures. The extent to which behavior determined by the survival constraint can differ from the profit maximizing behavior depends upon the market structure. Essentially the freedom that a firm has in determining its behavior is least in a competitive market and greatest in an unconstrained monopoly market. This attribute of markets is significant for the analysis of investment decisions.

Therefore, in order to have the theory of the firm in such a shape that we can use it to analyse investment decisions, we will expand the traditional theory in two directions. We will modify the cost curves of firms to specifically take into account the balance sheet. We will modify the demand curves confronting a firm so as to take the vulnerability of the firm into account. The firm will be assumed to maximize profits taking into account:

(a) the particular demand curve confronting the firm;
(b) the cost schedule of the firm;
(c) the effect of balance sheet structure (including risk) upon the cost schedule;
(d) the effect of market constraints upon a firm's particular demand curve.

The traditional theory of the firm takes only (a) and (b) into account.

The following analysis will be based upon profit maximization by firms under specified constraints. These constraints depend upon the market structure and the financing conditions. They can be represented by modifying the firm's demand and cost curves. Such constrained profit maximization as the rule of behavior of firms is closer to the conditional monopoly conception of Marshall than to the doctrines of modern price theory as derived from Robinson and Chamberlin.

NOTES

1. In addition, the assumption will be made that there does not exist a unique, externally (to a firm) determined interest rate for 'planning purposes'. This results in a modification of the cost curves.
2. When we take up the investment decision of a firm we have to allow for the process by which new firms come into existence or alternatively enter an industry.
3. See Viner (1936), p. 31.
4. See ibid., p. 34. Note that the volume or velocity of money is assumed to be independent of the prices of products or factors.
5. See ibid., p. 35.
6. See Viner (1936).
7. See ibid., p. 32.
8. See ibid., p. 37. In reference to the monopolist Viner says: 'Whether his price shall fall is, within limits subject to his own decisions rather than imposed upon him by market

conditions over which he has little or no control, as is the situation of the truly competitive producer'.

9. See ibid., p. 37.
10. See ibid., p. 38.
11. See ibid., p. 39. Note that active contra-cyclical fiscal policy, by changing the tax rate schedules contra-cyclically, introduces the appropriate type of price flexibility into the economy, whereas a passive contra-cyclical fiscal policy, such as leaving the tax rate schedule constant throughout the cycle, would not – except for the effect of fall in personal income upon marginal income tax rates with a constant tax schedule.
12. See Mitchell (1950), p. 170.
13. See ibid., p. 172.
14. This is independent of 'monopsony'.
15. Later on, we could define an efficient plant in terms of the survival conditions of a firm. Of course, efficiency depends upon a given set of factor prices.
16. A change in the price of existing 'plant' factors may occur so that the inefficient plant becomes 'efficient' – in the sense of normal profits at the market price. However, the new owner in maintaining and replacing this plant will have to pay market prices (which are compatible with the 'efficient' scale plant earning normal profits) for his plant factors. At the 'inefficient scale' he will be unable to earn normal profits upon such replacement investments. Unless the inefficiency is due to a factor that does not require replacement or maintenance (which is the Ricardian Rent case), in time, the effect of the deviation from optimum scale will be that the plant earns less than normal profits.
17. See Marshall (1920).
18. See Robinson (1933), p. 16.
19. See Hall and Hitch (1939) and Lester (1946).
20. A reason which can be advanced for the ready acceptance by economists of profit maximizing behavior, independently of the changes in the market structure which are introduced in their analysis, is that profit maximizing behavior leads naturally to mathematics in which derivatives of the difference between total cost and total revenue are set equal to zero. In this sense, under profit maximization the behavior of the competitive and the non-competitive firm are formally identical – the mathematical set-up is the same. The complexity added by non-competitive markets is resolved by the introduction of the demand elasticities confronting the firm at appropriate places in the analysis. In general equilibrium analysis the existence of monopoly does not lead to any adjustment in the equilibrium relations if profit maximizing is assumed; rather the effect of different degrees of monopoly is in the distribution of income and the allocation of resources. (See Lange (1944), chapter VII.). The passing over of the 'rationale' for profit maximizing in much of the analysis of monopoly can be imputed to the substitution of a tool of analysis for the problem.
21. See Chamberlin (1938) and Robinson (1933).
22. Installation costs, the cost of placing, or removing, an asset will make for differences in the going rate of return being currently earned by assets with the same market value used by different firms. Installation costs obviously set a limit to the difference in rates of returns that can be earned unless entry is not free. Also such installation costs will be 'short run' differences among asset owners, whereas the differences in return earned on assets which are imputed to industrial structure are long run.
23. See Reder (1947a), p. 456. 'The value of an organization is measured by the difference between the value of a firm when sold as an entity and the sum of the market values of its assets when sold piecemeal.'
24. 'Freedom of entry' is a difficult concept to define accurately. It obviously cannot be measured by differences in return earned upon assets by firms in the different industries: for these differences depend upon the freedom of entry (abstracting from growth and decay of an industry). Freedom of entry can be related to two factors: (1) institutional and legal barriers to entry; the patent and franchise type monopoly, long standing non-economic usages and (2) the lack of equivalence to consumers of 'technically equivalent' products. If technically equivalent products are not 'perfect substitutes' to consumers,

then entry is not free. Inasmuch as financing considerations are handled separately, the obstacles to entering an industry due to the size of the investment necessary is not considered here as a deviation from freedom of entry.

25. The discussion of 'freedom of entry' and of 'differential returns upon investment' among industries obviously refers to long run static equilibrium analysis. The utilization of such concepts in business cycle analysis involves difficulties, for the impact of changes in the level of income upon different firms may be inconsistent with the effect of the long run structural factors which determine the ranking of returns upon investment. However, in order to analyse the investment behavior of firms, we have to recognize that the investment threshold, in terms of earned returns, in the different industries depends upon the structure of the industries. In order to analyse investment behavior we also have to recognize that firms are interested in protecting their 'favorable' structural position and such a favorable structural position is a long run equilibrium concept. Obviously the transitional 'monopoly' of the innovator is something different from the structural deviation from competition.
26. The excess returns that an innovator earns is irrelevant to the problem under consideration. The situation under consideration here is of a 'stabilized' industrial structure that yields excess returns to a particular firm. The monopoly returns of an innovator are related to the process by which one production function is substituted for another in a particular industry.
27. See Klein (1950), p. 57. 'Economists have recently been claiming that business firms are as much concerned about the structure of their assets as about the size of their profits. In the words of Marshak, this means that firms behave so as to have the best possible profit and loss statement and the best possible balance sheet.'
28. In banking courses the solvency and liquidity constraints upon bank portfolios are concerned with much the same material as what I chose to call survival conditions in the theory of the firm.
29. Under unusual circumstances firms may sell assets to acquire cash to pay debts. This phenomenon is handled separately. The monetary obligations and receipts are both series of rates of flow over a price period.
30. A firm's liquidity may be maintained by new investment or by additional borrowings. This is of course what happens when a firm expands. Operating losses may also be covered by new funds. The survival condition as stated above is sufficient, not necessary.
31. In traditional analysis the imputed normal return upon owners' equity is considered a cost equivalent to any other cost. Hence, a firm in a competitive industry, where the maximum rate of return is the normal return, must profit maximize to survive. I prefer to consider any such positive 'normal' return as a voluntary exit rate and restrict the survival conditions to those conditions which would actually force firms to fail.

5. Cost curves and investment

The effect of a change in National Income, operating through a change in household income, is to shift some demand curves for consumption goods. The amount of induced investment (positive or negative) by business firms depends upon their reaction to the resulting shifts in the demand curves confronting them. This is true not only of the demand for investment goods by consumption goods producers but it is also true of the demand for investment goods by investment goods producers: realized induced investment depends upon the reaction of business firms to the shifts in their demand curves which are associated with a change of income.

It is necessary to distinguish between autonomous and induced investment. Autonomous investment is due to the introduction of new production functions and to changes in the supply conditions of factors of production. In our approach the influences upon a business firm are separated into demand conditions and supply conditions. Therefore, autonomous investment initially is due to changes in supply conditions, whereas induced investment initially is due to changes in demand conditions. In our work we are interested in induced investment, even though the 'autonomy' of autonomous investment may be questioned.

The immediate incentive to invest by a firm may be due to a change in the relative prices of factors of production. Such investment is induced if the change in the relative prices is due to the repercussions of a change in income whereas such investment is autonomous if changes in the relative prices of the factors of production are due to repercussions of a change in production techniques. For example, if during a business cycle expansion wages rise relative to capital costs, the least cost method of producing a given output changes to a more 'capital intensive' method. The investment which results is 'induced'. On the other hand, capital costs in a particular industry may fall relatively to labor costs, due to a change in the productive technique of capital goods. Such investment is 'autonomous'.

Independently of whether a particular investment is induced or autonomous, the investment decision of the firm depends upon a relation between revenues and costs. The firm's cost situation, with constant factor prices and constant technology, can be represented by a family of cost curves of the type which were first presented by Viner.[1] Investment can be related to movements along such cost curves.[2] To do this, it is first necessary to clarify the content of these cost curves and to transform them into tools which can be used to investigate investment behavior.

The average cost curve, as it is usually constructed in economic theory, is an unsatisfactory tool for the analysis of investment behavior. In traditional price theory, expansion or contraction of an industry (which we identify as investment or disinvestment) is linked to the average cost curve. The short run average total cost curve of a firm can either be defined as the set of prices and quantities of the output of the firm which, if realized, will result in neither expansion nor contraction of the firm or of the industry of which it is a part (in which case the relation between the average total cost curve and investment is tautological) or the average total cost curve of a firm can be defined, as it usually is in economic analysis, to include a normal profit.[3] This normal profit enables the average cost curve to act as a trigger to investment. If the analysis is to be based upon an externally determined profit rate it is necessary to have a unique 'cost' or 'value' independent of the price of the product for the investment in a firm. This standard approach leaves unsolved the determination of the normal profit rate.

As usually interpreted, the average total cost curve defines the set of prices and quantities of the product of the firm which yields a realized rate of return equivalent to the market rate of interest. In the usual investment analysis the short run average total cost curve of a firm is the loci of those prices and quantities of the product which result in the value of expected returns, discounted at the market rate of interest being equal to the cost of the capital goods. Such an interpretation of the average cost curve means that the essential discounting element in the analysis of investment behavior is incorporated into the average cost curve. It follows that qualitative analysis of the investment behavior of business firms can be handled in terms of the relation between market prices, outputs and the average cost curve.

The difficulty in using the average cost curve construction to analyse investment behavior centers around the content of the 'normal profit rate' which enters into the construction of the average cost curve. 'Normal profits' implies that those factors of production which are owned by the business firm (the capital factors) are included in the total cost curve on the basis of some 'normal' rate of return upon capital. In general theory such a normal rate of return upon capital is determined by the quantity of capital in the economy, the quantity of non-capital factors, and the prevailing technology. As such the rate of return upon capital which is utilized in the construction of the total cost curve is the same for all firms.

However, the 'investment threshold' for any firm in an industry involves not only the value productivity of capital and the cost of 'financing' the purchase of capital goods, but it also involves an estimate by the firm of the risk and uncertainty factors in the industry. If we interpret the average cost curve tautologically, then these 'risk or uncertainty factors' are buried in the average cost curve. If we interpret the average cost curve as being based upon an externally determined profit rate, the profit rate must be modified to allow for these factors. Therefore, the investment threshold average cost curve for a firm depends upon the evaluation of risk. To the extent that varying 'risk estimates' coexist in an industry, it is impossible to draw average total cost curves which are investment thresholds in an unambiguous manner. It is true that we might have an industry norm for the risk element, but even so, the need for a firm to conform to such an 'industry norm' is questionable.

In addition to the above, the normal rate of return upon capital becomes a hazy concept when the coexistence of various market structures in the economy is taken into account. Ignoring the effects of risk, a given quantity of capital in an economy with a given labor force and technology can result in different average realized rates of return upon capital depending upon the way in which it is distributed among industries which are competitive and industries which are non-competitive. Therefore, due to the existence of different market structures and varying risk elements in the economy, there does not exist a unique normal profit rate which acts as a threshold to investment. Rather we have that, for each industry, the cost and revenue conditions which imply investment or disinvestment by firms in an industry depend upon the market structure of the industry. Therefore the determination of the set of prices and quantities of the product which imply a change in capital becomes a function of the 'nature' of the industry.

As will be shown below, we can extend the concept of the short run average cost curve so that instead of positing the existence of a unique short run average cost curve for a given plant (identifying the given plant with a short run marginal cost curve) we can base our analysis of the behavior of a firm with an existing plant upon the existence of a family of short run average cost curves for each plant. This set of short run average cost curves associated with a given plant can be interpreted as iso-profit curves, for each of these curves is the loci of all prices and quantities of the product of the fixed plant which yields the same 'profit' or rate of return upon a fixed value of the plant. These iso-profit curves are in many ways equivalent to indifference curves and, in common with indifference curve analysis, positions of tangency between these average cost curves and the appropriate demand curve are the maximization position.

Assuming that the prices of the variable factors are given, we can identify each plant with a given short run marginal cost curve. Each such short run marginal cost curve therefore represents a fixed amount of durable capital goods as being under the control of the firm.[4] Each member of the family of average cost curves which is associated with a short run marginal cost curve represents a different rate of return upon the durable capital goods which are the plant. By allowing the quantity of plant factors associated with a given firm to vary, we have a family of short run average cost curves. We therefore can derive long run average cost curves which are envelopes of the particular rate of return short run average cost curves. These long run average cost curves enable us to consider changes in the plant in relation to either the rate of return being earned, or the rate of return which is anticipated, upon investment in the plant. These long run and short run average cost curves as generalized enable us to add a number of dimensions to the usual graphical analysis of the behavior of firms. For example, we can identify the effects of a number of the constraints operating upon firms, which we will discuss in detail in following chapters under the heading of vulnerability conditions, as elements of these families of average cost curves. That is, we can use elements from cost curves as parts of the effective demand curve. Also, we can modify the family of cost curves of a firm so that they more truly represent the financing constraints which operate upon firms. The traditional analysis of a firm is the special case in which no constraint but those which are derived from the prices of the factors of production and the production function (the family of cost curves) and the demand curve for the product of the firm are effective.

To recapitulate the above, the essential element in an investment decision relation is a threshold value to a variable, the return earned by the owners of a firm upon their investment in the firm. The average cost curve as it is usually interpreted in the theory of the firm is the set of prices and outputs of the firm which yields the investment threshold. In order to be able to use a set of prices and quantities as an investment threshold, we have to modify the average cost curves in two essential respects:

1. We have to recognize that the externally given normal profit rate is 'untenable' due to:
    (a) the existence of risk in the real world;
    (b) the co-existence of varying market structures.
2. We have to explicitly include the financing conditions confronting the firm.

By modifying the cost curves to take the above into account, we will be able to use them to determine the ideal plant for a firm under assumed conditions. The relation between this ideal plant and the given plant will imply either investment, disinvestment or no change in plant. This will explicitly depend upon the market structure and financing conditions. We will therefore be able to assert the manner in which the accelerator coefficient is dependent upon these conditions.

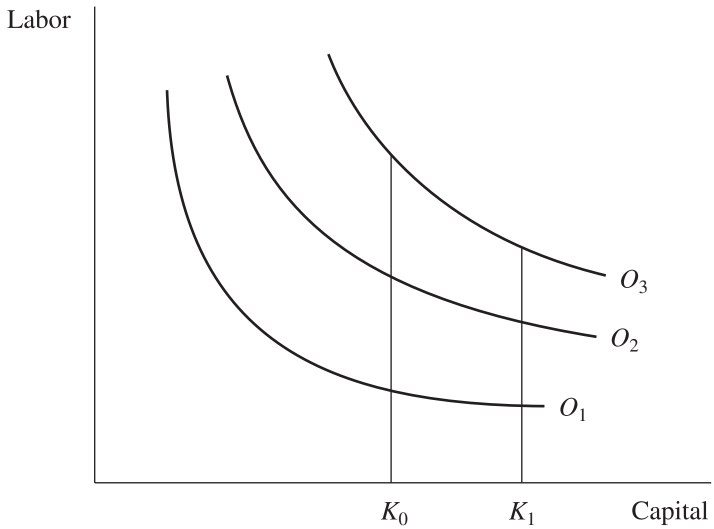


Figure 5.1

The cost conditions for a firm are transformations of the production function. If we assume only two variable factors, capital and labor (long run and short run variability), so that we can use indifference maps to represent the production function, we have that each output can be produced by varying the proportions of the two factors.

Assume that a plant represented by K_0 is in existence. By adding labor (short run factors) increased outputs can be produced. Because of (eventual) diminishing returns at a fixed scale of plant, increasing amounts of labor per increment of output are required. This will result, if the price of labor is constant, in a rising short run marginal cost curve. This short run marginal cost curve is associated with scale of plant K_0.

Similarly, if the plant K_1 were in existence, we would have another short run marginal cost curve. We therefore associate each short run marginal cost curve with a scale of plant, and a shift from one short run marginal cost curve to another will be interpreted as implying a change in capital. As investment is the time rate of change of capital, we can identify the shift from one marginal cost curve to another as investment by the firm.

The long run marginal cost curve is derived by varying both factors along an expansion path. The amount of both capital and labor that is used in production depends upon the relative prices of the two factors. As a purchase of capital factors entails present costs for future returns, the price ratio of the factors that defines the long run marginal cost curve is dependent upon a discounting process. This discounting process typically is based upon the interest rate. If we do not have an externally given interest rate, applicable to a firm in its investment behavior, there does not exist a unique long run marginal cost curve. Rather there exists a different long run marginal cost curve for each interest rate which can be used to determine the optimum manner of producing a given output.

Generalizing to more than two factors, there are as many 'runs' as there are factors. An 'n' factor production function $X = \varphi(a, b, c, \ldots n)$ and a particular set of prices of the factors of production $a, b, c \ldots$[5] will determine a minimum cost combination for producing a given output X_0, for example $X_0 = \varphi(a_0, b_0, \ldots n_0)$. If an output $X_0 + \Delta X$ is to be produced, we can rank the factors of production $a_0, b_0 \ldots n_0$ according to the time that it takes the firm to change the quantity used. For example, a_0 may be quickly variable, b_0 a bit slower, whereas n_0 may not be variable for much longer. (The ranking of the factors of production according to their variability in time may not be independent of whether $\Delta X \lesseqgtr 0$.) The long run marginal cost curve is the change in total costs of producing $X_0 + \Delta X$ instead of X_0 when all factors of production are variable (a through n). The shortest run marginal cost curve is determined by the change in cost associated with the change in output from X_0 to $X_0 + \Delta X$ by varying a only. For each intermediate time period, we can determine an intermediate run marginal cost curve by varying an appropriate set of inputs, say through g. For any time period therefore, there is a marginal cost curve which is based upon a defined subset of the factors of production being variable. Such a marginal cost curve, for all outputs greater than X_0, lies above the long run marginal cost curve, for, by definition, the point X_0 on the long run total cost curve is the minimum cost of producing X_0, the point $X_0 + \Delta X$ on the long run total cost curve is the minimum cost of producing $X_0 + \Delta X$. For outputs greater than X_0 the long run marginal cost curve lies below all the short and intermediate run marginal curves which are based upon the expansion of output from a scale of plant X_0, but it also follows that the longer the run of these short run marginal cost curves, the lower the marginal cost.[6]

In order to determine the optimum method for a firm to use in producing a particular output, the relative prices of the different factors of production have to be known. The 'price' of all except the shortest run factors is made up of three parts:

1. the cost of the capital good;
2. the 'cost' of the financing of the purchase of the capital good over the expected life of the capital good;
3. the risk and uncertainty evaluation of the business firm.

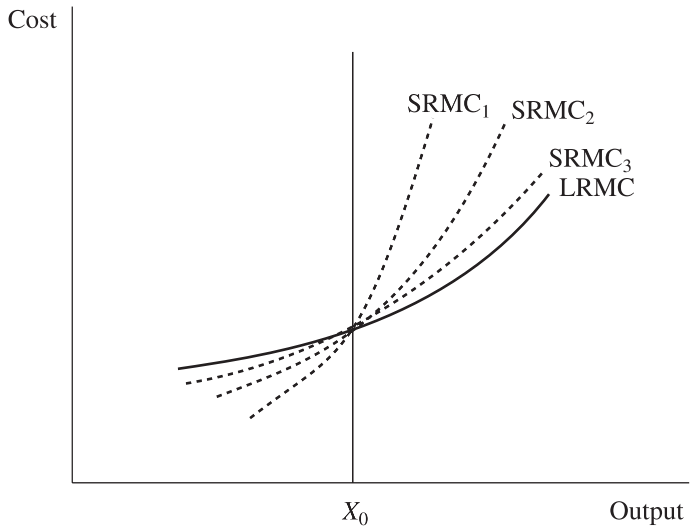


Figure 5.2

We could ignore factors (2) and (3) in the above list and draw an unambiguous long run marginal and long run total cost curve by using the cost of the capital goods as the determinant of their relative prices. This is equivalent to assuming that the cost of financing the purchase of capital goods is zero, and that risk and uncertainty do not exist. Alternatively, we could assume that the financing charges are the same for all firms. This will result in a unique price ratio between all factors of production so that a unique expansion path is determined. This would give us the unique long run marginal and total cost curves of standard price theory.

Rather than simplifying the problem as indicated above, we can assume that each firm is confronted with a set of long run marginal and total cost curves. Each element of the set of long run marginal and total cost curves is based upon a different interest rate and a different evaluation of risk and uncertainty. This will enable us to use an analysis which is based upon a set of price and quantities of the output of the firm resulting in a given rate of return upon a fixed investment (a given short run marginal cost curve). The firm's investment decisions are based upon comparing the returns being earned with a given plant with the returns which it could earn with different sized plants taking account of the terms upon which the firm can finance a change in plant size.

In the graphical analysis that follows, we will draw the demand curve for the firm with a negative slope. This assumes that the market under analysis is non-competitive and that the division of the market demand curve among the firms in the industry has been accomplished in an unambiguous manner. This does, for the time being, beg the most important question of the analysis of non-competitive markets; however, we need the generalization of traditional firm analysis which is derived in this chapter in order to be able to really see where the demand curve confronting the firm is necessary for our purposes.

With a given 'plant', a variable cost schedule is derived, and from the variable cost schedule, the marginal cost schedule. This is a straightforward metrical idea, and no imputed cost, aside from the 'user cost'[7] of capital equipment, need enter into the analysis. For the given plant and with a given demand curve we therefore are able to derive the familiar first order profit maximization condition: Marginal Cost equals Marginal Revenue, the price being determined by the demand curves. (Continuity and the increasing MC at MC = MR satisfies the second order condition for a maximum.)

The specification of the short run output cost relations for a fixed plant includes:

1. the value of the existing plant, K_0. This value is the present market price of the plant factors. It is not a capitalization of expected returns. If all of the capital equipment of a plant is reproducible, the value of plant factors can change only by means of either real investment (or disinvestment) or a change in the present market price of the plant factors. Non-reproducible capital equipment does not, of course, enter into aggregate investment; however, a particular firm may invest by purchasing non-reproducible capital equipment. We will assume that such non-reproducible capital has a market price which is given to the firm, and that such capital enters into the value of the fixed plant on the basis of this market price;
2. the total variable cost schedule, from which the marginal cost schedule is derived;
3. overhead costs which are those costs other than the imputed costs of capital which are independent of the scale of operations of the plant. Depreciation charges, other than user cost, which are part of variable costs, are included here;
4. various rates of return on plant – which we will designate $r_0, r_1, r_2 \ldots r_n$ so that for each r_i there is a unique cost element equal to the cost of capital for a given plant times the r_i $(K_0\, r_i)$. For different r_i this cost element changes. This element includes the imputed costs of capital. In traditional analysis one of the r_i would be used as a 'normal profit' rate. Obviously the period to which the rates of interest apply and the period of the cost curves are the same.

A family of short run average cost curves for each value of plant can be constructed. The average variable cost curve is the total variable costs divided by the output TVC/q, the marginal costs are the rate of change of variable costs dTVC/dq. To the total variable cost schedule we can add the overhead costs. This yields us another average cost curve, the average cost curve including overhead costs. To this total variable cost plus overhead cost we can add the $r_i K_0$ cost of capital for each particular rate of return. This yields us a distinct average cost curve for each rate of return on plant, $r_0, r_1 \ldots r_n$. All of these average cost curves will have their minimum point on the unique marginal cost curve which is identified with the plant. As r_i increases, the minimum point of the average cost points occurs at higher product prices and outputs. As r_i increases, the average cost curve for the higher r_i is completely nested within the average cost curve for the lower r_i. Inasmuch as the r_i can vary by an infinitesimal amount, the average cost curves are dense; each point in the region above the average cost curve including overheads represents some positive rate of return on capital.

The traditional equilibrium of the firm is where marginal cost equals marginal revenue, price determined by the demand curve. At the quantity at which marginal revenue equals marginal cost, the rate of change of total revenue and every total cost curve are equal. As the average cost curves are dense, at this quantity there exists one and only one average cost curve which is equal to the demand curve and it has the same slope as the demand

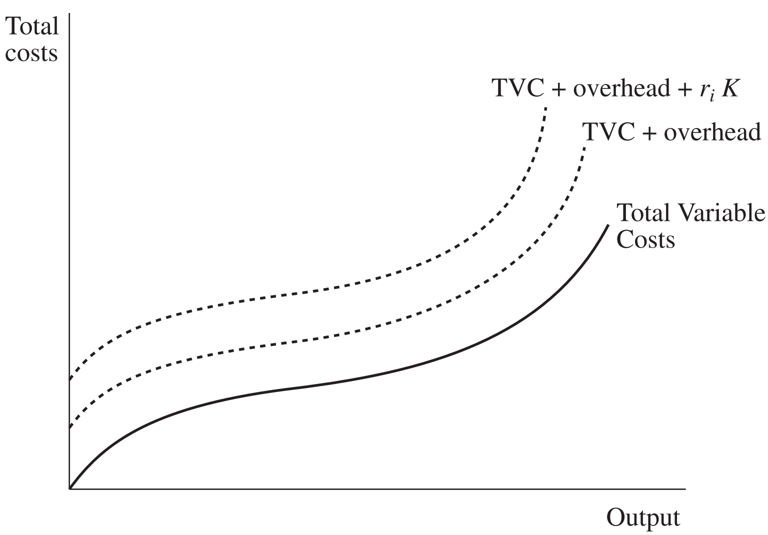


Figure 5.3.A Short Run Cost–Output Relations

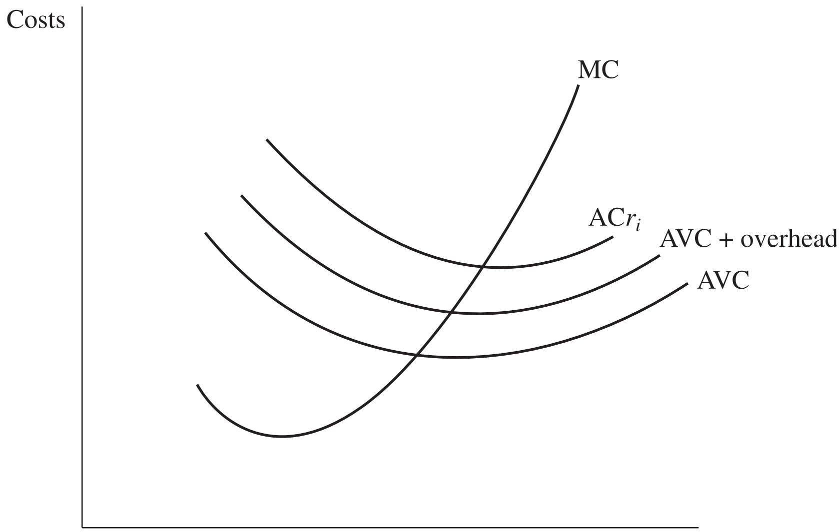


Figure 5.3.B Average and Marginal Costs

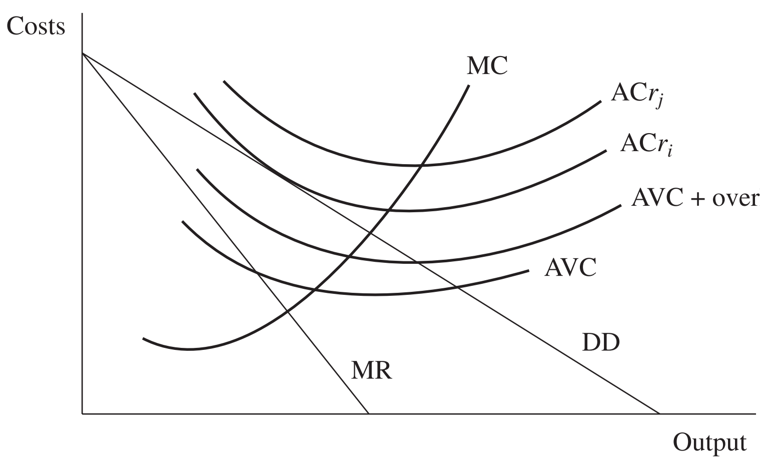


Figure 5.4 Traditional Profit Maximization Diagram

curve; the two are tangent. The rate of return which the profit maximizing firm earns upon its plant is given by the average cost curve which is tangent to the demand curve.

In the traditional analysis, the rate of return is built into the model independent of the operations of the firm, by means of the concept of a normal rate of return. In our analysis, the short run profit maximizing rate of return is given by the short run average cost curve which, for a given plant and demand curve, is tangent to the demand curve. The significance of this profit maximizing rate of return is that it represents the highest attainable return with a given plant for a given demand curve. No behavior on the part of the firm, given the demand curve for its product and its plant, can result in a rate of return upon its investment in the plant above this profit maximizing rate of return.

Professor Friedman finds the inherited average total cost curve to be essentially useless: 'It is more misleading than it is helpful'.[8] The difference between total revenue and total variable cost in his analysis is a price determined return to a fixed factor which he calls entrepreneurial capacity and the average total cost curve is drawn on the assumption that the rent would be the same at other outputs as it is in fact at the output where marginal cost equals marginal revenue. However, in drawing any short run marginal cost curves a set of production factors is held fixed at some level. The user cost of these fixed productive factors are elements of the total variable, hence marginal, costs. However, the fixed factors, which are in the longer run variable, have a market price. It is necessary if we are to analyse investment decisions to have a concept of total cost, short run and long run, based upon the various rates of return which can be earned upon these fixed factors. The positing of a rare and unique factor, called entrepreneurial capacity, as Professor Friedman does, which necessarily must earn a rent if it is to have a price and which serves as a limiting factor for firm size, is unnecessary. As will be shown in the next chapter, the effects upon the firm's costs of the limited equity of the firm will be such as to determine the size of the firm and variations in size among the firms in an industry.

In the case where the supply of 'plant' factors is fixed, so that no shifting of or change in the number of plant curves is possible, the rate of return upon the investment in the plant will be a parameter to which the market value of the plant adjusts. In this Ricardian case[9] the effect of changed demand conditions will be such that the market value of the fixed supply of plant factors will be adjusted so that the return earned by the investment in the plant remains consistent with an independently determined rate of interest, for example, the rate of interest on money loans. In the case where the supply of plant factors is infinitely elastic at a given price, the effect of changed market conditions will be that the price of the plant factors and the rate of return earned will be constant, the quantity of plant factors will vary. If the supply curve of plant factors is neither completely inelastic nor infinitely elastic, the result of a change in demand for the product is that both the quantity of the plant factors and their market price change.

The quasi-rent doctrine of Marshall fits into such a framework. If the supply of plant factors is inelastic in the short run, and infinitely elastic in the long run, the effect of a change in the demand conditions will first be felt as a change in the market price of existing plant factors. The short run behavior of the price of plant factors is essentially the same as the price of 'land'; in the long run, however, the price of these plant factors is fixed. If we price these plant factors on the basis of long run considerations, quasi-rent can be represented as the earning of a variable rate of return in the short run. The difference (positive or negative) between the short run rate of return and the long run rate of return on such plant factors can be interpreted as a quasi-rent.

If the long run supply curve of the plant factors is not infinitely elastic, then the long run equilibrium position involves a change in the market price of the existing plant factors: a long term capital gain or capital loss to some firms. In addition to the long term capital gain or loss, in the short run, the existing plant factors will earn returns which are different from the return that will be earned on these plant factors in the long run. These transitional variable rates of return upon plant factors can also be interpreted as quasi-rents. However, the pre-existing plant factors will, in equilibrium, have to be valued at a price consistent with their market price. Therefore, in the case of inelastic supply a capital gain or loss will be realized by some firms.

The construction of a family of long run average cost curves and long run marginal cost curves follows the model developed by Lerner,[10] excepting that we have a long run marginal and average cost curve for each different rate of return. That is, the long run average cost curve for a particular rate of return, r_i, is the envelope of all the short run average cost curves that yield this rate of return, r_i. The long run marginal cost curve which passes through the minimum point of this long run average cost curve is a long run marginal cost curve which is based upon the same rate of return as that which the short run average cost curve yields. The long run marginal cost curve, being dependent upon the relative prices of the factors of production, is independent of the rate of return which is contained in the average cost curve.

We therefore have a family of envelope curves and of long run marginal cost curves. We have to consider the interrelations among the members of these families. In effect, we have two types of iso-return curves: one, a set for each plant, and the other, a set for all possible size plants.

The expansion path represents the minimum cost of producing each output given the relative prices of the factors of production. But since the rate of return is used in determining the optimum plant for a given output, there is a different expansion path for each different rate of return. The long run marginal cost curve is the rate of change of total cost along

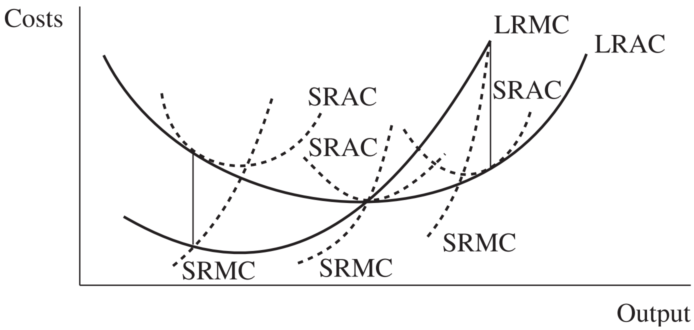


Figure 5.5 Lerner's Diagram

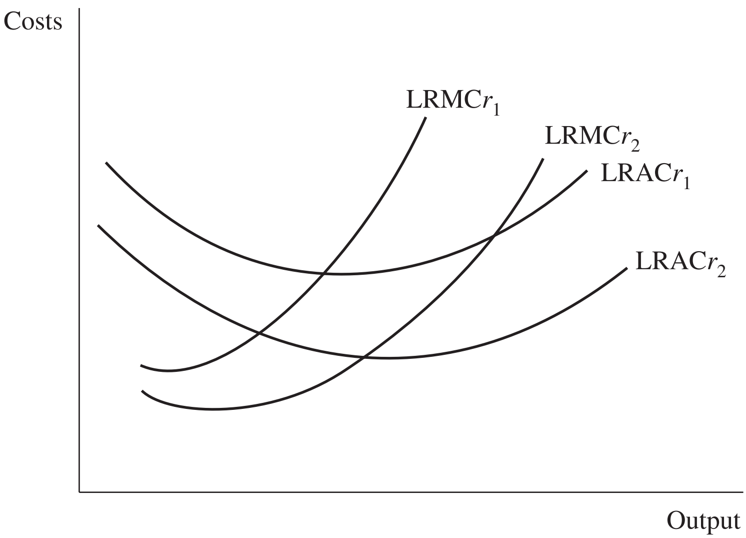


Figure 5.6 Relations among Long Run Average and Marginal Cost Curves

an expansion path, and for each different internal rate of return we have a different long run total cost curve. The lower the internal rate, the lower the total cost of producing an output, and as the area under the marginal cost curve is always equal to the total costs, at every output, the long run marginal cost curve associated with a lower interest rate cannot lie above the long run marginal cost curve associated with a higher internal rate. If the long run marginal cost curve for a lower internal rate always lies below the long run marginal cost curve for a higher internal rate, then the difference in total costs for the different interest rates will increase with output. It follows that the minimum points of higher return long run average cost curves occur at lower outputs than the minimum points of lower return long run average cost curves.[11] It also follows that the long run average cost curve for a lower rate of return is always below the long run average cost curves for a higher rate of return. The long run average cost curves are completely nested within one another.

Even though the long run marginal cost curve depends upon the rate of return used in its construction, the short run marginal cost curve is independent of the rate of return.[12] This is based on the assumption that the short run variable factors are truly short run. If we take a two factor production indifference map, we have that the short run marginal cost curve is based upon holding 'capital' constant while varying 'labor', whereas the long run marginal cost curve is based upon the expansion path where both are variable. In Figure 5.7, the amount of capital K_0 can be used to produce Output O_1 and Output O_2. The short run marginal cost schedule for capital K_0 is the rate of change in total cost as output is changed. This short run marginal cost schedule is independent of whether the rate of return used in the expansion path is r_2 or r_1. However, if the internal rate is r_2, and the planned output is O_2, then the combination L_2 of labor and K_0 of capital is the lowest cost combination. On the other hand, if the internal rate is r_1, and the planned output is O_1, then the combination of L_1 of labor and K_0 of capital is the lowest cost combination. If the amount of capital K_0 is 'costed' at r_1, then the long run total cost curve at output O_1, based upon the rate of r_1, and the short run total cost curve for plant K_0 are equal. Short run total costs, costing capital at rate r_1, for outputs both smaller and larger

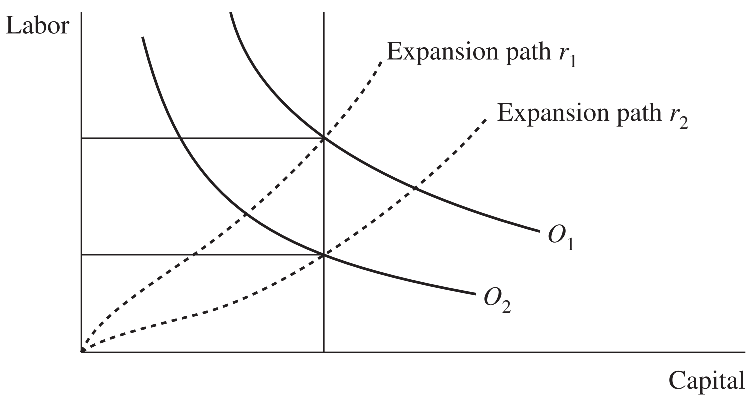


Figure 5.7 Iso-product Curve

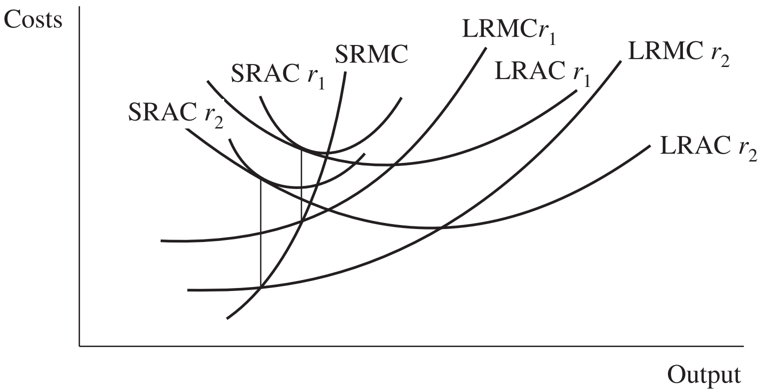


Figure 5.8 Relation between Short Run and Long Run Cost Curves for Different Rates of Return

than output O_1 are greater than long run total cost. Hence, the short run and long run total cost curves are tangent to each other at the output O_1. It follows that the long run average cost curve for a particular rate of interest is the 'envelope' of all the short run average cost curves which 'yield' that particular return. It also follows that the short run and the long run marginal cost curves intersect at the output at which the short run and long run average cost curves are tangent to each other.

As was mentioned earlier, with a given plant there always exists a short run average cost curve that is tangent to a given demand curve. If the firm is operating as a short run profit maximizer, it will operate so as to produce the output which yields this short run tangency relation. If this solution is achieved with a given demand curve and a given plant, the long run average cost curve which is the envelope of the short run average cost curve that is tangent to the demand curve may not be tangent to the demand curve. This means that there is another plant size which could yield a higher rate of return than the given plant. This plant size is given by the tangency of a long run average cost curve with the demand curve. The tangency of a long run average cost curve with the demand curve represents the highest rate of return that can be earned with a given demand curve. At the output where the long run average cost curve and the demand curve are tangent, the long run marginal cost curve for this rate intersects the marginal revenue curve. By building the plant so determined, the firm will earn the highest attainable rate of return given the demand curve.

If the firm uses a planning rate lower than the highest attainable rate ($r_j < r_i$) in determining its expansion path, then the maximizing plant will

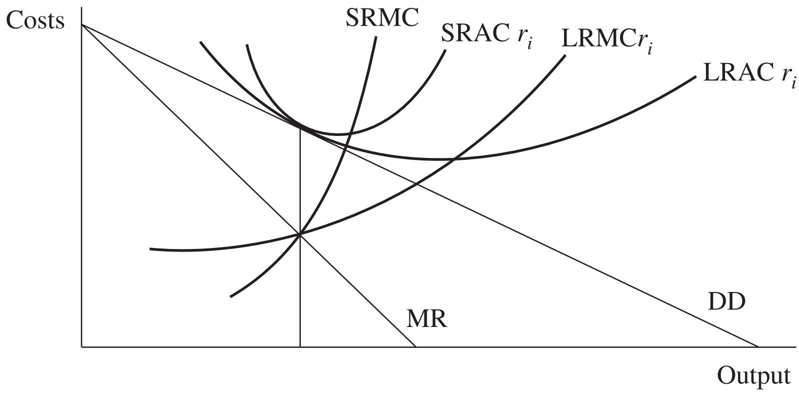


Figure 5.9 Profit Maximization Long Run Conditions

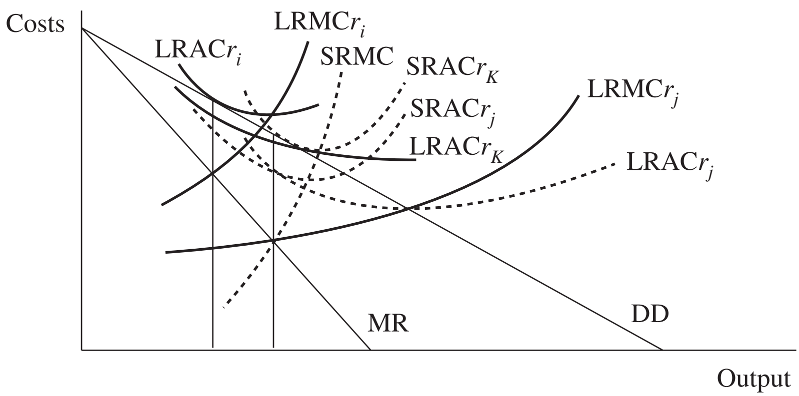


Figure 5.10 Profit Maximization: Planning Rate – Maximum Attainable Rate

be determined by the intersection of the long run marginal cost curve and the marginal revenue curve. At the output where they intersect SRMC = LRMC = MR and LRAC(r_j) = SRAC(r_j) The average cost curves for r_j will be parallel to the demand curve. However, if the firm maximizes profit with the plant so determined, the rate of return earned will be given by SRAC(r_k) where $r_i > r_k > r_j$, and the LRAC(r_k) will intersect the demand curve. The argument is simple if we use total cost and total revenue relations. With a given total revenue curve, there is a rate of return, r_i, which for some output O_i will make total revenue just equal to total cost

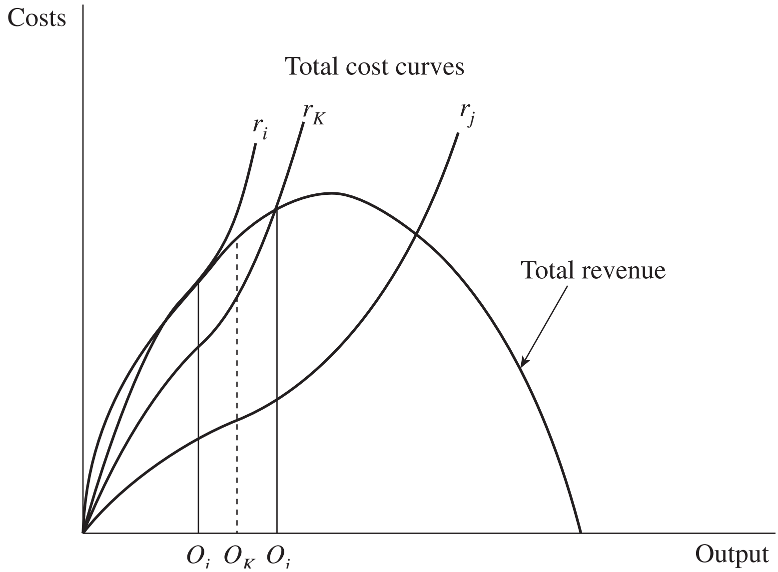


Figure 5.11 Profit Maximization: Planning Rate – Maximum Attainable Rate

and yield r_i on the plant. If a lower rate, say r_j, is used in determining the size of plant, then O_j will be greater than O_i and total revenue will be greater than total cost based upon r_j. However, r_k will be the actual yield and $r_i > r_k > r_j$.

Our conclusion is that if a firm is attempting to earn the highest rate upon its total investment, the plant decision is based upon the tangency of a long run average cost curve and the demand curve. If a firm uses a rate of return either higher or lower than the highest attainable rate in order to determine its long run marginal cost curve, it will earn a lower rate of return upon its total investment than the maximum attainable.

The history of such long run cost curves has been rather peculiar. The original article by Viner has achieved as much renown for the controversy between Professor Viner and Dr Wang[13] as for its content. However, after the clearing up of the interpretation of these curves,[14] the material now appears in price theory textbooks under the heading of planning curves,[15] and aside from its use as a classroom exercise, these curves, as far as I know, have not been used in further analysis. This neglect seems to me to be particularly unfortunate. However, for these curves to be truly useful, it is necessary to drop the assumption of a unique long run marginal cost curve for the firm, which we have done.

The distinction drawn in these curves between the short run and the long run refers to the scale of plant. 'The short run is taken to be a period which is long enough to permit any desired change of output technologically possible without altering the scale of plant, but which is not long enough to permit any adjustment in the scale of plant. It will arbitrarily be assumed that all of the factors can for the short run be sharply classified into two groups, those which are necessarily fixed in amount and those which are fully variable.'[16] Viner went on to define the increases of the scale of plant by asserting that 'each scale will be qualitatively indicated by the amount of output which can be produced at the lowest average cost possible of that scale'.[17] As our broadened concept of the average cost curve is inconsistent with the existence of a unique average cost curve for each scale of plant, we have to define the measure of the scale of plant in another way.

Each long run average cost curve has associated with it a long run marginal cost curve. The plant which a firm would build if it were maximizing the rate of return upon its total investment in plant is given by the tangency of a long run average cost curve and the demand curve. This is equivalent to the intersection of the long run marginal cost curve for that rate of return and the marginal revenue curve. The plant that will be built is given by this intersection and will determine a short run marginal cost curve. Each short run marginal cost curve therefore represents a scale of plant. If the firm is using a unique long run marginal cost curve in its planning, the optimum scale of plant for the firm can be defined in terms of this long run marginal cost curve's intersection with the marginal revenue curve. Inasmuch as the zero return on total investment has significance for survival (Chapter 6), we can define each scale of plant as the intersection of the short run marginal cost curve with the zero return long run marginal cost curve. The determination of the measure of the scale of plant is purely arbitrary: all we have to do is to be consistent in our usage.

'The long run is taken to be a period long enough to permit each producer to make such technologically possible changes in the scale of his plant as he desires, and thus to vary his output either by a more or less intensive utilization of existing plant, or by varying the scale of his plant by some combination of these methods.'[18] The ability of an existing producer in the long run to change the scale of his plant as he desires leads us to consider the relation between an existing plant and alternative plants. If we are really interested in the static long run in our attempt to analyse the generating factors for the coefficients in macroeconomic models, then we can ignore the possible existence of variations in the production relations other than along a uniquely defined long run production function. If we can restrict ourselves to a unique production function we can unambiguously define investment as a movement from one point to another on a particular return long run marginal cost curve.

However, if the time period which is relevant to investment decisions is short enough so that existing plant is a relevant factor in the incremental outlay[19] necessary to increase the scale of the plant, we cannot analyse investment in terms of a unique particular return long run marginal cost curve. We have to distinguish between the optimum manner of producing a given output at a given factor price ratio when no plant is in existence, and the optimum way of producing a given output at a given factor price ratio when the initial conditions include the existence of a plant. Of course, the second alternative, where the initial conditions include the existence of a plant, implicitly assumes that the relevant time period for our problem is short of that time period in which the plant can be reduced to a zero productive capacity, for in such a time period there is no difference between the choice with no existing plant and with an existing plant.

In an iso-product map, the long run marginal cost curve is the rate of change of total costs as you move along an expansion path. Short run marginal costs are the rate of change of total costs with plant factors held constant. If output is increased from O_1 to O_2, as shown in Figure 5.12, total costs (constant factor prices) are increased from L_1 to L_2 in the long run, from L_1 to L_2' in the short run. However, there may be an inter-

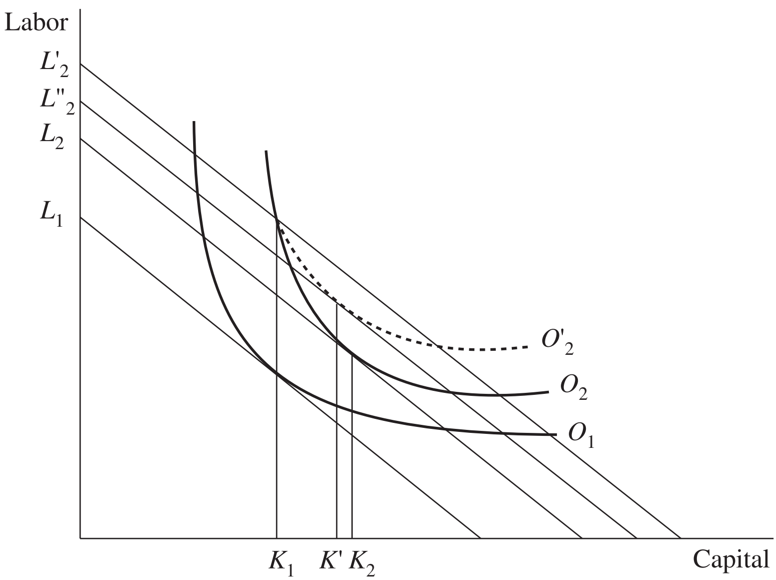


Figure 5.12 Production Conditions – Modification of an Existing Plant

mediate time period (let us label it an investment period) in which plant may be modified so that total costs fall from L_2' to L_2'' and nevertheless total costs are higher than they would be if, starting from scratch, a plant for output O_2 were built. The dotted iso-product curve O_2' is the relevant curve for plant expansion. It may be that in time a firm keeps on producing output O_2 , the O_2' iso-product curve approaches the O_2 iso-product curve; it may be that it never does. That is an element of the technique of production.

The cost curves as defined here are transformations of a production function. As such, once the factor price ratios are given (including the relevant earning rate), the long run marginal cost is given too. We can define the long run marginal cost curve as the marginal cost of a particular output starting from no plant as an initial condition. We can also draw a longer than short run marginal cost curve on the assumption that we are modifying an existing plant. If the existing plant is a relevant variable in determination of investment decisions (in business cycle analysis it is necessary to argue from where you are rather than in terms of long run timeless considerations), then the long run marginal cost curve using an existing plant as an initial condition is the relevant curve for the expansion of plant by a firm with an existing plant.

For a given firm, therefore, we can distinguish two types of long run marginal cost curves.[20] One is a long run marginal cost curve in which the firm completely ignores the existing plant: that is, a 'new plant' long run cost curve. Another is a long run cost curve which considers the existing plant as a given and modifies the existing plant. The modification long run marginal cost curve never lies below the zero plant initial condition long run marginal cost curve for outputs greater than the optimum output with the original scale plant. For a given output larger than the optimum with a given plant (factor prices fixed), the amount of capital which the modified plant will use is less than the amount of capital which the optimum plant for that scale of output would use. That is, in Figure 5.12, the modified plant uses K' of capital to produce the same output as a plant built to produce O_2 but starting from scratch (zero plant) and using K_2 of capital.

The relevant marginal cost curve for a firm with a given plant is different from, and it lies above (for outputs greater than the original output), the long run marginal cost curve drawn on the assumption that no plant exists. For the expansion of the plant, the curve labeled LRMC′ is the relevant curve, and it always entails a smaller amount of investment for a given output larger than O_1 than the long run marginal cost curve LRMC drawn on the assumption that no plant exists. However, for a firm with an existing plant, we have another long run marginal cost curve to consider – a long

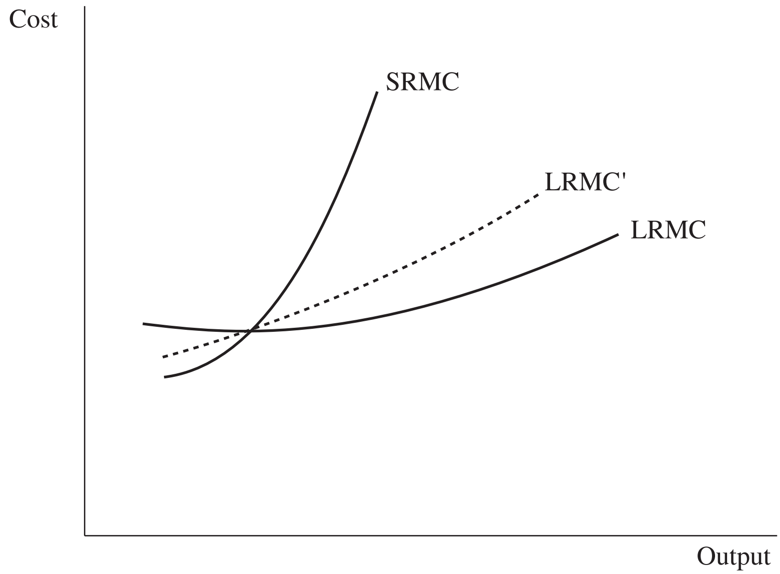


Figure 5.13 Cost Conditions with a Given Plant

run marginal cost curve which has its zero point at the scale mark of the existing plant and which is drawn on the assumption that no plant exists. If the production function is such that there are some large factor indivisibilities, so that the long run marginal cost curve has a negatively sloping portion, there would be a range of outputs at which the long run marginal cost curve (zero plant) would lie above the long run marginal cost curve based upon modifications of existing plant. This long run marginal cost curve is drawn in as $LRMC_2$ in Figure 5.14. The investment planning curve for the firm is then the lower of LRMC′ and $LRMC_2$. Until output O_2 is planned for, the optimum behavior of the firm would be to modify the existing plant; for outputs larger than O_2 the optimum behavior of the firm would be to build another plant. However, for a given planned increase in output, say from O_1 to O_3 in Figure 5.14, the investment needed to build a new plant is greater than for modifying an old plant. Financing conditions may modify the long run marginal cost curves in such a manner that the output at which it would be desirable to shift from one expansion scheme to another would change. It is obvious that if the plant is modified to let us say the scale O_2, the $LRMC_2$ would have to be drawn beginning at the output O_2, for the modification of capital equipment which is contained in $LRMC_2$ has taken place. To the extent that the useful life of capital equipment used in modifying plant is relatively short (a technical question),

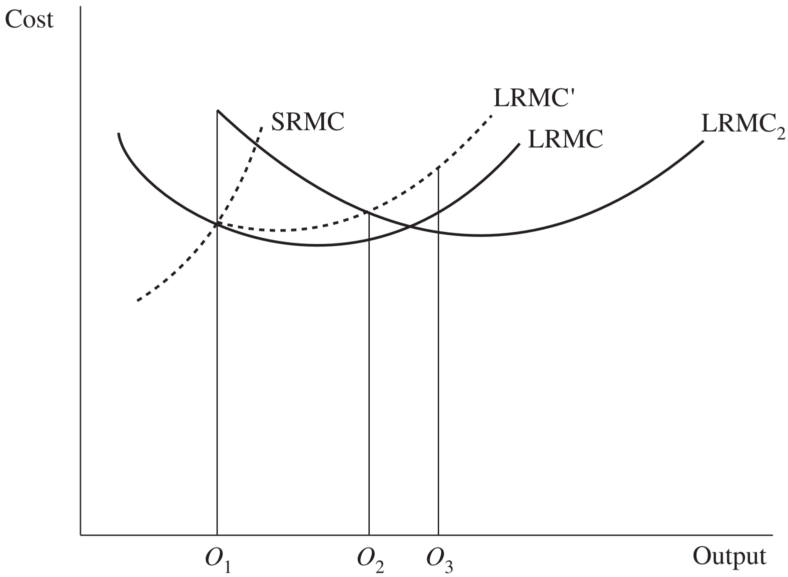


Figure 5.14 Long Run Cost Relations: Modification vs. Additional Plant

the firm may, after modifying its plant, still contemplate constructing a new plant meaning that the $LRMC_2$ begins at O_1 rather than O_2.

One aspect of the technological production function which is relevant to our analysis may be mentioned in passing. It is assumed in the literature that for any combination of inputs there exists one maximum output which can be produced. There is another aspect of a firm's production function which is relevant to firm planning – the modification possibilities of the plant. Plants built in expanding industries, plants built by optimistic entrepreneurs may have built-in modification or expansion possibilities. The purely technical input-output relations of the production function may be based upon plants which have steeply rising investment period marginal cost curves. The firm may choose a plant type which is inferior for a given planned output but which has easier modification possibilities. Even though for a given output the $LRMC_1$ defines the optimum plant, the plant which leads to $LRMC_2$ may be built because of its easier modification possibilities. The existence of such alternative production functions for a particular output is relevant to our analysis, for whereas the plant that can be modified easily entails a larger initial investment than the plant which cannot be easily modified, the amount of investment which is induced by an increase in demand is smaller when a plant is modified than when a new plant is constructed.

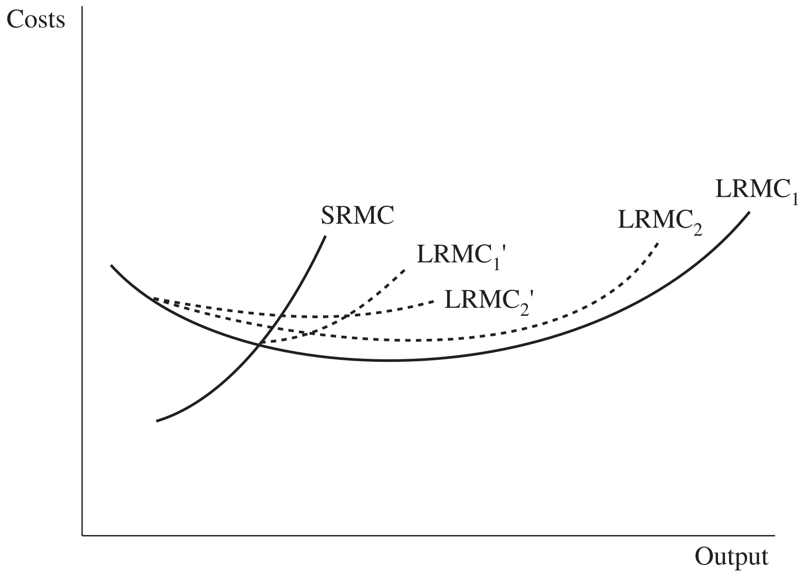


Figure 5.15 Long Run Cost Conditions: Flexible vs. Inflexible Plant

It is a question of fact, based upon technique, whether the plants in an industry are easily modifiable or not. If plants are easily modifiable so that the long run marginal cost curve for modifying plant does not lie appreciably above the long run marginal cost curve, assuming zero plant, increases of output in the industry would tend to be associated with increases in the scale of the plant. If plants associated with an industry are not modifiable, we would expect the increase in output in the industry to take place through an increase in the number of plants. These technological production conditions determine in part whether the industry will be an industry of few plants or many plants; they are insufficient to determine whether the many plant industry will be a few firm or a many firm industry. If plants are easily modified, we would expect that the industry, all other things being the same, would have fewer plants than an industry in which plant cannot be modified. However, the range of output for which long run marginal cost is falling or not rising would also be a determinant of the number of plants, and this attribute is independent of modification possibilities. That is, the long run zero existing plant marginal cost curve may be falling and still a plant once built may not be modifiable. In this case a new firm, entering the industry by building a new large scale plant, may succeed in destroying the value of existing plants.

The aim of this analysis is to identify movements along a long run marginal cost curve with the investment process in the economy. The long run marginal cost curve as it was originally drawn is a particular transformation of the production function, based upon constant factor prices, and upon the independence – because it is truly long run – of the long run marginal cost curve from the effect of existing plant upon the optimum way of changing output. We have shown how for firms with an existing plant we can draw a marginal cost curve based upon an investment period. If the investment period is short in relation to the long run, this investment period long run marginal cost curve is the relevant cost curve for some scale of output changes. Beyond a certain range of output changes, the possibility of constructing new plants dominates the modification of existing plant in the alternatives confronting a firm. The long run marginal cost curve which is relevant for investment decisions is not independent of existing plants for firms with plants in the industry, but is a cost curve which for some ranges of output changes determines that the firm will build new plants. For firms with no plants in the industry, the zero plant long run marginal cost curve is the relevant curve for investment decisions.

Any movement along the long run or investment decision period marginal cost curve implies investment in plant and equipment factors. The scale along the x axis is, however, an output scale. The ratio of plant and equipment per unit of output is a scale factor which transforms the change of output into a change of plant – that is, into investment. This output–capital scale factor may vary significantly along a given period cost curve with the extent of the change in output. We have already concluded that a given increase of output achieved by the modification of a given plant entails the use of a smaller increment of plant factors than the same increase of output would entail if it were achieved by building the optimum long run plant for that output. This solution is independent of the appearance of redundant plant factors if a large scale plant, along the zero plant marginal cost curve, is built, rather than achieving the same output by modifications of the existing plant. The complex of marginal cost curves we have derived enables us legitimately to identify movements along the long run marginal cost curve as investment. However, before we can take that up, we have to transform the long run marginal cost curves to take into account the financing conditions for investment.

NOTES

1. See Viner (1931).
2. A change in factor costs or in production technique results in a shift of the 'family' of cost curves. Such a shift in the family of cost curves may imply investment. This investment would be autonomous or induced depending upon the basis of the factor cost change which shifted the cost curves. Much of the later analysis will be carried on the basis of the

existence of alternative families of cost curves for a firm, due to the internal determination of at least one element in capital costs, the normal profit rate.

3. Normal profit for a given plant is a lump sum which is determined by multiplying a fixed value of capital by a normal profit rate.
4. The firm is a collection of plants. No firm is smaller than one plant, but a firm may have several plants.
5. For all except the shortest run factors, the discounting process enters into determining the price of the factors. The effect of this is smaller for the relatively short run factors than it is for the relatively long run factors. This is true for the following reasons: (1) 'interest charges' on the cost of the factors is a smaller ratio of their price the shorter the life of the factor; (2) short term interest rates are generally lower than long term interest rates;(3) forecasts are better (risk is smaller) for short time periods than for long time periods. In determining the relative prices of different factor combinations, a factor combination which substitutes short run factors for long run factors has to be 'costed' on the basis of lower interest charges. The structure of financing conditions and firms' risk evaluations are therefore relevant to the determination of the optimum-production process. This is taken up in detail in Chapter 6, 'The Survival of Firms.'
6. For a detailed exposition of the above, see Friedman (n.d.), p. 49. For outputs just less than X_0, the short run marginal cost curves lie below the long run marginal cost curve. The argument is symmetrical with the argument for outputs greater than X_0.
7. User cost is the difference in the loss of productive capacity of a plant which can be imputed to using the plant, as contrasted with the loss of productive capacity of a plant which occurs independent of its rate of use. Hence, a total variable cost schedule must include the total of such user costs. This is the only cost item in a variable cost schedule with a given plant that has to be estimated. In terms of the survival conditions, user costs do not in the short run imply that any payments must be made by a firm; therefore, a firm may remain liquid, and hence survive, even if total revenues are insufficient to cover total variable costs.
8. See Friedman (n.d.), p. 53.
9. See Viner (1931), pp. 30–32. This 'Ricardian Rent' case is not relevant to investment decisions as long as the plant factors are truly fixed. In such a case, for each 'industry' where such fixed factors are used, there will be a rate of interest which transforms earnings into the price of the factor. This rate at which the valuation of the fixed supply factors is made does not have to be the rate which is used in decisions to acquire reproducible plant factors by firms in the same industry.
10. See Lerner (1937).
11. Always assuming that the long run average total cost curve for each rate of return is U-shaped.
12. In this we are following Viner: that there are only two runs, a short run and a long run. 'It will be arbitrarily assumed that all of the factors can for the short run be sharply classified into two groups, those which are necessarily fixed in amount and those which are freely variable' (1931), p. 26.
    In addition, we assume that the price of short run factors is independent of the interest rate whereas the long run factors depend upon the interest rate.
13. See Viner (1931), p. 36, footnote.
14. See Harrod (1934) and Lerner (1937).
15. See for example: Stigler (1952), p. 141.
16. See Viner (1931), p. 26.
17. See Viner (1931).
18. See ibid., p. 38.
19. Implicit in this distinction in which the existing plant is relevant to the expansion technique is the financing problem which is discussed in the next chapter.
20. The argument is based upon the firm consistently using a fixed 'rate' in its planning.

6. The survival of firms

The apparatus that we have constructed, a family of average cost curves for each plant and of planning curves for each production function, can be modified to take into account the survival conditions for a firm. Survival conditions are an effective constraint upon the behavior of firms. Therefore these modified cost curves can be related to investment decisions. Survival conditions have been defined as requiring that total money expenses be less than or equal to total money receipts (ignoring whatever initial liquidity the firm possesses) for every time period from the initial position to the firm's horizon. The objective phenomenon related to the survival of a firm is its balance sheet structure. What we will do is construct cost curves which take into account the effects of the balance sheet structure of a firm upon its survival conditions.

We will first operate upon the cost curves to allow for the objective costs associated with a balance sheet structure. We will then allow for the risk associated by the firm with different balance sheet structures. These modified cost curves enable us to investigate how balance sheet structures and changes in financial markets affect firm's investment behavior.

The debts of a firm reflect the conditions which existed in the relevant financial markets at the date when the debts were assumed. The survival conditions therefore are measures of the effects that financial or money market conditions have upon the behavior of firms. Such a generalization of the theory of the firm to allow for money market phenomena is needed if the theory of the firm is to be useful in analysing the quantity of investment which takes place.[1]

A firm in which the owners' equity is greater than zero can make losses for some time without becoming insolvent. The owners' equity enables a firm to weather such an adverse period and still survive. Whether or not such an excess of costs over revenues impairs the liquidity of the firm depends upon its behavior with respect to maintaining plant and inventories. Depreciation and amortization allowances are included in costs. Therefore a firm may make losses and still have current cash receipts greater than cash expenses, and this may be used to increase the liquidity of the firm. On the other hand, if the firm maintains its productive capacity,[2] by maintaining and replacing plant and equipment, then the making of losses results in a loss of liquidity.

In such a case the maintenance of the ability to pay debts when due requires one of the following:

1. the firm begins with excess liquidity;
2. the firm sells assets to acquire liquidity;
3. the firm borrows (sells debt) to acquire liquidity.

Whether or not the firm maintains plant and equipment, the owners' equity is being reduced when the firm is making losses. This means that the survival conditions are deteriorating. These operating losses must be contrasted with the cash behavior of a firm while it is investing.[3] During a period in which a firm is investing, its total cash expenditures may be greater than its total cash receipts from its operations. The firm may, for example, have acquired this cash by means of stock issues. In this case, the total net worth of the owners is not being decreased. It is necessary to distinguish between the decrease in cash balances due to operating losses and that associated with investment. The above mentioned decrease of cash balances associated with investment can be split into two steps:

1. the increase in assets (cash) and in net worth or debt when the financing takes place; and
2. the transformation of cash into an investment good; exchange of one asset for another. The case under consideration occurs when, because of operating losses, no asset is acquired by the firm in exchange for its excess of cash spent over cash received.

A firm may have assets in its balance sheet which are not necessary for its production process (government bonds, idle cash and excess inventories are examples) or it may not. If it has such assets, a period of losses will not necessarily result in a decrease of plant size. A period of losses results in a decrease of owners' equity. In the case where the firm has no superfluous assets, losses imply that either investment is not being maintained or the firm's debts are increasing with a constant plant size. In the case where the firm has assets not utilized in production, the survival conditions of the firm should take this into account by transforming the firm's balance sheet into one in which all assets are necessary for the production process – eliminating from the balance sheet the liabilities which are most harmful to survival. Of course, it is possible for a firm to have so large a volume of assets superfluous to production that after all debts are canceled there still are more assets than are needed for production. Such a firm is acting as an investment trust for its owners.

What assets are necessary for the production process? How much of a cash balance is a 'productive' use of resources; how much of an existing cash balance is held by the firm for speculation or precautionary purposes? These questions are, in fact, difficult, if not impossible, to answer on the basis of the analysis of the balance sheet. When is an inventory a speculation and when is it of a size necessary for the smooth operation of the plant? All of these questions of fact in relation to the balance sheet position of a firm are relevant to our attempt to analyse the survival characteristics of firms. Such assets superfluous to production do improve the survival potentialities of a firm. For long run analysis, however, they are irrelevant. The making of losses over a long enough period will force a firm with superfluous assets into the position of a firm with no superfluous assets, so that any further losses will entail either a decrease of plant capacity or an increase in borrowing.[4] We will use the wholly owned plant as a tool in our analysis and compare survival potentialities of plants with debts to such wholly own plants. We can therefore conceive of three types of balance sheets.

Type 1

Productive assets	Owners' equity

Type 2

Productive assets	Debt Owners' equity

Type 3

Superfluous assets Productive assets	Owners' equity

The typical firm's balance sheet is type 2. Both the first and third type balance sheet are superior to it; the third to the second, from the point of view of survival. As we can divide the third type into two balance sheets, we will use the first kind of balance sheet as our 'zero' from which to evaluate balance sheet structure.

Type 3.1

Productive assets	Owners' equity

Type 3.2

Superfluous assets Productive assets	Owners' equity

The assets and liabilities of a firm – its balance sheet structure – require further analysis. Given the production technique, plant size and factor prices, each rate of output per time period requires that a specific value and distribution of productive factors shall be under the control of the firm's managers.[5] These are the factors of production that the management of the firm must have control over in order to carry on its operations. The firm may obtain control over these necessary productive factors by means of a number of different types of contracts. We need to distinguish between two types of contracts by which a firm may obtain control over a productive factor: either by purchasing or renting (or hiring) the factor. We will take up the effects upon the survival of a firm of these different types of contracts.

The total value of the assets that a firm can purchase is limited by the value of the liabilities of the firm. The liabilities of a firm can be divided into two classes: the firm's net worth and its various debts. If the total assets of a firm are increased, then either one or both of the two classes of liabilities must be increased. The acquisition of title to productive factors which is financed by an increase in net worth has different repercussions upon the survival attributes of a firm than the acquisition of title to a productive factor which is financed by an increase in the firm's debt. When the purchase of a productive factor is financed by debt, the firm undertakes two responsibilities which it does not undertake when it finances the acquisition by owners' equity: first, to repay the debt when due, and secondly, to pay the contractual service charges upon the debt instrument.[6]

A productive factor owned by a firm has a normal life span in the firm's production process, during which its value is reduced to zero or to a scrap value. The total cost of the firm's output over this time period includes the cost of this productive factor. The purchase of such a productive factor can be financed by means of a debt instrument. Ignoring service charges upon the debt for the time being, if the dating of the debt is the same as the normal life of the factor, then as long as total revenue is equal to or greater than total costs the firm will be able to repay the debt when due and hence to survive. Debt financing involves a firm in a fixed cash commitment over the life of the debt, which in part determines the revenue that the firm must have at each output if it is to survive. If the term of the debt is longer than the normal life of the productive factor the survival conditions are easier, and if the term of the debt is shorter than the normal life of the productive factor the survival conditions are harder than if their periods are the same length.

Renting or hiring does seem, on the surface, to be different from the purchase of a productive factor. The cost of the productive factor does not enter as an asset in the balance sheet of the firm and therefore no liability entry is necessary. However the rental contract does provide for a series of cash outlays by the firm. Except perhaps for the dating of the cash outlays a rental contract is equivalent, relative to the survival conditions of the firm, to a purchase financed by a debt instrument: that is, it results in a series of dated, fixed money payments.[7]

Let us consider a rental contract which provides for a number of fixed monthly payments. This lease enters the firm's calculations as a fixed money cost. Alternatively, the firm might have financed the purchase of the asset by a debt instrument. If the lease is for the same time period as the normal life of the factor, then the series of cash payments will add up to the value of the productive factor plus the interest and risk charges. A firm renting a factor of production for the factor's lifetime is confronted with the need to meet cash payments equal to the value of the asset plus interest charges, just like a firm which finances a purchase by means of a loan. Aside from risk premiums, the two different types of contracts impose the same cash requirements upon the firm. The distinction is not between rental or ownership of a productive factor; the relevant distinction is between equity financing as against both the renting or the purchasing, by means of debt financing, of a factor. If the firm finances the purchase of a productive factor out of its net worth, the firm is required only to meet the cost of the productive factor over the lifetime of the factor (no interest charges), and this is required only to maintain the scale of operations of the firm, or the size of the owners' equity, not to assure the short run survival of the firm.

There are differences between debt-financed purchases and rental contracts. For assets with a long life and many alternative uses, the type of debt instrument a particular firm can use to finance its purchase may either be short relative to the productive life of the asset or expensive. It is desirable for the firm to rent such a productive factor. The alternative uses of the productive factor make it a desirable asset for others to own. Such factors would therefore tend to have a rental market. On the other hand, factors which have specialized production attributes are undesirable factors for others to own. The above considerations are relevant to the determination of whether or not a particular productive factor will be one for which a significant rental market will develop. They are not relevant to the effect of rental versus ownership upon the survival of a firm.

As has been pointed out earlier, any given level of output, with a given scale of plant, can be interpreted as requiring that a specific value and distribution of productive factors be under the control of the firm.[8] The net worth of the firm determines what value of these necessary productive factors can be controlled by means of equity financed purchase. Control over the remainder necessary for a particular level of operation can be obtained by either rental contracts or by debt financed purchase. These forms of acquiring control over production factors involve fixed money costs per period of time which are independent of output; therefore, they are equivalent as far as the survival conditions of a firm are concerned, and we do not have to differentiate among firms as between these two forms of obtaining control over productive factors. The significant thing is the addition of fixed money costs independent of output, to the firm's total costs schedule.

There are, of course, factors of production which are usually hired, such as labor and short term bank loans that so rapidly enter into the production process that short run variations in output will result in changes in amount hired.[9] The commitment, for example, in hiring wage labor may be no more than a guarantee of a day's pay. Such extremely short dated rental contracts do not involve a significant fixed money cost commitment independent of output as longer run rental contracts do. On the other hand, schemes such as a guaranteed annual wage may involve the firm in fixed money costs commitments which are large enough to affect appreciably its survival possibilities.

In what follows we will assume except for the rental of such short run variable factors as labor, that a firm will acquire control over factors of production whose value is greater than its owners' equity by means of debt financed purchase. It must be pointed out, however, that where we derive a rising supply curve for factors of production to a firm because of the deterioration of the firm's financing conditions, as it increases its debts with a fixed equity base, the firm may, if there is a well-organized rental market for such factors of production, be actually confronted by an infinitely elastic supply curve of the factor. The development of specialized financing schemes, such as the equipment trust technique used by railroads, may be considered as due to the desire of firms to circumvent the deterioration of terms of straightforward debt financing by a firm. In these circumstances, such specialized debt instrument financing techniques may really be more like rental than debt financed ownership arrangements.

The scale of the plant will uniquely determine the value of the entry under the heading plant and equipment in the balance sheet of the firm. The level of operation of this plant will determine the size of the cash balance, the value of other liquid assets, and the value of the various types of inventories such as stockpiles of raw materials, goods in the process of production, and finished goods, that firms will need for the most efficient operation at the output level under discussion. This necessary level of inventories may be seasonally variable, witness the stockpiling of iron ore in the summer by steel mills, and the accumulation of inventories of finished goods for which the demand is highly seasonal in the hands of both processors and retail outlets. Therefore, given the seasonal pattern of either the production process or of demand, the level of inventories is a function of the scale of output, and with a given plant the optimum size of inventories will vary uniquely with the output of the plant.

The volume of cash which a firm has in its balance sheet will also be a function of its level of output. We can consider Keynes' three attributes of a cash balance as being relevant to the firm as well as the household: firms can hold cash for transactions, precautionary and speculative purposes. Transactions cash in a firm's balance sheet is determined by the level of operations of the firm and the price level. The need for precautionary cash arises from the dated nature of the liabilities which a firm has and the dated nature of the payments it receives. An excess of cash over that amount strictly necessary for transaction purposes facilitates bridging the time gap between receipts and expenditures and allows for the easy overcoming of trifling unforeseen occurrences which plague every real world operation. The precautionary cash can be likened to the minimum stockpile of raw materials that a firm tries to keep in hand so that production would be uninterrupted by minor interruptions in the transport system. The speculative portion of cash balance is irrelevant for our analysis of the firm as a production unit. As has been mentioned earlier, a firm may be used as an investment trust or as a vehicle for financial operations by its owners. Speculative cash balances are of that nature. Decisions on the part of management to speculate by increasing the cash balance, however, are relevant to the analysis of investment decisions by the firm.

The financial assets other than cash in the firm's portfolio, such as accounts receivable and government bonds, arise from two sources. Accounts receivable depend upon the level of operation of the firm in the recent past. Government bonds are a way in which the firm may hold its speculative or precautionary cash. The volume of accounts receivable which is extant at any time in a firm's asset structure is determined by the level of operations of the firm, the payment habits of its customers, and the firm's policy in respect to the discounting or factoring of its accounts receivable. If accounts receivable are discounted, this means that cash is increased and a contingent liability is accepted by the firm. This offsetting asset and contingent liability entry is a means by which the firm, at a price, can change its customers' payment habits to accommodate its own ends. The significant element in the above is that a given level of output per period of time results in a given volume of cash plus accounts receivable plus inventories of finished goods being entered into the asset side of the balance sheet.

The structure of the assets of a firm is determined by the following variables: the scale of plant, the level of operation of the plant over a time period, the temporal nature of both the demand for the product and the production process, the risk attached to the assets and the requirements imposed by the liabilities (which accounts for precautionary cash and other assets) and the firm's behavior with respect to the discounting of its short term assets. Given the values of the above, the volume and distribution of a firm's assets are determined. Any asset total greater than that so determined (as everything is scaled to the size of plant, the plant and equipment items are constant) involves the use of the firm by its management for something more than the production process: for 'speculative' or 'investment' purposes.

The liability side of the balance sheet also has a structure, but this structure cannot be so intimately related to the production process of the firm. The liability structure is more a matter of choice by the firm's management. As each size of plant and level of operations of a plant requires that a given value of assets be under the control of the firm (abstracting from those precautionary asset holdings which are determined in part by the structure of the liability side of the balance sheet),the total liabilities for each level of operation of a scale of plant are determinate. The money market institutions determine the alternative financing techniques available to a firm, and the conditions under which the firm can use each financing technique so as to acquire sufficient resources to obtain the assets necessary for a given level of operations.

The attributes of liabilities which are relevant to the behavior of a firm are the money costs which are attached to each liability, the date upon which payment of each liability is due, and the penalties under the law for non-payment of obligations as their due date occurs. The various liability forms can be scaled in relation to these attributes. In respect to these attributes equity financing is the ideal: the necessary costs are zero – all equity returns being contingent; there is no repayment date; and the penalty for non-payment of dividends is not necessarily a loss of control of the firm by the owner.[10] The other liability forms entail both a cost and a due date. The cheaper the costs, the more desirable the liability; and the further away the due date, the more desirable the liability. The money market behaves so that, for a given scale of operations, the larger the proportion of the assets of a firm which are financed by debt instruments, the less desirable the liabilities of the firm are as financial institutions assets, and therefore the less desirable are the liabilities which the firm can have. A scale of operations is here defined as the level of operation of a given plant size, for financing terms are not independent of the assets which a firm possesses. For example, two firms in the same industry may produce the same output even though the size of the plant is different: the smaller plant would be utilized more intensively in such a case. For the production of this output, the firm operating the smaller plant would have a larger proportion of its total assets in the form of inventories and cash and accounts due than the firm operating the larger plant. The money markets normally operate so that the terms on which a firm can borrow if it has short term assets are superior to the terms on which a firm can borrow if it has 'plant and equipment' assets. Therefore, the balance sheet of the firm which uses a small plant intensively will exhibit a more preferred class of liabilities than the balance sheet of the firm with the larger scale plant.

Balance sheets of firms which exhibit debt instruments on the liability side differ in the cost per unit of debt and in the average term of the debt. Both a higher cost per unit of debt, and a shorter average term of debt impair the survival chances of a firm, all other things being the same. A balance sheet of a firm is said to deteriorate as (1) the average cost of liabilities which the firm has increases; (2) the average term of the liabilities decreases (if the distribution of the asset side of the balance sheet remains fixed); and as (3) the ratio of equity to total liabilities decreases. The significance of the qualification added in regard to the distribution of assets is that there are limits within which the balance sheet cannot be said to deteriorate appreciably if the firm adds short term assets, such as inventories and accounts receivable, and short term liabilities simultaneously. The three criteria of balance sheet deterioration are not independent, for the ratio of equity to total assets is an important determinant of both the cost per unit of debt and the term of the debt. Therefore, the fundamental attribute of the balance sheet is the ratio of equity to total assets.

Balance sheet deterioration affects a firm in two ways: it does mean that a larger total revenue is needed at each level of operations if the firm is to survive, and the shorter term of the debt does make the liquidity conditions more stringent. From the standpoint of survival of the firm, the ideal balance sheet is one in which equity is the only liability entry. Therefore, we shall use this type of balance sheet as a zero point in our analysis.

The total assets needed by a firm with a fixed plant vary with the level at which the plant is operating. A given value of total liabilities is sufficient to finance the acquisition of assets needed to operate a given plant at a particular level. The operation of such plant at any higher level entails an increase in both assets and liabilities. The assets that would be increased would be entered under the headings of cash, inventories, or accounts receivable. The liabilities added would usually be of the short term variety such as payables and borrowings. If the level of operation of the fixed plant falls below that level which can be financed without borrowings, goods in the process of production and receivables decrease; and as the liability side of the balance sheet has a minimum value equal to the owners' equity, the asset side would show either an increase in the cash balance above that level necessary for current production, an accumulation of the finished product or an increase in those investments which are superfluous to the production process. This means that a fixed owners' equity is sufficient to operate a given plant at some particular level. The financing of any level of operations of the plant greater than this involves borrowing, and the operation of the plant at any level lower than this involves an increase in speculative and investment assets in the firm's balance sheet.

The cost curves as drawn up to this point are transformations of the production function. They are the cost curves of plants. If competition is assumed in the factor market, a firm may use any volume of the factors of production at unchanging terms. The assumption made in drawing such plant cost curves is that there is no financing problem involved in changes in the level of output. However, if we assume a given dollar value to owners' equity, there is a relation between the scale of plant and the level of operation of that plant which the owners' equity can finance. In what follows we will for expository purposes assume that there is a one to one correspondence between a firm and a planning curve. In the real world a firm may own a number of plants, which may be associated with different production functions. This assumption is a deviation from the reality of a 'conglomerate' firm.

A conglomerate firm has the same incremental relation between finances and changes in any output that we will derive. The finance terms are associated with the balance sheet of the firm rather than with any attribute of the plant. However, for such a conglomerate firm we do not have a unique relation between the level of operations of a scale of a particular plant and the owners' equity. Our assumption is made to elucidate the effect of debt financing upon costs. For firms which are financing a portion of their operations by debt, the relation between a change in output and a change in costs is independent of the identification of owners' equity with the financing of a scale of operations of a given plant. Therefore, for the incremental analysis we are not subject to a significant error by our assumed planning curve-firm relation.

Let us consider the relation between cost curves and the balance sheet of a firm. Each short run marginal cost curve represents a fixed plant. This plant has a value. Each output per price period that this plant can produce requires a determinate flow of short term factors, which, in turn, implies a value to the short term assets in the firm's balance sheet. For each size of plant we can therefore state the maximum output that the fixed owners' equity of a firm can finance. With each plant, any larger output implies a recourse to debt financing. In Figure 6.1, the fixed equity of a firm can finance output q_1 by the plant $SRMC_1$, output q_2 by plant $SRMC_2$ and output q_3 by plant $SRMC_3$. The fixed equity is just sufficient to purchase plant $SRMC_4$, so that no working capital can be financed without borrowing. The owners' equity is not sufficient to purchase plants larger than $SRMC_4$.

There is a change in the cost conditions confronting a firm at that output which can be financed wholly by owners' equity.[11] At that point, in addition

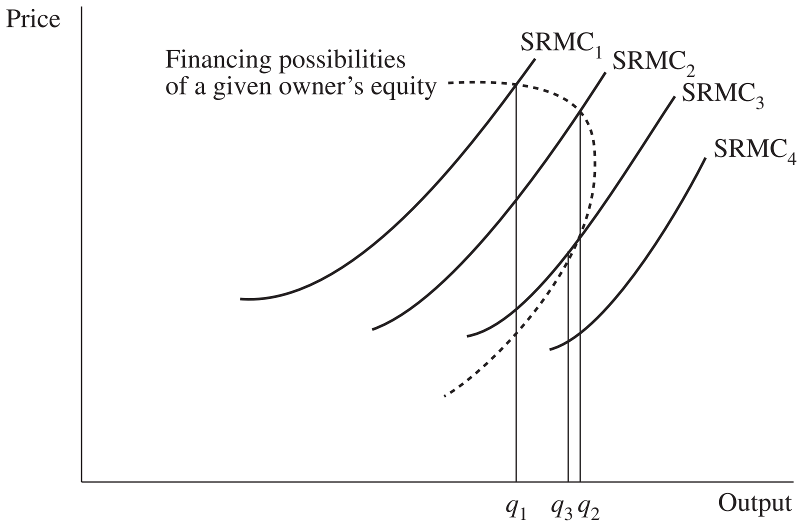


Figure 6.1

to the market costs of the factors of production, the firm has to add financing costs. The cost curves not only have to take the cost of factors of production into account; they have to take into account the cost of financing the outputs. With a given size of plant, the amount of working capital assets which a firm has to have in its balance sheet is a function of output and the prices of the variable factors. The appropriate total cost diagram not only includes the costs of the productive factor, it also has to include the costs of financing their acquisition. To the total variable cost diagram, for every level of output greater than that level which the firm can finance out of its own equity, the cost of financing that level of operation has to be added. If we assume a constant interest rate on such financing, and if the amount of such financing is a linear function of the level of output greater than that which can be financed by the firm's own resources, the total cost curve for this element of cost is a straight line with a slope equal to the interest rate times the financing necessary per unit of output. If outputs smaller than that amount which can be financed by internal funds are to be produced, the firm may earn a return on these superfluous funds. As the return which a firm can earn by investing such superfluous funds is generally lower than the cost of the funds it borrows, the cost of financing curve will have different slopes for outputs lower than the output which can be financed internally than it has for outputs which require borrowed funds. Hence, even if we take the opportunity costs of owners' equity to be greater than

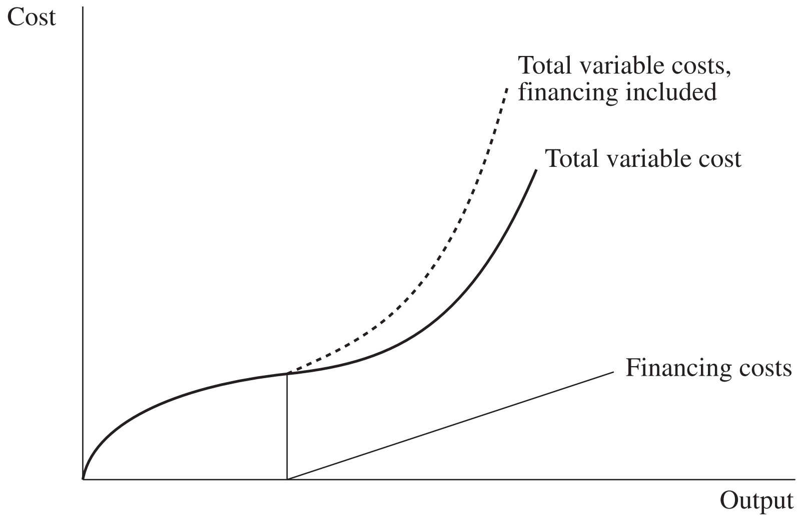


Figure 6.2

zero, the total cost curve has a sharp change in slope at the largest output that owners' equity can finance.

The effect of the addition of financing costs upon the marginal and average cost curves is to introduce discontinuities in these curves. The average cost curve has a sharp change in slope, the marginal cost curve a discontinuity at the output which can be financed internally. If the rate at which output can be financed changes continuously, there is only one such discontinuity in the marginal and average cost curves, whereas if these costs change discontinuously, there is a whole series of such discontinuities. In Figure 6.3, output q_1 is the output which can be financed internally, whereas any output greater than that involves the use of borrowed funds. AC_1 and MC_1 are average and marginal cost curves ignoring such financing charges; AC_2 and MC_2 are these curves including the financing charges.

If Figure 6.3 is for a firm in a competitive industry, any market price between p_2 and p_1 will result in the same output. At any price greater than p_2 the firm will produce less than it would have done if there were no financing costs. However, the profit earned on owners' equity will be greater than if the firm restricted itself to the level of operations that could be financed internally.

In the short run, with no superfluous assets in the original balance sheet, the survival condition is that total revenue is greater than the total variable cost associated with the particular plant (the total variable cost curve in this

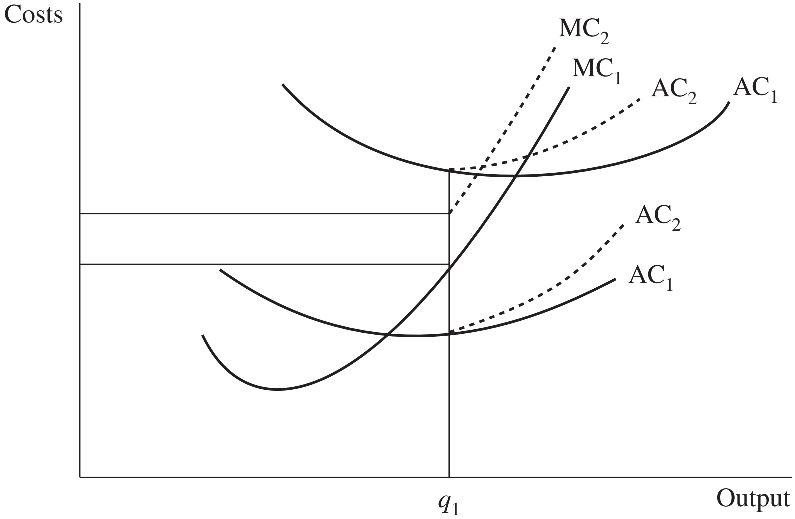


Figure 6.3

case does not even have to include the user cost of capital). In the long run (in this case the existence of superfluous assets is irrelevant), the survival condition is that total revenue yields not less than a zero rate of return on owners' investment. In the long run the firm can in this case maintain its plant intact, pay any debt charges, and, assuming the dating of its borrowing is appropriate, repay its debt when due and maintain its size by borrowing again. The zero rate of return to owners' investment has to be underlined. This means that the long run survival conditions depend upon the balance sheet structure of firms. A firm which has no debt – for example, wholly owned plant and working capital – has easier survival conditions than a firm which debt finances a portion of its plant or working capital. The zero debt firm can survive with a zero return on total plant and working capital, whereas a firm whose assets are financed in part by debt instruments can survive only by meeting its contractual interest payments. To survive, such a firm can earn a zero return only on that portion of its asset structure financed by owners' equity. Again it must be emphasized that a plant may satisfy the survival conditions and still may be earning so little that liquidation is desirable.

Therefore, for a particular plant, the average cost curve, which is identified as the survival average cost curve, must take into account the balance sheet position of the firm. This average cost curve asserts that the plant associated with the given marginal cost curve, and with the given structure of debt in its balance sheet, can survive over all time in the future only if the per unit revenue associated with any output is always greater than the per unit cost of that output which is shown by the survival average cost curve. As soon as the balance sheet changes, with the same plant, the survival average cost curve changes. If a firm that makes profits above dividends retires a portion of its debt, its survival average cost curve falls. The minimum survival average cost curve is the average cost curve associated with a zero return on the total investment necessary to operate a plant at a given level. It is the effective survival average cost curve only for a firm which, when operating that plant at that level, has zero debts. As a firm's debts approach zero, the survival average cost curve approaches this zero return on total assets average cost curve.

We can also draw long run cost curves which represent zero return on the total capital necessary for the operation of each size of plant at some particular level. Insofar as superfluous assets can always be held in the form of cash, and there are no costs attached to the holding of cash, this curve can be considered as the zero return cost curve for a firm which has an infinite equity. That is, the zero return curve is the same for all plants and levels of operation which can be financed by owners' equity. For each size plant we can also draw the short run survival conditions – that is, the average variable cost curve – excluding the depreciation cost on the plant and equipment. These two curves are the same for every firm in the industry with a given size plant. They are independent of balance sheet structure, being transformations of the production function. The larger the proportion of assets, for a given size plant at a given level of operation, that is financed by borrowed funds, the higher the survival average total cost curve; also, the less advantageous the terms upon which a particular firm can borrow, the higher the survival average total cost curve. We therefore find that in an industry where firms have identical size plants the survival conditions will vary for the different firms.

The addition of the balance sheet structure to a firm's attributes enables us to handle the effect of a deterioration in demand conditions for a product upon the forced exit of firms from an industry, and it enables us to give a more concrete meaning to the idea of a marginal firm in a competitive industry. Let us take a competitive industry, and assume that a shift of the market demand curve to the left occurs. This appears to each firm as a fall in its horizontal particular demand curve. If the firms in the industry have different balance sheet structures, the survival conditions will reflect these varying balance sheet structures. The total variable cost curves and the average and marginal cost curves in Figures 6.4 and 6.5 are the survival curves for firms with identical plants but having different balance sheets. One firm can finance only the level of operation given by

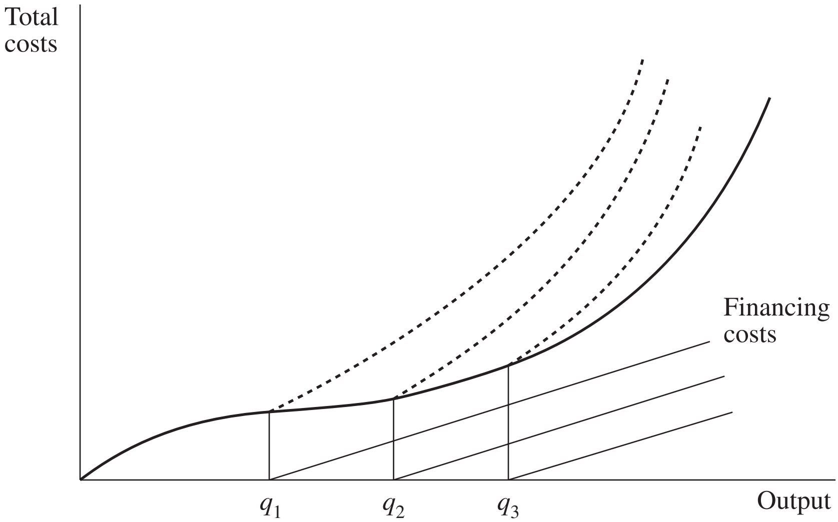


Figure 6.4

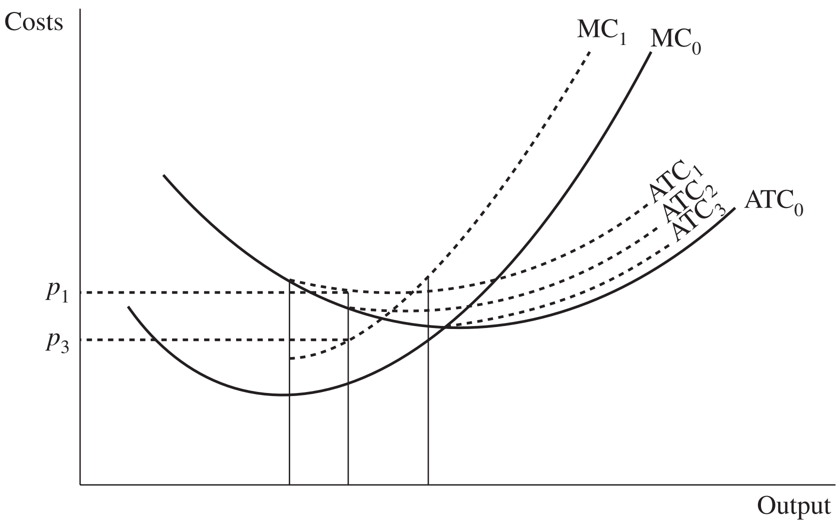


Figure 6.5

q_1 without borrowing; another firm, q_2; and a third, q_3. For all levels of output greater than q_1, the survival average total cost curve for the firm 1 lies above that for firms 2 and 3. If we assume that regardless of the amount of borrowing, the interest rate and the borrowing per unit of output remain the same, the marginal cost curves for all firms, once they begin borrowing, are the same as they are for that level of operations which can be financed out of internal funds.[12]

If r and β are the same for all firms and are also independent of the amount of borrowing a firm does, the marginal cost curves for all firms which borrow to finance a part of their operations are the same. If r is a function of the amount of borrowing a firm does such that $dr/d\beta q > 0$, so that the interest rate increases with the amount of borrowing, then the marginal cost curves for all firms which borrow to finance a part of their operations are not the same – the firm which borrows a smaller proportion will have a lower marginal cost for each output – or it can produce a larger output at a given marginal cost.[13]

A firm whose balance sheet has no debts, for example the firm whose survival curve is marked ATC_0 in Figure 6.5, could stand a fall in the market price to p_3 and still survive, whereas a firm whose balance sheet is such that it has the survival average cost curve marked ATC_1 as its survival curve, could not survive a fall in the price of the product below p_1. If we assume identical plant and equally efficient management, the survival conditions reflect the financial structure of the firms. A marginal firm therefore can be defined as a firm whose survival conditions are such that any fall in the market price of the product will cause an end to the firm's existence (or a decrease in the 'owners' equity').

The survival cost curves for a particular firm will shift with the occurrence of many events of business life which are to a firm fortuitous. A firm which finances a large part of its activities by means of bank loans will be adversely affected by a tightening of credit which raises interest rates or shortens the term of available financing. This will mean that a given level of activity will now result in a higher average cost curve being the survival average cost curve. Deterioration of the quality of its assets, such as occurs when a firm has a deposit in a bank which fails, also results in an upward shift in the survival curves for particular firms. Business history is full of otherwise healthy firms which did not survive as the result of such fortuitous occurrences to the particular firms (for example, events independent of its management or of the behavior of the market for its product).[14] A change in the price of its product will result in losses or gains by a firm. Such gains or losses will either raise or lower the survival average cost curve for a particular plant. A firm which has sizeable inventories will suffer a sharp rise in its survival average cost curve if the price of its inventories falls. The impact of a 'downturn' in business activities upon the failure of business firms operates through a rise in the survival average cost curve of those firms which finance a large part of their operations by means of debt instruments.

The long run survival average cost curve for a firm has to include the effects of the method by which plants of varying size would have to be financed. Let us examine this more closely, with the aid of figures.

Let us assume that a given firm can finance the operations of a fixed plant (represented by one short run marginal cost curve 1) at a level q_1 by its own funds (equity financing). Let us draw the zero return long run average and marginal cost curve for the production function of this firm. These zero return cost curves are based upon the expansion path in which the internal rate of return is zero. Let us assume that this firm plans its plant on the basis of a rate of return $\bar{r} > 0$ so that we can also draw the long run average and marginal cost curves for this rate.[15] The long run zero return, and the long run planning return $\bar{r}$ curves are the basic relations upon which is based an analysis of a firm's behavior when it has a finite equity.

Under the above conditions, if the firm had an infinitely elastic supply of equity financing, the optimum plant for the firm would be determined by the intersection of the marginal revenue curve and the long run marginal cost curve based upon the rate $\bar{r}$. In the above figure, the optimum plant for the firm is given by $SRMC_2$. The output it would produce is q_2, the price of the product p_2 . However, the owners' equity available to the firm, which

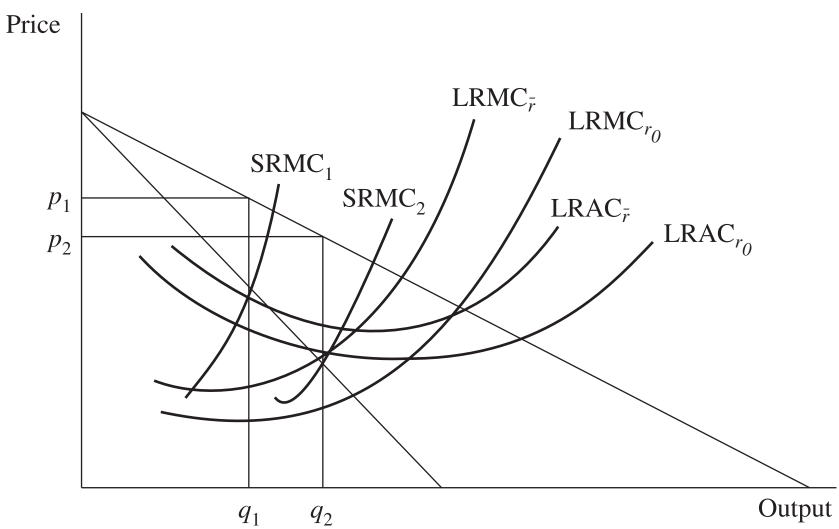


Figure 6.6

enables it to finance output q_1 by plant 1, is insufficient to finance output of q_2 by plant 2. The optimum plant for a firm to build must take into account the terms available to the firm to finance scales of operations greater than the output which its own resources can finance.

In order to build and operate a plant greater than size 1, the firm has to resort to 'debt' financing. For firms that equity finance, the short run average cost curve for plant 2, which is tangent to the long run average cost curve for zero return on total capital, represents zero returns on owners' equity. For firms which debt finance a portion of their assets, a higher short run average cost curve represents zero returns on owners' equity. This survival short run average cost curve indicates the prices and quantities of the output of the firm produced by the given plant which enables the firm to pay all its financing costs (including the repayment of the debt when due) without decreasing the owners' equity. Another higher SRAC curve defines the set of prices and quantities which enables the firm to pay all its financing charges and earn $\bar{r}$ on the owners' equity. The financing charges are greater than zero. However, they may be greater than, equal to, or lower than $\bar{r}$. If they are less than $\bar{r}$, then the SRAC curve which enables the firm to pay $\bar{r}$ on owners' equity lies below the LRAC curve for $\bar{r}$.[16] If the financing charges are greater than $\bar{r}$, then the SRAC curve which enables the firm to pay $\bar{r}$ on owners' equity lies above the LRAC curve for $\bar{r}$.

Given the owners' equity and the financing conditions confronting the firm, we can determine the short run average cost curves for each scale of plant which will yield zero and $\bar{r}$ on owners' equity.[17] The envelope of these zero and $\bar{r}$ returns on owners' equity cost curves determines a long run zero and $\bar{r}$ return average cost curve. This enables us to determine long run marginal cost curves for the firm that, together with the marginal revenue curve, determine the optimum plant.

In the above figures, it is assumed that output $\bar{q}$ can be financed by the firm's equity. Outputs greater than $\bar{q}$ have to be financed by debt. The long run total cost curve that yields $\bar{r}$ on owners' equity ($LRTC'_{\bar{r}}$) and the long run total cost curve that yields zero on owners' equity are parallel for outputs greater than $\bar{q}$. This means that there is a unique long run marginal cost curve (LRMC") for outputs greater than $\bar{q}$. This gives us the important result that the long run marginal cost curve which determines the optimum plant for firms that finance their operations by debt is independent of the planning rate that the firm may use. The above is based upon a 'heroic assumption' that the capital output ratio is the same no matter what the value of $\bar{r}$ and for short term and long term factors; that there is strict complementarity between such factors. If the capital output ratio is greater for long term than for short term factors, the output $\bar{q}$ would be produced

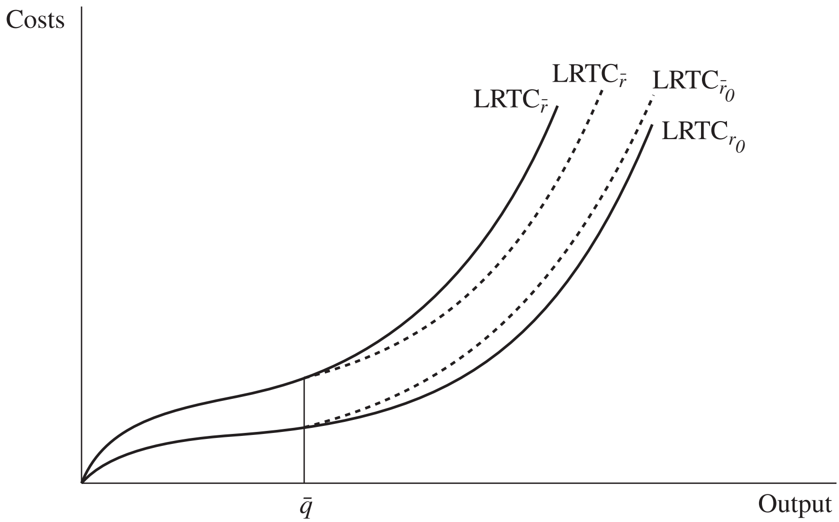


Figure 6.7

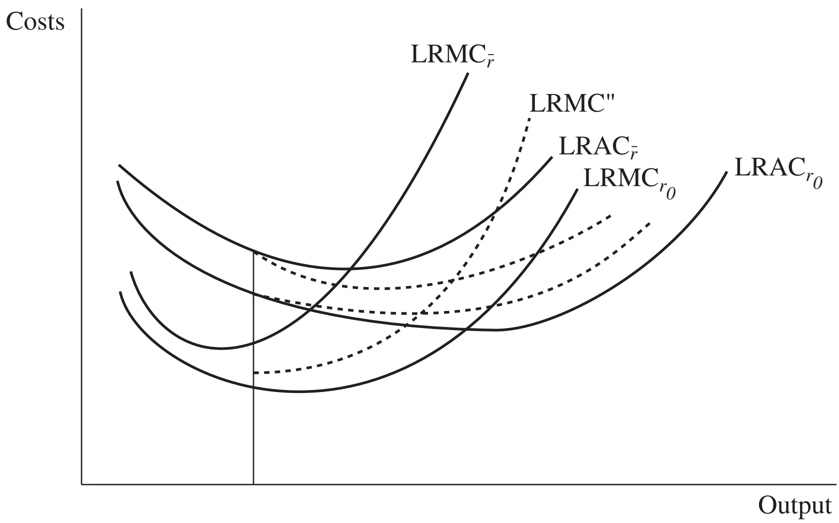


Figure 6.8

with a smaller quantity of capital, the higher the internal rate used in planning, and a given owners' equity would be able to finance a larger output. In this case the LRMC" would become the effective marginal cost curve at a greater output. The above also assumes that the same financing terms apply to long term and short term capital.

The difference between the total cost, zero returns and the total cost $\bar{r}$ returns for any output greater than that which can be financed by owners' equity is a constant, $\bar{r}$, times the owners' equity. Therefore, the $LRAC_{\bar{r}}$ approaches $LRAC_0$ as output increases beyond that which can be equity financed. Hence the difference between the $\bar{r}$ yield price and the zero yield price decreases with output. Therefore, given $\bar{r}$, the larger the debt financed output planned for, the smaller the fall in price which would result in a firm changing from making profits to making losses. The size of the fall in price which could be withstood without such losses can be considered as a measure of risk. The greater the fraction of output that is financed by debt, the greater the risk of losses.

We now have cost curves which take into account financing conditions. The financing constraint may be looked upon as one possible explanation for the finite size of firms where the production function does not exhibit decreasing returns to scale.[18] If financing conditions available to a firm deteriorate as debt increases, the long run marginal cost curve which reflects these financing conditions would be a rising curve, even though the long run marginal cost curves without these financial constraints taken into account are not rising. This device enables us to reconcile the existence of prosperous plants of different size in an industry where the same production technique is available to all firms. A firm in which the equity investment is large can build a larger plant and survive earning zero returns on that plant more easily than can a firm in which the equity investment is small.

We can investigate the effects upon plant size of different behavior patterns in the money market. If, as is often claimed, the large firm can tap financing sources not available to the small firm so that financing charges per unit of debt are smaller to the large firm, the result would be that with a production function that has constant returns to scale a given percentage of debt financing to equity financing will result in a smaller rise in per unit costs for the large firm than for the small firm. Therefore, the long run marginal cost curve for the large firm will lie below the long run marginal cost curve for the small firm, adjusted for the scale of owners' equity. On the other hand, if the money market is such that lenders weight the dollar value of loans more heavily than the relation of loans to equity, the larger firm's long run survival marginal cost curve will lie above the long run survival marginal cost curve for the small firm, again adjusted for the scale of owners' equity. The behavior of the money market, therefore, may be a relevant factor in determining the industrial organization of an economy. The money market may be neutral as between different scales of firms; it may support the growth of a few large firms; or it may abet the development of many small firms.

So far we have represented the survival conditions as a transformation of the original set of cost curves to take into account the costs of financing. This aspect deals with an objective market phenomenon, and it could just as well be interpreted in terms of a rising supply curve, to the particular firm, of a factor of production, finance. That is, with a given owners' equity, the greater the amount that a firm borrows, the higher the cost of borrowing; and the cost curves should take this change in the price of the factors of production into account. The money market operates so that each firm is confronted with a 'monopsonistic' supply curve of at least one factor of production – finance.

However, it is not only the quantity of debt financing and the incremental cost associated with it that is relevant to the analysis of a firm's continued existence. The survival conditions of a firm also depend upon the requirements imposed upon the firm by the nature of its liabilities. The relation between the amounts of particular types of liabilities and assets in a firm's balance sheet determines not only the cost of financing but also a set of constraints upon the behavior of the firm which are derived from the effect of different balance sheet structures upon the survival chances of a firm.

We can operate upon the family of average cost curves so as to take into account the subjective risk and uncertainty associated by the entrepreneur with different balance sheet structures. This subjective evaluation of balance sheet structure is a particular version of 'The Principle of Increasing Risk' associated with Kalecki.[19]

Let us assume that a particular output with a given size of plant can be financed by owners' equity. A larger plant can be financed only by resorting to borrowed funds. The cost curves as drawn earlier took into account the objective costs associated with the best method available to the firm of financing various outputs. In addition to this objective cost of debt financing, each dilution of equity which takes place as increasing proportions of the output are financed by debt instruments involves an increasing risk to the entrepreneur that he may lose his own investment, that he may lose his valuable organization. We might say that the increasing objective cost of financing as equity is diluted by borrowed funds represents the lender's risk, and what we wish to add now is the borrower's risk of loss of his valuable organization as his balance sheet deteriorates. This can be represented as a long run marginal and average cost curve, which lie above the cost curves that took financing charges into account.

Although this new long run marginal cost curve is a representation of the state of mind of the firm, we can construct it. Just as we drew zero and planning rate return on owners' equity long run cost curves, we can draw a long run average cost curve for each positive rate of return on owners' equity. These curves have to be adjusted as compared to the zero return on owners' equity for scales of output lower than the largest output which can be wholly financed by equity. For the zero return curve, superfluous equity caused no trouble. For outputs smaller than the maximum output which can be financed wholly by owners' equity, the fixed return on owners' equity curve will be drawn so as to yield this return on the minimum investment necessary to finance this scale of operations. That is, for outputs smaller than the largest output which can be financed by owners' equity, the fixed return average cost curves yield the return on the investment necessary and a zero return on any superfluous assets.[20]

In Figure 6.9 the firm can finance an output of q_1 by means of its equity; any larger output involves the firm in debt financing. The dotted lines drawn in the figure are the zero return long run average and marginal cost curves assuming that the operations are financed internally. For all outputs greater than q_1, the internally financed zero return average cost curve lies below the zero return average cost curve where a portion of the operations is financed by debt. The difference between the minimum price which a firm can stand and still survive with or without borrowing is a measure of the risk that the borrower carries. We assume that for each such increase in risk there is a

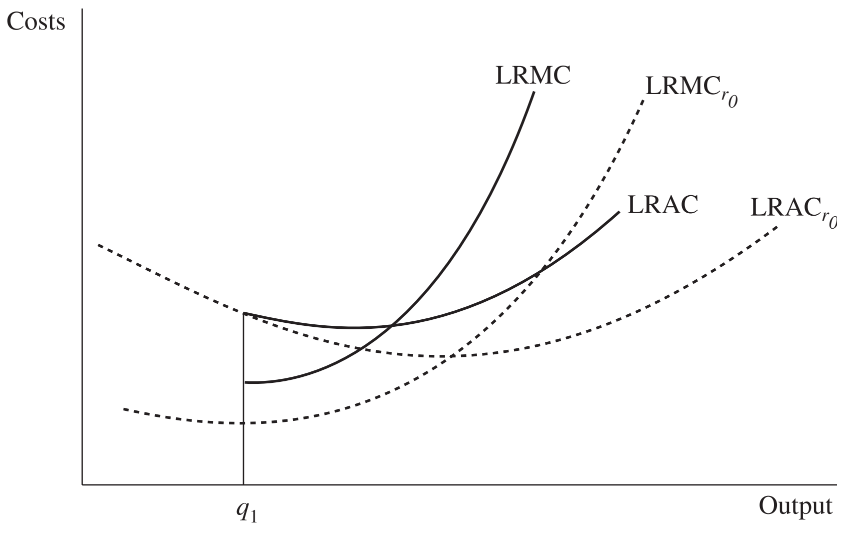


Figure 6.9

minimum gain which the entrepreneur will have to expect in order to take this risk, this minimum gain increasing with the risk. This minimum gain can be transformed into a positive return on the owners' equity. Therefore, the planning long run average cost curve which allows for risk involves, for each balance sheet which has debt in it, a return greater than zero, and this return greater than zero increases with the deterioration of the firm's balance sheet.

For each output greater than q_1, which can be financed internally, the firm in planning considers a larger return than $\bar{r}$ on owners' equity as compensation for the additional risk it carries by borrowing. A new long run average cost curve therefore can be drawn to replace the $\bar{r}$ return on owners' equity curve as the planning curve: this curve yields an increasing return on owners' equity as debt financing increases.

In Figure 6.10 the management is assumed to consider a return of r_1 on owners' equity as equivalent to $\bar{r}$ when it finances $q_2 - q_1$ by borrowing. A return of r_2 on owners' equity is regarded as equivalent to a return of $\bar{r}$ when it finances $q_3 - q_1$ by borrowing. This new average cost curve (LRAC[dr/dq]) based upon rising rates of return determines a marginal cost curve. This marginal cost curve (LRMC[dr/dq]) measures the incremental risk which the firm undertakes in addition to the incremental production and financing costs. The optimum scale of plant and the optimum proportion of debt financing for a firm is given by the intersection of this marginal cost curve and the marginal revenue curve.[21]

It is obvious that as the borrowers' risk premium increases, the slope

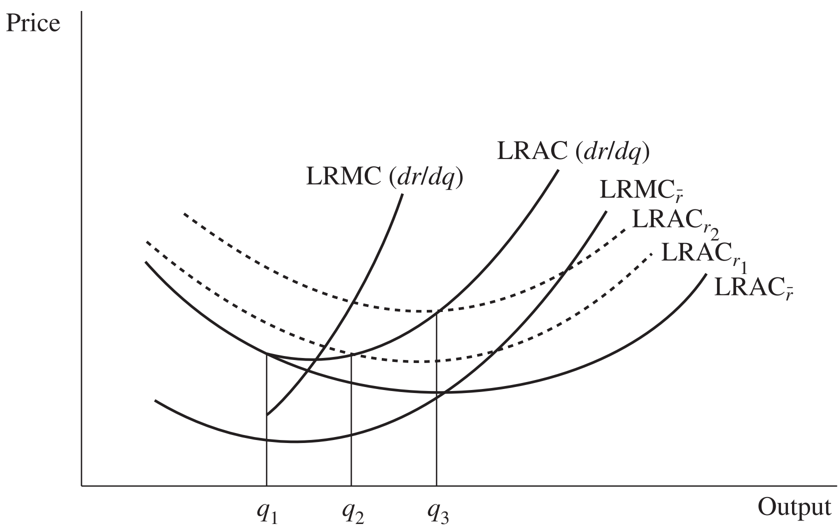


Figure 6.10

of the effective marginal cost curve increases more rapidly than the change in the objective financing rate indicates. If the risk premium increases rapidly, the planning marginal cost curve approaches a vertical line. In this case, no matter how much the demand for the product of the firm increases, the firm will not increase its borrowings, and therefore, no investment by the firm will result from a rise in demand. In such a situation, a necessary prerequisite for the accelerator to operate for the firm is that the equity interest increase. The risk factor implicit in financing investment by borrowing can act to break an accelerator expansion.

The optimum scale of plant and level of operations of a scale of plant by a firm depends upon two factors in addition to those usually considered in price theory. These are: (1) the charges which a firm has to pay in order to finance outputs greater than its equity can finance; and (2) the evaluation by the firm of the risks that it bears as a result of financing part of its operations by debt instruments. The impact upon firms of developments in financial markets centers around these two elements.

To a firm that is financing a part of its operations by debt, a rise in the market rate of interest raises (upon refunding) the survival short run average cost curve. Prior to the rise in interest rates, if the industry was in equilibrium, the expected returns on owners' equity were just sufficient to compensate the firm for the risk it was running due to its balance sheet structure. As a result of the rise in interest rates, the gap between the survival price and the market price is reduced. The marginal cost of the output is now greater than the price of the output. The firm has too great a debt load for its expected earnings and the entrepreneur can:

1. reduce the scale of operations (and, in time, the scale of its plant), thereby decreasing the firm's debt load. This will lower the required difference between the market price and the survival price;
2. increase the equity base, which also lowers the survival average cost curve, and also decreases the required difference between the market price and the survival price.

In Figure 6.11 the firm has an equity base sufficient to finance the production of q_1. At a market rate of interest r_1 and a price of the product p_1 the firm builds the plant given by SRMC_1. A rise in the interest rate to $r_2 > r_1$ raises both the $\text{LRAC}_{q_1r_1}$ and the $\text{LRMC}_{q_1r_1}$. This also raises the LRAC'_r and LRMC', which take into account the risk borne by the firm due to debt financing. Therefore, with a given equity base the optimum scale plant is reduced. If the firm still desires to retain the scale of plant given by SRMC_1, an increase in the equity base to the amount that is sufficient to finance q_2 is called for.

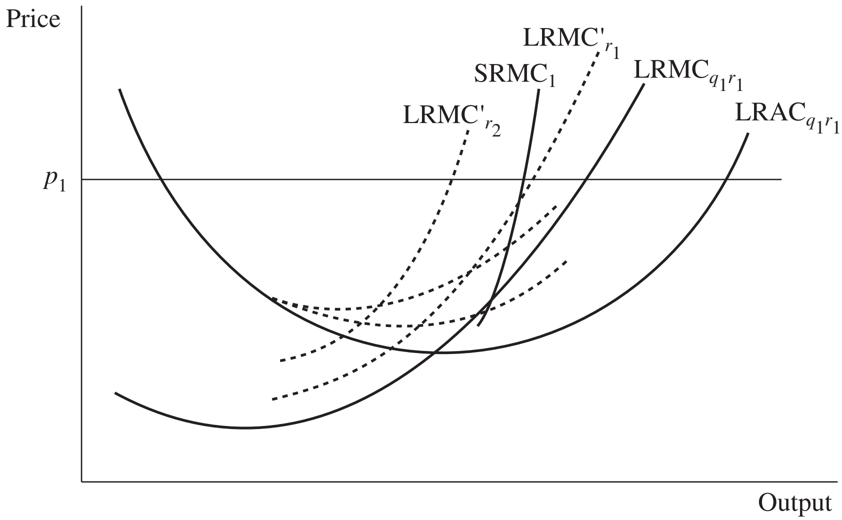


Figure 6.11

The impact of a stiffening of financial market conditions is to raise the desired ratio of equity to total assets. In a later chapter we will systematically consider the relation between shifts in firms' demand curves and induced investment as it is affected by changes in the financial markets.

With a given plant, and a given debt structure, the firm's survival short run average cost curve is determined. This determines the minimum price of the product consistent with the survival of the firm. The debt structure also imposes a series of due dates upon the firm. A fall in the price of the product below this minimum price means that the firm is not earning enough to repay the debt when due and to pay the service charges on the debt. A firm in such a case may either sell assets or borrow more. In either case, the survival average cost curve rises.

A reaction of an existing firm to the constraint imposed by its debt structure is to attempt to 'pay off' debt. An entrepreneur, whenever the price of its product is greater than the price which yields zero return on its own investment, may use this income to:

(a) pay dividends;
(b) increase his equity base and use this larger equity base to finance plant expansion;
(c) increase his equity base and use the retained earnings to repay debt. What combination of (a), (b), or (c) the entrepreneur uses depends

> upon his forecast as to the expected market price of his product. If his forecast is that market conditions are favorable, the optimum use of earnings is to expand plant; if his forecast is that market conditions are unfavorable, the optimum use of earnings is to lower debt.

The use of retained earnings as a basis for plant expansion means that the savings which are taking place are being used as an equity base. If the ratio of borrowings to equity is λ, an expansion of investment by $(1 + \lambda)$ savings is possible without a deterioration in the ratio between equity and debt financed capital. If a firm uses retained earnings to pay off debt, then, for the firm, the ratio of equity to debt increases. The savings of the firm have been channeled into 'institutions' which own debt. If the institutions which finance investment by owning equities and which finance investment by owning debt are strictly compartmented, then the retained earnings of the firm have increased the debt financing ability of the economy. If lenders look at the ratio of equity to debt in the balance sheet of firms as the standard to be used in their lending activities, such growth in debt financing ability, unless there is an increase in equity financing from some other source, would imply that a deterioration of the ratio of equity to debt financing for other firms is necessary if the earnings of the firms that expand equity to reduce debt are to be transformed into investment demand.

A period of expansion of firms' plant by means of debt financing of investment can be followed by a period in which firms are saving in order to repay debt. Such savings may result in a 'relative shortage' of equity financing or of 'venture capital', especially if the debt financing institutions are limited by law or custom in their portfolios. The rationale for such behavior of firms is that repaying debt lowers the product prices which are consistent with the survival of business firms. As business firms may be considered to be as interested in surviving as in maximizing profits, any generalized adverse forecasts can lead to a deficiency in equity financed investment and therefore to a fall in aggregate investment.

APPENDIX 6.1: AN ANALYSIS OF FINANCING

Assume that a production function uses only two inputs, labor and capital. Substitution between labor and capital is possible. Financing costs, and equity limitations, are assumed to be relevant only to the amount of capital used. Assume that all capital has a life of n production periods. The firm's survival expansion path is based upon a zero internal rate upon owned

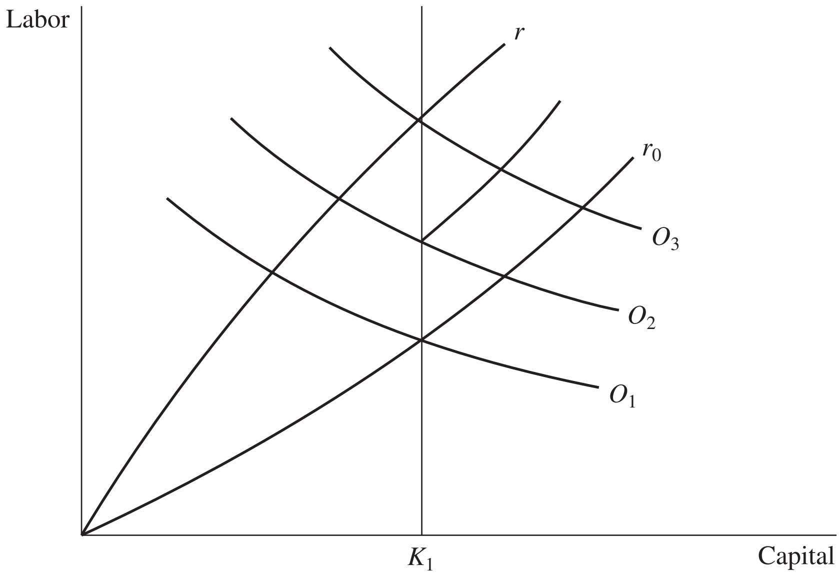


Figure 6.A.1

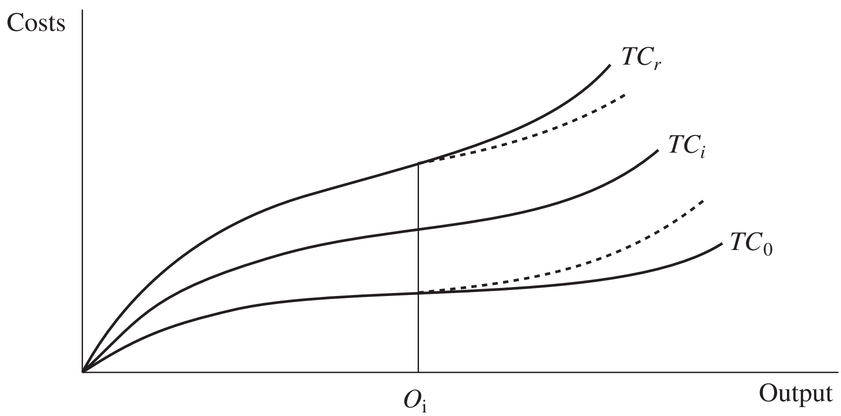


Figure 6.A.2

capital. The firm's planning expansion path is based on an $\bar{r} > 0$ internal rate (the price of the factor, capital in the planning expansion path is $(1 + \bar{r})$, the price of capital in the survival expansion path). The firm's equity is sufficient to finance a plant of size K_1.

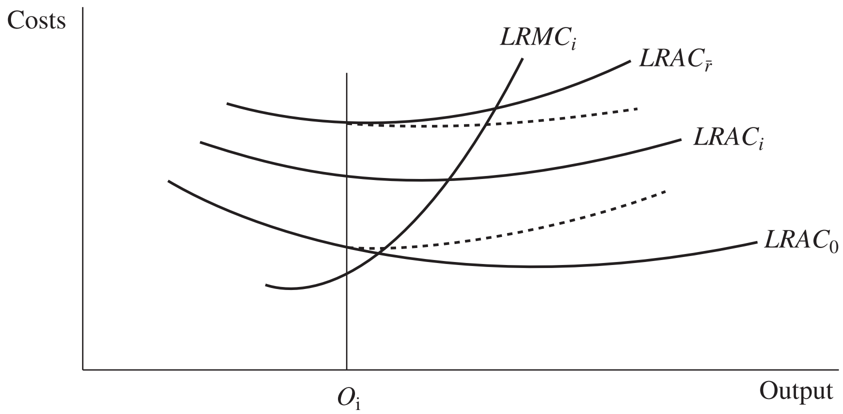


Figure 6.A.3

The financing rate is i_1 and $0 < i_1 < \bar{r}$. The survival long run total cost curve will be based upon a zero cost of capital for K_1 of capital, and an r_1 rate for any capital greater than that. This will result in a discontinuity in the total cost curve, leading to a jump in the long run marginal cost curve. The average cost curve, for a zero return on owners' equity, will, as output increases (ratio of borrowed to own capital rises), approach the long run average cost curve for i_1 interest rate.

Similarly, if $\bar{r} > i_1$, then the planning LRTC will be based upon an $\bar{r}$ rate of return for K_1 of capital, and i_1 for any capital greater than at O. The long run marginal cost curve will fall, becoming the LRMC curve appropriate to i_1, and the long run average cost curve will approach LRAC$_i$, as output increases. The result is that the difference between the price of the product which yields $\bar{r}$ and the price of the product which yields zero upon equity decreases as the ratio of debt financed to equity financed output increases.

We can assume that the difference between the survival (zero return) price and the expected price is a measure of the firm's risk of failure. A firm may protect itself against this risk by setting a minimum difference between the survival price and the 'expected' price of the product. This is equivalent to moving to higher internal rates on owned capital as the ratio of debt financed to equity financed capital increases. The resultant average cost curve yields a long run marginal cost curve which rises more rapidly than LRMC$_i$ as debt increases.

If the borrowing rate is above $\bar{r}$, then the LRAC$_{\bar{r}}$ approaches the higher LRAC$_i$ as output increases, as does the survival LRAC curve. Again an attempt to keep the distance between the expected LRAC and the survival

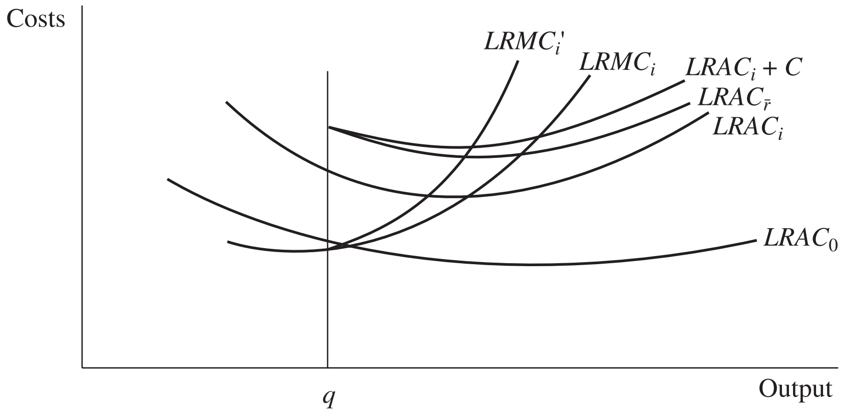


Figure 6.A.4

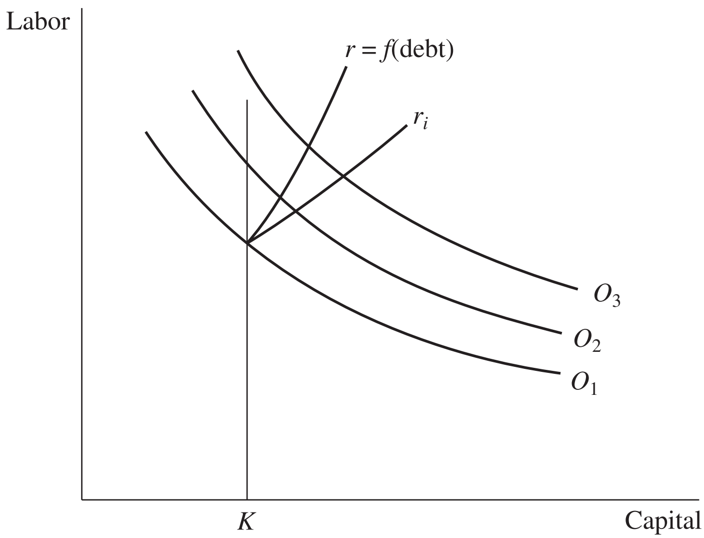


Figure 6.A.5

LRAC at some minimum value results in a rise in the effective planning marginal cost curve.

In addition to the above, the interest rate applicable to loans may rise as the amount borrowed with fixed equity base increases. In such a case, the LRTC curve rises more rapidly than the constant interest rate LRTC rises.

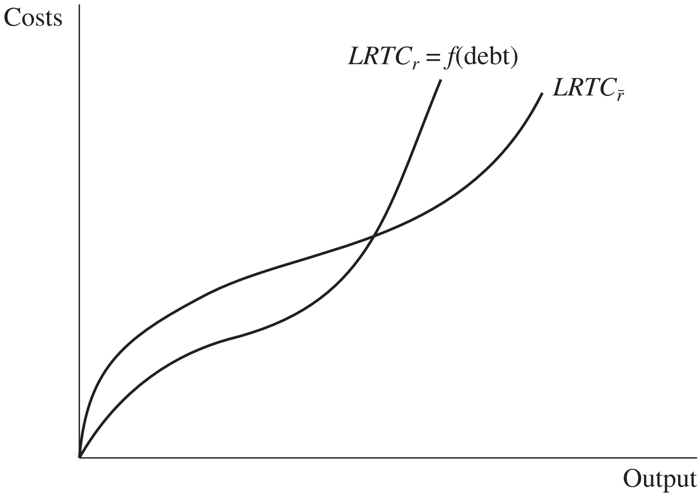


Figure 6.A.6

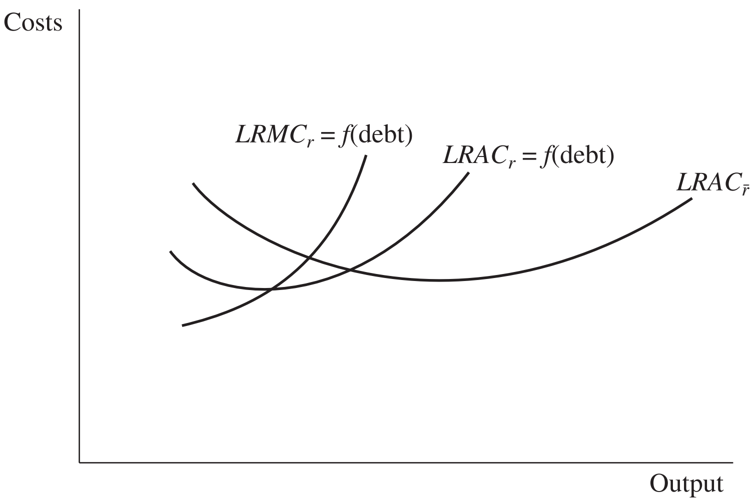


Figure 6.A.7

As the ratio of debt financed to owned plant increases, the long run average cost curve which yields $\bar{r}$ on owners' equity and the long run average cost curve which yields zero on owners' equity asymptotically approach the $LRAC_{r\,=\,f(q)}$. If the firm desires to keep the difference between the price which yields $\bar{r}$ and the price which yields losses constant, the effective long run marginal cost curve again rises more rapidly than the objective cost curves indicate.

In addition, as the ratio of debt to equity increases the firm may feel constrained to increase the difference between the planning return and the zero return price. This will result in an even more rapid rise in the effective long run marginal cost curve.

In all cases, the effect of the risk element due to borrowing is that the planning curve rises more steeply than it otherwise would. The effect of the risk element therefore is to decrease the size of the optimum plant as compared to the optimum plant in a riskless world, and also, by increasing the slope of the planning curve, to decrease the amount of investment induced by a given rise in demand.

NOTES

1. Bankruptcy is a striking attribute of business cycles. A theory of business cycles must be consistent with the bankruptcy observations. By considering the survival characteristics of firms, we are in a position to relate bankruptcy to the cyclical pattern in product and financial markets.
2. It may be technically possible that the output capacity of a firm can be maintained for a period even though the firm is not reinvesting its depreciation allowances. This is due to either the 'errors' of the depreciation technique as far as economic analysis is concerned or to a 'discontinuity' in productive capacity, and the allowances are a technique for averaging the discontinuity in productive capacity. This chapter in many ways borders, if not encroaches upon, accounting problems. The usages of the professional accountant are no doubt at variance with my use of terms. What I am aiming at is a theory of firm behavior with respect to financial constraints, and the exact way in which these financial constraints are estimated is not particularly significant.
3. When a firm is disinvesting, symmetric changes in assets and liabilities occur.
4. The analysis is not long run if it is to be relevant to business cycles. However the decrease in owners' equity takes place when a business firm with superfluous assets makes losses, and the firm technically survives although for a period of time it does not satisfy the survival conditions.
5. If $x = \varphi(a, b, \ldots n)$ is a firm's production function, and factors $h \ldots n$ are fixed at some level $\bar{h} \ldots \bar{n}$ ($\bar{h}$ through $\bar{n}$ represent the fixed plant) so that short run variations in output can only be achieved by varying, say, factors $a \ldots g$, then for the firm to produce an output $X_\alpha = \varphi_\alpha(a_\alpha, \ldots g_\alpha, \bar{h} \ldots \bar{n})$ per time period, it requires control over $a_\alpha, b_\alpha, \cdots g_\alpha$ of the short run variable inputs per time period in addition to the $\bar{h} \ldots \bar{n}$ of the fixed inputs.
6. In general we will assume that debt instruments call for the payment of periodic service charges and lump sum repayment. The various alternative debt instruments are equivalent in their effects upon the business firm. For example, a discount of a note can be interpreted as the firm borrowing the amount credited to its account, with the due date for the service charge and the repayment date coinciding.

7. Rental contracts which vary the payments with the use of the rented factors have to be split into two contracts: a fixed money cost contract, independent of output, and a variable money cost contract which is a function of output.
8. In note 5, it was pointed out that the production of X per time period required a flow of the variable factors $a \ldots g$ in addition to control over $\bar{h} \ldots \bar{n}$ of the fixed factors. If each factor a through n has a price $p_a \ldots p_n$ then we know that the firm must spend on the variable factors a through g a sum $a_\alpha p_a + \ldots + g_\alpha p_g$ and control fixed factors of the value $\bar{h} p_h + \ldots + \bar{n} p_n$. To the total short run expenses $a_\alpha p_a + \ldots + g_\alpha p_g$ and a level of output X_α there corresponds a value of inventories g_{μ_α} and a quantity of transactions cash γ_α. If we assume that the firm owns its fixed factors, we have the total assets are $\mu_a + \gamma_\alpha + \bar{h}p_h + \ldots + \bar{n}p_n$. If the firm's equity is η, then $\mu_\alpha + \gamma_\alpha + \bar{h}p_h + \ldots + \bar{n}p_n - \eta$ has to be financed by debt instruments.
9. The difference between piece rates and time rates here becomes the rapidity with which the firm can 'discharge' workers.
10. The distinction between owners and managers is significant. Although owners are not penalized when a firm does not pay dividends, the management may be. The management of a firm which is characterized by a separation of ownership and management may operate under a constraint to maintain dividend payments. This institutional constraint upon the behavior of the firm may be just as effective as a legal constraint in regard to debt. Management may pay dividends in order to retain control when protection of the owners' investment may indicate that dividends should be passed. Such behavior by firms would indicate that what is considered 'net worth' is in truth a form of debt. Recognizing that accounting categories do not conform to our model's distinctions in every case, we will nevertheless assume that the virtues imputed to equity financing in the body of this volume really exist for equity financing in the world.
11. In deriving the survival conditions, the rate of return used for owners' equity is zero. It has been pointed out to me by Professor Borts that this is equivalent to assuming that the opportunity cost of owners' equity is zero. In reality we have to distinguish between a 'survival rate' and a 'voluntary exit' rate. The voluntary exit rate depends upon the opportunity costs of owners' equity. The assumption that underlies the analysis is that the fear of bankruptcy is a real influence over firm behavior, and that voluntary liquidation possibilities do not exert the same control upon investment decisions. After all, a firm that voluntarily liquidates is more successful in maintaining owners' equity than a firm which fails.
12. The marginal cost curve is the derivative of the total cost curve independent of financing plus the derivative of the financing charges.

    Total factor cost $= f(q)$ r = interest rate
    Financing charges $= r\beta(q - q_1)$ β = capital per unit of output
    $\mathrm{MC} = f'(q) + r\beta$ q = output
    q_1 = output that can be financed by owners' equity.

    Due to the survival condition being a zero rate of return on owners' equity and due to the fact that the minimum borrowing rate is finite and positive, there is of necessity a discontinuity in the cost curves at the maximum output that can be equity financed. This is just one of the discontinuities that occurs in cost curves. The significance of financing costs in the survival conditions and as a determinant of investment leads to its being emphasized here.
13. If β (the capital per unit of output) is a function of output, then $\mathrm{MC} = f'(q) + r[\mathrm{d}\beta/\mathrm{d}q] + r\beta$ again is independent of the equity base.
14. The universally recognized need for the supervision by the state of commercial banking may be interpreted as the recognition of the widespread effects that bank failures have upon the balance sheets and hence survival conditions of firms.
15. $\bar{r}$ is the rate of return which is equivalent for the firm to the freely available rate, for example, $\bar{r}$ is the opportunity cost of owners' equity.
16. $\bar{r}$ would be greater than the borrowing rate when ownership risks are large.

17. A firm earning more than zero and less than $\bar{r}$ on owners' equity is in the voluntary exit range.
18. See Kalecki (1937), p. 443.
19. See ibid., p. 442.
20. Here we use the convention of segregating superfluous assets into an 'investment' trust.
21. Total factor costs $= f(q)$ $\beta q =$ total financing necessary
    E = equity of the firm $\bar{r}$ = the planning rate (or the zero rate on owners' equity)
    r = the market rate for finance

    Total Costs $= f(q) + \bar{r}E + r(\beta q - E)$

    Marginal Costs $= f'(q) + r\beta$, assuming that r is independent of $\beta q - E$. If we assume that $\bar{r}$, the planning rate, for levels of operation such that $\beta q > E$ increases to allow for the risk borne by the debt financing firm, then $\bar{r} = \varphi(\beta q - E)$ such that when $\beta q - E > 0$, $\bar{r} > 0$ and that $d\bar{r}/dq > 0$. Then

    $$\text{Total Costs} = f(q) + \varphi(\beta q - E)E + r(\beta q - E)$$
    $$\text{Marginal Costs} = f'(q) + \frac{d\varphi}{dq} \cdot E + r\beta$$

    As $d\bar{r}/dq > 0$, the marginal cost which allows for the risk associated with the debt is greater, for every output, than the marginal cost of ignoring such risk.

7. Market constraints upon firms: vulnerability

In this chapter we wish to take up the effect of alternative market structures upon the behavior of firms. Our aim is to develop a theory of the behavior of the firm which enables us to determine whether or not the amount of investment induced by changes in either cost or demand conditions depends upon alternative market conditions. Our problem is how these market structures affect the value of the accelerator coefficient. For our purposes market structures can be divided into three classes: competition, monopoly and the region in between, which is generally called oligopoly. In addition to the demand curves confronting a firm or an industry, we have available the modified cost curves (the iso-profit curves) which were derived in Chapters 5 and 6. This apparatus enables us to use cost curves in a more meaningful and systematic manner than they have been used to date in the study of investment behavior.

The aim of this chapter is to prepare the ground for an investigation of the relation between these alternative market structures and the effects of shifts in product demand curves upon investment activity. We are not particularly interested in problems such as the relation between the price of products and the quantities produced under these alternative market structures, nor are we concerned with the question as to whether or not the equilibrium conditions derived for these alternative market structures closely approximate or are widely divergent from the principles for the optimum allocation of resources as derived in the new 'welfare economics'.[1] Our aim is to see whether or not a change in market demand conditions for a product will affect investment behavior of firms differently when the firms are units in industries with different market structures.

We have not much to add to the inherited analysis of the equilibrium of the firm for the competitive case and for monopoly. Our perspective on the region between competition and monopoly, oligopoly, is derived from the comments on conditional monopoly by Marshall.[2] In the literature of the last two decades, since the discovery and application of marginal revenue curves, this perspective has been neglected. The general statement of the perspective is that all significant non-competitive elements in the economy are really conditioned, or constrained, by what can be broadly considered as external market conditions. Our contribution, such as it is, is that we deal with these external market conditions constraining non-competitive firms by means of the various cost curves we have derived. We can show that by using these cost curves we can effectively analyse the behavior of these non-competitive firms by substituting profit maximization within these constraints for unconstrained profit maximization as typically expressed in the phrase marginal revenue equals marginal cost.

As is true of many ideas derived from Marshall, the content of the concept of conditional monopoly can best be determined by referring to his writings:

> Similarly it will be found, generally speaking, that the ownership of exclusive facilities for production or trade in the modern world does not always suggest to a man of sound judgement that he should pursue a severely monopolistic price policy. On the contrary he will keep a watchful eye on the sources of possible competition, direct or indirect. If it appears that these sources are likely to prove large and strong, and that the pace at which competitive supply runs, is likely to become considerable before long: then he will not make full use of his power, but will adjust his price to a firm hold in the market before he can be caught by competitive supply 'following quickly at his heels'.[3]

This suggests that an effective upper limit upon the price that a firm will set for a product is the price at which the firm's management believes that some potential rival could market the same, or a similar product.

> It will in fact presently be seen that, though monopoly and free competition are ideally wide apart, yet in practice they shade into one another by imperceptible degrees: that there is an element of monopoly in nearly all competitive business and that nearly all the monopolies that are of any practical importance in the present age hold much of their power by an uncertain tenure; so that they would lose it ere long, if they ignored the possibilities of competition, direct and indirect.[4]

We shall show that the effect of regulation, as a constraint upon the behavior of a firm, is in its formal aspects equivalent to the effect of possible competition in determining the behavior of a monopolistic firm, so that in a formal sense (for example, in the similarity of the analysis) an element of competition enters even into the behavior of such legal monopolies.

> Stress must be laid on the fact that absolute monopolies are of little importance in modern business as compared with those that are 'conditional' or 'provisional': that is which hold their sway only 'on condition' that or 'provided' that, they do not put prices much above the levels necessary to cover their outlays with normal profits.[5]

This suggests that the iso-profit curves we have derived are significant in determining the upper limits to firm's prices.

> But many monopolies, which seem absolute, are yet to some extent liable to be assailed by indirect routes; and are incomplete and subject to the 'condition' that the monopolist makes no such extreme use of his power as will induce others to force their way through obstacles and set up effective competition.[6]

Again the inducement will be a profit rate: the possibility of shifting the demand curve of the product of the 'monopolist' by the action of potential competitive firms is limited by the profits they can expect to make and the cost of shifting the demand curve.

> These obstacles to competition are mainly of two kinds. The first is the necessity for sinking much capital and effort in setting up the plant and organization, suited for competing on nearly even terms with a strong business, already in possession of a field.[7]
>
> The second obstacle to the setting up of efficient competition with a business, that has acquired a conditional monopoly, is the *vis inertiae*, the opposition to change, which is inherent in human nature and in human conditions. It is being continually diminished by the influences of modern technique, no less than by those modern habits of thought and life, and accordingly some monopolies, so strongly fortified by large capitalistic resources, advanced methods, high ability, and large business connections, that they would have been practically impregnable not long ago, are now often quickly impaired.[8]

The first obstacle deals with production function indivisibilities and financing charges. Financing charges are the concern of Chapter 6. Production indivisibilities – technological conditions – which abet monopoly, it can be said, are very few indeed in areas of the economy which are not regulated. One of the leading phenomena of recent business evolution has been the development of multi-plant firms as substitutes for single giant plant firms, indicating that the technological basis for monopoly is severely restricted in its effectiveness. The inertia seems smaller today in the United States than it was in Britain prior to World War 1: and even then Marshall emphasized the weakness of this obstacle.

Our picture of a firm in the region between competition and monopoly is that the firm has a negatively sloped demand curve for its product: and yet its price cannot be so high that it is either making too large profits or that other firms could produce the product so that they would be willing to sell it at a price significantly lower than the price set by the 'monopoly' firm. The behavior principle which we will assume as operating for these firms that possess constrained monopolies is that they will maximize profits, taking into account the effective constraints upon their position.

Inasmuch as most so-called competitive firms may be considered as having some business connections not freely available to all firms in the industry, the conditional monopoly may be considered as the basic market structure.

The conditions which the market imposes upon the monopoly behavior of firms can be related to cost conditions. The negatively sloped demand curve for the product of the firm (with its associated marginal revenue curve) and the firm's incremental cost curve are not sufficient to determine the price of the product. In addition to these relations we need to know the cost of production of the same or similar products by other, perhaps not yet existing (potential) firms and the rate of return on investment which will be sufficient to cause competing enterprises to be formed.

Earlier we defined a vulnerability concept. A firm is vulnerable when it possesses some advantages of market position which result in the rate of return on the owner's investment in the firm being greater than the rate of return generally available in the economy; and when there exist prices of the product of the firm or rates of return on the investment in the firm which will induce other economic units to behave so as to eliminate or decrease the foundation of the firm's advantageous position. A firm which is vulnerable[9] is a conditional monopolist in the Marshallian sense.

In addition to the vulnerability constraint we have taken up a survival constraint. The survival constraints were transformed into balance sheet and financing conditions which can in turn be interpreted as objective or subjective costs. The complete analysis of firms' investment behavior depends upon both the vulnerability and survival constraints; that is, upon the modified cost curves and the market constraints. All of the cost curves in this chapter take the financing (survival) condition into account.

To recapitulate, the conditional monopoly set up to be used below has the following attributes: (a) each firm has an unambiguous demand curve which has a negative slope; and (b) the market structure adds to this demand curve a constraint which may be either objective or subjective. The firm is assumed to maximize profits, taking the constraint, the market demand conditions and its own cost conditions into account. We therefore retain profit maximization as the rule for firm behavior but we add to the traditional elements the effective constraint upon firm behavior which is determined by the structure of the market. We have also expanded the tools available for analysis so that we can include the rate of return being earned upon investment in the firm as a variable.

The by now traditional analysis derived from Joan Robinson[10] completely defines the equilibrium of the monopolistic firm in Figure 7.1. The negatively sloped demand curve confronting the firm is a datum.[11]

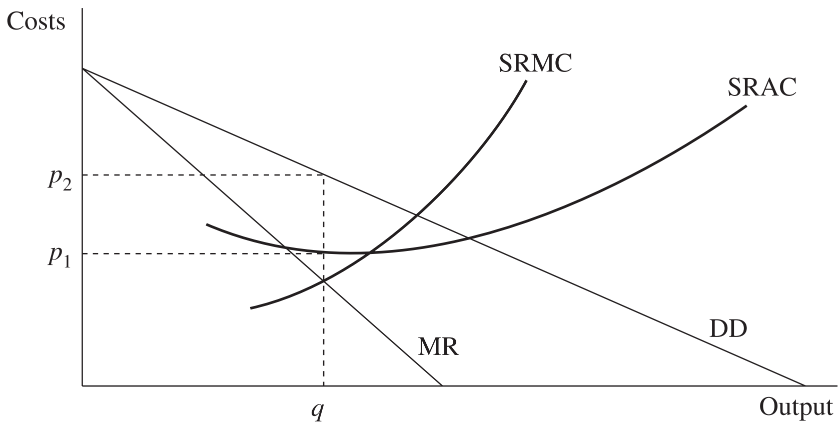


Figure 7.1

The intersection of the marginal revenue curve and the marginal cost curve determines the quantity the firm will produce; the price is of course determined by what that quantity can be sold at as given by the demand curve. 'Monopoly profit is the difference between average cost and average revenue, multiplied by output.'[12] In order to operate in such a manner, that is, to fully exploit the demand curve confronting the firm, the firm must be free of any constraints, as to price, quantity produced, or rate of return earned upon investment. These additional conditions upon the behavior of the firm are derived from, and therefore are transformations of, the structural characteristics of the market within which the firm operates. Each specific type of market structure which lies between competition and monopoly can be transformed into a condition upon either the price of the product, the quantity of the product to be produced or the rate of return upon investment which the firm can earn. We will show how a number of these market structures can be so transformed.

The famous tangency solution, between the demand curve confronting a firm and an average total cost curve, which Professor Chamberlin[13] derived as the solution to the large group monopolistic competition problem, can be interpreted as a special case of conditional monopoly in which the rate of return upon investment operates as both a minimum and a maximum constraint upon the firms in the 'industry'. In Figure 7.2 the firm under consideration is earning, under profit maximizing assumptions, a rate of return upon investment given by ATC(2). The maximum rate of return upon investment which a firm can earn while still preventing entry is given by ATC(1). If each firm in the monopolistically competitive group

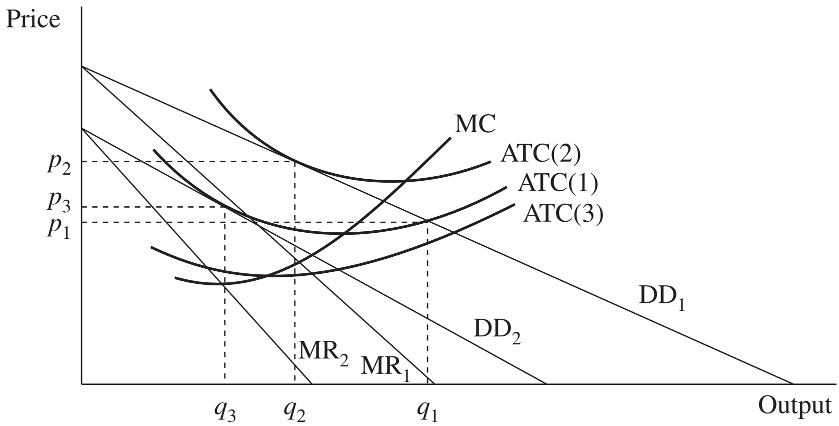


Figure 7.2

produced q_1, and sold it at a price p_1 (average cost equals price), the industry would be in equilibrium and there would be no entry.

However, Professor Chamberlin assumes unconditional profit maximizing behavior by the firms in the industry. Each firm, therefore, produces $q_2 < q_1$ and sells it at a price $p_2 > p_1$ which results in a profit rate which is sufficient to induce entry into the industry. The effect of entry is to shift the demand curve for each firm downward (Chamberlin's symmetry assumption). Entry will continue until the demand curve for each firm shifts to a position where the marginal cost equals marginal revenue solution yields but a 'normal' profit (DD_2 in Figure 7.2). In this case, 'normal profits' have meaning; it is a rate of return earned upon investment by the firms in the industry which does not induce entry into the group or industry. The argument also holds for exit of firms if the rate of return earned is less than that indexed by ATC(1). If the price in this monopolistic competitive industry is to be the same for the output of each firm, it is necessary to assume identity of the cost conditions of the different firms or that the division of the market is such that the elasticity of demand for the different firms varies in the appropriate manner.

On the other hand, if the firms in the group recognize that theirs is a profit constrained conditional monopoly and if the rate of return which induces entry (ATC(1)) is greater than the rate of return which induces exit (ATC(3) in Figure 7.2), and if entry is permitted, it may result in a rate of return lower than ATC(3) being earned. Then the firms in the industry can protect their advantage by using the ATC(1) curve as an upper bound to their price. We can even say that the effective demand curve confronting

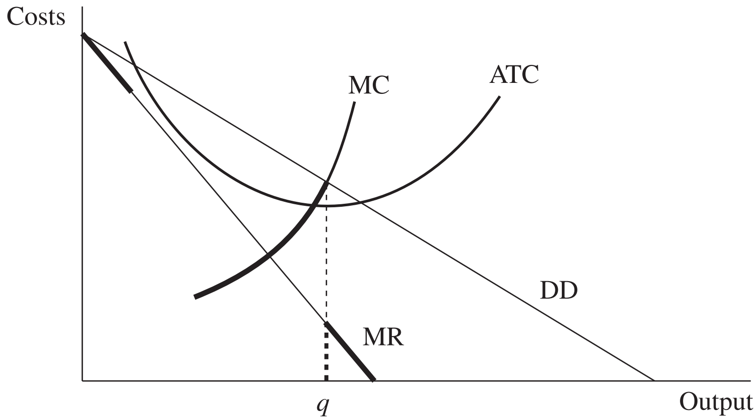


Figure 7.3

the firm is the lower of ATC(1) and the demand curve. This results in a discontinuity in the effective 'marginal revenue' curve for the firm. Until the output q_1 is reached, the effective marginal revenue curve is the marginal cost curve; at the output q_1 the effective marginal revenue curve jumps to the marginal revenue curve derived from the demand curve. Similar discontinuous 'marginal revenue' curves are easily derived for all of the conditional monopoly cases.

In order for a constraint to be effective it must set a limit in a variable which is different from the value that would be derived from unconstrained profit maximization with the given cost curve and the given negatively sloped demand curve confronting the firm. Therefore, given the demand curve for the firm, the attainable profit position for a profit constrained conditional monopoly, that is the average cost curve which represents the achieved earnings of the firm is lower than the average cost curve which represents the equilibrium derived from unconstrained profit maximization. The derivation of the effective constraints from market structures is not in all cases as straightforward and unambiguous as it was in the Chamberlin case handled above. This is true because the market structures in the real world are not as well defined as the Chamberlin large group monopolistic competition market structure.

In any particular industry which is not characterized by competition or monopoly, the particular form of the conditional monopoly which develops will depend upon a number of factors, not all of which are susceptible to formal economic analysis. The effective constraints in a market at any moment of time are due to the peculiar historical development of the industry, the legal limitations (both in terms of law and of its effective application) upon firms' cooperation, as well as upon the purely economic and technological relations between demand conditions and cost conditions.

In product markets in which a few firms are in existence, the demand curve that confronts a particular firm is not independent of the behavior of other firms in the industry. Each firm's demand curve can be interpreted as predominantly depending upon the prices charged by the other firms which sell the same product. In such cases, the effect of vigorous and active price competition among the various firms leads to an instability in each of the particular firms' demand curve. This instability of the firms' demand curve is due to its dependence upon the behavior of other firms in the industry. As long as marginal cost is below price, and as long as each firm is not certain of the reactions of other firms to any change, such instability of the demand curve confronting each firm results.

If a particular firm is uncertain of the reaction of the firms which produce the same or closely related products to a change in the price which it sets for its own output, the firm is confronted (subjectively) with a number of alternative demand and marginal revenue curves. For the demand curve confronting each firm depends upon the prices being charged by the other firms in the industry. Therefore for each set of prices of the outputs of the other firms in the industry there is a different marginal revenue curve confronting a particular firm. As compared with a given profit position, profit maximization with respect to some of these alternative marginal revenue curves represents an improvement; other marginal revenue curves represent a deterioration of the profit position of the firm. A price policy decision in this context involves an estimate by the firm as to the reaction of the other firms in the industry to a change. As the reactions of the other firms are uncertain, a change of price determined upon such a basis will in many cases lead to reactions by other firms which were, in making a firm's decision, deemed unlikely. Such instability of the firm's demand curve, which can be traced to the internal reactions among the firms within an industry, is historically but a temporary development. Such uncertain reactions are replaced by rules of operation which make precise to a particular firm the expected behavior of other firms. The reason for the substitution of a rule of behavior for conjectures as to the behavior of other firms is that the uncertain situation results in active price competition which, in turn, leads to unsatisfactory profit positions for each firm, whereas the profit position attained by each firm under the rule is satisfactory. For a small group of firms the imputation of an unsatisfactory profit position to the behavior of members of the industry is straightforward.

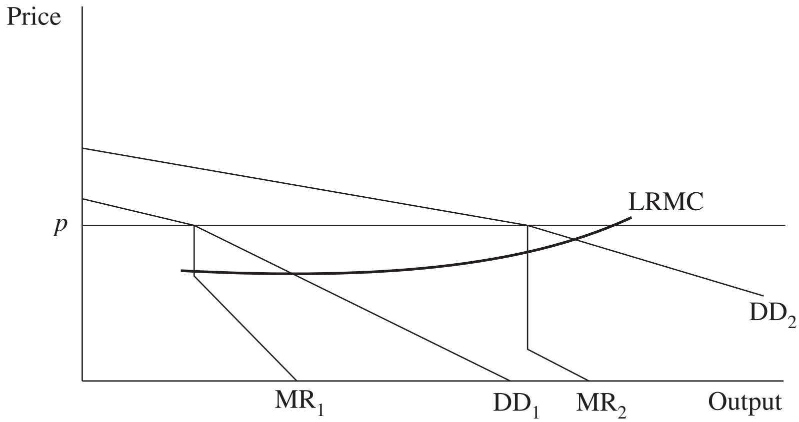


Figure 7.4

The above argument can be made more precise by using the form of the discontinuous oligopoly demand curve.[14] The demand curve confronting a firm in the original discontinuous oligopoly demand curve argument is assumed to have a sharp change of slope at the prevailing price, for the assumption is that if a firm raises its price the other firms will not follow, whereas if a firm lowers its price other firms will follow (Figure 7.4). The effect of a rise in price therefore is a sharp fall in the quantity sold. The result is a 'jump' in the marginal revenue curve, and the conclusion of the analysis is that for a whole series of cost and demand conditions the price of the product would be the same: and particularly that shifts in the demand curve, within limits, would not result in a change in price.

The above conclusion depends upon a specific assumption as to the reaction of other firms. Such specific assumptions are unwarranted for an industry which has not determined or evolved a rule for coexistence of semi-independent price determining units. A firm in an oligopoly situation where no rules exist is confronted by a whole series of conjectural demand curves, each demand curve being based upon specific assumptions as to the behavior of the other firms in the industry. In Figure 7.5 point *A* represents the price quantity that a particular firm is selling. If it lowers its price to p_2 it may sell the quantities represented by b_1, b_2, b_3, b_4, and so on depending upon the reaction of the other firms. The firm's alternative demand curves can be considered as lines connecting points *A* and b_1, b_2, b_3, b_4, and so on and each such demand curve is associated with a marginal revenue curve. With the particular marginal cost curve drawn in the above figure it would

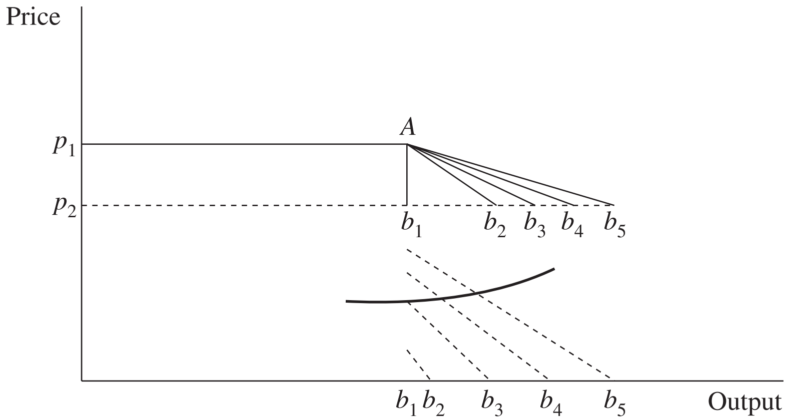


Figure 7.5

pay for the firm to lower its price if b_5 or b_4 is the true demand curve, not if b_1, b_2 or b_3 is the true demand curve.

In such cases the firm would have to weigh the possible gains from a change as against the possible losses from a change in price. Firms would often be wrong and as long as the marginal cost less than price situation exists, firms could continue to be wrong. The end result would be a price so low that the profit position of each firm in the industry deteriorates. Such instability is inherent in any situation where a few firms share a product demand curve and where either no formal organization or no accepted way of life of the firm in such an industry exists.

The solution often adopted for such internal instability is the use of a common pricing rule by the firms in the industry. Some of these pricing rules are price leadership, market sharing, cost plus pricing, and break even point pricing. Such rules prevent the inherent internal instability of the industry from causing unfavorable profit developments for all the firms in the industry. In cases where such pricing rules are adopted there often is no need in the analysis for the market demand curve to be divided into particular firm demand curves. The rule of behavior determines an 'offer price' for each firm, the demand curve determines the amount that will be taken by the market and the amount taken from each firm is not determinate. The acceptability or inacceptability of the resultant position for each firm depends upon its achieved profit. If a resultant position is unacceptable profitwise to a firm and the firm imputes the unfavorable profit position to the behavior imposed upon it by the rule, the rule of behavior for the industry breaks down.[15]

A maximum price as the effective limiting condition upon a firm can be derived from a number of different market structures. Price leadership can be interpreted as in part a maximum price and in part a maximum rate of return conditional monopoly. There are a number of firms producing a product, or a group of products, which closely substitute for each other. The price of the related products is the dominant determinant of how much a firm can sell at any price. In price leadership one firm takes the initiative in determining price. For all the firms in the industry except the price leader, the demand curve confronting the firm is dominated by the price set by the price leader. That is, the price is a parameter to a non-leader firm. However, we can even assume the existence of some sort of product loyalty or differentiation, or geographical dispersion of plants (with freight on board (f.o.b) pricing), and still put the price leadership phenomenon for all except the leading firm into the mold of constrained profit maximization. In Figure 7.6 *a–a* and *b–b* are partial demand curves for the firm, *a–a* at prices higher and *b–b* at prices lower than that set by the leader – *a–a′* and *b′–b′* are the marginal revenue curves associated with these partial demand curves. If the firm's marginal cost curve intersects *a–a′*, the firm charges a price higher than the leader; if the marginal cost curve intersects *b′–b′*, the firm charges a lower price than the leader. If the marginal cost curve intersects neither *a–a′* nor *b′–b′*, the firm charges the leader's price.

For the price leader the problem is different than for the follower firms. The price leader determines the market price on the basis of the amount that will be sold by each firm at each price he might set, and the market demand conditions. This could be interpreted as leading to an unconstrained profit maximizing behavior by the price leader. However, the price leader's decision has to consider the elements which condition his 'monopoly'. He must take into account the conditions under which the

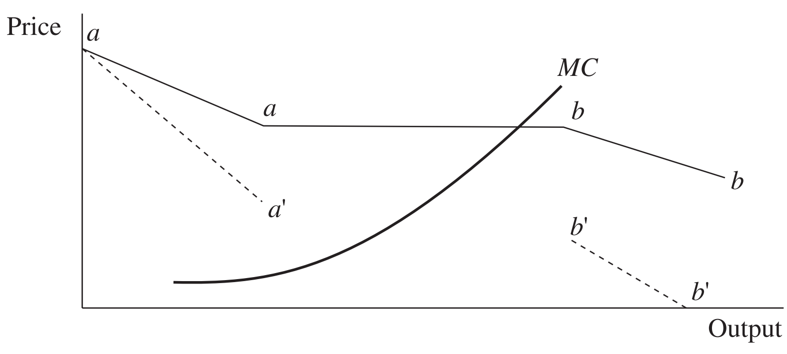


Figure 7.6

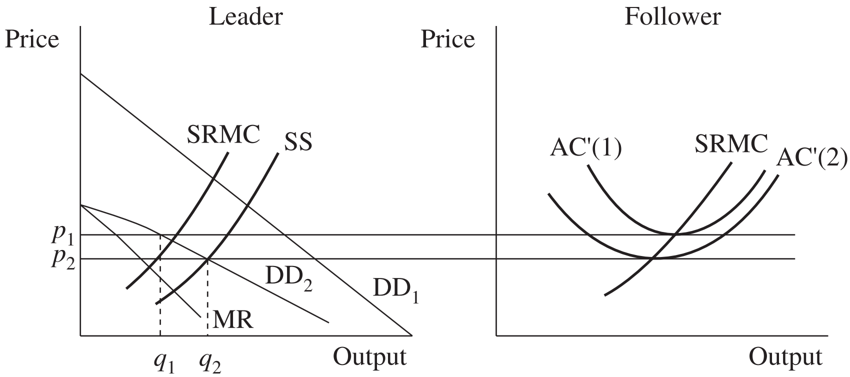


Figure 7.7

follower's supply curve would shift to the right. The price leader is interested in preventing the growth in number and size of the firms which are his followers. The growth of existing firms and entry of new firms into the industry depends upon the rate of return on investment which the followers earn. Therefore as the follower firms treat the price set by the price leader as a 'parameter', the price leader is constrained to set a price of the product which results in a rate of return being earned by the follower firms that is not so large as to induce them to expand or alternatively to induce entry of new firms into the industry.

In Figure 7.7, p_1 is the optimum price from the viewpoint of price leader profit maximization, and AC(1) represents the highest rate of return upon the total investment that the price leader can earn. However, if the price is p_1 the followers' earnings will be high enough to induce entry into the industry or expansion of follower firms. A price of $p_2 < p_1$ reduces the followers' earnings sufficiently so that there will be no entry into or expansion of the industry. If the price leader sets such a price he will supply q_2 rather than q_1 of the output. The price leader will be rational in behaving in this manner if a return of AC′ (1) by the follower firms induces entry into the industry and if the return which the leader can earn and still not induce entry is a higher return than the leader expects to make if entry or expansion takes place.

It is to be noted that the formal analysis cannot determine which firm is to be the price leader. The price leader is often not the same firm under different contingencies. If there is sufficient product differentiation the price leader, when a fall in demand takes place, may be a firm which follows when a rise in demand occurs.

A monopolist will be conditioned by a maximum price when there exist alternative ways of producing the product. We can, for example, assume a mass of production, large plant technique and a small scale technique as existing for the production of a particular product. The small scale technique involves but a small investment by a firm; therefore this would be an industry with relatively easy entry. The price at which the small scale producer could market the product would set an upper limit on the price at which the large scale producer could consider the demand curve for the product as being less than infinitely elastic. In Figure 7.8, as an unconstrained profit maximizer, the firm would produce q_1 and charge p_1. Due to the existence of the alternative production function, the maximum price the firm can charge is p_2. The firm will produce q_2 and earn a return of AC_2 on its investment.[16]

Break even point pricing, cost plus pricing, or full cost pricing have all appeared as names in the literature for the pricing technique adopted by many firms. Attempts have been made to reduce these pricing techniques to marginal cost pricing. In our perspective, such rules of price behavior are really techniques by which an inherently unstable situation, that of a few firms 'dividing up' a product demand curve, can be stabilized. The 'rule' becomes an offer price: and the quantity sold by a firm becomes a 'resultant' of the market demand curve and the division of the market demand curve among the firms which results from non-price competition elements that are allowed to operate.

The equivalence between break even point pricing and cost plus or full cost pricing is easily shown. The usual technical assumption for large scale manufacturing is that, at least for some ranges of output, the marginal cost curve is horizontal.[17] Break even point pricing is the fixing of price so that

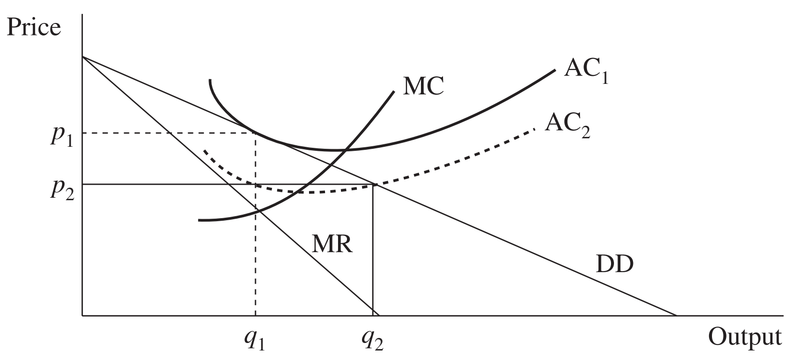


Figure 7.8

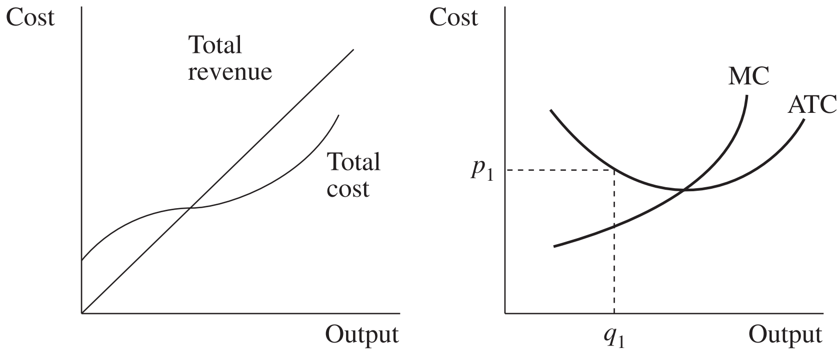


Figure 7.9

at some 'normal' output the firm breaks even: at this output total cost equals total revenue. At the quantity q_1, price p_1, the firm breaks even. Inasmuch as marginal cost is a constant for these outputs, the same price could result from a percentage mark-up over marginal cost. This $P_1/P_2 \cdot$ MC pricing rule is exactly equivalent to a break even point price if marginal cost is constant over the relevant range of outputs.

These pricing rules can be analysed by our tools. The price line's intersection with the demand curve for the firm's output determines the quantity it will sell, and the average cost curve which passes through this point indexes the rate of return that the firm earns. If the firm judges this to be satisfactory, the pricing rule will remain in effect. If the firm judges that the profit position is unsatisfactory, there are a number of alternative courses which it may follow. The firm may attempt to shift its particular demand curve to the right by 'selling', and the firm may attempt to lower its costs; or the firm may break the industry rule and begin price competition. What the firm does depends as much upon the history of the particular industry and the peculiarity of the firm's management as it does upon the objective economic attributes of the industry.

A detailed analysis of the behavior of business firms is beyond our aim. The significant element for our analysis is that firms that follow these rules can be interpreted as being price constrained conditional monopolists. The only difference between the alternative production function case, or the price follower case and these cases, is that the maximum price is obtained from internal data in the latter cases, whereas in the other cases it is determined by external data.

The part played by the other constraint, the ceiling rate of return on investment, in price constrained conditional monopoly is as a signal to the

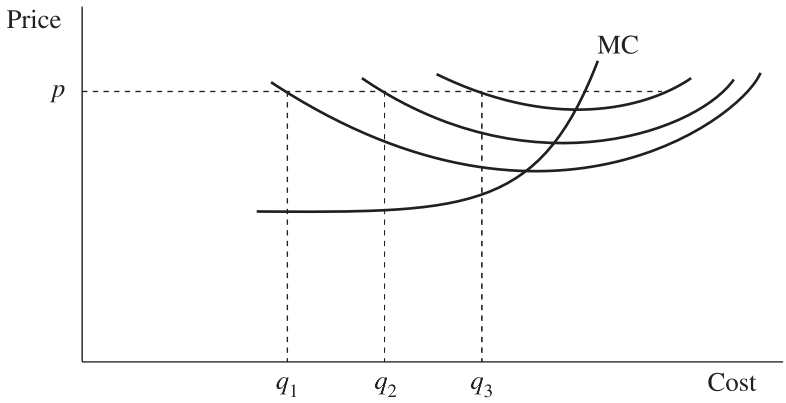


Figure 7.10

entry of new firms or expansion of existing firms in the industry. In Figure 7.10, the AC_1 associated with sale of q_1 by the firm represents a satisfactory profit position. If the quantity demanded of the firm increases, profits rise. Such a rise in demand and the resulting increase in profits earned generally occur in price constrained conditional monopoly markets. As a result new firms may enter the industry. Further analysis of this situation, however, had better await the systematic analysis of the effects upon investment of upward shifts of product demand curves.

The price policy imposed upon a regulated monopoly, such as railroads and public utilities, centers around an average cost concept, the fair rate of return upon investment. This means that the price set by the 'regulator' must be such as would yield a 'fair rate' of return upon the investment. The regulatory body must determine the fair rate of return and the value of investment. Once this is done, and the variable costs are known, we have a unique average cost curve as the loci of prices and quantities sold which will yield the required rate of return. The estimate of demand at each price yields a price for the product which will yield the desired rate of return. We have a case where the average cost curve, when it is below the demand curve, is the effective demand curve for the product. This again is a 'conditional monopoly' case, equivalent to the entry profit rate case which we handled earlier.

Sufficient to say, that for each firm which has a conditioned monopoly of a product, the condition can be represented as either a maximum price or profit rate.[18] It may very well be that both constraints are effective, for different ranges of output.

The above is based upon short run profit maximizing behavior for a firm which is confronted by a set of constraints derived from the structure of the market within which it is selling. Its short run cost curves contain the relevant survival material; its vulnerability to a loss of its favorable market position is represented by a modification of the demand curve confronting the firm to allow for the price and profit values which constrain the firm. Our analysis of induced investment will be based upon the long run cost curves, as modified to allow for the existence of plant and for the financing constraints. We will use the vulnerability material to modify the market demand curves. We will investigate the change in the optimum plant as the demand curve confronting a firm shifts under alternative market conditions, and we will see under what conditions a change in the optimum plant will result in an attempt to achieve the optimum plant. We identify any realized change in plant for a firm as investment by the firm, and barring the existence of unused capacity in the economy which is made available to the firm, such investment by the firm implies investment for the economy as a whole.

NOTES

1. See Reder (1947b).
2. See Marshall (1919).
3. See ibid., pp. 396–7.
4. See ibid., p. 397.
5. See ibid.
6. See Marshall ibid., p. 398.
7. See Marshall (1919).
8. See Marshall ibid., pp. 398–9.
9. The analogy between vulnerability as defined above and being vulnerable in contract bridge is not accidental. A vulnerable firm and a vulnerable bridge partnership are both in an advantageous position, and in both cases the penalty for a miscalculation (being set in bridge) is more severe than the penalty would have been if the firm or bridge team had not been vulnerable.
10. See Robinson (1933).
11. See Stigler (1949). Professor Stigler correctly points out the weakness of the Robinson analysis which assumed away the most significant aspect of the monopolistic or imperfect competition problem: the derivation of the demand curve confronting the firm.
12. See Robinson (1933), p. 56. The content of the average cost curve in Robinson is vague. As monopoly profit is an excess over normal profit the average cost curve must contain normal profits. But in 'A World of Monopoly' (book X, p. 305) how can you define normal profits?
13. See Chamberlin (1938).
14. See Sweezy (1939) and Bronfenbrenner (1940).
15. See Lester (1946). If the firm does not impute the unsatisfactory profit position to its obeying of the rule, it will react by 'becoming more efficient' either in production or in selling. This makes 'economic sense' out of some questionnaire replies.

    Also, a break even point price rule is based upon total cost being equal to total revenue at some output lower than the output the firm expects to sell. The offer price is set independently of the expected sales by a firm.

16. An equivalent constraint upon a firm is the possibility of the purchaser producing the commodity himself. This sets a maximum price for the product. The continued existence of firms which produce a part of their needs themselves, purchasing the rest, indicates the existence of such a constraint.
17. This usually ignores financing conditions.
18. The condition may also be that the firm must supply a minimum quantity of the product. For example, if the major foundation of a monopoly is customer loyalty, the firm must attempt to satisfy such customers' requirement even though the incremental cost of output is greater than the price which it can for some reason or other charge.

8. The effect of market structure upon induced investment

1. INTRODUCTION

'Direct'[1] induced investment by the firms in an industry is the result of an upward shift in the demand curve for the product. Such induced investment can take the form of either a change in stocks, a change in 'work in progress', or a change in plant and equipment. With a fixed plant the level of stocks and work in progress are determined by the level of output. This stock-flow relation has been put into the form of difference equations.[2] There is little doubt of the validity and wellnigh automatic nature of this relation. The questionable relation is between changes in output and changes in plant and equipment. The inventory cycle can follow an accelerator pattern without investment in plant and equipment conforming to the pattern.

Major business cycles can be imputed to variations in plant and equipment investment. Hence, if models based upon induced investment are to be the core of business cycle theory, the accelerator must refer to changes in plant and equipment. Such investment can take one of two forms: either existing firms expand or new firms enter the industry. The expansion of existing firms can be identified as the change in the short run marginal cost curve along a planning curve. For all firms the planning curves have to contain the effects of both the financing conditions and the evaluation of risk associated with debt financing. For existing firms the effect of existing plant upon the expansion possibilities has to be included in the planning curves. In this chapter we will examine what the theory of the firm indicates the relation between investment and changes in income to be under different market conditions. The alternative market structures we will consider are competition, monopoly and conditional monopoly.

We will analyse induced investment by means of the familiar comparative static technique. We will begin with a situation in which a given plant is considered to be a satisfactory plant. We assume that neither net entry nor net expansion of existing firms is taking place. By shifting the firm's demand curve we will see whether or not investment is induced. The relation between the change in both output and expenditures upon the product to investment will be derived.

As a comparative static analysis, the following tells us nothing about the rate at which the new equilibrium will be approached. To truly fit into the accelerator model form, this induced investment has to be transformed into an 'investment per unit of time'. In order to do this we have to compare this 'demand' situation with a supply situation.[3] This leads us to consider 'savings' and 'financing' which will be done in the next chapter. In addition the 'capacity' of the investment goods industries is relevant to the time rate at which desired plant can be attained.

The sum for all firms of the difference between desired and present plant is induced investment. This cannot be measured, as this sum does not necessarily add up to the aggregate investment in the economy. Aggregate net realized investment which can be measured is for any time period the difference between national income and consumption. An individual firm may increase its productive capacity by acquiring title to existing capital goods. The process by which the individual investment decisions are aggregated to yield net realized investment involves a rationing mechanism, due to short run inelasticity of the supply of capital goods, the volume, productive capacity and price of existing capital goods, and the availability of financing.

When a firm invests by acquiring title to existing capital goods no demand for productive factors results. In this case the demand for capital goods results in changing the portfolios of the previous owners of the capital goods. The accelerator process depends upon the inducement to invest resulting in factor income. The existence of redundant capital can abort the income effect of induced investment. Hence, the effective induced investment, under given market conditions, is the sum over all firms of the difference between desired plant and realized plant, minus whatever effect redundant capital equipment has.

Effective induced investment is a sum of individual firm decisions, the decisions being under given conditions. The given conditions take into account the supply and financing conditions for the specific capital goods which are needed to change the present plant into the desired plant. We can therefore derive a schedule of induced investment. This schedule is the relation between the effective induced investment and the interest rate and the prices of the particular capital goods required by the firm. As a rise in the price of capital goods is equivalent to a rise in the rate of interest, we will be mainly concerned with the interest rate-induced investment schedule.

The investment demand schedule is derived by adding to the schedule of induced investment the schedule of autonomous investment (as a function of the interest rate). Aggregate realized investment is the sum of the difference between desired capital equipment and present capital equipment (minus the effect of redundant capital) at the equilibrium conditions in the financial and particular capital goods markets. This equilibrium is determined by the investment demand schedule and the supply schedule of finance and of particular capital goods. In this chapter we are interested in determining the 'schedule of induced investment'.

We have distinguished between the accelerator as a coefficient of induced investment under given conditions and as a coefficient of induced investment where conditions are allowed to vary.[4] The first is derived from technological considerations and we will here investigate whether or not it is independent of (1) market structure; (2) the behavior principle of firms; and (3) the stage of the business cycle at which the income induced change in demand occurs.

The changes which are relevant to investment are in the demand conditions for the product, in the relative prices of the factors of production and in the production functions. As our main interest is in the accelerator coefficient, we will take up the effects upon investment of changes in the demand conditions confronting a firm. To investigate the effects of changes in production functions upon investment requires an analysis of innovations. If there is any meaning to a distinction between induced and autonomous investment, such investment has to be considered as 'autonomous'. Changes in the relative prices of the factors of production can be the repercussions of innovations or of changes in income.[5] Investment due to changes in factor prices can be considered as indirectly induced or indirectly autonomous. One change in price is of particular significance for investment decisions – the change in financing conditions.

The satisfactory plant of the initial conditions is determined by the intersection of a long run marginal cost curve and a marginal revenue curve. As was emphasized earlier, the long run marginal cost curve that is the effective determinant of the optimum plant for a firm depends upon the financing conditions and the evaluation of risk of a particular firm. However, the internal rate of return used in planning may be independent of the market borrowing rates – particularly for firms in a non-competitive industry. It is also true that for both competitive and non-competitive industries the market does not constrain the firm completely in its evaluation of risk.

The assumption that the initial conditions are satisfactory implies that the effective planning curve for the firm is given. For competitive markets this planning curve is, within limits which are due to variations in the evaluation of risk among firms, the same for all firms in the industry. We will show how the limits within which the planning curve may vary among firms in a competitive industry are determined by market processes. For firms which are in non-competitive industries the planning curve may change during a business cycle with constant financing terms.

2. COMPETITIVE INDUSTRIES

With each decision making unit in a competitive industry we associate a family of U-shaped average cost curves. The rising long run marginal cost curves are derived from either the technological production function assuming constant factor prices other than financing, or they are due to the deterioration of financing conditions as the firm increases its borrowings.[6] As is well known, the impact upon a firm of an upward shift in the product demand curve results in a rise in its infinitely elastic demand curve. Assume that p_1 (Figure 8.1) is an equilibrium price. The firm is producing q_1 and is earning r_1 upon total investment. This is represented by $SRAC_1$ and its envelope long run average cost curve ($LRAC_1$). The rate of return earned upon the owner's investment is greater than r_1. It is just sufficiently greater to compensate for the borrower's risks involved in financing an output of q_1. Therefore the effective long run marginal cost curve, based upon financing and borrower's risk evaluation passes through the point (p_1q_1). The shift in the product demand curve results in a rise in the market price to p_2. To an existing firm in the competitive industry this results in the rate of return that it earns upon its investment. In Figure 8.1, this is indexed by $SRAC_2$ and $LRAC_2$. Such a rise in the rate of return implies that investment will be induced in the industry.

Three alternative reactions by an existing firm to a rise in the price of its product, and its profitability can be identified. These are:

1. the firm may alter its plant to that size plant which will yield the highest rate of return upon total investment at the new price;

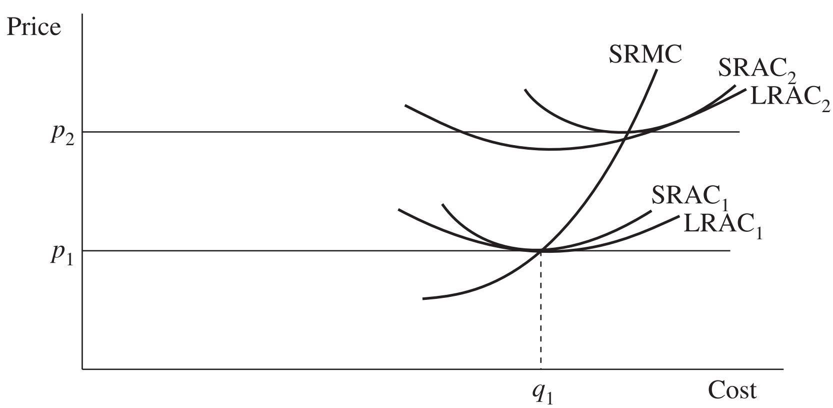


Figure 8.1

2. the firm may alter its plant to that size determined by the planning curve and the new price;
3. the firm may passively enjoy the higher profit rate.

These alternative reactions are based upon differences in what the firms are maximizing and in their interpretation of the meaning of the rise in price. The latter element is usually labeled 'expectations'. The usual technique in accelerator analysis is to assume that the expectations of all firms at any moment of time are the same. There is no need to make this assumption. Therefore, we have to take up the reaction of a firm to the rise in price under different assumptions as to the significance of the change.

The literature on capital theory contains a controversy between the advocates of two alternative maximization principles that a firm may follow. One school maintains that a firm maximizes its own net worth (or equivalently the return earned upon its own capital) and the second that the firm maximizes its internal rate of return.[7] The market processes under non-competitive conditions do not so constrain the firm that it does not have some freedom as to what it wishes to maximize; hence, some firms may maximize their internal rate of return while others maximize their net worth. Under competitive conditions market processes tend to make the long run equilibrium result of the two maximization principles the same. In the short run, however, these two alternative behaviors lead to a different size plant as the ideal plant.

The plant size that will yield the highest rate of return upon the total investment is given by the long run average cost curve, independent of financing costs, whose minimum point is at the market price. Let us assume that the firm's equity is insufficient to finance and operate a plant of this optimum size. Also, we will assume that the objective (ignoring borrowers' risk evaluations) financing rate is lower than the maximum internal rate of return.[8] In Figure 8.2, I_0 is the firm's own investment, I_1 is the investment at which the internal rate of return is maximized, I_2 is the investment at which the marginal internal rate of return is equal to the marginal objective borrowing rate. Unless the subjective evaluation of risk by the borrower is zero, the optimum plant on the assumption that the firm maximizes the rate of return on owners' equity, will be smaller than I_2. As is obvious from Figure 8.2, there exists an evaluation of borrowers' risk which will equate the maximization of the return upon owners' equity with the maximization of the internal rate of return. Are there market processes which would tend to make the risk evaluation which achieves this result the typical risk evaluation?

If we assume no risk (either borrower's or lender's) and that the market rate of interest is equal to the maximum internal rate of return, then the

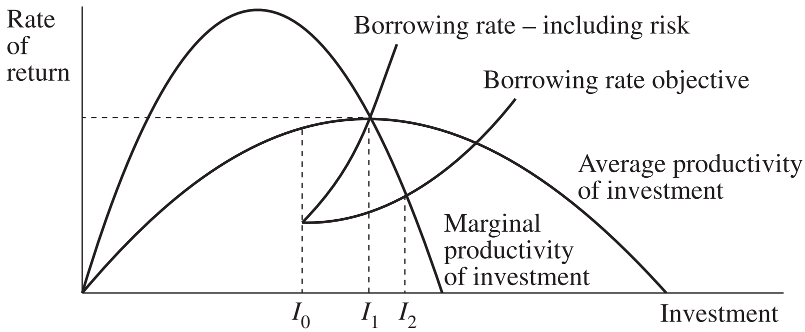


Figure 8.2

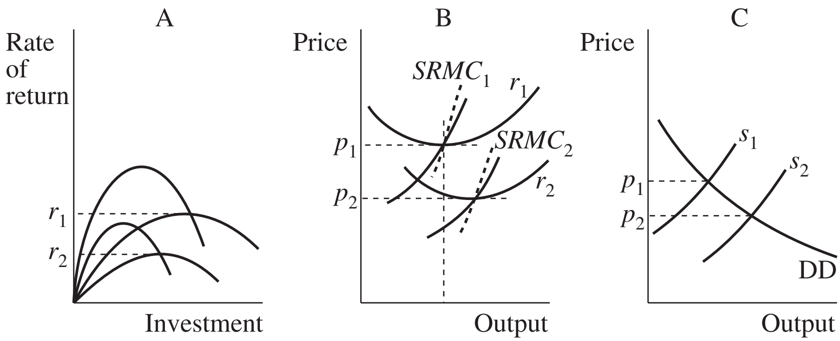


Figure 8.3

investment is the same if the firm follows either behavior principle. The equilibrium of an economy in which resources are freely transferable (risklessly) among industries (recalling that the average and marginal rate of return relations depend upon the price of the product) is such that the investment which maximizes the internal rate of return is the same as the investment which maximizes the rate of return upon owners' equity. Also, if there is a unique borrowing rate and a unique borrower's risk premium, then the two behavior principles again lead to the same total investment.

This 'riskless' competitive world can be taken up in greater detail. Assume that a given industry is in equilibrium at a price p (Figure 8.3B). A plant of size $SRMC_1$ will yield the highest attainable rate of return r_1 for a firm which produces q_1 with a total investment of I_1. If the borrowing rate (the equilibrium rate) is $r_2 < r_1$, then at the market price of p_1 firms which either maximize the internal rate of return or which maximize the rate of return upon owners' equity, will be earning more than r_2. This implies entry into the industry which shifts the supply curve to the right and lowers the market price. Equilibrium will be achieved in the industry when the same investment results from firms either maximizing the rate of return upon total investment or maximizing the rate of return upon owners' equity.

The same relation between the marginal and average rates of return and the market price of the product exists in a world of borrower's risk. Assume that firms maximize the rate of return upon owners' equity by setting the marginal rate of return equal to the marginal borrowing rate (including risk). If at a particular price of the product the returns earned by firms in the industry are sufficient to induce entry, then the price of the product will be lowered. This will lower the maximum internal rate of return upon total investment which can be earned and increase the size of the investment which yields the maximum attainable rate of return at the new price. Such an expansion of the industry will end when the rate of return earned by the typical firm in the industry upon their own investment is just sufficient to compensate for the risk associated with borrowing.

The rise in the objective borrowing rate and the increase of the borrower's risk premium are both functions of the ratio of borrowings to equity. Allow the owners' investment to be a variable. The highest rate of return upon owners' equity is achieved when the own investment is such as to result in the marginal debt cost, including borrower's risk, being equal to the marginal rate of return at the output where it is equal to the average rate of return. The equilibrium position assuming that firms maximize the rate of return is under competitive conditions the same as the equilibrium position assuming that firms maximize the internal rate of return.

This, however, is an 'equilibrium' equality. For periods short of the time necessary to achieve equilibrium there is a difference between the two. Just as firms are free to interpret the meaning of the rise in price, so they are free to determine their plant as:

(a) the plant which maximizes the internal rate of return;
(b) the plant which maximizes the return on the owners' equity.

For investment by new firms, and in the long run, the plant size which maximizes the internal rate of return will tend to dominate, as the ratio of equity to borrowing is over the long run more flexible than the size of plant. However, an existing firm, with a given equity base, will tend to expand by building the plant which maximizes the rate of return on owners' equity.

Alternative reactions to a rise in the product price are illustrated in Figure 8.4. At the price p_1, assume that the firm has a plant of the size

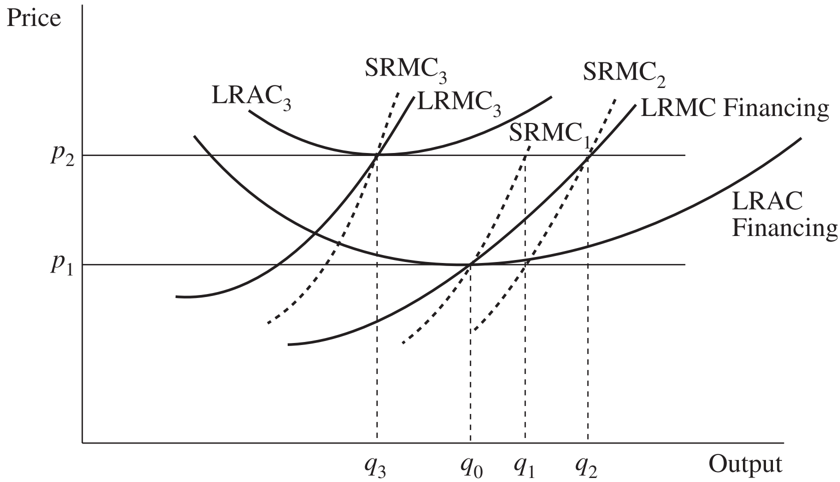


Figure 8.4

$SRMC_1$, which is consistent with both profit maximizing criteria. As a result of a rise in price the firm has its choice between three different plants and outputs. If the firm chooses not to change its plant, it would increase its output to q_1. If it chooses to maximize the rate of return earned upon owners' equity, the firm would expand along the financing curve and construct the plant labeled $SRMC_2$, producing q_2. Or it may choose to maximize that rate of return upon the total investment by constructing a plant of size $SRMC_3$ and producing q_3.

Each reaction of a firm to a change in the price of its product is rational on the basis of some additional hypothesis. If the firm assumes that the price rise is temporary, that p_1 is the 'normal' price, then it will keep the same plant. If the firm assumes that the higher price of the product is permanent (or that it will last long enough) then it may either debt finance an expansion or build that size of plant which yields the highest obtainable internal rate of return.

If the production function involves decreasing returns to scale, then a rise in the price of the product implies disinvestment by a firm which maximizes the internal rate of return. However, an existing firm has a fixed equity, and of course, the time rate at which it can disinvest is given by the depreciation characteristics of its plant. Ignoring changes in borrowing rates and the risk premium of firms, the optimum behavior is either standing still or expanding. However, new entrants can adjust their equity base. Therefore the construction of the highest rate of return plant will be associated with the entrance of new firms into the industry and with the 'expansion' of multiple plant firms.

What is the meaning of the assumption by a firm that there exists a normal price of the product? It can be interpreted as a belief that as long as firms in the industry are making a return on total investment greater than the maximum attainable at that price, industry will expand. This implies that the long run supply curve of the product is a horizontal line at the normal price. Existing firms that do not expand their plant as a result of a rise in price assume that expansion of the industry takes place easily whereas those that expand assume that the industry cannot be expanded easily. The expanding firms assume that the particular skills or attributes necessary for successful operation in the industry belong to the existing firms. This becomes, in the limit, a 'Ricardian Rent' case where expansion of plant is a way in which a scarce resource is utilized most efficiently. The equilibrium higher realized rate of return is a payment of rent to those scarce skills.[9]

A business cycle expansion results in a series of upward shifts in the demand curve for the product of the industry. As long as no entry into the industry occurs and no expansion of existing plant takes place, the price of the product will rise, and as a result the rate of return earned upon investment will also rise. If the original rate of return upon total investment is that rate which acts as a threshold to entry in the industry and existing firms do not expand their plant, expansion by means of the entry of new firms into the industry will occur. If each new firm builds the optimum plant for the original price and if the financing conditions and the production function are the same for the new and the old firms, then each new firm will build the same size plant as the existing firms have. Under these strict assumptions the number of plants will increase in the same proportion as the output of the industry increases and the amount of investment induced by an upward shift in demand for the product is proportional to the increase in output.[10]

The above result may be formalized. Let I_μ be the dollar value of induced investment in the μth industry, $\Delta q\mu$ the change in industry output, p_μ the market price of the product. We will distinguish two coefficients of induced investment. One coefficient $\overline{\beta}_\mu$ relates investment to the change in output, while the other coefficient β_μ relates investment to the change in expenditures upon the product. In the case where the investment decisions are based upon an assumption that p_μ is an invariant normal price we have that

$$I_\mu = \overline{\beta}_\mu \Delta q_\mu$$
$$I_\mu = \overline{\beta}_\mu \Delta q_\mu p_\mu$$

As p_μ is a constant, we have that $\overline{\beta} = \beta_\mu p_\mu$. The only difference between the physical output and the expenditure accelerator is a change of scale constant p_μ.

We can also define a capital coefficient, ν_μ. Prior to the rise in demand the total investment in the industry K_μ could be written as a constant ν_μ times the output q_μ.

$$K_\mu = \nu_\mu q_\mu$$

In this competitive model $\beta_\mu = \nu_\mu$ the accelerator coefficient is a constant, and aside from problems which arise due to the investment period, the accelerator is equal to the capital coefficient. To the extent that industries are of this abstract competitive type, a linear accelerator business cycle model is a meaningful model.

However, some firms may react to the rise in the market price of the product by building that size of plant which maximizes the rate of return upon owned capital at the new price. They will build the size of plant determined by the planning curve which includes financing costs and borrowers' risk and the new price.[11] If some firms react in this manner, then the amount of investment induced by the cyclical rise in demand will not be a linear function of the increase in output. Assuming that any return greater than the original return is sufficient to induce entry, then the long run supply curve is a horizontal line at the original price. At this long run equilibrium price, the firms which expanded on the assumption that the higher price is the equilibrium price will be making less than the entry rate, and as a result, the equilibrium output will be produced with more plant facilities than are necessary.

Formalizing the above result, we would have that $I_\mu = \overline{\beta}_\mu \Delta q_\mu$; however, $\nu_\mu < \beta_\mu$. For example we could still formally relate investment to the change in output but the induced investment would be a larger proportion of the change in output than the previous capital stock was to output. To the extent that the scale of plant is expanded during a business cycle expansion, investment will be greater than the capital coefficient indicates.

Alternatively, a firm may decide to construct that scale of plant which yields the highest internal rate of return. This will be a smaller plant than the optimum for the original price. Again, as long as the original price is a true equilibrium price, the firms which build these smaller plants will earn less than the equilibrium rate of return upon investment when price is again at the original price.

Formalizing the above result, we would again have that $I_\mu = \overline{\beta}_\mu \Delta q_\mu$; however, $\nu_\mu < \beta_\mu$ per plant. The realized induced investment per plant is less than the previous capital output ratio would indicate. However, the number of plants built will be larger than if all plants had been designed to produce at the normal price. This excess number of plants will result in more induced investment than the original capital coefficient indicates; therefore, for the industry as a whole $\overline{\beta}_{\mu} > \nu_{\mu}$.

In any competitive industry we have identified three alternative reactions by firms to a change in demand. Each reaction is rational under alternative hypotheses by the firm. For existing firms we can compare firms that 'stand still' with firms that expand. An existing firm is assumed to have a fixed 'own investment' during the period under analysis. The firms that 'stand still' will earn a lower return upon their own investment when the price is above normal than the firms which expand. On the other hand, firms which expand their plant will earn a lower rate of return on total investment and upon their own investment when the price is once again normal. If the price fluctuates in a range over the cycle, then the optimum size of plant is indeterminate. The merits of the different sizes of plant depend upon the duration of the cyclical peaks and troughs. The plant a firm builds depends, within limits, upon the subjective outlook of the firm as much as on the production function.[12]

Two remarks of general interest can be made in light of the above. If the life of an investment is short as compared to the cycle, then all firms will expand that type of investment. In particular the life of working capital and of inventories is short as compared to the duration of a cyclical expansion. Therefore every firm can be expected to expand within the limits given by their plant facilities, their use of such short term capital. The accelerator inventory and working capital cycles can therefore be considered as valid for competitive industries.

Secondly, to the extent that some firms in an industry plan their plants on the basis of maximizing profits at periods of peak demand, for all except such periods, there will be excess capacity[13] in the industry. Also, if some firms constructed plants for capacity output at a boom price, and if investment in the industry continues until the price is at the equilibrium price, then investment will be taking place even though some firms in the industry have excess capacity. And if some firms construct their plants on the basis of a median price, then such firms, in investing or in replacing plant and equipment when the market price is lower, will be constructing excess capacity.

We can compare the rates of return earned upon own and total investment under the three alternative reactions of firms to a rise in the price of the product. Assume that at p_1, the plant which yields the highest rate of return upon total investment and the plant that yields the highest rate of return upon own investment are the same. This plant is labeled $SRMC_1$. In Figure 8.5, the solid lines represent the return upon total investment, whereas the broken lines represent the return upon a fixed owners' investment. At the price of p_2 the firm may choose among three alternative

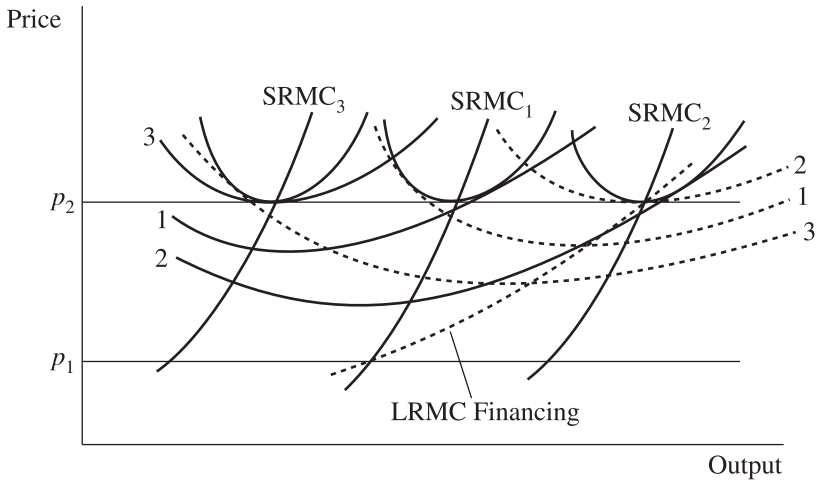


Figure 8.5

plant sizes: $SRMC_3$ which maximizes the rate of return upon total investment, $SRMC_2$ which maximizes the rate of return upon own investment and $SRMC_1$ which is the result of the firm 'standing still.' At the price of p_2, $SRMC_1$ yields a higher rate of return upon total investment than $SRMC_2$.[14] Given that the owners' investment is fixed, $SRMC_3$ results in the lowest rate of return upon own investment. However, if the firm is free to vary the owners' investment, then there exists an owners' investment which yields plant size $SRMC_3$ as the plant that maximizes the rate of return upon owners' investment. The rate of return that is earned upon this particular owners' investment, will be the highest attainable rate of return upon owners' investment at price p_2.

The real significance of the highest internal rate of return is that for a plant of this size there exists an owners' investment which, given the financing conditions, will yield a higher return on owners' investment than is available from any other size of plant. The highest internal rate of return plant in a competitive industry can therefore be associated with the entry of new firms. The effect of these new entrants, however, will be to lower the price. This will result in lowering the realized rate of return upon owners' equity. At the 'normal' price, such a firm will be earning a lower rate of return upon total investment than the plant based upon the normal price. This could result in 'losses' and bankruptcy when price returns to normal.

The effect of a cyclical downswing is that the demand curve for a product is shifted downward. The impact of this change upon firms in a competitive industry is that the price of the product is lowered. This results in lowering the profits of existing plants. For firms which are financed by means of debt, the fall in the market price of the product results in a decrease in the scale of plant which yields the highest rate of return upon investment. If the effect of the use of plant is to decrease the size of plant (depreciation transforms a large plant into a small plant) then the return being earned upon owners' investment will rise when plant is not maintained. This will reduce the disinvestment necessary in order to achieve the best returns upon investment that is possible with the given market price. An excessive reduction in the scale of plant will lower the rate of return below the 'best obtainable'.

Once the market demand curve is stabilized, the sequence of events in the industry is as follows: there is a period in which the return earned upon total investment is rising while the scale of plant is being reduced through non-replacement of plant and equipment; this is followed by a period in which plant is maintained. This occurs once the plant size is achieved which yields the highest attainable rate of return upon owners' investment at the market price of the product. The firms will still be making 'subnormal' profits, but they will be maintaining their scale. Plants that are making below normal profits which are 'maintained' are not unusual during business cycles. Even if the industry has a horizontal long run supply curve in the upswing, in the downswing it can have a positive slope. This acts as a stabilizer: reducing the quantity of disinvestment to zero before the industry reaches the size which yields 'normal' returns upon investment.

This tendency to under-depreciate during cyclical downswings results in another non-linearity in the induced investment coefficient. If the 'normal' return is the return that induces investment, as the industry is stabilized at below normal returns on the cyclical downswing, there is a minimum upward shift in the demand curve of the product which is necessary to induce investment. Therefore, in a competitive industry, the accelerator coefficient becomes zero on the downswing sooner than the change in quantity produced indicates, and the accelerator coefficient is zero for some range of shifts in the demand curve on the upswing.

$$\overline{\beta} < \nu \text{ when } \Delta q < 0$$

and

$$\overline{\beta} = 0 \quad \text{when } 0 < \Delta q < \theta \text{ and } q_{\mathrm{t}} < q_{t-\lambda}.$$

θ is the maximum rise in output at which existing plant facilities will earn no more than normal profits.

The model of firm behavior which we have constructed yields a non-linearity in the accelerator coefficient. This non-linearity is independent of the time rate of change of capital which is possible on the cyclical upswings and downswings. This non-linear accelerator can be explained by recognizing that the survival rate for firms is lower than the normal entry rate. During a period of poor business, firms will use more capital intensive techniques that are implicitly associated with the survival rate. The effect of a rise in demand until such time as the normal rate of return is being earned will be that less capital intensive techniques will be employed. At the entry rate, the techniques of production are stabilized, and expansion of the industry takes place under conditions of fixed proportions.

We can now summarize the behavior of a competitive industry during a business cycle. During the expansion phase, some existing firms expand their plant, other existing firms remain unchanged in size, and new firms enter the industry. Assuming that the new firms tend to build the highest rate of return plant at the boom price, and that old firms tend to expand along their financing curves, the new firms will tend to be too small and the expanding firms too large for the normal price. They will both make lower returns upon total investment and upon owners' equity than those firms which did not change, at this normal price. Even though the 'smaller firm' used less capital per unit of output than the 'normal return' firm, the net effect of the expansion of the industry is that more investment has taken place than the capital coefficient of the normal return plant would indicate. During a strong boom the accelerator is greater than the capital coefficient.

Once a downswing occurs, the rate of return earned by the 'expanding' and the 'new entrant' firms will tend to fall below the rate of return earned by the 'equilibrium firm.' As a result of debt financing on the basis of boom prices (by both the expanding and the new firms) their survival average cost curves are higher than those for the non-expanding firms. If the cyclical downswing results in a price of the product lower than the survival price for these firms, redundant plant appears. During a strong decline such redundant plant will not be used. This is equivalent in its effect upon the industry supply curve to a rise in capital consumption. As a result, the rate of return upon solvent firms' investment does not fall as far and their disinvestment becomes zero sooner than otherwise. For a given fall in demand, the time period during which disinvestment takes place is shorter because some firms fail.

However, the productive capacity of the firms that fail during the downswing is available for use during an expansion. Once the industry is stabilized, such productive capacity begins to be utilized. Such cyclically redundant capacity acts as a damper on new investment when demand rises. To the extent that the downswing is halted by the rapid fall in active productive capacity due to failure, an upward shift in the demand curve for the product will not result in investment until such capacity is reabsorbed. Therefore both the failure of firms and the rise of the capital coefficient during a depression tend to make the accelerator zero for some range of increases in demand from the cyclical trough.

The non-linearities in the accelerator coefficient are due to a number of factors. One is that expansion of an industry – unless all expansion takes place on the basis of an 'equilibrium price' – involves more investment than the capital coefficient at the equilibrium price indicates. Secondly, the effect of a cyclical decline is an increase in capital intensity. Thirdly, during a cyclical decline the appearance of redundant plant in effect accelerates depreciation. As a result the accelerator coefficient is zero during a cyclical upswing until all redundant plant is absorbed and until the return earned upon investment reaches some normal rate.

We therefore have, for competitive industries, unless expansion takes place on the basis of an equilibrium price, and unless the production function is one of constant returns to scale over the significant range, that:

(a) $\overline{\beta} > \nu$ during an upswing and the extent to which $\beta > \nu$ depends upon the strength of the boom.
(b) $\overline{\beta} \rightarrow \nu$ as the boom tends to come to an end (when market price is lowered toward the equilibrium price).
(c) $\overline{\beta} < \nu$ during a downswing due to the implicit increase in capital intensity during the downswing.
(d) $\overline{\beta} = 0$ for some rises in output from the cyclical trough.

The greater the amount of redundant plant, the greater the increase in demand that is necessary to induce investment. The result of the accelerated depreciation due to redundancy on the cyclical downswing is that induced investment is reduced in the initial stages of an upswing.

3. MONOPOLY

The effects of shifts in product demand curves upon investment by unconditional monopolies can be studied by applying the modified cost curves. We will ignore the effects of changes in financing conditions. Modifications of the behavior of monopoly firms will be introduced in order to eliminate the postulate of continuous response. This analysis applies unconstrained profit maximizing behavior to a situation in which the negatively sloped demand curve for the product of a firm shifts.

For competitive markets we solved the problem of 'what does the firm really maximize?' by showing how the two maximization principles (that firms maximize the rate of return upon their own investment and that firms maximize the rate of return on total investment) are equivalent in equilibrium. This is not true in the same sense in monopoly markets. It is, of course, trivially true, that there exists an owners' equity such that maximization of the rate of return upon owners' equity and upon total investment are the same. However, in general, we cannot posit that a firm will be characterized by exactly this owners' equity, and as long as the 'freely' available rate of return is significantly lower than the rate which the monopolist can earn in his firm, there is no reason for the monopolist to shift his equity from his own firm to other firms.

Under the assumption that the monopolist is maximizing the rate of return upon the total investment in the firm, the optimum position is determined by the tangency of the demand curve and a long run average cost curve. In Figure 8.6, the LRACr_1 is tangent to the demand curve at the output q_1. Obviously the long run marginal cost curve for r_1 and the marginal revenue curve intersect at this output. If we take another long run marginal cost curve, say $\text{LRMC}_{r_2}[r_2 < r_1]$, we have that at the output where it intersects the marginal revenue curve the demand curve and the average cost curve have the same slope. But the average cost, earning r_2 upon the total investment, is lower than the price; symmetrically, for a long run marginal cost curve based upon a rate of return r_3 which is greater than r_1. The standard monopoly argument, with the resultant monopoly profits, is based

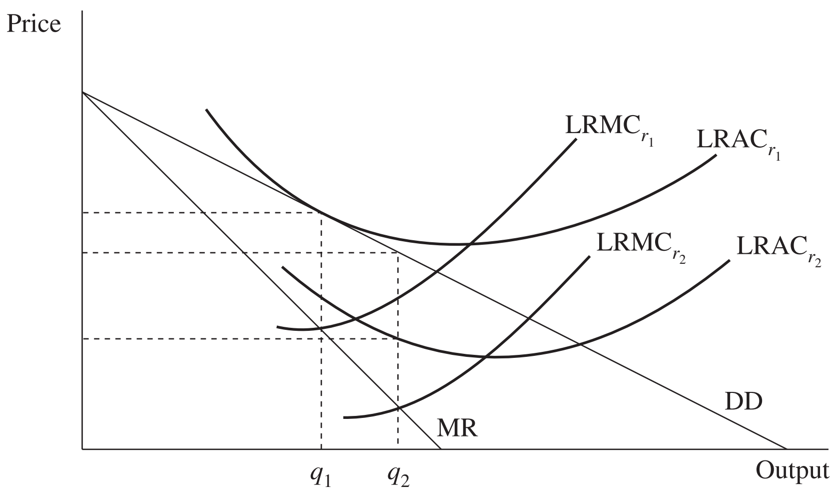


Figure 8.6

upon the firm using a rate of return, such as r_2, which is lower than the highest attainable rate of return, in its planning.[15] The resultant difference between price and average costs, multiplied by output, is called monopoly profits. Without some basis for the firm using a planning curve such as $LRMC_{r_2}$, the usual argument is false. Such a basis can be the use of 'market financing' by the firm, and the relation between the freely available rate of return and the return which can be earned within the firm. The arguments which are necessary to determine the 'optimum' size of operations are relevant to the 'investment decision' problem.

If a monopolist maximizes the rate of return upon total investment then the firm will produce that quantity at which a long run average cost curve is just tangent to the demand curve. An upward shift in the demand curve raises the rate of return which the firm earns with the given plant. However, this plant does not yield the maximum possible rate of return upon total capital. This maximum available return is given by the plant determined by the long run average cost curve which is tangent to the shifted product demand curve.

There is no way of knowing whether the new optimum plant will be larger or smaller than the original plant. Therefore there is no way of knowing whether profit maximization in the sense used here will imply investment as the result of an upward shift in demand. If the upward shift in demand results in an increase in the elasticity of demand then the new optimum plant may be larger than the old plant. If the shifted demand curve is parallel to the original demand curve, then the plant which maximizes the rate of return with the new demand curve is smaller than the original plant.

It is obvious that an unconstrained monopolist can choose to transform a rise in demand into a 'Ricardian Rent' case. In this case the plant remains fixed – induced investment is zero. For both the 'rent' case and the maximization of the rate of return upon total investment case the accelerator phenomenon breaks down – a rise in demand may not induce investment. Therefore, to be able to use an accelerator, the economy must not be characterized either by monopolists who are 'lazy' and transform a rise in demand into a rent or by monopolists who maximize the rate of return upon total investment.

With a fixed equity base, a monopolist maximizing the rate of return earned upon its equity will have an invariant planning curve based upon the market financing terms and the evaluation of the risk of debt financing, for all plants greater than the largest plan which the owners' equity can finance. In this case an upward shift in the demand curve results in an increase in the volume of debt financing.[16] This implies that investment takes place; therefore the accelerator can be used.

In allowing for the entry of new firms into a competitive industry we implicity allowed the equity base of the industry to increase. What is the optimum equity base for a firm that is a monopolist? With a given demand curve, the highest rate of return upon investment is earned by that plant that results in the tangency between the LRAC curve, the SRAC curve and the demand curve. Both larger and smaller plants result in a lower rate upon total investment. This can be represented by an average and marginal return upon investment curve. An upward shift in demand shifts these curves upward.

For a particular monopolist we define a rate of return which is freely available ρ. With a given demand curve, the maximum equity investment in a monopoly firm will be given by the condition that the marginal rate of return equals the freely available rate of return.

This means that the plant determined by the intersection of the long run marginal cost based upon ρ and the marginal revenue curve will be built. This plant will earn more than ρ upon the total investment. If the 'equity' base of the monopolist is smaller than that necessary to build this plant, the monopolist may either debt finance or build the plant which the equity will finance (this is equivalent to a completely inelastic 'risk' premium). For the debt financing firm, the long run marginal cost curve that determines the optimum plant is based upon the borrowing rate and risk premium of the firm.[17]

For a monopoly firm which borrows or which has an infinitely elastic supply of equity funds, a rise in the demand curve of the product implies

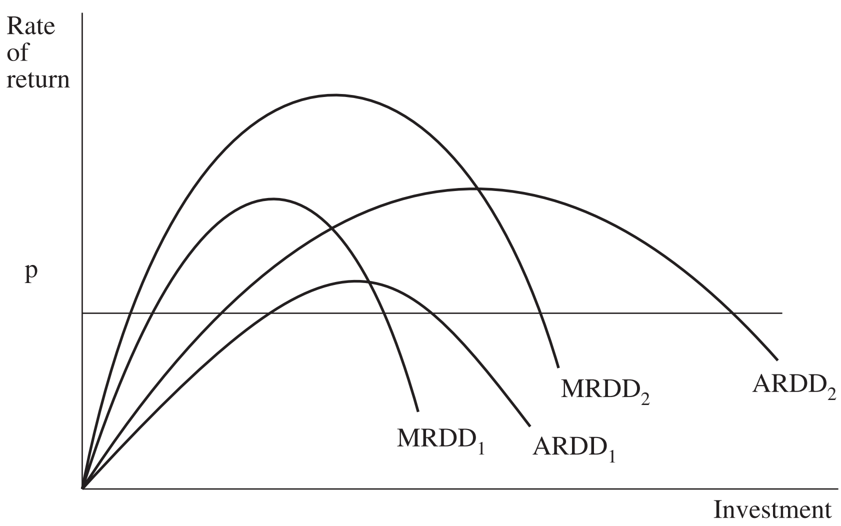


Figure 8.7

investment. The behavior of such a monopolist is consistent with the accelerator.

If a plant is continuously and instantaneously variable the short run marginal cost curves lose their significance. A firm would always be able to adjust its scale of plant to a shift in the demand curve. The relation between the quantity of investment induced and the change in the quantity of the product produced depends in these circumstances upon the nature of the firm's production function. Unless the production function is linear and homogeneous of the first degree, successive increases in output will involve, at constant factor prices, changing proportions of the factors of production. Therefore, the construction which we used for a competitive industry, which enabled us to derive a linear homogeneous production function for the industry, does not apply to monopolies.

For a monopoly that profit maximizes with relation to a fixed planning curve we can still write that

$$I = \overline{\beta}\Delta q$$

but as price may change

$$I = \beta(q\Delta p + p\Delta q + \Delta q\Delta p)\Delta p > 0$$

The expenditure coefficient of induced investment is less than the output coefficient of induced investment, whereas in the competitive industry the relation was the same aside from a scale constant. This difference between competition and monopoly is significant for the efficacy of a rise in money income in inducing investment. In a competitive industry, none of the rise in income[18] is absorbed by a rise in price, whereas in general a monopoly absorbs some of the rise in income as a price rise. Therefore the quantity of investment induced by a rise in money will be lower if the affected demands are for the outputs of monopoly industries than if the affected demands are for competitively produced products.[19]

Expansion of a monopoly is along a production function, whereas expansion of a competitive industry can take place by adding production functions. For those monopoly firms which have production functions that are not linear

$$K_\mu = \nu_\mu q_\mu$$

However, $\nu_\mu = \varphi(q)$. If the production function is one of increasing returns, $\bar{d}\nu_\mu/\bar{d}q_\mu < 0$; if the production function is one of decreasing returns $\bar{d}\nu_\mu/\bar{d}q_\mu > 0$. For a monopoly a change in output may result in changing proportions of investment to output due to the characteristics of the production function.

For a monopolist who is financing his expansion by debt, the relative price of the factors of production does not remain constant as the ratio of debt financing to equity financing changes. With a given equity base, the borrowing rate (including borrower's risk premium) rises with the upward shift in the demand curve. This is equivalent to raising the price of capital. This results in the firm's optimum plant for a given output being smaller than if the borrowing rate had not increased. To the extent that expansion is debt financed it will result in a less intensive utilization of capital. Therefore the amount of investment induced is less than if output had been expanded along the same 'expansion' path as the original plant was on. The resultant accelerator coefficient β became a function of the borrowing rate and risk premium: $\beta = \varphi(r)$.[20]

In the case of a downward shift in the demand curve, the analysis for monopoly is symmetric with that for an upward shift in the demand curve. The effects of a downward shift in the demand curve are two: first the rate of return upon investment in a given plant is reduced, secondly a smaller plant yields the highest rate of return upon owners' equity. If depreciation transforms a large plant into a small plant, the rate of return upon owners' investment will rise if, as plant is being depreciated, with a stabilized demand for the product, the plant becomes that size determined by the intersection of the long run marginal cost curve based upon the financing rate and the marginal revenue curve. When this occurs, the firm will have the plant which yields the highest attainable rate of return. The relation between the amount of capital consumption and the change in the quantity produced depends upon the technical characteristics of the firm's production process. There is no reason inherent in the set-up for the unconstrained monopolist why there should be any time gap between the shift in the demand curve for a product and the resultant change in the capital used. Any time gap which occurs, any deviations from the long run production function, arise from the limitations due to the attainable time rate of change of capital.

If limitations as to the time rate of change of capital exist, then a cyclical movement in the level of income may be reversed prior to the achievement of the optimum plant. For example, if a cyclical downswing in the level of income is reversed prior to the achievement, through depreciation, of the optimum size plant for the lower level of demand, the initial upswing of the level of income may not induce investment. Similarly, the reversal of a cyclical upswing may not be effective in inducing disinvestment.

The assumption of a continuous response on the part of a profit maximizing monopolist to shifts in its demand curve is too stringent. It is true that to each demand curve confronting a monopolist there corresponds a unique plant size which will yield the highest returns. Ignoring the financial problems which arise in changing the size of plant, the firm may still choose not to alter the plant scale for each shift, however small, in its product demand curve. This is a relaxation of the unconstrained profit maximization assumption. For such firms the accelerator coefficient is zero for some range of changes in the demand curve. It also follows that shifts in the demand curve which are sufficient to trigger investment may result in responses which are out of proportion to the incremental shift.[21]

Let us assume a sequence of upward shifts in the product demand curve. With DD_1 as the demand curve the plant indexed by $SRMC_1$ is the optimum plant, that is, it is the plant which yields the highest attainable return on owners' investment. A shift of the demand curve DD_2 may not be sufficient to induce a firm to build the plant $SRMC_2$; rather the firm will continue to operate plant $SRMC_1$. However, a further shift of the demand curve, say to DD_3, will be sufficient to induce the firm to construct the optimum plant for the demand curve DD_3. The amount of investment induced by the shift of demand from DD_2 to DD_3 is greater than this incremental shift in demand alone warrants.

If a profit maximizing firm expects a cyclical swing in its demand curve between DD_1 and DD_3, and if the time rate of change of capital is long

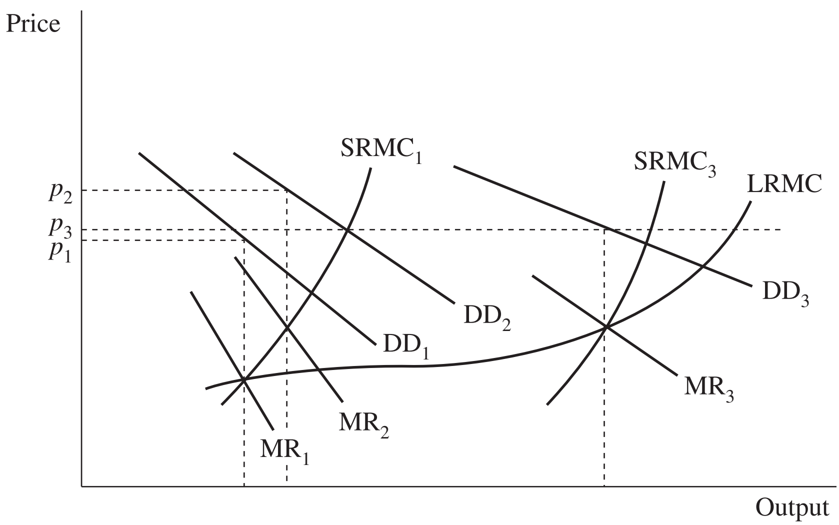


Figure 8.8

relative to the phases of the business cycle, then the firm will have to choose a plant somewhere between $SRMC_1$ and $SRMC_3$. Once such an unconstrained monopolist makes his choice, the usual cyclical swings of the demand curve will not induce any investment. For firms of this sort where the cyclical accelerator is zero, only shifts in demand which are greater than the cyclical pattern will induce investment.

Monopolized industries exhibit a number of investment characteristics different from competitive industries. One is that 'profit maximization' in the sense of a maximum rate of return on total investment may induce disinvestment as the result of an upward shift in demand, so that the coefficient of induced investment is negative. For monopoly firms which finance debt, and for monopoly firms which are characterized by owners who 'distribute' their portfolios, an expansion involves a rise in the effective planning rate. Therefore independent of changes in the market rates of interest, the expansion is taking place on the basis of more expensive financing which lowers the accelerator coefficient. In addition, as a monopolist firm expands, the laws of return to scale of the production function affect the amount of investment which takes place. For decreasing returns monopolists, this increases the ratio of investment to output.

In addition, for a monopolist, we can expect that a rise in market price will take place as the demand curve shifts independently of changes in factor prices. As a result, a portion of a change in money income is transformed into a change in price, making the quantity of investment induced by a rise in expenditures smaller than if the industry had been competitive.

A monopoly industry is similar to a competitive industry in that, from the trough of the business cycle, we can expect that the initial increases in demand will not be effective in inducing investment. In fact, for monopoly firms which expect cycles, the cyclical accelerator may be zero. Investment will be induced only by increases in demand beyond previous peak demands.

Monopoly firms also have the alternative of remaining passive when demand changes: not varying their plant with a change in market demand. The passive behavior filters out the effect of small changes in demand. Such firms would be characterized by spurts of induced investment. Induced investment will be zero for small changes in demand and a disproportionate increase in investment will take place for large changes in demand. Whereas in a competitive industry a strong boom tends to increase induced investment somewhat, for a monopoly a strong boom may be a necessary condition for investment to be induced.

The workings of the accelerator coefficient seems much more certain in competitive industries than in unconstrained monopoly. Therefore the cyclical behavior of an economy should be affected by the proportion of industries characterized by these market structures.

4. CONDITIONAL MONOPOLY

Conditional monopolies have been divided into two classes: those in which the effective constraint upon the firm is a maximum price of the product and those in which the effective constraint is a maximum rate of return upon the investment in the firm. In the maximum price constraint case the optimum size of the plant for a particular demand curve is determinate in some cases only if an additional behavior hypothesis is added.

The maximum price constraint can be considered as a transformation of the demand curve confronting the firm into a kinked demand curve. The demand curve confronting the firm can be split into a horizontal line at the constraint price (this is the effective demand curve for all quantities less than the largest quantity which can be sold at the constraint price), and a negatively sloped demand curve (this is the effective demand curve for all quantities greater than the largest quantity which can be sold at the constraint price). It follows that the marginal revenue curve is a horizontal line at the constraint price where the demand curve is horizontal and where the demand curve is negatively sloped the marginal revenue curve is the curve marginal to the demand curve. The demand side of our analysis is based upon this discontinuous marginal revenue curve. The discontinuity occurs at that quantity where the demand curve confronting the firm intersects the constraint price line. A shift in the product demand curve changes the quantity at which the discontinuity in the firm's marginal revenue curve occurs.

Again we have to distinguish between the two profit maximizing rules for a firm: maximization of the rate of return upon total investment and maximization of the rate of return upon owners' equity. For a firm that maximizes the rate of return upon total investment, the highest attainable rate of return long run average cost curve either is tangent to the horizontal price constraint line or it is tangent to the negatively sloped portion of the firm's demand curve. The first situation is not an equilibrium situation if the tangency occurs at the left of the discontinuity of the demand curve, as at the constraint price the quantity demanded will be greater than the quantity supplied. Assuming that the constraint price is effective, and that non-price rationing does not take place, new firms will enter the industry. This results in a shift of the negatively sloped portion of the demand curve to the left. The industry will be in equilibrium when the cost curve is tangent to the price constraint line at the discontinuity of the demand curve.

This leads us directly to the effect of an upward shift in the demand curve for the firm which is maximizing the rate of return in the 'total' sense and which is in the equilibrium position stated above. Such a firm will not expand as a result of the rise in demand. However, new firms will enter the industry. The price will remain at the constraint price. The amount of investment, barring differences in the production function of the new entrants, will be proportional to the change in output. The price remains constant, none of the rise in expenditures upon the product will be absorbed by a rise in price. Just as for the pure competitive model we constructed we have:

$$I = \bar{\beta}\Delta q$$
$$I = \beta\Delta qp$$

Also if $K = vq$ in the original situation then $I = v\Delta q$; for example, the capital output ratio remains constant.

Alternatively, the highest attainable rate of return upon total investment may be achieved along the negatively sloped portion of the demand curve. In this case the price constraint is not effective. Such a firm can behave as an unconstrained monopolist for some range of upward shifts in demand. An unconstrained monopolist, depending upon the elasticity of demand for the product, may invest or disinvest due to an upward shift in demand. However, for a price constrained monopolist there exists a terminal upward shift in demand which results in the plant that maximizes the rate of return being tangent to the horizontal price line. For shifts in demand short of that, a portion of the change in expenditures will be absorbed in the rise in price. Also if, in the initial situation $K = vq$, v will be a function of output as the movement will be along a particular production function. In addition the 'planning' rate of return will rise which results in a change of the expansion path: for example $v = f(q, r)$ and r is rising. This will tend to reduce the coefficient of induced investment: in fact $\bar{\beta}$ and β may be negative.

Another behavior open to a conditional monopolist is to remain passive. As a result of the rise in demand the firm may choose not to change its plant. Such a 'rent case' is treated in detail later for a conditional monopolist who maximizes the rate of return upon owners' investment. For a conditional monopolist the rent alternative has a natural limit.

The more interesting conditional monopoly is where firms maximize the rate of return upon owners' equity. Their plant is determined by the intersection of the effective marginal cost curve and the marginal revenue curve. Given the maximum price constraint, we can distinguish three different cost-marginal revenue situations for such firms. In the first situation the effective long run marginal cost curve intersects the price line to the left of the discontinuity; in the second the long run marginal cost curve passes through the discontinuity in the marginal revenue curve, and in the third the long run marginal cost curve intersects the negatively sloped marginal revenue curve to the right of the discontinuity. We will determine the optimum scale of plant for the firm to build subject wherever necessary to additional behavior hypotheses.

If the effective long run marginal cost curve intersects the constraining price line, the optimum size of plant for a firm depends upon the additional behavior hypotheses. The firm may choose to supply the entire quantity that the market will take at the constraint price rather than to maximize profits. This results in the construction of a larger plant and in a lower rate of return being earned upon the owners' investment than if the firm maximized profits.

If the firm chooses to maximize profits some demand at the constraint price is unsatisfied.[22] The industry will be in equilibrium when new firms enter and shift the firm's demand curve to the point where at the constraint price market supply equals market demand. In this case, an upward shift in the demand curve implies investment by the new firms and there would be no change in the capital-output ratio in the industry barring production function differences between the old and new firms. For both profit maximizing rules, when the price constraint is effective, expansion of the industry takes place through the entry of new firms. In this case the accelerator generation relation is similar to the relation for a competitive industry.

Alternatively a firm may feel constrained to supply the entire quantity that the market will take at the constraint price. The optimum plant is determined by the long run average cost curve which passes through the intersection of the price constraint line and the firm's demand curve. Such a firm has a larger plant than if it maximized profits, and an upward shift in demand will increase the size of the optimum plant. If the firm's production function is one of decreasing returns to scale, then the ratio of the increment of investment to the increment of output will be larger than the ratio of investment to output with the previous demand curve.

We have for this type of behavior that

$$I = \overline{\beta}\,\Delta q$$
$$I = \beta \Delta qp$$

There is no absorption of the inducement to invest by a rise in product price. Also, in $K = vq$ we have that $v = f(q)$ and the change in q associated with the upward shift in demand results in increasing v. For the price constrained conditional monopolist that attempts to satisfy the entire market, the accelerator coefficient is larger than the accelerator coefficient associated with the expansion of output through new entrants into the industry.

As was pointed out earlier, profit maximization in the case where unsatisfied demand exists results in the development of alternative sources of supply. Therefore a firm which maximizes profit is willing for alternative sources of supply to develop and to have the negative portion of its demand curve shift to the left. If the demand curve shifts to the left beyond the profit maximizing position, a deterioration in the profit position of the firm occurs. A firm would rationally allow an excess demand for its product to arise if it did not fear the effects of the development of alternative sources of supply or if it expects that the rate of return which it will earn with alternative sources of supply in the market is equal to or greater than it could earn if it supplied the entire market. If the firm's set of cost schedules remains unchanged, each shift of the demand curve to the right lowers the returns earned upon investment for such a firm. Each such fall in the rate of return upon investment makes the alternative of profit maximizing more appealing.

On the other hand, a firm that supplies the entire market at the constraint price is acting rationally if it believes that the development of new sources of supply will so adversely affect the firm's earnings that it is better off trying to supply all that the market will take at the constraint price rather than allow alternative supply sources to expand. For such an attitude upon the part of the firm to be rational, the earnings that such a firm achieves upon its own investment when it is supplying all that the market will take must be appreciably greater than the returns freely available.

Two comments can be made about this 'supplying a market' behavior. First that with each increase in the demand for the product the rate of return earned by the firm is lowered. Successive upward shifts in demand may result in such a decrease in the earnings of the firm that it no longer becomes an advantageous return. This could lead to the collapse of the firm. Secondly, as marginal cost is in excess of price, a firm following such a behavior rule will be greatly interested in the discovery of alternative production functions which can lower its long run marginal cost curve. This pressure upon profits from an increase in demand would be conducive to investment in research in seeking alternative production techniques.

In the second case, where the firm's long run marginal cost curve passes through the discontinuity in the marginal revenue curve, supplying the market and profit maximizing coincide. In Figure 8.9, the highest attainable rate of return upon investment from selling q_1 at p_2 is given by the long run average cost curve which passes through the point p_1q_1.

The short run average cost curve which is tangent to this curve at the point p_1q_1 is a short run average cost curve for the plant size determined by the condition that long run marginal cost equals short run marginal cost at quantity q_1. In this case the optimum size plant is that which would be built to produce the quantity given by the discontinuity.[23]

In the third case, where the long run marginal cost curve intersects the marginal revenue curve to the right of its discontinuity, the maximum price constraint is not effective. The optimum size plant on the basis of profit maximizing behavior by the firm is determined just as in unconstrained

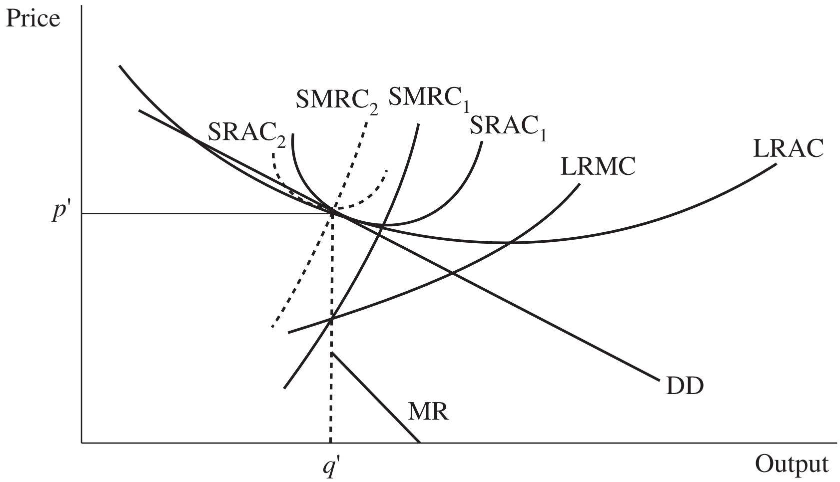


Figure 8.9

monopoly. In this case the firm behaving as if it were a profit maximizing monopolist produces a larger output which it sells at a lower price than if it sold all that the market would absorb at the constraint price. The firm also has a larger plant and is earning a higher rate of return upon investment than if it followed a policy of selling at the constraint price. The effect of upward shifts in demand will be the same as for a profit maximizing monopolist unless the demand curve shifts sufficiently to have the long run marginal cost curve pass through the discontinuity in the marginal revenue curve.

In order to take up the investment that is induced in price constrained conditional monopolies where the firms are maximizing the rate of return upon owners' investment, we will assume that the effective long run marginal cost curve will be unchanged. We will investigate the effects upon the optimum size plant for a firm if the discontinuity in the marginal revenue curve shifts to the right as a result of shifts in demand. When this discontinuity shifts, the relation between the cost curves and the marginal revenue curves may change from one of the situations as described above to another.

If we assume: (1) that the long run marginal cost curve is rising in the relevant quantity ranges, (2) that the demand for the product of the firm is always elastic at the constraint price, and (3) that there are no significant variations in elasticity as demand shifts, then we can rank the marginal cost-marginal revenue situations. Using upward shifts in the demand curve for our classification we have that for 'low demand' the marginal cost curve intersects the negatively sloped segment of the marginal revenue curve, that for 'intermediate demand' the marginal cost curve passes through the discontinuity in the marginal revenue curve and that for 'high demand' the marginal cost curve intersects the constraining price line.[24]

If, starting with the marginal cost curve intersecting the negatively sloped portion of the marginal revenue curve, the demand curve shifts upward and the cost curve still intersects the negatively sloped portion of the marginal revenue curve, then the price constraint is not effective. The result is identical with that of unconstrained monopoly.

If as a result of an upward shift in the demand curve, the long run marginal cost curve passes through the discontinuity in the marginal revenue curve, then the change in the optimum size plant is determinate. In Figure 8.10, as an unconstrained monopolist the optimum size plant was $SRMC_0$, upon which the firm earned a return indexed by $LRAC_0$. As a result of the shift in the demand curve to the right the firm becomes a constrained monopolist, and the optimum size plant is $SRMC_1$ upon which the firm earns a return indexed by $LRAC_1$. In this case, the price rises to $\bar{p}$ so that we have

$$I = \bar{\beta}\Delta q$$

and

$$I = \beta(\Delta qp + q(\bar{p} - p) + \Delta q(\bar{p} - p))$$

where $\bar{p}$ is the constraint price. As the expansion takes place along a production function $v = f(q)$ and as the financing rate increases along the marginal cost curve $v = f(q, r)$: the capital output ratio can be expected to change. For this case the following hold: (1) a portion of the inducement to invest is

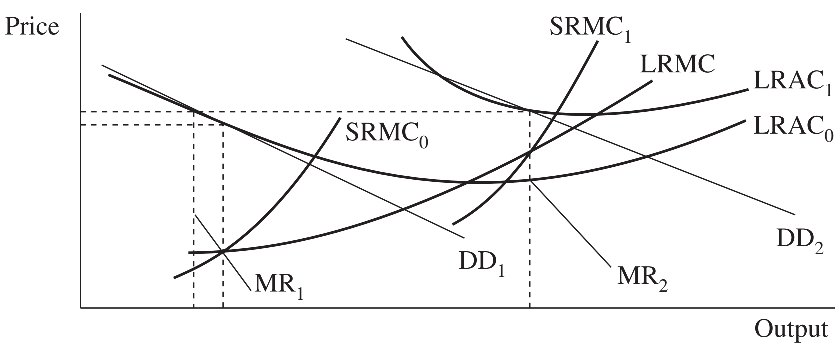


Figure 8.10

absorbed by the rise in price, (2) the characteristics of the production function may induce investment that is a larger proportion of the change in output than the original capital output ratio indicates, and (3) the rise in the financing rate will tend to reduce the capital output ratio. Inasmuch as this firm supplies all that is demanded at the constraint price, no new entry takes place.

If the demand curve shifts to the right, for a firm whose long run marginal cost curve passes through the discontinuity, the result may be that the long run marginal cost curve intersects the constraining price line. In this case we have to distinguish between the firm's two alternative behavior patterns – that of satisfying the market demand at the constraining price or of maximizing profits. If the firm chooses to maximize profits, it will build the size of plant determined by the long run marginal cost curve's intersection with the constraint price. This size of plant will yield the highest rate of return upon owners' equity available to the firm given the market situation. However, the firm may rationally choose to satisfy the market demand at the constraint price. The most effective way of producing this output is by constructing the plant which passes through the intersection of the demand curve and the constraint price. In both these cases price does not rise, but the production function and rate of return changes which were taken up earlier occur.

If the demand curve shifts to the right when the long run marginal cost curve already intersects the constraining price line, there will be no change in the size or profitability of the profit maximizing firm. If the firm is attempting to satisfy the market, each upward shift of the demand curve implies an increase in the optimum size of plant and a decrease in the rate of return earned upon owners' investment.

The amount of investment that is induced by a given upward shift in the demand curve for the product depends upon the cost situation which a firm is in and upon the behavior principle it has adopted. If the firm is in the unconstrained position, the amount of investment which takes place depends upon the relation between the change in output and capital in the production function. However, as long as the firm is unconstrained, the upward shift of the demand curve will result in a price increase, and this price increase results in the change in quantity resulting from a given upward shift in the demand curve being smaller than the change in quantity at a constant price.

For that range of outputs for which the firm's long run marginal cost curve passes through, the discontinuity in the marginal revenue curve, the amount of investment induced depends upon the investment output relation in the production function. In this case, as in the competitive case, the price does not change. However, this differs from the competitive case in that we are not expanding output by increasing the number of identical units; rather we are moving along a production function. Therefore, whereas in the competitive case β was independent of q, in this case β is a function of q, $\beta = \varphi(q)$ and again $\beta = \varphi(q, r)$.

In the case where the long run marginal cost curve intersects the price constraint line, and the behavior of the firm is profit maximizing, an upward shift of demand will result in no additional output, and no additional investment will be induced by the firm. In this case $I = 0$ for the firm. If the firm behaves so as to supply all that the market will take at the constraint price, then the optimum plant size will increase with the upward shift in demand, and $I = \beta\Delta q$ and $I = \beta\Delta q\bar{p}$. β again is a function of q; however, as the long run marginal cost curve for this case must be rising, there is a presumption in favor of asserting that the amount of capital required per unit change in output is increasing, that is $d\varphi(q)/dq > 0$.

If the effect of an upward shift of the demand curve is to move the firm from one marginal revenue cost situation to another, the induced investment is a sum of the amount induced in one situation plus the amount induced in the other. It is possible for a rise in demand to induce disinvestment by a price constrained conditional monopolist. If a firm is in the situation where long run marginal cost intersects the price constraint and is guided by a supplying of the market rule, successive shifts of the demand curve lower the rate of return upon investment. As a result the firm may be forced to give up its attempt to supply the market. This may cause the firm to change its rule to profit maximization, which will imply a decrease in the size of its plant.

As long as we are not allowing for the effects of time, then the effect upon investment of a downward shift in the demand curve is symmetric with the effects of an upward shift in the demand curve.

If we discard the assumption of instantaneous and continuous changes in plant and allow for the time taken in the construction of plant and the unwillingness – aside from financial considerations – on the part of a firm to change plant immediately and minutely with every change in demand, we have to make a behavior assumption for the firm with regard to plant changes. Let us make the plausible assumption that the firm will not expand its plant as long as the rate of return earned on the fixed plant at a price of the product equal to or less than the constraint price, rises when the demand curve shifts upward. This leads to a discontinuous response on the part of the firm to upward shifts in the demand curve. With a given plant no investment takes place for some range of shifts in the demand curve.

In the case where a firm is a profit maximizing monopolist due to the ineffectiveness of the price constraint, no investment will be induced under the above behavior assumption until the short run marginal cost rises to the constraint price. In Figure 8.11 $SRMC_1$ is the existing plant. The negatively sloped portion of the marginal cost curve (MR_4) intersects the short run

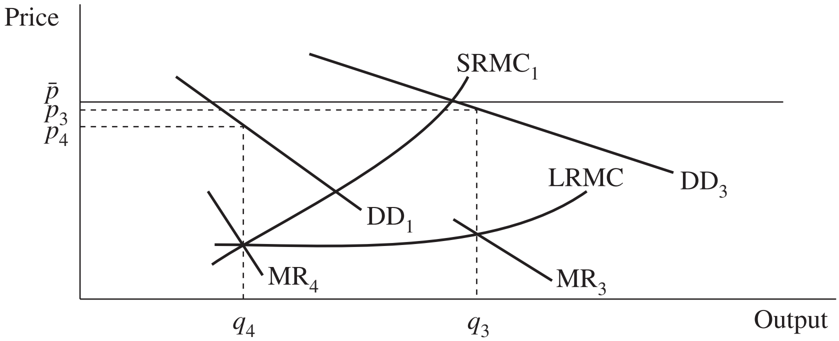


Figure 8.11

marginal cost curve leading to a profit maximizing price of p_4, output q_4. As the demand curve shifts to the right, the short run marginal cost curve passes through the discontinuity in the marginal revenue curve, which results in a profit maximizing price of $\bar{p}$, the constraint price. Further shifts in the demand curve to the right result in marginal cost becoming greater than the constraint price so that the rate of return being earned in the given plant falls – assuming that the firm behaves so as to satisfy the market demand for the product at the constraint price. This will lead to a decision to build a new plant, and as we have drawn the long run average cost curve in Figure 8.11, the profit maximizing position with the optimum plant $SRMC_2$ for demand curve DD_3 leads to the unconstrained monopoly position with its price lower than the constraint price.

Earlier we rationalized the behavior of firms which produced an output so large that their marginal costs were greater than the constraint price, on the ground that their profit rate depended upon the maintenance of the market structure. In this case we have another rational reason for the short run maintenance of the constraint price even though it involves a temporary fall in the profit rate. If there is a significant gestation period for building a plant, a firm would not be willing to suffer any deterioration in its market position during the gestation period, knowing that the new plant will significantly reduce costs (increase capacity). Hence it will supply all the market will take at such a price so as to prevent the development of alternative sources of supply. The example of the behavior of the 'unconstrained' price constrained monopolist results in a fall of the price of the product after the construction of the new plant, without the need to assume any technological change. It may be true that much of the lowering of price of products which is imputed to technological change induced by rise of demand is really the effects of the construction of the optimum plant for higher demand conditions.

The case of a firm whose short run marginal cost curve passes through the discontinuity in the marginal revenue curve is similar to that of the firm whose marginal cost curve intersects the negatively sloped portion of the marginal revenue curve. No investment will take place until short run marginal cost equals the constraint price. For this range of shifts in demand the rate of return earned upon investment will be rising. Again the investment decision will be discontinuous.

Until such time as the long run marginal cost curve intersects the constraint price, the rate of return earned upon investment in the price constraint monopoly case will be rising. Any further rise in the demand curve will result in either a fall in the rate of return upon investment due to the attempt on the part of the firm to supply the market demand at the constraint price or the development of alternative supply sources, with a possible deterioration in the firm's conditional monopoly position.

For price constrained conditional monopolists we have derived alternative investment responses to upward shifts in demand. Such an industry may expand by means of an increase in the size of existing firms or by the entry of new firms. We have shown that if existing firms expand, the ratio of investment to the change in output may be different from the ratio of capital to output that existed prior to the change in demand. This is due to two factors: one that the investment takes place along a production function which may exhibit changing returns to scale, the other that the planning rate of return changes.

In addition, there is a presumption of a discontinuous response by existing firms. Until the demand curve becomes such that the marginal cost curve of a given plant approaches the constraint price, no investment need be induced. When the price approaches the price constraint, a small incremental rise in demand can induce a large change in plant.

In terms of the systematic variation of the accelerator coefficient over a business cycle, this all or nothing response of the price constrained conditional monopolist can result in a large accelerator coefficient when income is substantially higher than the trough income. On the other hand, small increases in income tend to be absorbed by the increasing profitability of existing firms. If the investment is imputed to the incremental change in income, the accelerator coefficient will be large. However, the optimum plant built in these circumstances will be such that further rises in income will again tend to be absorbed by the increasing profitability of the existing firms.[25]

A downswing will shift the negatively sloped portion of the firm's demand curve to the left. If originally the firm's long run marginal cost curve intersected the price constraint line or passed through the gap in the marginal revenue curve, a fall in demand may result in the long run marginal cost curve passing through the gap. In these cases the price remains constant at the constraint price. The result of the optimum size plant decreasing will be that firms will tend to consume capital. To the extent that such disinvestment takes place, the firm's earnings will rise. With a stabilized low income demand curve, this decrease of plant size will halt when the optimum plant for the low level demand has occurred. As the movement is down a marginal cost curve, in which financing enters, independently of the cyclical fall in interest rates, the fall in the borrower's risk premium as the borrowings of the firm are reduced means that the optimum capital output ratio rises. As a result the disinvestment will not be carried as far as the change in output indicates. However, as the decrease in size is along a production function, to the extent that decreasing returns exist, capital consumption will be large in relation to the fall in output.

If the fall in demand is such that the short run marginal cost curve for an existing plant intersects the negative sloped portion of the marginal revenue curve, then the firm will lower the price below the constraint price. Such price cutting will tend to maintain output, absorbing a portion of the decrease in demand in the price change. However, if the long run marginal cost curve passes through the gap in the marginal revenue curve, the price cutting phase will end as the firm succeeds in decreasing plant size. The 'price cutting phase' of the business cycle will be associated with the initial sharp fall in income. Of course, during the downswing, the fall in demand may be at a faster rate than plant can be reduced (a donkey and carrot affair), so that the 'price war' lasts. However, once the demand curve is stabilized, the tendency to establish the normal price will begin to dominate. The establishment of the normal price situation will mean that, for some range of increases in demand, no investment will be induced.

The conditional monopolist who supplied the entire market will, as he succeeds in disinvesting, earn larger returns upon his own investment unless demand falls too far. If the quantity demanded at the constraint price is less than the profit maximizing quantity at the constraint price, then the returns earned fall. As this firm will have tended to intensify its use of capital at the boom, the disinvestment will tend to be a larger ratio to the change in output than the 'boom' capital output ratio would indicate.

As is true of the competitive case, the fall in demand may result in some firms falling below the survival limit. This would tend to be particularly true where the 'price constraint' is due to the existence of alternative production functions and the expansion of the industry took place through the entry of firms with this alternative production function. In the case of bankruptcy, either the firms are reorganized and the capital equipment is used, or the phenomenon of redundant capital will occur. The effect of the appearance of redundant capital is that the demand curve of the surviving firms is stabilized at a higher level, therefore decreasing the disinvestment by such firms. During the period in which capital equipment is rendered redundant, the 'induced disinvestment' is greater than the capital output coefficient would indicate. However, during the recovery phase of the business cycle, expansion of the industry will take place, in part, through the absorption of this redundant capital. This will tend to reduce induced investment.

Price constrained conditional monopoly therefore tends to result in non-linearities in the induced investment coefficient in both the upswing and downswing. On the upswing we would expect that induced investment would be small for some range of output changes and very large for other ranges of output changes. To the extent that the 'gap' remains the effective determinant of plant size, the construction of that plant involves the construction of 'excess capacity'. Therefore some further rises in demand would tend not to induce investment. The accelerator effect operates intermittently, but strongly, when it operates, upon firms in such an industry. If the economy operates so that for the different industries the high accelerator coefficient occurs at the same time, and the low accelerators occur together, the economy would tend to have 'strong booms' followed by periods in which the accelerator operates so as to dampen the economy down. The effect upon the downswing of the 'price wars' and the appearance of redundant capacity also tends to result in a rapid fall in income which will be followed by a damping accelerator. And on both the upswing and the downswing, the behavior of the price constrained conditional monopoly tends to dampen out the effect of small changes in demand.

Our other model of conditional monopoly is where the firms are limited by a maximum rate of return upon total investment. This may be due to a legal limitation as in the regulated monopolies, or it may be due to the firm's desire to maintain a market position: for example, to prevent the entry of new firms. In this case the effective constraint upon the firm consists of the long run average cost curve for the 'limitational' rate of return or the demand curve, whichever for a particular quantity yields the lower price.

If the product demand curve lies below the constraining long run average cost curve, the firm is an unconstrained monopolist. If the demand curve is tangent to the rate of return constraint, then depending upon the long run marginal cost curve, the firm either behaves as an unconditional monopolist or charges the price given by the tangency. Similarly for the case where the demand curve intersects the constraining cost curve, the firm either behaves as an unconditional monopolist or charges the price given by one of the intersections. The second is the interesting case.

Unless the constraint is ineffective, the firm cannot operate so as to maximize the internal rate of return. Therefore, we will assume that there exists a unique long run marginal cost curve based upon the financing changes; and that the firm maximizes the rate of return upon a fixed owners' equity.

As mentioned earlier, the interesting case is where the demand curve intersects the constraint long run average cost curve. In this case, the effective demand curve consists of two parts. The demand curve for the product to the left and to the right of the intersection with the constraining average cost curve is part of the effective demand curve. The other part of the effective demand curve consists of the average cost curve between these intersections. To the left of A, the marginal revenue curve is the curve marginal to the demand curve. Between A and B, the marginal revenue curve is the long run marginal cost curve for the constraining rate of return. To the right of B, the marginal revenue curve is the curve marginal to the demand curve. Unless the firm's long run marginal cost curve is such that the firm behaves as if it were an unconstrained monopoly, the equilibrium price would be either p_1 or p_2. This depends upon which gap the marginal cost curve passes through.[26]

Let us assume that the firm's planning long run marginal cost curve passes through the high output gap in the long run marginal cost curve. The optimum plant will be determined by this intersection. In Figure 8.12, the price charged will be p_1, the $SRAC_1$ will index the yield being earned upon investment. The yield in total investment will be less than the constraint yield, although the yield on equity may be greater. The plant built will have excess capacity in the following senses: (1) marginal cost will be less than price and (2) for some increases in the quantity produced at the given price the return earned upon investment will increase.

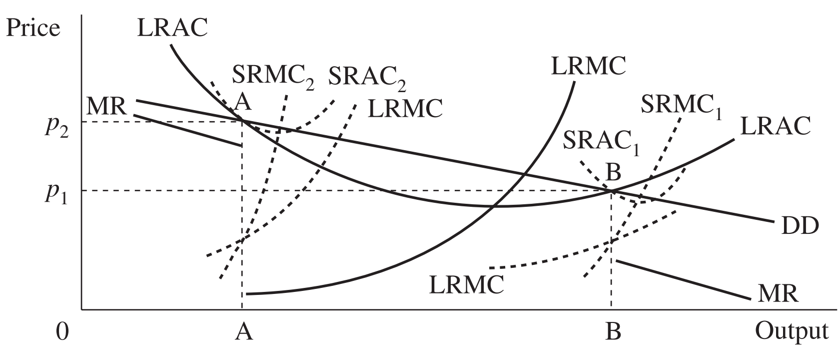


Figure 8.12

Some range of upward shifts in demand, with a given plant, will result in a rise in earnings. Until the rise in demand is such that marginal costs are greater than price there will be no pressing reason for the firm to expand – there is a natural discontinuity in the inducement to invest. When this investment threshold is reached, the optimum plant for the firm to construct is again given by the intersection of the firm's long run marginal cost curve with the gap in the marginal revenue curve. The new plant will again have excess capacity.[27] Therefore, the inducement to invest will again not be pressing for a range of upward shifts in demand.

On the downswing, given this lag, the existing plant will be equal to a smaller than the optimum plant for some range of demand curve shifts. For this range of demand curves there will be no inducement for the firm to disinvest; the induced disinvestment coefficients will be zero.

In general for a 'rate of return' constrained monopolist there is no inducement to invest or disinvest for small changes in demand, whereas large changes in demand result in a large volume of induced investment.

If the planning marginal cost curve passes through the 'upper gap' in the effective demand curve, each upward shift in demand results in a rise in price and a fall in quantity.

If the planning marginal cost curve passes through the 'upper gap' in the effective demand curve, then an upward shift of demand results in a higher price and a lower output. When the existing plant yields the constraining returns upon investment, then a further increase in demand may induce disinvestment. In this case a fall in demand implies investment. This perverse cyclical response of the 'high price' conditional monopolist with an effective rate of return constraint is due to the planning marginal cost curve lying above the constraint rate marginal cost curve. This may hold for some capital shortage cases. Its primary interest however is as a 'novelty' in which the accelerator coefficient is negative.

For the rate of return case expansion of a firm takes place along a firm's production function. Again the effect of the laws of return operate; if the production function exhibits decreasing returns, then investment will be more than proportional to the change in output, and so on. Also, the rate of return upon which the planning curve is based may rise during the expansion of the firm, resulting in a less intensive use of capital.

5. SUMMARY

Earlier we showed that the aggregate coefficient of induced investment is a sum of coefficients for the various firms. Aggregate induced investment is equal to $\beta\Delta Y$ and $\Delta C_\mu = \alpha_\mu \Delta Y$, where α_μ is the marginal propensity to consume particular goods. Also $I_\mu = \beta\Delta C_\mu$ and $\Sigma I_\mu = I$ so that

$$\beta = \frac{\Sigma\beta_\mu\alpha_\mu\Delta Y}{\Delta Y}$$

$$\beta = \Sigma\beta_\mu\alpha_\mu.$$

The aggregate induced investment coefficient is the weighted average of the induced investment coefficients of the different industries, where the weights are the marginal propensities to consume the different goods.

In this chapter we have shown that the investment reaction of a firm to a rise in income depends upon market structure. We have shown that it is necessary to make either strict competitive model assumptions or to assume that the industry is a conditional monopoly in which the price constraint is effective and existing firms profit maximize if the behavior of firms is to be consistent with a constant valued accelerator coefficient. For other market structures we have shown that the relation between a change in income and investment will vary over the business cycle.

Therefore, the aggregate accelerator coefficient depends upon the industrial structure of the economy. If the weight in the economy of the different types of markets changes, we can expect the business cycle pattern to change.

In particular we have shown that for price constrained conditional monopolies there is a range of increases in demand for which induced investment is zero, and that when such a firm expands it constructs excess capacity. During a business cycle expansion the capacity of such conditional monopolies is related to the demand for their product at the previous peak income. Hence we expect that for these industries induced investment will tend to be zero until income approaches its previous peak. When income is equal to or greater than this previous peak all of the 'conditional monopolies' will tend to invest and the investment will be a large ratio to the incremental change in demand. There is, therefore, a tendency for a 'lumping' of the investment of conditional monopolies and an economy characterized by a large proportion of conditional monopolies would be characterized by periods of 'stagnation' and 'inflation'.

6. THE INTEREST RATE – INVESTMENT RELATION

In order to deal with aggregate induced investment we have to sum the individual firms' investment induced by a rise in income. This aggregate depends upon a number of factors, in addition to the change in income. Among these other factors, the most useful for analytic purposes is the market rate of interest. The market rate of interest is the objective rate at which borrowing begins; it is the prime rate of banking practice. Of course, for a particular firm the effective rate depends upon borrower's and lenders' risk. However given the particular firm's financial condition, the effective rate varies with the prime rate. This effective rate is a determinant of the firm's planning curve. We are interested in investigating how the elasticity of investment with respect to the rate of interest depends upon market structure.

The market rate of interest can affect induced investment in two ways. The first is via the substitution effect; a low rate of interest to the price of non-capital factors will tend to make the firm substitute capital factors for other factors in producing a given output. The second is through a scale effect; a low rate of interest, by lowering the planning curves of firms, will tend to make the optimum output larger. The substitution effect depends upon the nature of the firm's production function. The marginal rate of substitution between factors in production may be high or low.[28] The greater the marginal rate of substitution among factors in production, the greater the interest elasticity of demand for investment. However, there is no reason to believe that the marginal rate of substitution among factors is correlated with market structure. Therefore, the interest elasticity of demand for investment due to the substitution effect will be independent of market structures in the community. The change in optimum output due to a lowering of the firm's planning curves is not independent of market structure. We are interested in seeing how the interest elasticity of demand for investment due to the scale effect depends upon market structure.

The optimum plant for a firm in an industry depends upon the following factors:

1. the demand curve for the product;
2. the market structure of the industry;
3. the firm's behavior principle;
4. the firm's balance sheet structure;
5. financing conditions.

For each firm, factors 1 through 4 can be assumed as fixed. As the financing conditions can be represented by the prime market interest rate, the aggregate investment demand curve is the sum of the investment desired by each firm at each interest rate. The elasticity of this aggregate investment demand curve due to the scale effect depends upon the weight of the different market structures and different behavior principles in the economy.

The market rate of interest is a relevant variable in determining the optimum plant only for those firms that maximize the rate of return upon owners' capital. The optimum plant for firms that maximize the rate of return upon total investment is not affected by changes in the market interest rate. Firms that transform shifts in demand into changes in rent also are not affected by borrowing rates.[29] Therefore, for these firms aggregate investment is independent of changes in market interest rates.

The planning curves of firms that maximize the rate of return upon owners' equity will shift with changes in the market rate of interest. The lower the rate of interest, the larger the scale of plant desired by each firm in a competitive industry at a particular market price of the product. Also, the price at which the firm can earn any given return on owners' investment is lower, the lower the market rate of interest, and if we assume that the market price of the other factors used by the firms in such an industry is independent of the output of the industry, then the infinitely elastic long run supply curve of the industry is also lower. The change in aggregate output due to a fall in interest rate, and therefore the amount of investment induced by the scale effect, is greater as the price elasticity of demand for the product is greater, and therefore the elasticity of the investment demand curve depends upon the elasticity of demand for the product.

For the profit maximizing monopolist, a fall in the financing rate lowers the firm's planning curve. This increases the optimum output, and therefore induces investment. The amount of investment induced by the fall in interest rates depends upon the price elasticity of marginal revenue. In Figure 8.13, a fall in financing rates from r_0 results in increasing the optimum output and the optimum scale of plant of the firm. A fall in interest rate from r_0 to r_1 increases the optimum output from q_1 to q_2. This will induce $I_\mu = \beta_\mu(q_{\mu 2} - q_{\mu 1})$ of investment. This induced investment of course

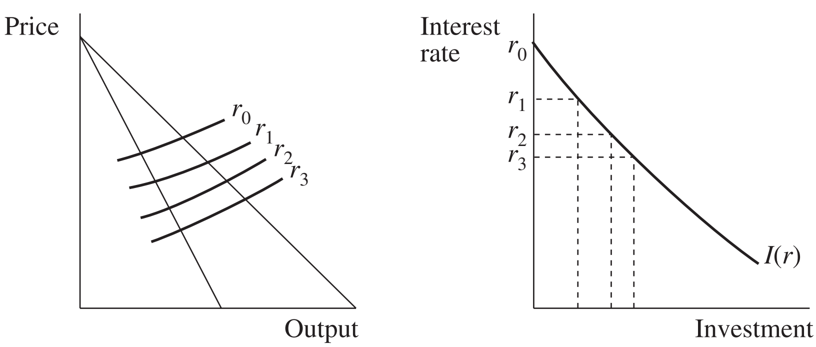


Figure 8.13

depends upon the firm not having 'excess capacity' at the output q_1. As the marginal revenue curve is steeper than the demand curve, a given fall in the planning curves will induce less investment in a monopoly industry than in a competitive industry with the same demand elasticity.

In the case of the price constrained conditional monopolist, the price constraint is either effective or not effective. If the constraint is not effective then the firm behaves as if it were an unconstrained monopolist. If the constraint is effective then either the planning curve intersects the constraint price or the planning curve passes through the gap in the marginal revenue curve. If the planning curve passes through the gap in the marginal revenue curves, then the set of changes in the interest rate which results in the planning curve still passing through the gap will not change the optimum output. Therefore, the scale effect elasticity of investment with respect to interest is zero. If we assume that the initial interest rate is r_0, a fall in the interest rate to r_1 will induce investment $I_1 = \beta_\mu(q_2 - q_1)$, whereas interest rate changes between r_1 and r_4 will not induce investment.

In Figure 8.14, a fall in the interest rate from r_4 to r_5 will induce $I_2 - I_1$ of investment $((I_2 - I_1) = \beta(q_3 - q_2))$. In this case, for a range of interest rates, the interest elasticity of investment is zero. If r_4 is lower than some liquidity considerations minimum and r_1 is considered 'too high' by the monetary authorities, then in an economy with a heavy weight of price constrained monopolists, the monetary authorities may have to consider the scale effect interest elasticity of investment as zero. The case of the rate of return constrained conditional monopoly is equivalent to the price constrained conditional monopoly.

If the planning curve of a price constrained conditional monopoly intersects the effective price constraint, the firm may supply all that the market will take at the constraint price or it may profit maximize and allow

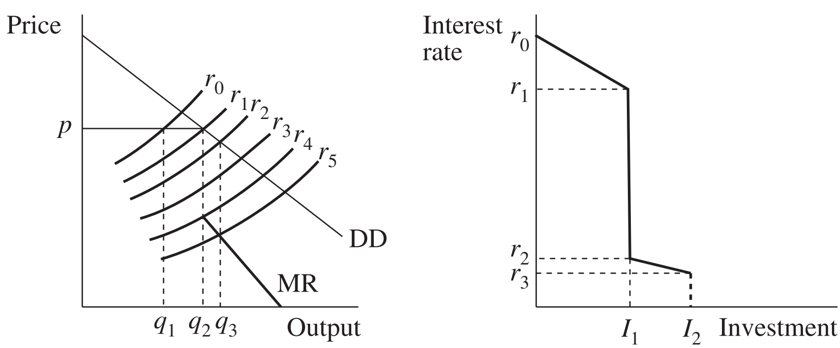


Figure 8.14

other firms to enter the market. If the firm chooses to supply the market, a reduction of financing rates will increase the owner's earnings with the given plant but the scale effect induces no investment. If the firm is profit maximizing, then investment is taking place in the industry through the expansion of other firms. A reduction of the market rate of interest increases the scale of plant that the given conditional monopolist will desire. However, as far as the scale effect is concerned, with a constant price constraint no change in total output will occur as a result of the lowering of interest rates. Therefore, the demand schedule for investment due to such firms is inelastic.

The induced investment of the acceleration doctrine is a point on an induced investment-interest rate schedule. This is obvious if we recognize that such induced investment is determined on the basis of a given planning curve. The planning curve actually used by a firm that maximizes the rate of return upon owners' equity depends upon the market interest rate. This rate is the result of the supply conditions of financing as well as the induced investment schedule.

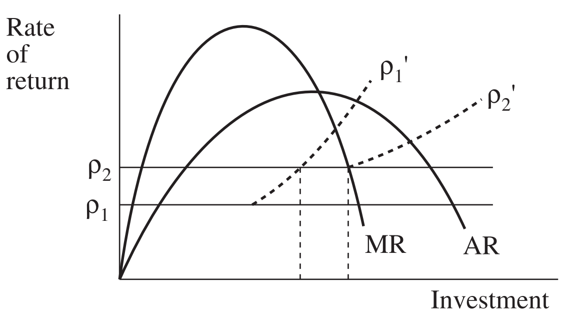
The relation between induced investment and the interest rate derived here is due to the change in optimum output that results from changing the planning curve. This investment does not depend upon any change in the capital output ratio. As such, the interest rate-investment relation can just as well be interpreted as a price of capital goods-investment relation.

Hence, the greater the weight of competitive industries in the economy the more effective changes in the rate of interest will be in affecting investment. The greater the weight of conditional monopoly, the smaller the expected interest elasticity of demand for investment. This relation between the elasticity of demand for investment and market structure has obvious policy implications, especially with respect to the efficacy of monetary policy in raising income. This is particularly relevant to economies in which liquidity considerations result in a minimum value for the prime rate of interest. In such cases the availability of financing at favorable terms may not be sufficient to induce investment. This is in addition to the existence of a 'rent' transformation of increases in demand. The efficacy of monetary policy in restricting investment is due as much to credit rationing by financial institutions as it is to changes in posted prime rates. The non-availability of credit, due to the inelasticity of the money supply, is equivalent to an infinite interest rate. Such an interest rate obviously will restrict investment.

NOTES

1. It is necessary to emphasize that 'direct' induced investment is being taken up here. Systematic changes in the relative prices of the factors of production, changes which may be associated with changes in income, can also 'induce' investment.

2. See Metzler (1941).
3. Of course, the gestation period of plant and equipment is a determinant of the time rate at which the desired plant and equipment can be achieved.
4. The accelerator as a coefficient of realized investment, of course, depends upon the accelerator as a coefficient of induced investment.
5. Ignoring population and taste changes.
6. If the production function results in constant returns to scale, then the long run average cost curves are 'flat bottomed'. Then at a given price, the 'optimum' scale plant is indeterminant. In such conditions a plant which earns the maximum attainable rate of return upon total investment at one price may also earn the maximum attainable at another price. The effect of such conditions is to make the success or failure of a firm during an expansion or contraction depend solely upon the balance sheet structure of the firm.
7. A summary of the controversy is given by Lutz and Lutz (1951); see particularly chapter II, 'Criteria of Profit Maximization', pp. 16–48.
8. Although the borrowing rate may be greater than the lending rate, if the borrowing rate is significantly greater than the maximum internal rate, no owners' equity will be invested in the firm.
9. See Viner (1936), p. 30. To be truly a 'rent' case that results in a higher rate of return these scarce factors must be uniquely linked to the existing firms. Otherwise the bidding for these factors will raise costs so that the original scale of plant and rate of return will be the equilibrium rate of return at the higher price.
10. By such a construction the industry production function becomes one of constant returns to scale. As emphasized by Wilson (1953), pp. 67–8 some form of linear returns to scale is assumed by the acceleration theorists: quoting Hicks' remark that 'it is a good rule in economic theory to stick to (the assumption of) constant returns to scale until you have some reason for giving it up' (Hicks 1950, p. 58.) Wilson wrote that 'it must be made clear that by constant returns we mean constant long run average costs, not constant short run average costs'.
11. That the risk estimation may decrease during good times is well known. However, we will not deal with such effects.
12. See Stigler (1939).
13. Excess capacity here means that the output for which the plant is designed is greater than the output that the plant is producing.
14. The return earned upon total investment at p_2 with $SRMC_2$ may be lower than the return earned upon total investment at p_1 with $SRMC_1$.
15. See Harrod (1952). He points out in Essay 8, 'Theory of Imperfect Competition Revised', that the traditional monopoly theory assumes a given plant, or a given long run marginal cost curve and that the use made of these tools is incorrect: see especially pp. 140–57, 'Doctrine of Excess Capacity'.
16. Of course, assuming that the planning curve is not inelastic.
17. In Figure 8.7, we drew the freely available rate as a horizontal line. A rational owner of wealth may seek to distribute risks among a number of different assets. Assume ρ is the



freely available rate. Assume that after a certain investment (perhaps a percentage of the owners' total worth) the value of a distributed portfolio rises so that the equivalent return on investment in the monopoly increases. At the same time you have the 'riskless borrowing rate', which is assumed to be greater than the freely available rate of return ρ. The rising ρ due to the value of the distributed portfolio will intersect the 'borrowing rate'. As a result an owner of a monopoly may be borrowing funds for his monopoly firm at the same time as he is investing, at lower rates of return than he pays for borrowing, in other assets. In a world of limited liability, this is rational.

18. From a cyclical trough, assuming that the survival rate is lower than the entry rate, a rise in demand will not result in investment in a competitive industry. In the monopoly case such an absorption of a rise in demand by a rise in price holds at all levels of income.
19. It is necessary to add: with the same supply conditions.
20. If the equivalent freely available rate rises due to the value of a balanced portfolio, then $\beta = \varphi(\rho)$ in the same manner as $\beta = \varphi(r)$ and with the same effects.
21. For conditional monopolies a standard can be set for such discontinuous responses. For monopoly firms the shift from the rent case to the profit maximizing case cannot be imputed to any single objective phenomenon.
22. This is the point at which Bronfenbrenner (1947) began his analysis.
23. Note that it is not typically the plant which has its minimum point at the discontinuity on the price line: contrast this with the equilibrium condition when the firm maximizes the rate of return upon total investment. This is, of course, the familiar point made by Harrod (1934).
24. By making assumptions about the elasticity of demand we can change the order of the cost situations. If the high level demand is more elastic than low level demand, then the discontinuity would decrease in size as demand rose. The size of the gap is $p_c - p_c(1 - [1/\varepsilon(p_c)] = p_c/[\varepsilon(p_c)]$ where p_c is the constraint price and $\varepsilon(p_c)$ is the elasticity of demand at p_c. If $\varepsilon(p_c)$ increases as demand shifts to the right, the size of the discontinuity decreases.
25. This is an alternative explanation of the phenomenon taken up by Chenery (1952).
26. It is not possible for the effective marginal cost curve to pass through both gaps. If at the output OA, the financing costs are greater (including borrower's risk premium) than the constraint rate, then the effective long run marginal cost curve always lies above the LRMC. If the borrower's risk premium rises rapidly (or the equity base is so small that borrowing rates rise), then the effective LRMC may intersect the constraint rate of return LRMC from below. In this case the firm will either make more than the constraint return upon total investment or have to resort to non-price rationing. Both these alternatives are inconsistent with our assumption that the rate of return is an effective constraint. In this case the conditional monopoly will break down.
27. Again see Chenery (1952).
28. It is an unresolved question of fact whether or not the substitution effect in production is significant. The usual arguments by which the interest elasticity of the investment demand curve is deduced center around the substitution effect.
29. It has been argued that when the yield of firms falls below the entry rate of return, as it does in depressions, firms will not invest for some upward shifts in demand curves. For the set of demand curves which do not induce investment, such firms are transforming shifts in demand into rent. Therefore, we expect that for low income levels, the elasticity of demand for investment with respect to interest will be more heavily weighted with rent cases than during periods of high income.

9. Monetary behavior and induced investment

1. INTRODUCTION

Business cycle models based upon the interaction of the accelerator and multiplier have been mainly concerned with the implications of the formal model. That is, the pure model generates the cyclical time series, and the expository elements in the writings of the accelerationists have been primarily directed at the interpretation of the formal results. The processes by which investment is induced have not been considered in detail. We have taken up the transformation of the change in aggregate demand into the demand for investment goods. We now have to relate the demand for investment goods and the volume of realized investment; that is, we have to consider the supply side of the investment goods markets.

Accelerator and multiplier models have considered the supply side of the investment market by introducing either lags, due to the gestation period of the capital goods, or ceilings, due to either full employment or the limited productive capacity of the investment goods industries.[1] It often seems as if such factors are introduced mainly to get around the embarrassing results of the linear model. It is true that a non-linear accelerator-multiplier model is not limited to the unsatisfactory alternative states of a linear model. The non-linearity can be introduced in the formal model, as Goodwin and Hicks do, or in a model which determines the value of the accelerator coefficient. Neither Goodwin nor Hicks satisfactorily rationalizes the non-linearities they assume. In particular their analysis of the supply side of the investment goods markets is fragmentary.

The fundamental problem on the supply side of the investment goods market is how resources are made available to the investing unit. This is essentially a savings and a money market phenomenon. It may be true that capital goods differ from consumption goods in the nature of the factors required to produce them, and therefore the short period supply curve of capital goods may have an inelastic range even though financial resources are available for the investing unit. Unemployment may occur when aggregate demand is divided between consumption and capital goods in a manner inconsistent with the structural production rigidities. Assuming aggregate demand is sufficient for full employment such sectoral unemployment may occur in two ways:

1. excess demand (rising prices) in the consumption goods industries with unemployment in the capital goods industries; or
2. excess demand (rising prices) in the capital goods industries with unemployment in the consumption goods industries.

According to the accelerationist doctrine, excess demand in the consumption goods industries implies a rise in the demand for capital goods. Therefore the first type of structural unemployment is typical of the expansion phase of the business cycle. On the other hand, according to the accelerationist doctrine, sectoral unemployment in the consumption goods industries with excess demand in the capital goods industries can occur (a) if there has been a shift in the nature of consumption demand and there exist structural rigidities in the production of consumption goods;[2] (b) if the demand for capital goods lags behind the rise in income and there has been a rise in the savings ratio. Excess demand in the capital goods industries induces investment in the capital goods industries. Only if structural unemployment is due to a rise in the savings ratio does it affect the accelerator-multiplier process, and it affects the business cycle process through the money market. Therefore, the structural rigidities do not enter into the aggregate cycle analysis aside from their effect upon the price level.

Savings are made available to investing units through various financial intermediaries which are inexorably intertwined with the banking (money creating) system. Changes in the money supply or in its velocity of circulation can make resources available to investing units. Therefore the supply of resources to investors is tied up with the behavior of the monetary system. However, the analysis of financial markets, while necessary for the understanding of the process by which investment is realized, is not sufficient. The available financing (plus the consumption expenditures of households) may exceed (or fall short of) the possible output of the investment goods industries (of the entire economy) at the current price level. This can result in price level changes. The effect upon the price level of the relation between aggregate effective demand and supply will be considered in terms of the demand and supply of labor.[3] Essentially, therefore, the monetary aspects of the business cycle will be considered as dealing with the process by which resources are made available to the investing units. The price level phenomenon of the business cycle will be handled as the derived demand for and the supply of labor.

2. MONETARY BEHAVIOR

2.1 Introduction

The monetary aspects of business cycles are primarily associated with two phenomena:

1. the financing of investment during an expansion; and
2. the survival conditions of firms on the downswing.

As was shown earlier, the financing of a firm's expansion may result in a deterioration of its survival conditions. Therefore, the liquidity crises of the downswing can be imputed to the developments of the expansion. Business cycles are both monetary and real phenomena.

> Economic action cannot, at least in capitalist society, be explained without taking account of money, and practically all economic propositions are relative to the *modus operandi* of a given monetary system. In this sense any theory of, say, wages or unemployment or foreign trade or monopoly must be a 'monetary' theory, even if the phenomena under study can be defined in non-monetary terms.[4]

However, there has been a tendency for business cycle theorists to emphasize one or the other aspect in their models. The problem as to whether the business cycle is monetary or real has a deeper aspect. Given the observance of cyclical behavior on the part of the economy, is this behavior due to an 'inherent' instability in the economic process, or is it due to the imposition upon a stable process of outside stimuli? The inherent stability school, in its primitive form, takes the guise of a harvest, weather, or even of a war cycle. The more sophisticated writers of this school have emphasized monetary instability. Their position is that the economy is inherently stable and that the perverse behavior of the monetary system is the fundamental cause of observed cycles.[5]

Among the 'real' school of business cycle theorists are the accelerationists. To these authors the economy is inherently unstable (explosive). Due to the existence of 'floors or ceilings' this results in a cyclical movement of income. As Professor Haberler has pointed out, the non-linear accelerator business cycle theorists have returned to a 'cumulative process' and 'turning point' analysis of the cycle.[6] Due to the asymmetry of investment and disinvestment, Hicks has felt it necessary to rely upon monetary phenomena to achieve the observed rapidity of the downturn. By identifying monetary phenomena as the supply side of the investment relation, we will integrate the monetary and real aspects of the business cycle.

'The main function of the money and capital market is trading in credit for the purpose of financing development. Development creates and nourishes the market.'[7] For our purposes, for 'development' read the expansion phase of the business cycle in the above quotation from Schumpeter's fundamental work, and the approach to monetary phenomena which is necessary in business cycle analysis follows. The essential difference between alternative monetary systems, with respect to business cycles, is in the way they finance the expansion and how they react to falls in income. Both dynamic (growth) and cycle theory have to allow for the financing of investment in excess of voluntary (planned or ex-ante) savings. In addition cycle theory has to allow for the technique by which realized savings in excess of ex-ante investment can be affected.

> The essential function of credit in our sense consists in enabling the entrepreneur to withdraw the producer's goods which he needs from their previous employments, by exercising a demand for them, and thereby to force the economic system into new channels. Our second thesis now runs: insofar as credit cannot be given out of the results of past enterprise or in general out of reservoirs of purchasing power created by past development, it can only consist of credit means of payment created *ad hoc* which can be backed neither by money in the straight sense nor by products already in existence. It can indeed be covered by other assets than products, that is by any kind of property which the entrepreneurs may happen to own. But this is in the first place not necessary and in the second place it does not alter *the nature of the process, which consists in creating a new demand for, without simultaneously creating a new supply of goods*.[8]

Schumpeter's 'insofar as' can be interpreted as 'to the extent that'. Therefore, the funds available for investment consist of savings (the result of past enterprise) and liquid (cash) holdings (reservoirs of purchasing power) and credit means of payment created '*ad hoc*'. Investment in excess of savings must take place for economic development to occur – and the sources of such investment are two: liquid cash holdings and a net expansion of the money supply. We will investigate the relations between such sources of financing and the accelerator models. The effect of financing sources upon realized income depends in part upon institutional arrangements; and we will specify the institutional limitations upon the different financing sources.

> The creation of purchasing power characterizes, in principle, the method by which development is carried out in a system with private property and division of labor. By credit, entrepreneurs are given access to the social stream of goods before they have acquired a normal claim to it. It temporarily substitutes, as it were, a fiction of this claim for the claim itself. Granting credit in this sense operates as an order upon the economic system to accommodate itself to the purposes of the entrepreneur, as an order on the goods which he needs: it means

> entrusting him with productive forces. It is only thus that economic development could arise from the more circular flow in perfect equilibrium. And this function constitutes the keystone of the modern credit system.[9]

> If new credit means of payment, new purchasing power in our sense, are created and placed at the entrepreneur's disposal then he takes his place beside the previous producers and his purchasing power its place beside the total previously existing. Obviously this does not increase the quantity of productive resources existing in the economic system. Yet 'new demand' becomes possible in a very obvious sense.[10]

The Schumpeterian view of the monetary phenomenon is that it makes possible the 'innovationally created' demand for investment goods. The natural view of the role of money for accelerator business cycle theories is that the monetary system makes possible the 'accelerator created' demand for investment goods. Just as in Schumpeter's schema development could not take place without a permissive monetary system, so the accelerator process depends upon the existence of a permissive monetary system. We will consider as monetary changes both changes in velocity (or its equivalent, liquidity) and in the quantity of money. We will take up the reaction of alternative monetary systems to the accelerator process in both the expansion and contraction. The money market is the most significant other market for the accelerator process, as its operation implies that certain changes occur in financial markets which in turn affect the behavior of the acceleration process. This is, of course, a feedback mechanism.

As a result of the economic system accommodating 'itself to the purposes of the entrepreneur' the volume of realized savings differs from the volume of planned savings – savings are 'forced'. The excess of realized savings over planned savings may be real – that is, it may be a change in the quantity of capital goods produced – or it may be monetary, in which case it results in a change in the price level. If the economy has unused resources, the forced savings can be readily realized as real investment. If the economy has no unused resources, the realization of forced savings as real investment depends upon the reaction of saving units to changes in income when the price level is rising. If a rise in the price level does not alter the aggregate volume of real savings, then the effect of monetary changes is to redirect a fixed volume of investment goods among investing units, rather than to alter the volume of realized real investment.

Schumpeter, in his volume on business cycles,[11] considered the sources which can finance an act of expenditure in detail. From a long list he emphasized: (1) previous receipts, (2) overspending ('allowing one's balances to fall below the amount appropriate to the requirements in the previous neighborhood of equilibrium')[12], (3) selling assets, (4) borrowing from banks, and

(5) using one's own, or borrowing, other households' or firms' uninvested savings or accumulations. In our analysis we divide previous receipts into consumption and ex-ante savings. The other four of Schumpeter's list of financing sources can be classified as either changes in velocity or changes in the money supply. Hence there are three sources which can finance investment: ex-ante savings, changes in velocity and changes in the money supply. Of course, in the real world, consumption expenditure may be financed by velocity and quantity of money changes. However, given the consumption function assumption, the volume of ex-ante savings is net of households' utilization of monetary sources for financing consumption, so that the financing sources for investment are as given above.

Hicks, in his volume on the trade cycle, does not consider in detail the relation between the monetary system and the business cycle. 'We have, of course, not denied that the cycle has monetary repercussions; but we have only invoked these repercussions to explain one simple characteristic of the real cycle – the rapidity of the downswing. Excepting in the one connection, the monetary aspect has been kept firmly in the background; the monetary system has been given nothing more than a passive role.'[13] The use of the monetary system to create the rapid downswing implies that during the upswing the position of firms has been affected.

> For consider what must happen if the boom has proceeded without effective monetary check until the real downturn is reached. Output then begins to fall, and effective demand to fall; sales become difficult and fixed costs oppressive; the rate of bankruptcy rises; all these things are inevitable even in the absence of monetary strain, but they *breed conditions in which there is bound to be a sharp rise in liquidity preference*. The rise in liquidity preference is itself the monetary reaction, or what, in common speech, is called the crisis.[14]

Hicks identifies the 'monetary phenomenon' as a rise in liquidity preference due to what, to the monetary system, is an exogenous change in income. The reason for a downturn in income and employment leading to a 'sharp rise in liquidity preference' is not stated: the necessary argument would involve the financing techniques of the expansion. Inherent in accelerator process analysis is the necessity of financing the boom; and the 'effective monetary check' involves the limitation set by a monetary system in realizing induced investment. The expansion phase of the cycle naturally leads to Schumpeter's view of money's role, that expansion is financed in part by debt and therefore implies that the rise in liquidity preference occurs when income turns down.

To Schumpeter's original view of the monetary process we have to add a specific consideration of the liquidity phenomenon. The rise in liquidity preference that Hicks relies upon for the rapidity of the downturn has a counterpart in the expansion phase – liquidity preference falls. This entails (1) that business firms are willing to go into debt – to see their balance sheet deteriorate, and (2) that owners of liquidity are willing to become illiquid – to spend (or invest) their liquidity. These monetary phenomena are essential permissive elements for an accelerator expansion.[15]

As the monetary phenomena are essential permissive elements to an accelerator expansion, we note that the relation between a rise in income and realized investment is due in part to the behavior of the monetary system. The 'accelerator mechanism' can determine the resources that the entrepreneurs who desire to expand desire to command. Financing conditions may make available to such expanding firms a quantity of resources which differ from the accelerator demand. Therefore, we will have to be concerned with the relation between induced and realized investment. The achieved level of income depends upon realized investment. Rather than rely upon monetary phenomena for a single observable phenomenon of the business cycle, as Hicks does, monetary behavior enters into the essential nature of the realized business cycle.[16]

2.2 Ex-ante and Ex-post Savings

In accelerator business cycle models that are written in difference equation form investment is greater than savings during the upswing, and during the downswing savings are greater than investment. Such a Robertsonian savings lag is basic to these models. Savings enter these accelerator models by way of the consumption function. In the Hicks model $Y_t = \alpha Y_{t-1} + \beta(Y_{t-1} - Y_{t-2})$; savings are $(1-\alpha)Y_{t-1}$. If $Y_t > Y_{t-1}$ then $\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$; if $Y_t < Y_{t-1}$ then $\beta(Y_{t-1} - Y_{t-2}) < (1-\alpha)Y_{t-1}$. During an expansion, how is the investment in excess of savings financed? Typically the accelerator theorists ignore this problem. Similarly, the effects of savings in excess of investment on the downswing are ignored.

As correlations to the financing problem we have: how are the resources needed to produce the investment goods made available to the investing units and what happens to the 'price level' during such an expansion? On the downswing savings are not all utilized for the purchase of investment goods. What happens to such 'excess savings' and how are they related to the falling price level of a downswing?

We know that there exist resources in the community, sufficient to produce, at the present price level, Y_{t-1}. By means savings $(1-\alpha)Y_{t-1}$ of resources are available for investment. The demand for investment, given by the accelerator, is $\beta(Y_{t-1} - Y_{t-2})$. $\beta(Y_{t-1} - Y_{t-2})$ is greater than $(1-\alpha)Y_{t-1}$ during the expansion. If all of the induced investment can be financed, then $\beta(Y_{t-1} - Y_{t-2}) - (1-\alpha)Y_{t-1}$ worth of resources in excess of the known resources in the community are demanded for investment. If at the price level of $t-1$, p_{t-1} these resources are available, then all of the induced investment can be realized. The price level does not change. If at p_{t-1} this quantity of resources is not available, and if all of $\beta(Y_{t-1} - Y_{t-2})$ is financed, then p_t will be greater than p_{t-1}. There does not exist any guarantee that $\beta(Y_{t-1} - Y_{t-2})$ of investment can be affected without repercussions upon the *ceterus paribus* assumptions implicit in the value of β.

Symmetrically on a downswing $(1-\alpha)Y_{t-1} > \beta(Y_{t-1} - Y_{t-2})$, savings equal to $(1-\alpha)Y_{t-1} - \beta(Y_{t-1} - Y_{t-2})$ have to find an outlet other than in investment. If such an outlet exists, then resources at p_{t-1} equal to $(1-\alpha)Y_{t-1} - \beta(Y_{t-1} - Y_{t-2})$ are redundant. The repercussions of both the 'savings outlets' and the redundant resources may affect the *ceterus paribus* assumptions implicit in the value of β. The 'excess demand' and 'excess supply' phenomena are dealt with in the second part of this chapter.

The Swedish ex-post and ex-ante language[17] can be used to clarify the relations between savings and investment in such models. Pure accelerator models assume that all of the relevant investment is induced investment, which can be identified as the ex-ante investment. In the determination of ex-ante investment the long run planning curves of the affected firms are a given. The money market adjustment process may affect the long run planning curves. Ex-ante savings are given by the consumption function. During an upswing ex-ante investment is greater than ex-ante savings.

Realized (ex-post) investment is always equal to realized (ex-post) savings. Given the inequality of ex-ante savings and investment this equality of realized savings and investment depends upon changes in either the income level or in the firms planning curves (or both). The ex-post investment and savings can be considered as a *mutatis mutandis* concept where the ex-ante parameters that change are either the level of income or the planning curves of firms. If ex-ante investment is greater than ex-ante savings and realized investment is equal to realized savings, then one of the following must be true:

1. Realized Investment = Ex-Ante Investment and Realized Savings > Ex-Ante Savings. This implies that income rises.
2. Realized Investment = Ex-Ante Savings and Realized Investment < Ex-Ante Investment. This implies that firms' planning curves rise.
3. Realized Investment < Ex-Ante Investment and Realized Savings > Ex-Ante Savings. This implies that both income has risen and that firms' planning curves have risen.

If ex-ante investment is greater than ex-ante savings and there exists no source by which investment can be financed except ex-ante savings, then the ex-ante savings have to be rationed among the investors. The market in which this rationing process takes place is the money market. The excess of demand over supply at the price results in a rise in financing rates and terms. This shifts the planning curves of firms upward. For firms which maximize the rate of return upon owners' equity (aside from certain conditional monopolies) such an upward shift reduces the scale of the optimum plant and thereby lowers investment to that level at which realized I is equal to ex-ante S. In Figure 9.1 ex-ante investment is based upon the interest rate r_1. The rise in income shifts the demand curve for investment, so that the induced investment curve is I'. The inability to finance more than I_1 of investment results in a rise in the interest rate to r_2. Such a 'monetary system', in which savings are the only source by which investment can be financed, leaves no room for an 'accelerator' cycle. A necessary condition for the accelerator process to function is that there exists a source of financing of investment in addition to ex-ante savings.[18]

If in addition to ex-ante savings there exist other sources by which investment can be financed, then such additional financing is added to the ex-ante savings. This may result in realizing all of the ex-ante investment. Then income must rise to such a level that realized savings are equal to ex-ante investment, and the demand curves confronting firms shift upward, again inducing investment. In this case, the accelerator process continues.

If ex-ante investment is greater than ex-ante savings and no financing ability in addition to ex-ante savings is available, then ex-ante investment is greater than realized investment. If, with stable product demand curves, the

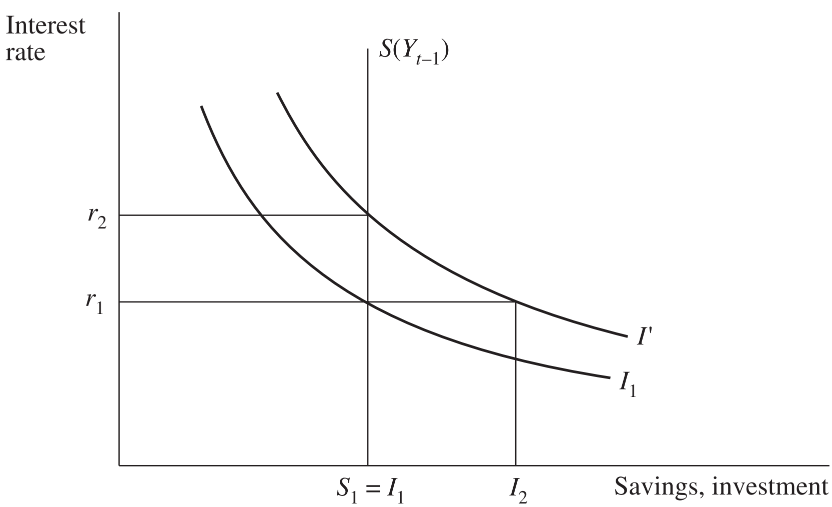


Figure 9.1

financing conditions return to the conditions upon which the original planning curves were based, then investment equal to ex-ante investment minus ex-post investment may be desired. The original induced investment is divided into an 'immediate' investment and a 'lagged' investment by the operations of the financing market. Also, if the financing conditions become more favorable to investment than originally, more investment than the original ex-ante investment will take place. Investment is therefore susceptible to inducement by the operations of the financial markets. Changes in the period over which an investment stimulus operates may transform an 'explosive' movement into a 'damped' movement of income.

On the downswing, ex-ante investment is less than ex-ante savings. As realized (ex-post) investment is always equal to realized savings, adjustments must occur in either savings or in investment. The adjustment in realized savings depends upon changes in income; the adjustment of realized investment depends upon changes in planning curves. If there exists no way in which savings can be utilized but in investment, then the terms upon which firms can finance investment must change so that realized investment is greater than ex-ante investment. This equality of ex-ante savings and realized investment stabilizes income, halting the 'inducement to disinvest'. If there exists a way in which savings can be utilized aside from investment, then income can fall so that realized savings equals ex-ante investment. The existence of a minimum set of financing terms depends upon the existence of a way to utilize savings other than in real investment. Keynesian theory is based upon the existence of a way to utilize savings aside from investment.

2.3 Monetary Changes

The purchase of investment goods by firms can be financed from two sources: one is ex-ante savings; the other is 'monetary' changes. These two sources of financing investment are mutually exclusive and exhaustive. However, as the fractional reserve banking system acts as a financial intermediary as well as a money creating institution, and as changes in the portfolios of financial intermediaries can affect velocity, the identification of the source of the financing of a particular investment is impossible. Ex-ante savings can be utilized in two ways: one is to finance investment; the other is for monetary changes. When realized investment is greater than ex-ante savings, the monetary changes are inflationary; when ex-ante savings are greater than realized investment, the monetary changes are deflationary.[19]

Monetary changes are either a change in the quantity of money or a change in its velocity of circulation. If investment exceeds ex-ante savings some combination of a rise in the quantity of money or in its velocity must occur. If ex-ante savings exceeds investment some combination of a decrease in the quantity of money or a fall in velocity must occur. All monetary phenomena, such as the change in interest rate, associated with business cycle experience are the result of the following:

1. the change in the demand for financing (ex-ante investment);
2. the supply of ex-ante savings;
3. the behavior of the monetary system.

The accelerator gives us the demand for financing. We combine the supply of savings with the monetary changes into a 'supply of financing' schedule. The supply and demand schedules for financing determine realized investment and the interest rate. Realized investment is the essential varying element in an accelerator business cycle model.

In modifying the theory of the firm, the distinction between debt financing and equity financing of investment for a particular firm was introduced. At the end of Chapter 6, it was shown how the sources of financing may affect the ratio of debt to equity financing for the economy as a whole. Total productive assets of business firms rise when investment takes place. This is offset in the balance sheets of firms by one of the following:

1. a rise in either equity or debt on the liability side of the balance sheet;[20]
2. a decrease in either cash or superfluous assets on the asset side of the balance sheet.

The first is consistent with the financing of business investment by the savings of households, by a decrease in the liquidity of households, by the retained earnings of business firms and by the creation of bank money. The second balance sheet change is associated with a decrease in the liquidity of business firms. No matter how financed, the result of net investment is an equal rise in the productive assets of firms and in the net worth of households.

When economic units save, they can finance investment either directly or through financial intermediaries by acquiring either equities or debt assets. If the money supply is increasing, the increased money supply can only finance investment by means of an increase in debt.[21] If velocity is increasing (or liquidity decreasing) the investment that results from the utilization or economizing of cash balances by non-firms may result in either a rise in debt financing or a rise in equity financing. If velocity is increasing because of the utilization of cash balances by business firms neither equity nor debt need rise on firms' balance sheets; the increase in the net worth of households is represented by a rise in their cash balance.[22] In Figure 9.2, the

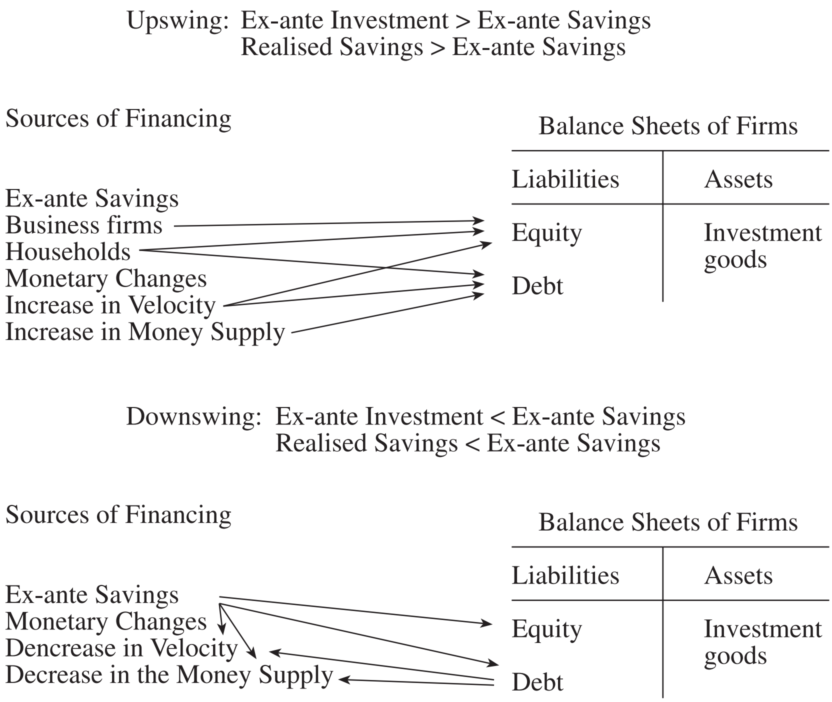


Figure 9.2

arrows indicate how particular financing sources affect the liability structure of business firms.

During the upswing of the business cycle, the excess of realized investment over ex-ante savings is financed by means of an increase in the money supply or by an increase in velocity. During the downswing a portion of ex-ante savings will be absorbed by a decrease in velocity and by a decrease in the money supply. These monetary phenomena are not the cause of the business cycle. Depending upon the nature of the monetary system, the monetary phenomena associated with the business cycle are but the image of the savings and investment phenomena. However, the way in which the monetary system reacts to the changes in savings and investment relations is an important determinant of the realized cyclical pattern.[23]

In what follows we will consider three alternative monetary systems:

1. one in which only velocity changes;
2. one in which only the quantity of money changes

(a) without limit: an infinitely elastic monetary supply;
(b) at a pre-determined rate of growth;
(c) one in which there exists a ceiling to the quantity of money.

Cases 2b and c, where the absolute or time rate of change of the money supply has a limit lead naturally to a monetary system

3. where both the velocity and the quantity of money change.

We can identify the different cases with alternative historical or suggested monetary systems. Case 1, where the quantity of money is constant, is a world in which only the velocity of circulation can change. This is either a world of 100 per cent money, or a world in which at the 'initial point' there exists excess liquidity. Case 2a, where the money supply is infinitely elastic, is a world of a paper money authority which ignores price level considerations, or a world in which a central bank follows a 'needs of business' rule. Case 2b, where the quantity of money has a determinate rate of growth, is a gold standard world where gold mining determines the rate of growth of the money supply. Case 2c, where there exists a ceiling to the money supply, is a gold standard world where the supply of gold is fixed. Aside from Case 1, we assume that there exists a fractional reserve banking system, and that the money supply is changed by either the creation of deposits in exchange for business firms' debts, or the destruction of deposits by business firms' repayment of bank debt.

The model of the banking system that we have adopted is that banks lend to business firms at the initiative of the business firm. Therefore, our model banking system is a 'commercial banking system' rather than a banking system that deals in government and other securities. The reason for doing this is that bank purchase of securities, which creates money in exchange for government bonds, affects business firms only through affecting the liquidity of firms and households. Such open market type operations result in increasing the liquidity of non-bank economic units. Bank lending to business firms enables the business firm to effect an investment plan, and such bank lending decreases the liquidity of business firms. The rise in the liquidity of firms due to open market operations by the banking system affects investment only if the increased liquidity affects financing conditions. The effect of increasing the quantity of money by bank lending to business is direct: the inducement to invest is there and the bank's behavior is necessary to realize the induced investment.

Some of the differences between the Classical Quantity Theory of Money and the Keynesian Liquidity Preference Theory of Money stem from the model of bank operations that is used. The Quantity Theory approach follows from bank lending to business. In this case the lending by the bank enables the business firm to affect a decision made on the basis of assumed financing and income conditions. The impact upon the economy of a rise in the money supply that comes about in this way is direct and immediate: the borrower purchases goods and services. Therefore, propositions such as that a net increase in the quantity of money raises money income follows. The open market model of bank operations involves a substitution of one asset, bank money, for another asset, bonds, in the portfolios of households and firms. There is no immediate and direct impact upon investment. Any effect which such operations have upon investment depends upon the reaction of business firms to the improved liquidity and perhaps lower borrowing rates that follow. In such a world the liquidity preference relation, which is a shorthand for the substitution relation between money and other assets, becomes the appropriate tool to use in the analysis of the behavior of the monetary system.

In all that follows the central bank's relations with the commercial banks are integrated into the 'monetary system'. For example, an infinitely elastic money supply can be affected by a central bank lending to commercial banks, or by a central bank purchasing open market paper. Also in a monetary system we include the specialized financial intermediaries.

2.4 Quantity of Money Constant: Change in Velocity

The velocity of circulation of money and the liquidity preference relation can be characterized as mirror images of each other.[24] When velocity rises, then the liquidity of the economy falls; and vice versa. A useful construction is to assert that for each level of money income Y, there exists a minimum quantity of money M_T which is necessary to sustain the volume of payments associated with Y. If M_T is the total quantity of money in existence, then there is no money available for portfolio use. Therefore, we have a maximum income velocity of money $\bar{V}$ such that for each Y, M_T, $\bar{V} = Y$. If M is greater than M_T then the actual velocity, V, is less than $\bar{V}$. The difference between M and M_T is M_L, the amount of money which is held as a liquid asset. The use of such portfolio money to finance investment increases actual V. If the quantity of money is constant, M_L, portfolio money must fall when actual V rises.[25]

If $V < \bar{V}$ then $M_L > 0$. With $M_L > 0$, the interest rate is determined by the demand curve for investment; ex-ante savings and the terms upon which holders of liquidity are willing to substitute earning assets for money and money for earning assets. If $M_L = 0$, then the interest rate is determined by the demand for investment, the supply of savings and the terms upon which individuals are willing to increase their holdings of portfolio cash. With a given money supply in excess of M_T there exists a rate of interest at which households and business firms in balance are not willing to either increase or decrease their holdings of money. Any market interest rate but this interest rate involves either an increase of cash balance, so that savings are utilized to increase liquidity, or a decrease of cash balance, so that investment is financed by the reservoir of purchasing power.

If we assume a constant money supply, then realized investment can differ from ex-ante savings only if there is a change in the velocity of circulation of money. At any given level of money income, there exists a minimum quantity of money which is necessary for transactions. If the quantity of money in the community is greater than this, it is used as an 'asset'. This asset, held by business firms and households, has unique virtues as to its liquidity and behavior in respect to certain contingencies. This is, of course, the Liquidity Preference rock upon which Lord Keynes founded his monetary theory. No further comment is necessary upon the asset properties of money. We only need to note that during an upswing, the rise in money income implies a rise in transactions money. The utilization of money which had been held as an asset for the purchase of an investment good is the way in which the rise in income is affected. For this to happen, the rise in the demand curves for the product must result in making the giving up of the virtues of liquidity desirable. Either there was an 'excess liquidity' à la the Keynesian liquidity trap or the interest rate, which is a measure of both the portfolio alternatives of individuals and the investment alternatives for business firms, must rise.

Ignoring the liquidity trap, a rise in transactions money as income rises means that, with a constant money supply, portfolio money becomes scarcer. The price at which cash can be withdrawn from portfolios into the income stream becomes higher. The market rate of interest rises as money is withdrawn from the asset fold to finance investment in excess of savings. The effective planning curves of firms which are financing their expansion by debt rises, and the amount of investment induced by a given shift in demand, are reduced.

With a fixed quantity of money and a rise in income the balance sheets of households and firms show a smaller ratio of cash to total assets. The liquidity of the public decreases. The decrease in liquidity raises the survival limits of firms. This in turn raises the effective planning rate of firms more than the rise in interest rates alone indicates. This acts to decrease the amount of investment induced by a given upward shift in aggregate demand.

Alternatively, on the downswing, ex-ante investment is smaller than ex-ante savings. If the supply of money is constant, this excess savings is absorbed by a reduction in velocity. Money available for asset purposes increases as it is withdrawn from the income stream. The interest rate falls and the liquidity of the community rises. The improvement in the liquidity of firms in itself improves the survival position of firms. These changes tend to decrease the amount of disinvestment induced by the given downward shift in demand. On both the upswing and the downswing, the monetary system which is based solely upon changes in velocity acts as a stabilizer of realized induced investment.

Here, however, we must recognize the validity of Keynes' position that there exists an effective floor to the interest rate which is greater than zero. A fall in income may be so great that the money released from transactions purposes lowers the interest rate to the floor rate. In this case the stabilizing effect upon aggregate investment that results from the fall in the planning curves of firms will not take place. Once the interest rate hits the Keynesian floor, the planning curves tend to be stabilized.

The financing of investment and the absorption of savings by means of cash balances are illustrated in Figure 9.3. At the interest rate r_1, and an income Y_{t-1}, the velocity of circulation of money remains constant. This is illustrated by the L_1 curve which intersects the zero change in cash balances line at r_1. At higher interest rates, cash assets would be used to finance investment; at lower interest rates savings will be absorbed by cash balances. The amount of financing available at any interest rate is equal to the sum of ex-ante savings and the change in cash balances. Assume that the accelerator effect shifts the investment curve to I_2. The *ceteris paribus* induced investment is I_2'. With a constant money supply, however, only $I_2'(2)$ of investment can be financed, and the realized

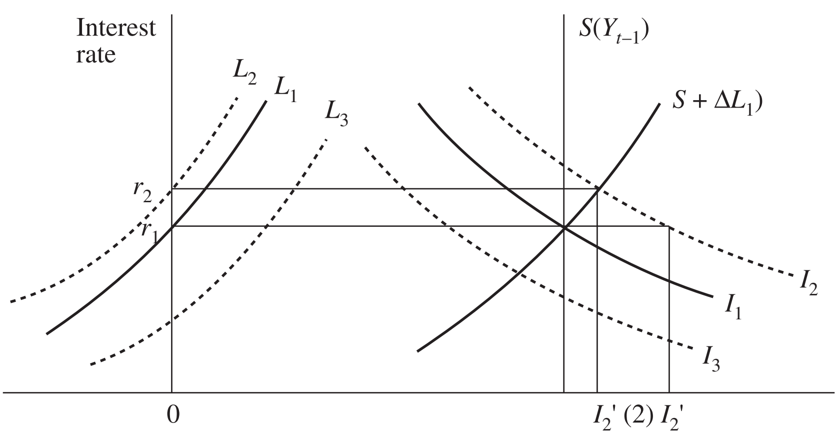


Figure 9.3 The Supply of Finance: Fixed Money Supply

induced investment is $I_2'(2)$. Of the realized investment, OM_1 is financed by a decrease in liquidity and the interest rate will rise to r_2. As $I_2'(2)$ is greater than ex-ante savings, income will rise. The rise in income will increase the transaction demand for cash. This will raise the schedule of the change in cash balances. At this higher level of income the rate of interest at which investment will be financed by a fall in liquidity will be higher.

If a fall in income shifts the investment demand curve to I_3, then the *ceteris paribus* induced disinvestment is I_0'. With a constant money supply, however, the excess of ex-ante savings over induced investment will depress the interest rate, and realized investment will be $I_3'r > I_3'$. ΔM_2 will be added to cash balances. As $S > I_3(r)$, income will fall. This will shift the liquidity curve downward so that cash balances can be used to finance investment at an interest rate lower than r_2.

The Keynesian Liquidity Trap situation can also be illustrated in Figure 9.4. If the investment curve is I_1, there will be no change in income or in cash balances. If the investment curve is I_2 and income is Y_1, $I_2' - S$ of investment will be financed by a decrease in cash balances. If the investment curve is I_3, $S - I_3'$ will be added to cash balances. No change in interest rates will occur; the planning curves will remain fixed; all induced investment will be realized. In a Keynesian liquidity trap the behavior of the financial markets dampens down neither the 'boom' nor the 'crash'. On the boom side, this will continue until the transactions cash absorbs a sufficiently

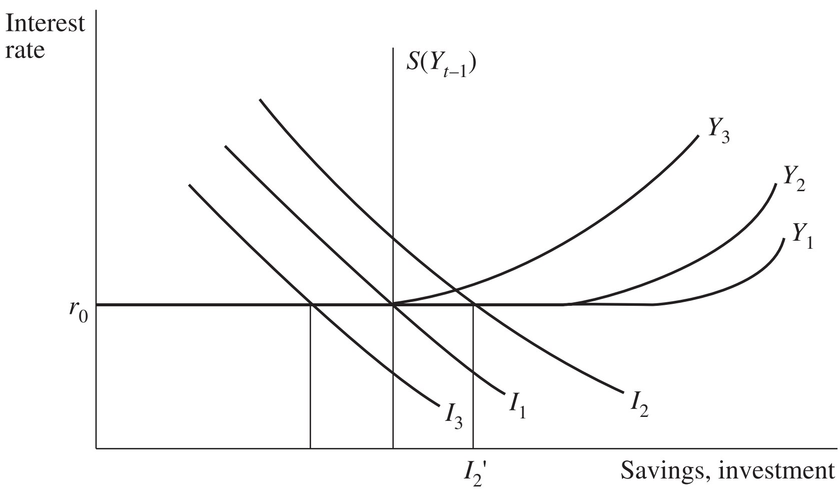


Figure 9.4 The Keynesian Liquidity Trap

large portion of the available cash so that the Keynesian liquidity trap comes to an end. There is no end to the liquidity trap on the downswing. In some ways a great investment boom, such as the post-war boom of 1946 to 1953 can be interpreted as an accelerator boom with a permissive liquidity trap foundation.

Realized investment greater than ex-ante savings raises money income. The rise in money income, in turn, induces investment. The cumulative rise in income which results increases the quantity of money needed for transaction purposes. This results in progressively smaller asset holdings of money which are available to finance investment in excess of savings. The highest attainable level of money income is that level of income at which all of the available money supply is used for transaction purposes.[26] At that income level realized investment cannot exceed ex-ante savings. Realized investment equal to ex-ante savings results in a constant income under accelerator assumptions. A constant income level induces zero investment. Independently of what happens to interest rates as asset holdings of money decrease, and of the effects of interest rate and liquidity changes upon the planning curves, a monetary system with a fixed quantity of money has a money income ceiling.[27] This ceiling is not determined by full employment, nor by the capacity of the investment goods industries. The ceiling is determined by the limited financing ability which exists due to changes in velocity.

The fall in investment that occurs when the money income ceiling is reached results in a fall in money income. Transaction money falls, and savings which are not realized in investment take place by adding asset money to portfolios. If the price level does not fall, the ceiling real income remains fixed. If the price level falls, then the ceiling money income remains fixed, but ceiling real income rises.

Net investment implies an increase in productive capacity. With a constant money supply and a maximum velocity the larger real incomes can be realized only if the price level falls. The accelerator inducement to invest is large only when quantity demanded is at least approximately equal to productive capacity. Hence for the accelerator to be a source of strong cycles with a constant money supply, the price level must be falling.

The effect of expectations that the price level will fall is to increase the pay off periods of investment. This is equivalent, in its effect upon firms' planning curves, to a rise in interest rates with a constant price level. The falling price level will therefore tend to lower the value of the accelerator coefficient. This will mean that the business cycle will be characterized by weaker booms than alternative monetary systems would provide. Such a monetary system will be associated with a tendency toward relatively stable income; the accelerator being damped down, the investment booms will tend to be short lived. Long periods in which realized investment exceeds ex-ante savings will not occur.

2.5 Quantity of Money Changes

Realized investment may differ from ex-ante savings due to a change in the quantity of money which may vary (a) without limit, (b) at an independently determined rate, or (c) with a ceiling upon the quantity of money.[28] If the quantity of money varies without limit then no matter what the difference between ex-ante investment and ex-ante savings, the difference can be made up by a change in the quantity of money. The terms upon which the banking system lends do not change. These terms will determine the interest rate upon which the effective planning curves of firms maximizing the return upon owners' equity will be based. Such a monetary system is consistent with the existence of an explosive accelerator process as it permits a cumulative rise in money income to take place. Is there anything inherent in the operations of such a monetary system which will lead to a dampening of the accelerator process? In taking this question up we will ignore the political repercussions of a cumulative rise in prices which is implicit in a full employment situation where the rate of growth of money income is greater than the rate of growth of productive capacity. In such a model of necessity, all of the variables and parameters refer to money income.

Assume that commercial banks create money by lending to business firms. The increase in the money supply is equal to the difference between induced investment and ex-ante savings:

$$M = I \text{ ex-ante} - S \text{ ex-ante} = \Delta Y$$

Assume that $V = Y/M = \Delta Y/\Delta M = 1$. The increase in the money supply in the hands of households or of business firms, other than the borrowing firms, constitutes the unexpected savings which result in ex-post savings being equal to ex-ante investment.[29] As the velocity is 1, there will be no net change in the quantity of money that individuals wish to hold as assets. This is equivalent to assuming that the interest rate at which banks lend to business is the interest rate at which money and earning assets are substituted in portfolios.[30] If at the higher level of income and higher net worth due to the realized investment, households desire to change their cash balances, let us assume that open market operations by the central bank neutralize the effect of cash balance changes. The only relevant monetary change in this model is in the quantity of money.

The supply of money is infinitely elastic at a given interest rate and the liquidity curve is such that the change in cash holdings at the rate is zero.

Ex-ante savings are assumed to be interest inelastic. If ex-ante investment is greater than ex-ante savings, the money supply will increase by the difference. If ex-ante investment demand is less than ex-ante savings then the money supply will decrease by the difference. Because of our special velocity assumptions the change in income and the change in the money supply offset each other, so that the liquidity curve does not shift.

During an expansion, the increase in the money supply takes place through investing business firms adding bank debt to their liabilities. Assuming that the percentage distribution of ex-ante savings between debt and equities is constant, a cumulative explosive expansion on the basis of the creation of money will result in a fall in the ratio of equity to debt in the balance sheet of firms. Even if the terms upon which firms can borrow are unchanged by the deterioration of their balance sheets, borrowers' risk will rise. This will raise the effective planning curves of firms which will lower the amount of investment induced by a given rise in income.[31] Even assuming that the monetary system permits all of ex-ante investment to be realized, the financing of investment by bank debt can result in lowering induced investment which in turn lowers the rate of increase of income.

The higher ratio of debt to equity raises the survival limits of business firms. The smaller rate of increase of income decreases induced investment. With a fall in income, the excess of ex-ante savings over induced investment will be utilized to reduce debt. Also, the failure of firms whose survival limit has risen during the expansion will result in the substitution of equity for debt in balance sheets. Both changes during the downswing raise the ratio of equity to debt in firms' balance sheets,[32] which in turn lowers their effective planning curves. This acts as a stabilizer on the downswing. The endogenous limitational factors upon an explosive accelerator process in the absence of restrictions in the money supply are the deterioration of firms' balance sheets which occurs when debt is used to finance investment on the upswing, and the improvement of firms' balance sheets during the liquidation process that occurs on the downswing.[33]

As ex-ante savings are a decreasing portion of total investment in such an explosion process, an increase in the ratio of savings that flow to equities may, for a time, prevent a deterioration in the balance sheets of firms. However, in time the accelerator process will begin to so dominate the balance sheet ratios that the balance sheets of firms must deteriorate. If cumulative price level inflation is 'politically possible' such a deterioration of balance sheets need not occur. Business firms are borrowers and the burden of a debt decreases with a rise in the price level. If the assets of business firms are valued at their current replacement costs, then the rising price level raises the equity account. Such capital gains improve the balance sheet of firms and they occur generally in an inflation. The price level rise, plus the flow of ex-ante savings to equity investment may be sufficient to keep the debt equity ratio constant: thereby preventing any deterioration in the balance sheets of firms. However, this requires an increasing rate of change in the price level.[34] Nevertheless, if an explosive inflation is politically tolerable, there is no endogenous reason why an accelerator process with an infinitely elastic money supply need come to a halt.

Therefore, at least two monetary situations allow full scope to an explosive accelerator process. These are the Keynesian liquidity trap and an infinitely elastic monetary supply. If we accept Hicks' intuitive judgment as to the value of the accelerator coefficient (the coefficient of ex-ante investment) then the realization of induced investment depends upon the existence of an elastic monetary system. It is perhaps no accident that his volume on the trade cycle appeared at a time when the high volume of government bonds outstanding and their support by central banks made the money supply in fact infinitely elastic. An era of tight money, on the other hand, naturally leads to an examination of the monetary prerequisites to the operation of the accelerator phenomenon.

A monetary system in which the rate of growth of the money supply is exogenously given, for example, a fractional reserve banking system based upon a gold standard, is equivalent to an infinitely elastic monetary supply unless the difference between ex-ante investment and ex-ante savings exceeds the rate of growth of the money supply. The only limitation to expansion in this case comes from the deteriorating liquidity of business firms, as is true with an infinitely elastic money supply. The interesting alternative is when the difference between induced investment and ex-ante savings is greater than the rate of growth of the lending ability of banks.

Let us consider a Hicks type accelerator model:

$$Y_t = \alpha Y_{t-1} + \beta(Y_{t-1} - Y_{t-2})$$

Assume $\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$ so that $Y_{t+1} > Y_t$ if all of the induced investment is realized. Assume a constant arithmetical increase in the money supply per time period: $\Delta M = \sigma$. Then the total amount of financing available assuming no excess liquidity in the community is $(1-\alpha)Y_{t-1} + \sigma$. Only if $\beta(Y_{t-1} - Y_{t-2}) \geq (1-\alpha)Y_{t-1} + \sigma$ is the monetary constraint effective, and in this case realized investment is $(1-\alpha)Y_{t-1} + \sigma$. Then $Y_t = \alpha Y_{t-1} + (1-\alpha)Y_{t-1} + \sigma = Y_{t-1} + \sigma$ and $\beta(Y_t - Y_{t-1}) = \beta\sigma$; income rises at the arithmetic rate σ. Eventually $\beta(Y_{t+n} - Y_{t+n+1}) = \beta\sigma < (1-\alpha)(Y_{t-1} + n\alpha) + \sigma$, so that $Y_{t+n+1} < Y_{t+n} + \sigma$. Therefore $\beta(Y_t - Y_{t-1}) < \beta(Y_{t-1} - Y_{t-2})$ and the accelerator process will turn down.

If the money supply increases at a constant arithmetic rate, then income will grow at this rate until ex-ante savings increase so that induced investment is less than the available financing. The increase of income will be smaller than the previous rise; therefore lowering induced investment. The cumulative upward process proceeds at a rate given by the increase in the money supply, and the downturn occurs when voluntary savings catch up with the expansion process so that all of the investment induced by the constant arithmetic rate of growth of income can be realized without using all of the newly available credit. After this, income turns down and the constant arithmetic increase in the lending ability of banks increases liquidity. This, of course, may shift planning curves downward by improving firms' financing conditions, which would tend to stabilize income.[35] (In addition, a fall in interest rates may stabilize the economy.)

Alternatively consider a money supply which increases at a constant rate of growth, say at a constant ratio to savings: $\Delta M = \mu(1-\alpha)Y_{t-1}$.[36] We again use a Hicks model $Y_t = \alpha Y_{t-1} + \beta(Y_{t-1} - Y_{t-2})$ and $\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$ so that if the induced investment is realized, income will rise. The autonomous increase in financing ability of banks is $\mu(1-\alpha)Y_{t-1}$, so that $(1+\mu)(1-\alpha)Y_{t-1}$ of investment can be financed. If $(1+\mu)(1-\alpha) Y_{t-1} \geq \beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$ then the monetary constraint is not effective. The model behaves as if the monetary system were infinitely elastic. If $\beta(Y_{t-1} - Y_{t-2}) \geq (1+\mu)(1-\alpha)Y_{t-1}$ then $(1+\mu)(1-\alpha)Y_{t-1}$ is the realized investment so that $Y_t = \alpha Y_{t-1} + (1+\mu)(1-\alpha)Y_{t-1} = [1+\mu-\alpha\mu]Y_{t-1}$.

Therefore, for t+1, induced investment is $\beta(\mu(1-\alpha))Y_{t-1}$. If $\beta\mu(1-\alpha) Y_{t-1} \geq (1+\mu)(1-\alpha)Y_{t-1}$ implies that $\beta \geq [1+\mu]/\mu$, then realized investment is constrained by the financing ability of the community and income increases at the geometric rate $(1+\mu-\alpha\mu)$. If $\beta < [1+\mu]/\mu$, then all of the induced investment can be realized as with an infinitely elastic money supply. The maximum rate of growth of income is the rate of growth of the money supply.[37]

Two points are worthy of notice. First, as long as $\beta < [1+\mu]/\mu$ and the increase in the money supply is truly autonomous, then lending ability accumulates in the banks. If β increases so that $\beta > [1+\mu]/\mu$, then realized investment can exceed $(1+\mu)(1-\alpha)Y_{t-1}$ until such time as the reservoir of lending ability is exhausted. At that time realized investment will fall to $(1-\alpha)(1+\mu)Y_{t-1}$. This results in a fall in the rate of growth of income which may lower β, so that income falls. By means of such a mechanism, a backlog of lending ability may lead to a fall in income.[38]

The second point is that the rate of growth of income is not determined by the properties of the accelerator process but by the properties of the monetary system. This throws some light on the 'gold production' theories of the business cycle advanced by Cassel and others.

In the case where the money supply is growing at a constant rate, the deterioration of the firms' balance sheets during an expansion, which we noted for an infinitely elastic monetary system, will occur if the financing of the expansion is inconsistent with the balance sheet structure of firms prior to the expansion. However, as the rate at which the money supply can increase is limited, then, as long as a major portion of ex-ante savings is used for equity financing, the balance sheets of firms cannot deteriorate as far or as fast as in the case of an infinitely elastic money supply.

If the change in the quantity of money is the effective constraint upon expansion, is there any attribute of the process which can lower the rate of growth of income? The limitation upon the quantity of money implies that the interest rate will rise so that realized investment is equal to the available financing which is sufficient to keep the expansion going. Unless there is a lag in the shift of the planning curves, the rise in interest rates is not sufficient to halt the expansion. However, the combined effect of balance sheet deterioration and the rise in interest rates may lower induced investment so that induced investment is less than the available financing at the institutional minimum financing terms. These factors do not on 'a priori' grounds seem very strong. A constant rate of growth money system seems conducive to a business cycle with lengthy expansions and short contractions.

The monetary system may be such that the quantity of money has a ceiling. Given an explosive upward movement of income, the money supply will expand, permitting the financing of investment in excess of income. When the monetary ceiling is reached, investment will fall to the quantity that can be financed by savings. This will stabilize income; therefore the next period's induced investment will be zero. Such a monetary system will result in a sharp break in the level of income. Table 9.1 indicates how an accelerator model with a ceiling to the money supply behaves. It is assumed that realized investment is equal to induced investment as long as the money supply is available to finance the investment. When the money supply reaches its ceiling the only source available by which investment can be financed is ex-ante savings, and realized investment falls to ex-ante savings.

If the rate of change of the ceiling money supply is very small, for example, if new gold is but a small percentage of the reserve gold, the banking system will have an essentially fixed peak to its lending ability. Such a financing system may break an explosive boom very rapidly. In terms of interest rates, the development until period 3 would have been at unchanged financing terms, period 3 would have seen a rise in financing terms, and period 4 would be a period of 'panic' money market rates. The inducement to invest is still very strong, but the available financing is current savings. The

Table 9.1

Accelerator Process $\alpha = 0.8$ $\beta = 4$					Monetary System +10 may change Money Supply		
Time	Y	C	S	Induced $\beta(Y_t - Y_{t-1})$	I (realized)	Financed by Δ in Money Supply	Available Financing
0	10	8	2	–	3	1	9
1	11	8.8	2.2	4	4	1.8	7.2
2	12.8	10.2	2.6	7.2	7.2	4.6	2.6
3	17.4	13.9	3.5	14.4	6.1	2.6	0
4	20.0	16.0	4.0	10.4	4	0	0
5	20	16	4.0	0	0	−4	4
		etc.					

behavior of a system with an explosive accelerator and a monetary ceiling would include a financial panic. In many ways the breaking of strong booms in the period of the international gold standard conforms to the pattern of such a monetary system.[39]

2.6 Changes in Quantity and Velocity of Money

In taking up the case where investment in excess of ex-ante savings is financed by an increase in the quantity of money, we assumed that the velocity of circulation of money was 1. We can now drop this assumption. If velocity is greater than 1, and if an excess of investment over ex-ante savings is financed by an increase in the quantity of money, then excess liquidity results. This excess liquidity can be utilized to finance investment.

Assume that the excess liquidity resulting from an initial bank financing of investment is used to substitute business debt or equities to the public for business debt to banks. If $\Delta M = Y_t - Y_{t-1}$ and $V > 1$, then transaction cash is $\Delta M/V$, and asset cash is $\Delta M - [\Delta M/V] = (1 - [1/V])\Delta M$. The net increase in debt to banks is $1/V(Y_t - Y_{t-1})$ and investment is $Y_t - \alpha Y_{t-1}$.

We therefore have that:

$$\frac{\Delta \text{BankDebt}}{\Delta \text{TotalAssets}} = \frac{[Y_t - Y_{t-1}]/V}{Y_t - \alpha Y_{t-1}} = \frac{1}{V}\left[\frac{Y_t - Y_{t-1}}{Y_t - \alpha Y_{t-1}}\right].$$

As an explosive accelerator process takes hold, the ratio $[Y_t - Y_{t-1}]/[Y_t - \alpha Y_{t-1}]$ rises, the limit (as Y_{t-1}/Y_t approaches 'zero') being one. Therefore, the ratio of change in bank debt to change in total assets approaches $1/V$. If the distribution of the public's investments in business firms is unchanged as between debt and equity assets, during an expansion the balance sheets of business firms will deteriorate. As the weight of bank financing is less than in the case of unit velocity, however, this will not be as rapid as in the case where bank creation of money is the sole technique by which investment in excess of ex-ante savings can be financed. Therefore, the possibility that the deterioration of firms' balance sheets will reduce the efficacy of a rise in income to induce investment will be lower.

In the same expression $1/V([Y_t - Y_{t-1}]/[Y_t - \alpha Y_{t-1}])$ we note that a rise in velocity reduces the ratio of bank financing to total change in assets, and that a rise in the propensity to consume increases the dependence upon bank financing of investment. Therefore, autonomous or cyclically induced changes in these parameters alter the ratio of debt to equity financing. Also as bank debt of firms is generally shorter term than firms' debt to the public, a change in these parameters will affect the dating of business firm liabilities. This can operate so as to affect the survival constraint of firms. Note that a rise in velocity operates so as to decrease the rise in the planning curves due to the deterioration of firms' balance sheets that occurs during a business cycle expansion.

In the case of $V > 1$, just as in the case where $V = 1$, a ceiling to the quantity of money, either in terms of its absolute level or rate of growth, will break the explosive accelerator boom. The only difference will be that whereas when $V = 1$, $\Delta Y = \Delta M$, when $V > 1$, $V\Delta M = \Delta Y$. The effect of hitting the ceiling in the money supply will be a downturn in income. However, the possibility exists that velocity changes will counteract the effect of the ceiling in the money supply.

Autonomous or cyclically induced changes in velocity, or liquidity preference, can change the dependence of an expansion upon changes in the money supply and therefore affect the ratio of bank debt to total assets in firms' balance sheets. If liquidity preference decreases, the excess of investment over ex-ante savings may be financed by withdrawals from cash balances at an interest rate lower than the bank interest rate. Such an 'autonomous' decrease in liquidity preference can, by both improving financing terms and decreasing the dependence of business firms upon bank financing, lower the planning curve of firms. A stock market boom, such as in the late 1920s, may be interpreted as a lowering of liquidity preference; as a result the financing of business expansion independently of the banking system was easily achieved. This succeeded in lowering planning curves and therefore strengthened the 'investment boom'.

Alternatively, an autonomous rise in liquidity preference may lead to business borrowing from banks increasing the liquidity of households rather than financing investment. A liquidity preference curve, as drawn in

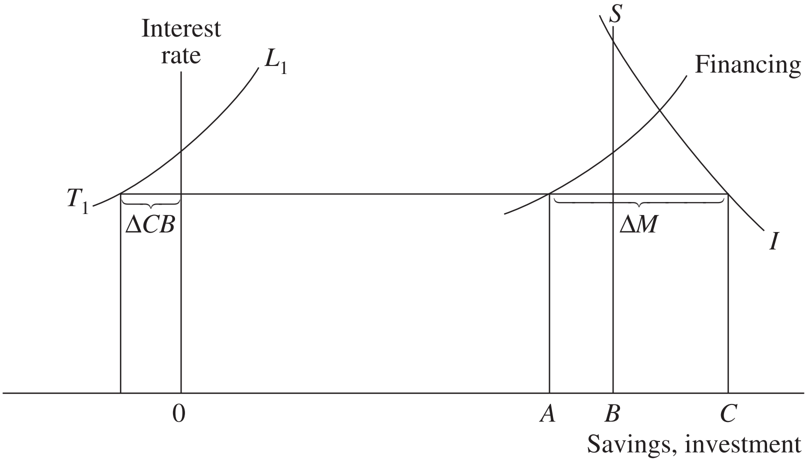


Figure 9.5

Figure 9.5, would result, assuming the money supply is infinitely elastic at the interest rate r_1, in business borrowing ΔM from banks, of which AB will end up as 'liquid hoards' of households and BC will end up as unexpected cash balances of households. This borrowing by business firms in excess of the difference between ex-ante savings and realized savings will increase the rapidity with which firm balance sheets deteriorate. The rise in income BC, may be insufficient to yield an investment demand curve in the future to the right of I_2 because of the increase in risk elements. Therefore, an explosive accelerator process may be broken by such changes in liquidity preference.

Such changes in liquidity preference have been labeled autonomous. There exist plausible mechanisms by which the upward movement of an explosive accelerator process would lead to a fall in liquidity preference. However, there does not exist any equally plausible mechanism by which a rise in liquidity preference can be considered as endogenous during an expansion. During a downswing there exists a plausible mechanism which can raise the liquidity preference curve of households. This can force a deterioration of firms' balance sheets, and thereby, through its effect upon planning curves, a fall in investment.

If the money supply is not infinitely elastic, but is an increasing function of the interest rate, then as the financing from cash balances relation is an increasing function of the interest rate, the available financing curve is the horizontal summation at each interest rate of the three sources of financing: savings, changes in velocity and change in the cash balance. The effects of a ceiling on the money supply can, of course, be eased by a rise in velocity. If velocity is a function of the interest rate, then at the ceiling money supply, or when induced investment exceeds the financing possibilities inherent in the limited rate of growth of the money supply, the growth of income can be maintained by increasing velocity. As this takes place at rising interest rates, the efficacy of an inducement to invest gradually declines. However, the limitation that $Y_t = Y_{t-1}$ when the ability to increase the quantity of money is exhausted no longer holds. This transforms a sharp break in the level of income into a gradual decrease of the rate of increase of income.

Government deficit financing with an infinitely elastic money supply increases the cash holdings of the public without an equal increase in business debt to banks. Such a situation results in improving the liquidity of both households and firms. As a result of deficit financing, the ratio of bank debt to total assets of business firms falls, and household ownership of cash and government bonds increases. Hence, if the investment demand curve is shifted upward so that 'induced investment' is greater than ex-ante savings, no financing difficulties should lie in the way of realizing the firms' investment plans, and this financing does not entail a deterioration of firms' balance sheets.

2.7 Summary: Money and the Accelerator Models

The scope that an explosive accelerator process has depends upon the monetary system. The accelerator process requires a permissive monetary system. This permissive monetary system may exist due to either a 'liquidity trap' or an infinitely elastic money supply. Such monetary systems do not impose any constraint upon the explosive accelerator process. If the monetary system is neither completely elastic nor one with an infinite excess cash balance, then the monetary system either determines the maximum rate of growth of income or the maximum level of income. In such cases the business cycle of experience is neither real nor monetary, but a combination of the two.

For infinitely elastic money supplies, the only endogenous limitational factor upon the operations of the accelerator process is the deterioration of the balance sheets of business firms due to the increased weight of bank financing to total investment during the expansion. This may result in shifting planning curves upward through the operation of the survival conditions. However, such a break to the explosive process is neither as certain nor as quick as the break in expansion that operates by way of the ceilings on money supply and on money velocity.

2.8 The Price Level and the Savings Coefficient

As a corollary to the emphasis upon the mechanical properties of the models in the accelerator-multiplier formulations of business cycle theory, the content of the accelerator coefficient is ambiguous. In particular whether the model is to be construed as referring to real or to money demand for investment is in general hazy. If a model of the business cycle is to be complete it must be possible to interpret the formal results as leading to a pattern of price level changes consistent with observed values. Therefore, in the accelerator models, it is necessary to be able to interpret the coefficient of induced investment as a relation between money investment and aggregate money demand. Such an interpretation of the investment coefficient is an additional break in the relation between realized investment and production techniques. As it is necessary to be able to interpret the accelerator process in money terms, the savings coefficient can no longer be abstracted from the effects of price level changes.

We know that at the price level of a particular time, p_{t-1} there exist resources sufficient to produce a money income Y_{t-1}. During the expansion phase of an accelerator business cycle $\beta(Y_{t-1} - Y_{t-2}) > (1 - \alpha) Y_{t-1}$. If p_t is to be equal to p_{t-1} then there must exist in the community at time t resources sufficient to produce $Y_t - Y_{t-1}$ of output, which resources were unused during time period $t - 1$. Such resources available to produce Y_t and not used to produce Y_{t-1} have two sources:

1. any change in productive capacity between period t and $t - 1$;
2. unused resources of period $t - 1$.

Productive capacity of period t can differ from that of period $t - 1$ due to the following factors:

1. any change in production technique;
2. any change in population;
3. the investment that took place in period $t - 1$ and prior periods which is first available for productive use in period t.

The first two we will ignore. In order to abstract from complex lag patterns, assume that the addition to productive capacity in period t due to investment is all due to the investment of period $t - 1$. This increase in productive capacity can be written as γI_{t-1}. We can also write the productive capacity of unused resources in period $t - 1$ as $\bar{Y}_{t-1}$.

If p_t is not to be greater than p_{t-1}, then $\beta(Y_{t-1} - Y_{t-2}) \leq (1 - \alpha) Y_{t-1} + \gamma I_{t-1} + \bar{Y}_{t-1}$. If β and γ are in real terms, then β is the capital per unit of output and γ is the output per unit of capital. Therefore, $\beta = 1/\gamma$. As I_{t-1} is, in accelerator models, $\beta(Y_{t-2} - Y_{t-3})$ we have: $\beta(Y_{t-1} - Y_{t-2}) > (1 - \alpha) Y_{t-1} + Y_{t-2} - Y_{t-3} + \bar{Y}_{t-1}$. If β, α are such that they lead to an explosive accelerator process then in time

$\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1} + Y_{t-2} - Y_{t-3} + \bar{Y}_{t-1}$. Therefore, $\bar{Y}_{t-1}$ will be absorbed, and in time $\bar{Y}_{t-1} = 0$. If $\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1} + I_{t-1}$; and $\bar{Y}_{t-1} = 0$, then $p_t > p_{t-1}$; therefore, an explosive accelerator process cannot continue without leading to a rise in the price level. That is, the model leads to a position at which consumption demand plus investment demand is greater than the productive capacity of the economy. This is the full employment ceiling which Hicks uses to break the explosive boom. We are willing to consider a model of the explosive accelerator process that results in $p_t > p_{t-1}$, so that $p_t = p_{t-1}$ during an expansion until $\bar{Y}_{t-1} = 0$, $\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1} + (Y_{t-2} - Y_{t-3})$, at which time $p_t > p_{t-1}$.

If the unused resources of period $t-1$, $\bar{Y}_{t-1}$,[40] are represented as a quantity of unemployed labor, the price level phenomena can be considered as a labor market relation. The effect of investment with full employment is a substitution of capital for labor in production, which has been called a deepening of capital, whereas the effect of investment with less than full employment is a 'complimentary' expansion, which has been called widening of capital. A problem in the interpretation of the accelerator coefficient is whether or not the assumption of the investing firm is that the expansion of plant will result in a proportionate increase in the labor force, a widening type of expansion, or whether the increase in output will take place with the same labor force, a deepening type of expansion. If there is full employment, an expansion of plant can increase output only by means of the substitution of capital for labor in production. Therefore, the coefficient γ, output per unit of investment, can be expected to fall as the economy approaches full employment.

Assuming that $Y_{t-1} > Y_{t-2}$ and that Y_{t-1} is a full employment period, Y_t will be less than Y_{t-1} if $(1-\alpha)Y_{t-1} > \beta(Y_{t-1} - Y_{t-2})$. As a result of this, and adding the productive capacity newly available at time t due to the investment of period $t-1$, γI_{t-1}, resources sufficient to produce $(1-\alpha)Y_{t-1} - \beta(Y_{t-1} - Y_{t-2}) + \gamma I_{t-1}$ at the price level of period $t-1$ are unused. These unused resources may or may not result in a fall of the price level. This depends upon the nature of the supply curve of labor. If the supply curve of labor is the extreme Keynesian type at the wage rate ruling when the price level is p_{t-1}, then the unused resources $\bar{Y}_t$ will not lead to a fall in wages and output will fall. This will lead to excess productive capacity and induced investment will, in strict accelerationist terms, become negative.

If the supply curve of labor is classical, and if the time rate of change of wages is fast enough, then there will be a rise in real output in period Y_t, and no unused resources will appear. This, however, depends upon an assumption in regard to the elasticity of aggregate consumption demand and aggregate investment demand with respect to a fall in the price level. If the elasticity is equal to or greater than 1, then total expenditure will be constant or rise with a fall in prices. The excess supply of factors leads to a fall in wages. This lowers marginal cost curves. Assuming αY_{t-1} will be spent on consumption goods, output of consumption goods will rise. As a result of the increase in output of consumption goods, the optimum scale of consumption goods plant will be larger. If the time rate of changes in wages is sufficiently rapid, then p_t will be sufficiently lower than p_{t-1} so that $Y_t/p_t = [Y_{t-1} + \gamma I_{t-1}]/p_{t-1}$.[41] Due to the existence of debts of firms based upon $p_{t-1} > p_t$ the fall in the price level raises the survival limits of firms. This raises the planning curves of firms so that the coefficient of induced investment will fall. This will again tend to result in $(1-\alpha)Y_t > \beta(Y_t - Y_{t-1})$, at the price level of p_t. This process will continue, lowering the price level still more. To the extent that cumulative deflation is possible, the accelerator depression is a price level rather than a real income depression.

If the time rate of change of wages is greater than zero, and less than that rate sufficient to result in $Y_t > Y_{t-1}$ in real terms, then the downswing will have both employment and price level attributes. As a result of the sticky wage levels, excess capacity will appear. With excess capacity, the inducement to invest becomes zero or negative. This is to be contrasted with the completely flexible wage system where real income increases while the price level is falling, and therefore is consistent with positive induced investment. The result of such sticky wages is intermediate between the classical and Keynesian case in the first instance; however, as a 'type' it more nearly approximates the Keynesian case due to the effect upon induced investment of excess capacity.

Assume $\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$ and that the monetary system is such that all of the induced investment is realized. The ratio of realized investment to income is greater than the ratio of ex-ante savings to income. That is $(Y_{t-1} - Y_{t-2})/Y_t > (1-\alpha)$. If Y_t is greater than Y_{t-1} in real terms, so that αY_{t-1}, the consumption of period t, is greater than αY_{t-2}, the consumption of period $t-1$, the difference between realized savings and planned savings does not lower consumption. On the other hand, if p_t is greater than p_{t-1}, then the forced savings can result in lowering the level of real consumption during an expansion. Assume that Y_{t-1} is a full employment level of income. Ignore the effect of I_{t-1} upon productive capacity during period t. If all of $Y_t - Y_{t-1}$ is due to a rise in prices, then the realizing of all of $\beta(Y_{t-1} - Y_{t-2})$ lowers the absolute level of consumption in period t below the level of period $t-1$. On the other hand if $Y_t > Y_{t-1}$ in real terms, then $C_t > C_{t-1}$ in real terms.

We can include this effect in our model by assuming that the full employment level of income ratio of consumption to income is the minimum ratio

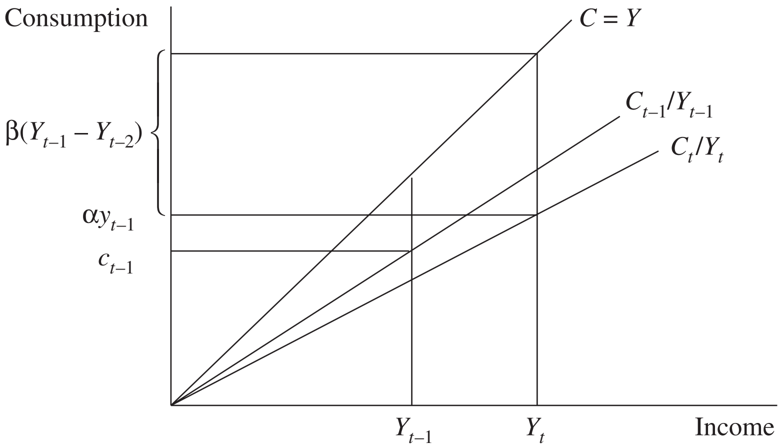


Figure 9.6

of C/Y that can be affected, and that any expansion of income in money terms beyond this will not lead to an increase in the ratio of investment to income.

During an expansion $\beta(Y_{t-1} - Y_{t-2}) - (1 - \alpha) Y_{t-1}$ of unplanned savings take place. If these savings can be realized without lowering the level of planned consumption, for example if $Y_t = \alpha Y_{t-1} + \beta(Y_{t-1} - Y_{t-2})$ can all be produced at a constant price level, let us assume that no change in consumption behavior takes place. However, if Y_t cannot be produced, if prices rise, then as long as only αY_{t-1} is spent on consumption goods, the real level of consumption falls. In this case forced savings in a real sense takes place. If we assume that consumers resist forced savings then money income will rise so that all of realized savings are wanted savings. In this case $(Y_{t-1} - Y_{t-2})/Y_t' = 1 - \alpha$ or $Y_t' = \beta(Y_t - Y_{t-1})/[1 - \alpha]$, $Y_t' > Y_t$.[42] A resistance on the part of the workers to a fall in their real standard of living will tend to accentuate the rise in the price level that takes place at full employment, and will tend to prevent inflation from resulting in an increase in realized real investment.

Such unexpected increases in consumption expenditure due to the change in consumers' behavior at full employment will increase monetary investment. The marginal equal average propensity to consume behavior will tend to raise the rate at which income increases; it is equivalent to a rise in β. As a result the explosiveness of an accelerator process once full employment is reached increases rather than decreases as Hicks assumed. Unless the monetary system acts as a constraint, either in terms of the financing available or of the effects of the expansion upon planning curves, there is no reason why such an explosive expansion need break.

NOTES

1. See Hicks (1950), p. 75: 'But I assume (2) that there is a direct restraint upon upward expansion in the form of a scarcity of employable resources'. Also Goodwin (1946).
2. Innovations may result in this second type of structural unemployment. Such a situation would occur in an innovation cycle expansion.
3. 'Labor' is a shorthand symbol for factors which have an essentially fixed supply over the period of the business cycle. The labor market is chosen because the pervasiveness of labor as a cost makes labor market developments directly affect the price level.
4. See Schumpeter (1939), p. 548.
5. See Mints (1950) and Fisher (1933), pp. 337–57.
6. See Haberler (1951), pp. 379–80: 'Practically every theory of the cycle describes the cumulative process of expansion as a mutual stimulation for investment and consumption; in modern terminology we speak of an interaction of multiplier and acceleration principle, but the essence of the matter is contained in pre-Keynesian theory, for example, in the "Wicksellian Process" '.

   'The cumulative process is always essentially the same, but we cannot be sure that the turning point is always brought about by the same factors (even apart from possible disturbances from outside the economic system) or that the same system of difference equations will satisfactorily describe the upswing as well as the upper turning point.'
7. See Schumpeter (1934), pp. 126–7.
8. See ibid., p. 106; emphasis added.
9. See ibid., p. 107.
10. See ibid., p. 108.
11. See Schumpeter (1939), pp. 578–81.
12. See ibid., p. 579.
13. See Hicks (1950), p. 160. Hicks has 'not denied', merely ignored, the monetary prerequisites for his model to function.
14. See ibid., p. 160; emphasis added.
15. The second attribute 'that owners of liquidity are willing to become illiquid – to spend their liquidity' is strictly speaking necessary only if an expansion is to take place with a fixed quantity of money.
16. See Haberler (1951), p. 381: 'Expansible monetary and credit supply is undoubtedly an indispensable condition for the business cycle. But very few writers would be ready today to attempt a complete explanation of the cycle in terms of monetary and banking arrangements and policy.'
17. See Ohlin (1937). Reprinted in *Readings in Business Cycle Theory*, Philadelphia: Blakiston Co., 1944, pp. 87–130.
18. In addition to a rise in interest rates reducing realized investment, a fall in the price level may result in S_1 of monetary savings being able to finance I_2 of real investment. Conversely a rise in the price level will lower the amount of real investment that a given amount of money savings can finance. In Figure 9.1 the savings curve can be read as the supply curve of investment goods, and the $I(1)$ and $I(2)$ curves can be read as demand curves of investment goods. Then reading r_2, r_1 as price levels, the accelerator phenomenon becomes a 'price level' of investment goods determinant. A rise in the price level of investment goods, with no change in the price of the product, also results in raising the effective planning curves of firms.
19. Schumpeter, by emphasizing the role of money creation in financing development, tended to ignore the ability to finance investment which is implicit in ex-ante or voluntary

savings. Of course, on the level of theoretical abstraction that he was using, beginning with a stationary circular flow model, he was completely justified. The consumption function assumption of the accelerator models implies that we cannot begin our analysis of the role of finance at the same level of abstraction that Schumpeter used; we have to include voluntary savings. Whether the institution of 'voluntary' savings has its origin in th 'forced' savings of economic development is a deep question of the interpretation of economic evolution. In a developed capitalist society voluntary savings take place. Such voluntary savings can finance 'autonomous' or 'innovational' investment, but during the expansion they cannot finance all of 'induced' investment. Reason: induced investment depends upon a rise in money income, and as long as only voluntary savings are used to finance investment, money income cannot rise.

20. This includes business savings and the resulting rise in surplus accounts.
21. Banks may create money by lending to households which in turn acquire equities in businesses. This may occur through a 'margin' purchase of securities by individuals. In such a case the individual's debt increases, reducing his liquidity. However, the equity position of firms improve; therefore, improving their liquidity. Through such a technique an increase in the money supply may finance equity investment. This is, of course, equivalent to banks functioning as 'Capital Banks'. Nevertheless the model of bank lending used here restricts bank lending to the acquisition of debt by the bank, and the entry of bank debt on the liability side of business firms' balance sheets. To the extent that bank lending does result in the acquisition of equities by banks, the model requires modification. The institutional specification is that banking and money creation are essentially in the 'Anglo-American Tradition'.
22. Recall that in Chapter 6 we used the device of segregating superfluous assets in an 'Investment Trust'. In such cases business firms would not be able to finance investment by reducing their liquidity. The following balance sheets indicate how such savings and investment takes place. If the 'investment trusts' return to their original cash holdings, households would own the equity or debt of business firms used to finance the investment.

Firms

+Investment goods	+Equity +Debt

Investment Trusts

−Cash balance +Equity or debt of firm	

Households

+Cash balance	+Net worth

23. See Schumpeter (1939), p. 548: 'That economic analysis cannot . . . abstract from money is a truth which is useful only if supplemented by the other truth that monetary processes never carry their explanation in themselves and cannot be analysed in monetary terms alone.'
24. See Pigou (1951), p. 17.
25. Schumpeter distinguishes between the 'velocity of any unit that is actually sent over its path' and 'another component of the velocity figure which refers to the proportion of existing units so sent'. See Schumpeter (1939), p. 546.
26. Behavior of a changing velocity monetary system: assumptions: constant money supply, velocity changes

Accelerator Process	Monetary System
$\alpha = 0.8 \quad \beta = 4$	Money Supply 10,
$Y_{t-1} = 10$	Maximum Velocity 2,
	Assume no change in r.

Time	Y	C	S	$(Y_t - Y_{t-1})$	I (realized)	Money Supply	Financed from ΔV	Realized r
0	11	8	2		3	10	1.0	1.1
1	12.8	8.8	2.2	4	4	10	1.8	1.28
2	17.4	10.2	2.6	7.2	7.2	10	3.6	1.74
3	20.0	13.9	3.5	6.1	6.1	10	2.6	2.00
4	20	16	4	4	4	10	0	2.0
5	16	16	4	0	0	10	−4	1.6

27. See Haberler (1951), p. 381. In discussing the factors which tend to end cyclical upswings Prof. Haberler mentions: '4. Under the gold standard in older times the ceiling might be purely monetary'. This Wicksellian case holds for constant money supplies as taken up here. As will be shown later, a money supply that increases at an exogenously determined rate need not break a boom.
28. See Schumpeter (1939), pp. 546–7: 'In any case, however, we cannot here consider quantity of "existing" or "circulating" or "available" money as an independent variable because . . . it also varies in response to the other variables of our process, entrepreneurial activity in particular'.
29. The balance sheets of business firms, the public and the commercial banks are an aid to assessing the effects of banking lending to business. The change in net worth of the community is equal to ex-ante savings plus the unexpected savings which are equal to the increase in demand deposits. This is equal to the increase in investment goods by the investing firms. The household assets increase by the value of the debts and equities of the firms and the rise in the quantity of money.

Banks

Assets	Liabilities
Debts of firms ΔM	Demand deposits ΔM

Investing Firms

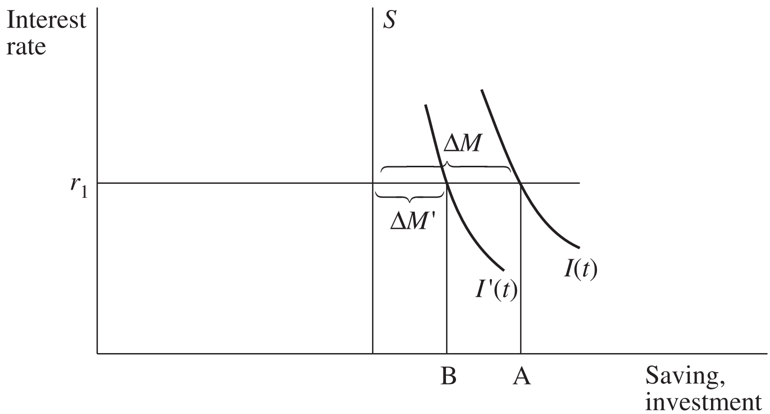
Assets	Liabilities
Investment goods I ex-ante	Debt or equity to households* S ex-ante Debt to banks ΔM

Households (public)

Assets	Liabilities
Debt & equities of firms S ex-ante Demand deposits $\Delta M/S$ ex-post	Ex-ante savings + S ex-ante Unexpected savings + $\Delta M/I$ ex-ante Change in net worth

* Includes business retained earnings.

30. Alternatively if r_1 is the liquidity trap rate of interest, even if $V > 0$, the rise in the quantity of money in excess of transaction needs can all be absorbed by households' portfolios without lowering the interest rate. However, in this case any rise (virtual) in interest rates would imply a substitution of earning assets for money in the portfolios of households. This then becomes a case of financing investment from cash balances.
31. On the interest rate-investment graph below, assume that the money supply is infinitely elastic at r_1. If investing firms are free of debt, the change in income during periods $t-1$, $t-2$ would result in the investment demand curve $I(t)$. As a result of bank debt and deteriorated balance sheets, an equal change in income will result in $I'(t)$. This will appear as if induced investment has fallen from OA to OB. The increase in lender's and borrower's risk raises planning curves as an expansion takes place even if the prime market rate of interest is unchanged.



32. The balance sheets of the three sector changes as follows:

Banks

Demand deposits (S ex-ante $-$ I ex-ante) $= -\Delta M$	Business debt (S ex-ante $-$ I ex-ante) $= -\Delta M$

Firms

Capital equipment + I ex-ante	Debt & equities to households: + S ex-ante Debt to banks $-$ ($-$ S ex-ante$-I$ ex-ante)

Households

Demand deposits $-$ (S ex-ante $-$ I ex-ante) Business assets + S ex-ante	Net worth + I ex-ante

If failures occur then, in the account of households labeled business assets, equities will be substituted for debt and in the account of business firms labeled debt and equities to households, equity will be substituted for debt. Also, as business firms fail, banks acquire titles and debts which are considered unsuitable for bank portfolios. The sale of such assets to the public results in the substitution of business assets for demand deposits in the public portfolios, and in a net reduction of demand deposits. These changes obviously do not affect the net worth of households and the capital equipment of firms' accounts. In addition, as the value of productive capacity may be reduced during a downturn, the value of the capital equipment account of firms and the net worth account of households will be reduced. This typically occurs through equity liabilities of firm and equity assets of households losing a part or all of their value.

33. Total induced investment is $\beta(Y_t - Y_{t-1})$. Ex-ante savings are equal to $(1-\alpha)Y_t$. Assuming that a constant proportion of ex-ante savings are used for equity financing, equity financing is $\lambda(1-\alpha)Y_t$. The ratio of equity to total investment therefore is:

$$\frac{\lambda(1-\alpha)Y_t}{\beta(Y_t - Y_{t-1})} = \frac{\lambda(1-\alpha)}{\beta(1-[Y_{t-1}/Y_t])}$$

The general solution to the second order accelerator process is of the form $Y_t = A_1\mu_1^t + A_2\mu_2^t$ where $\mu_1 > \mu_2$. Therefore we can write:

$$\frac{Y_{t-1}}{Y_t} = \frac{A_1\mu_1^{t-1} + A_2\mu_2^{t-1}}{A_1\mu_1^t + A_2\mu_2^t} = \frac{1+[A_2/A_1](\mu_2/\mu_1)^{t-1}}{\mu_1 + [A_2/A_1]\mu_2(\mu_2/\mu_1)^{t-1}}$$

The limit of $(\mu_2/\mu_1)^t = 0$, ∞ therefore $[Y_{t-1}/Y_t] - 1/\mu_1$. Hence the ratio of $\lambda(1-\alpha)Y_t/[\beta(Y_t - Y_{t-1})]$ approaches as a limit $\lambda(1-\alpha)/\beta(1-1/\mu_1)$. In the early stages of an explosive accelerator process the ratio of $Y_t/Y_{t-1} > 1/\mu_1$. Therefore the ratio of equity financed to total investment decreases as the accelerator process continues.

The following table illustrates the argument:

Infinitely Elastic Money Supply
Accelerator Process: $\alpha = 0.8$ $\beta = 4$ Monetary Phenomena $\lambda = 1$

Time	Y	C	S	$\beta(Y_{t-1} - Y_{t-2})$	I (realized)	ΔM	equity financing / total investment
0	10	–	–	–	–	–	–
1	11	8	2	–	3	+1	0.67
2	12.8	8.8	2.2	4	4	1.8	0.55
3	17.4	10.2	2.6	7.2	7.2	+4.6	0.36
4	28.3	13.9	3.5	14.4	14.4	10.9	0.24

As the explosive accelerator process takes hold, the ratio of ex-ante savings to total investment falls sharply.

34. In the arithmetical example given in note 33, in time period 3 only 0.36 of total investment was financed by savings. If the price level rose so that the value of capital goods doubled then the balance sheets of firms would show a one to one ratio of equity to debt liabilities. In time period 4, the price level would have to rise enough to increase

the value of capital goods by 7.4 if firms are to show a one to one ratio of equity to debt financing.

35. The following arithmetical example illustrates the argument:

Accelerator Process $\alpha = 0.8$ $\beta = 4$					Monetary System $\sigma = +1$ per time period	
Time	Y	C	S ex-ante	Induced I $\beta(Y_{t-1} - Y_{t-2})$	Realized I	Investment financed by increased money supply
0	10	8	2			
1	11	8	2	–	3	+1
2	12	8.8	2.2	4	3.2	+1
3	13	9.6	2.4	4	3.4	+1
4	14	10.4	2.6	4	3.6	+1
5	15	11.2	2.8	4	3.8	+1
6	16	12	3	4	4	+1
7	16.8	12.8	3.2	4	4	+0.8
8	16.64	13.44	3.36	3.2	3.2	−0.16

In time period 7, S ex-ante $+\Delta M > I$ ex-ante; therefore $Y_7 - Y_6 < \Delta M$. As a result in time period 8 the accelerator expansion is broken.

36. As long as we assume that α and V are constants this is equivalent to an increase in the money supply at a constant geometric rate. Let μ_1 be the percentage increase in the money supply per period. We know that $Y_{t-1} = VM_{t-1}$ and that $Y_t = V(1+\mu_1)M_{t-1}$. We also have that $Y_t = Y_{t-1} + \mu(1-\alpha)Y_{t-1}$; therefore $\mu(1-\alpha)Y_{t-1} = \mu_1 VM_{t-1}$ and $\mu(1-\alpha) = \mu_1$. If the money supply increases by 10 per cent per annum and $\alpha = 0.90$, then μ, the ratio of monetary financing to ex-ante savings, is 1.

37. Note that $Y_{t-n}/Y_t = [1 - \mu(1-\alpha)]^n$. The Harrod-Domar case of the accelerator as a generator of steady growth can be achieved by means of an explosive accelerator coefficient and a constraining money supply.

38. The argument also applies to an arithmetic rate of growth of lending ability. If β is a function of the rate of growth of income, the fall in the rate of growth of income may lead to a sharp fall in income. This is consistent with the stagnation-inflation alternatives of Chapter 2.

39. See note 27, quoting Haberler on the gold standard as a ceiling to business cycle expansion.

40. If the unused resources of period $t-1$ are a labor and capital combination, then $\beta(Y_{t-1} - Y_{t-2})$ has to be interpreted as the demand for capital *net* of the productive facilities idle in period $t-1$ which are absorbed in period t. Such absorption of capital dampens the investment effect of a rise in income.

41. In real terms: $Y_t = \alpha Y_{t-1} - \beta(Y_{t-1} - Y_{t-2}) < Y_{t-1} + \gamma I_{t-1}$. Aggregate money demand at period t is $p_{t-1}\, Y_t$. There exists a price level $p_t < p_{t-1}$ such that $p_{t-1}\, Y_t = p_t(Y_{t-1} + \gamma I_{t-1})$. Therefore: $Y_t/p_t = [Y_{t-1} + \gamma I_{t-1}]/p_{t-1}$.

42. Given $(Y_{t-1} - Y_{t-2}) > (1-\alpha)Y_{t-1}$

$$Y_t = \alpha Y_{t-1} + \beta(Y_{t-1} - Y_{t-2})$$

$$Y_{t-1} = \frac{Y_t}{\alpha} - \frac{\beta}{\alpha}(Y_{t-1} - Y_{t-2})$$

$$\beta(Y_{t-1} - Y_{t-2}) > (1-\alpha)\frac{Y_t}{\alpha} - \frac{\beta}{\alpha}(Y_{t-1} - Y_{t-2})$$

$$\beta(Y_{t-1} - Y_{t-2}) + \frac{(1-\alpha)}{\alpha}\beta[(Y_{t-1} - Y_{t-2})] > \frac{1-\alpha}{\alpha}Y_t$$

$$\frac{\alpha\beta + \beta - \alpha\beta}{\alpha}[Y_{t-1} - Y_{t-2}] > \frac{1-\alpha}{\alpha}Y_t$$

$$\beta\left[\frac{(Y_{t-1} - Y_{t-2})}{1-\alpha}\right] > Y_t$$

10. Conclusion: business cycle theory and economic policy

It has been shown that both linear and non-linear accelerator models, standing by themselves, are not sufficient for business cycle analysis. However, it was assumed that such flow of income models can be used as a core, a skeleton structure, for cycle theory. This use of accelerator models focuses attention upon the determinants of the parameters of the models, and it is shown that the parameters of these models depend upon the behavior of elementary economic units. Accepting the work that has been done on the relation between households and the propensity to consume, the relation between individual firm behavior and the accelerator coefficient has been investigated. In particular, the relationships between induced investment and financial constraints and market structures have been studied.

It cannot be claimed that the findings are definitive; in fact, the major purpose of setting out on this track was to test the plausibility of the approach. In addition, the tool of analysis selected, a modification of Marshallian price theory, may be too weak and the factual basis for the transformation of the findings into a value of the accelerator coefficient is lacking. Certainly the economic theory of the investment behavior of individual firms and the testing of the theory of the investment are among the major problems in economics. Nevertheless, some results, which may be viewed as the most tentative of hypotheses, have been obtained.

It has been shown that the investment response of conditional monopolies to a change in demand is discontinuous, whereas the response of firms in competitive markets is continuous. This result should be testable in the real world. If true, then a world in which market structures are dominated by conditional monopolies would tend to exhibit inflation and stagnation as alternative stable states. In such a world the business cycle disappears, being replaced by these alternative stable states and a rapid movement from one state to another.

It has also been shown that as far as the accelerator process is concerned, the division between real and monetary phenomena is misleading. The inflation state of an explosive accelerator model depends upon a permissive monetary system. Without an appropriate monetary system the explosive inflation cannot exist and the result is a stagnant low employment equilibrium. There is no endogenous mechanism which necessarily will end the stagnant state. It has also been shown that an explosive accelerator process can be transformed into a steady rate of growth by means of an appropriate monetary system.

It was also shown that the balance sheets of firms can affect investment behavior and hence the accelerator process. Models were constructed in which, even assuming a permissive monetary system, the willingness of firms to go into debt limits the expansion. This may be of particular importance for economies in which a large portion of savings are channeled through financial intermediaries with limited portfolio possibilities. That is, the growth of savings by classes that hold only debt may increase the tendency toward stagnation of an economy.

The relation between industrial structure and the accelerator process which has been derived leads to an interesting evolutionary hypothesis. If the structure of markets in an economy changes as an economy evolves, then the business cycle contours change. A competitive society would tend to have short cycles with relatively small amplitude to the fluctuations. An economy in which conditional monopolies dominate would tend to have a longer cycle, with a long expansionary inflationary boom and a long flat-bottomed trough or stagnant state. The behavior of the economic system depends upon the behavior of individual economic units so that if the weight in the economy of different market structure changes, the cyclical behavior of the economy changes. This changing composition of the economy may result in either a continuous gradation of cyclical patterns, or there may be a few patterns and few critical mixtures such that the behavior of the economy changes sharply with a slight change in the proportion of different market structures. If the hypothesis as to the relation between market structures is valid, then areas of economic policy such as anti-trust and anti-cyclical policy cannot be treated in separate compartments. In particular, a 'reasonable' anti-trust policy may negate a 'vigorous' stabilization policy. It is also true that if cyclical behavior evolves, then the evidence of times series data is suspect. If the evolution of the economy is rapid enough, then the runs of data based upon a given economic state may be too short for fruitful statistical analysis.

If monetary behavior regulates the excess of realized investment over ex-ante savings on an upswing, then the rate of growth of the economy can be determined by monetary mechanism: in particular, monetary policy can transform an inflationary explosion into a steady growth of the economy. The efficacy of such a policy primarily depends upon the availability of finance and only indirectly upon the interest rate. When ex-ante savings exceeds ex-ante investment then monetary policy can only affect the liquidity of the community. Unless investment is interest elastic, such a monetary expansion on the downswing will have no effect. It was shown that the presumption is that the interest elasticity of investment depends upon market structure – that a world of conditional monopolies would tend to have an interest inelastic investment demand curve. This means that the efficacy of monetary policy also varies with the structure of the economy.

Investment which is due in the first instance to changes in production functions has not been taken up. That is, the effect of innovations upon aggregate investment has not been discussed, except to note that the supply curve of financing, and of investment goods, relates to both induced and innovational investment. As was mentioned in the text, such innovational investment may be a significant element in business cycles by supplying 'energy' to a damped flow of income process. The effects of innovations may be represented by introducing a stochastic factor into the model. Perhaps innovational investment can be integrated into our approach by considering the demand curves and financing conditions confronting an innovating firm.

The material presented here is primarily intended to demonstrate that a generalization about the behavior of an economy as a whole is meaningless unless it is integrated with the known or accepted material about the behavior of individual economic units. Rather than be content with spawning additional formal-mathematical aggregate models, the simplest flow of income model with an investment coefficient was used as a framework with which to explore the relation between the behavior of firms and aggregate induced investment. It is believed that the resulting variable coefficient accelerator model can serve as an adequate basis for business cycle analysis, especially as the variations in the coefficients are related to the behavior of individual economic units. If this approach is to be used in further work, then empirical research has to be undertaken to determine whether the significant homogeneous classes of firms with regard to investment behavior are, as is assumed here, the homogeneous classes of firms derived from price theory. If this turns out to be true, then predictions as to the behavior of the accelerator coefficient can be made which lead to predictions as to the cyclical behavior of the economy.

References

Alexander, S.S. (1949), 'The Accelerator as a Generator of Steady Growth', *Quarterly Journal of Economics*, **63**, 174–97.

Bronfenbrenner, M. (1940), 'Applications of the Discontinuous Oligopoly Demand Curves', *Journal of Political Economy*, **48**, 420–27.

Bronfenbrenner, M. (1947), 'Price Control under Imperfect Competition', *American Economic Review*, **37**, 107–20.

Chamberlin, E.H. (1938), *The Theory of Monopolistic Competition*, Cambridge, MA: Harvard University Press, 3rd edition.

Chenery, H.B. (1952), 'Overcapacity and the Acceleration Principle', *Econometrica*, **20**, 1–28.

Domar, E. (1946), 'Capital Expansion, Rate of Growth and Employment', *Econometrica*, **14**, 137–47.

Domar, E. (1947), 'Expansion and Employment', *American Economic Review*, **37**, 34–55.

Duesenberry, J.S. (1948), 'Income Consumption Relations and their Implications', in Lloyd A. Metzler, *Income, Employment and Public Policy, Essays in Honor of A.H. Hansen*, New York: W.W. Norton and Company, pp. 54–82.

Duesenberry, J.S. (1949), *Income Savings and the Theory of Consumer Behavior*, Cambridge, MA: Harvard University Press.

Edgeworth, F.Y. (1925), *Papers Relating to Political Economy*, London: Macmillan.

Fisher, G.H. (1952), 'Some Comments on Stochastic Macro-economic Models', *American Economic Review*, **42**, 528–39.

Fisher, I. (1933), 'The Debt Deflation Theory of Great Depressions', *Econometrica*, **1**, 337–57.

Friedman, M. (n.d.), 'The Relationship between Supply Curve and Cost Curves', unpublished, University of Chicago.

Friedman, M. (1949), 'The Marshallian Demand Curve', *Journal of Political Economy*, **57**, 463–95.

Frisch, R. (1933), 'Propagation Problems and Impulse Problems in Dynamic Economics', in Gustav Cassel, *Economic Essays in Honor of Gustav Cassel*, London: George Allen and Unwin, Ltd., pp. 171–206.

Goodwin, R.M. (1946), 'Innovations and the Irregularity of Economic Cycles', *Review of Economic Statistics*, **28**, 95–104.

Goodwin, R.M. (1948), 'Secular and Cyclical Aspects of the Multiplier and the Accelerator', in Lloyd A. Metzler *Income, Employment and Public Policy, Essays in Honor of Alvin H. Hansen*, New York: W.W. Norton and Company, pp. 108–32.

Goodwin, R.M. (1951), 'The Non-linear Accelerator and the Persistence of Business Cycles', *Econometrica*, **19**, 1–17. Hectographed, University of Chicago.

Haavelmo, T. (1944), 'Probability Approach to Econometrics', Supplement to *Econometrica*, **12**.

Haberler, G. (1941), *Prosperity and Depressions*, Geneva: League of Nations.

Haberler, G. (1951), 'Business Cycles in a Planned Economy', Conference on Business Cycles: Universities National Bureau Committee, New York: National Bureau of Economic Research, pp. 375–86.

Hall, R.H., and C.J. Hitch (1939), 'Price Theory and Business Behavior', *Oxford Economic Papers*, No. 2, 12–45.

Harrod, R.F. (1934), 'Doctrines of Imperfect Competition', *Quarterly Journal of Economics*, **48**, 442–70.

Harrod, R.F. (1936), *The Trade Cycle: An Essay*, Oxford: Clarendon Press, chapter 2, p. 3.

Harrod, R.F. (1939), 'An Essay in Dynamic Theory', *Economic Journal*, **49**, 14–33.

Harrod, R.F. (1948), *Towards a Dynamic Economics*, London: Macmillan, chapter 2.

Harrod, R.F. (1952), *Economic Essays*, London: Macmillan.

Hicks, J.R. (1950), *A Contribution to the Theory of the Trade Cycle*, Oxford: Clarendon Press, chapter 1, p. 3.

Kalecki, M. (1937), 'The Principle of Increasing Risk', *Econometrica*, New Series 4, 440–47.

Keynes, J.M. (1936), *The General Theory of Employment, Interest and Money*, New York: Harcourt Brace and Company, chapter 1, p. 5.

Klein, L.R. (1946), 'A Post Mortem on Transition Predictions of National Product', *Journal of Political Economy*, **54**, 289–308.

Klein, L.R. (1950), *Economic Fluctuation in the United States 1921–41*, New York: John Wiley and Sons.

Lange, O. (1944), *Price Flexibility and Employment*, Bloomington, Ind.: Principia Press.

Leontief, W.W. (1947), 'Introduction to a Theory of the Internal Structure of Functional Relationships', *Econometrica*, **15**, 361–73.

Lerner, A.P. (1937), 'Statics and Dynamics in a Socialist Economy', *Economic Journal*, **47**, 253–70.

Lester, R.A. (1946), 'Shortcomings of Marginal Analysis for Wage-Employment Problems', *American Economic Review*, **36**, 63–82.

Lundberg, O. (1940), *On Random Processes and their Applicability to Sickness and Accident Insurance*, Uppsala: Almquist and Wicksells, Boktryckeri, A.B.

Lutz, F. and V. Lutz, (1951), *The Theory of Investment of the Firms*, Princeton: Princeton University Press.

Marshall, A. (1919), *Industry and Trade*, London: Macmillan.

Marshall, A. (1920), *Principles of Economics*, New York: Macmillan, 8th edition.

Marshall, A. (1925), 'The Present Picture of Economics', reprinted in A.C. Pigou (ed.), *Memories of Alfred Marshall*, London: Macmillan, first published, 1925.

Metzler, L. (1941), 'The Nature and Stability of Inventory Cycles', *Review of Economic Statistics*, **23**, 113–29.

Metzler, L. (1948), 'Keynes and the Theory of Business Cycles', in S. Harris (ed.), *The New Economics*, New York: Knopf, pp. 436–49.

Mints, L. (1950), *Monetary Policy for a Competitive Society*, New York: McGraw-Hill.

Mitchell, W.C. (1950), *What Happens During Business Cycles*, New York: National Bureau of Economic Research, chapter 1, p. 3.

Modigliani, F. (1949), 'Fluctuations in the Savings-Income Ratio: A Problem in Economic Forecasting', *Studies in Income and Wealth*, **11**, New York: National Bureau of Economic Research, pp. 370–440.

Ohlin, B. (1937), 'Some Notes on the Stockholm Theory of Savings and Investment', *Economic Journal*, **47,** 53–69, 221–40.

Patinkin, D. (1948), 'Price Flexibility and Full Employment', *American Economic Review*, **38**, 543–664.

Pigou, A.C. (1943), 'The Classical Stationary State', *Economic Journal*, **53,** 343–51.

Pigou, A.C. (1951), *Keynes's General Theory*, London: Macmillan.

Pilvin, H. (1952), 'A Geometric Analysis of Recent Growth Models', *American Economic Review*, **42**, pp. 594–99.

Reder, M.W. (1947a), 'A Reconsideration of the Marginal Productivity Theory', *Journal of Political Economy*, **55**.

Reder, M.W. (1947b), *Studies in the Theory of Welfare Economics*, New York: Columbia University Press.

Robinson, J. (1933), *Economics of Imperfect Competition*, London: Macmillan.

Samuelson, P.A. (1939), 'Interaction between the Multiplier Analysis and the Principle of Acceleration', *The Review of Economic Statistics*, **21**, 75–8.

Schumpeter, J.A. (1934), *The Theory of Economic Development*, Cambridge, MA: Harvard University Press.

Schumpeter, J.A. (1939), *Business Cycles*, New York and London: McGraw-Hill, 2 volumes.

Stigler, G. (1939), 'Production and Distribution in the Short Run', *Journal of Political Economy*, **47**, 305–27.

Stigler, G. (1949), *Five Lectures on Economic Problems*, London: London School of Economics and Political Science.

Stigler, G. (1952), *Theory of Price*, New York: Macmillan.

Sweezy, P.M. (1939), 'Demand under Conditions of Oligopoly', *Journal of Political Economy*, **47**, 568–73.

Tinbergen, J. (1938), 'Statistical Evidence on the Acceleration Principle', *Economica*, New Series 5, 164–76.

Tinbergen, J. (1951), 'Reformulation of Current Business Cycle Theories as Refutable Hypothesis', Conference on Business Cycles: Universities National Bureau Committee, New York: National Bureau of Economic Research, pp. 131–40.

Tsiang, S.C. (1951), 'Accelerator, Theory of the Firm and the Business Cycle', *Quarterly Journal of Economics*, **65**.

Viner, J. (1931), 'Cost Curves and Supply Curves', *Zeitschrift für Nationalökonomie*, No. 3, 23–46.

Viner, J. (1936), *Business Cycle Theory: Can Depressions be Avoided or Tempered?*, U.S. Department of Agriculture Graduate School.

Wilson, T. (1953), 'Cyclical and Autonomous Inducements to Invest', *Oxford Economic Papers*, **5**, 65–89.

Index

accelerator
and change in income 39
as a coefficient of induced investment 20, 21, 64, 66–7
as a coefficient of realized investment 20, 21, 64, 66
conditions necessary for operation 67–8
and firm behavior 68
production function 21
as a random variable 53
as a structural parameter 20, 21, 64–6
accelerator coefficiency, non-linearities 191
accelerator coefficient 179, 198
aggregate 64–5
and conditional monopoly 208
constant 69–70
determination of value 64, 70
and financing of investment 72
generation of 64–84
negative 81, 82
accelerator mechanism 226
accelerator or multiplier coefficient 10
accelerator process
explosive 246
and industrial structure 259
and real and monetary phenomena 258–9
accelerator-multiplier mechanism 7
accelerator-multiplier models 9–10, 15, 20–63
alternative non-linear models 37–47
capital stock in 28
Goodwin model 27–37
Hansen-Samuelson model 22–5
Hicks type of linear model 25–7
linear models 22–7, 258
non-linear models 27–47, 258
stochastic models 47–60
accelerator-multiplier type analysis 6
floors and ceilings in 5
aggregate induced investment 212–13
aggregated models 6
parameters 6
Alexander, S.S. 27, 61
alternative market structures 160
alternative non-linear models 37–47
behavior of firms 37–8
change in income 39, 45
consumption function in 39
level of income in 38
asset properties, of money 234
autonomous changes, in velocity 244
autonomous investment 81, 82, 102, 179

balance sheet deterioration 135, 242
balance sheet structure 74, 85, 89, 97, 98, 140, 142
and firm behavior 98
and firm survival 127
balance sheets 129, 230, 238, 239, 253, 254
of firms which exhibit debt instruments 135
and investment behavior 259
relation with cost curves 136
banking system model 232–3
bankruptcy 127, 157
behavior of households 10
break even point pricing 172–3
Bronfenbrenner, M. 168, 175, 201, 219
business cycle 177
alternative theories 1–2
damping of 55
'energy source' 48
and industrial structure 88
monetary aspects 222

business cycle analysis 1–19
  methodological division 2
  methodology 14–18
business cycle models, and linear relationships 5
business cycle theory 1, 222
  and economic policy 258–60
  industrial structure variant 86
  in relation to theory of the firm 85–101

capital, demand for 75
capital market 91
  imperfect 98
capital stock, in accelerator-multiplier models 28–31
capital theory 181
'capitalization of earnings' 94
cash balance 133
ceilings 5
  money supply 242–3
central banks 233
Chamberlin, E.H. 78, 91, 100, 164, 175
Chamberlin's symmetry assumption 165
Chenery, H.B. 20, 68, 83, 212, 219
Classical Quantity Theory of Money 232–3
common pricing rule 169–71
competition 79, 86, 90
  imperfect or monopolistic 91
  and monopoly 77–8, 85, 93
competitive industries 180–91
  behavior during business cycle 190
  market demand curve 140–42
  reactions to change in demand 187–8
  reactions to rise in price of products 180–81
conditional monopoly 160–61, 163, 174, 199–212
  and the accelerator coefficient 208
  demand curve 199, 205, 206, 208
  investment response 258
  long run marginal cost curve 202–5
  maximum price constraint case 199
  planning marginal cost curve 212
  and price constraint 208–10, 216
  'rent case' 200
'conglomerate' firms 136
constants 5
consumer good production 68, 69
consumption classes 77
consumption demand 87
consumption-income relation 80
'continuing inflation' 46
contracts 130
cost conditions, plant 122
cost curves 85, 98–9, 127
  average 103, 104
  and investment 102–26
  of plants 136
  relation with balance sheets 136
  short run and long run, for different rates of return 116
  short term average 104, 105, 151
  to investigate investment behavior 103
  as transformations of a production function 121
  U-shaped 89
  see also long run average cost curves; long run cost curves; long run marginal cost curves
cost plus pricing 172

damped fluctuations 48, 49
damping, of business cycle 55
deficits 46, 47
demand curve 77, 78, 79, 85, 86, 177, 185
  conditional monopoly 199
  discontinuous oligopoly 72, 168
  for firms 90, 92, 93
  firms 108–9
  horizontal 92
  for investment 233
  objective 90
development 223, 224
direct induced investment 177
disaggregation 4
discounting process 107
disposable income, and national income 79
Domar, E. 17, 19, 21, 60
Duesenberry, A.S. 10, 14, 18, 19, 77, 84

Econometrica 68
economic policy, and business cycle theory 258–60
Edgeworth, F.Y. 7, 18

'energy source', in business cycles 48
entrepreneurial capacity 112
entrepreneurs 29, 32, 70, 79, 89, 91, 149, 151–2
envelope curves 113
'equilibrium price' 191
erratic shock-damped oscillation hypothesis 47–8
erratic shocks 47, 48
error approach 47–9
ex-ante savings 226–9, 239–40
  and realized investment 259–60
ex-post savings 226–9
'excess capacity' 68
expansion path 113–14
'expectations' 181, 237
explosive accelerator process 246

factors of production 89, 107, 132
  price changes in 102
financing
  analysis 152–7
  conditions 85
  constraint 146
  of investment, fixed money supply 235–6
financing sources 224–5
  and liability structure of firms 230–31
firm behavior 37–8, 68–9, 70, 91, 95–6
  and balance sheet structure 98
firm survival, analysis of financing 152–7
firms
  accounts receivable 133
  assets 93, 128–9, 133–5
  balance sheet constraints 97
  best attainable position 93
  cash balance 133
  'conglomerate' 136
  debt 93, 97, 127, 128, 130
  debt load, reducing 150–51
  definition of 88
  demand curve 90, 108–9
  equilibrium 110
  equity base 89, 90
  financing 89, 96, 105
  financing sources and liability structure 230–31
  government bonds 133
  internal instability 169
  investment alternatives 98
  investment decisions 102
  'investment threshold' 103
  liabilities 134, 147
  liquidity and survival 98
  long run survival 139
  long run survival average cost curve 143
  market constraints 160–76
  non-competitive position 94–5
  optimum methods of production 107–8, 120
  owners' equity 14, 97, 127, 136–8, 147, 200
  price-quantity value 90
  retained earnings 152
  and rises in income 79
  survival conditions 96–7, 127
  survival of 127–59
  theory of, in relation to business cycle theory 85–101
  total assets needed 135–6
  vulnerability 95, 160–76
  see also plant
Fisher, G.H. 47, 48, 62
Fisher, I. 80, 222, 251
floors and ceilings 24, 222
  in accelerator-multiplier type analysis 5
  in Hansen-Samuelson model 24
  interest rate floor 235
  money income ceiling 237
  money supply ceiling 242–3
free entry assumption 91
'freedom of entry' 100–101
  industries ranked by 94
Friedman, M. 15, 19, 107, 112, 126
Frisch, R. 24, 47, 61, 62
full cost pricing 172
function indivisibilities 162

general theory, Keynes 80
'generating relation' 53
gold 232
'gold production' theories 241
Goodwin model 27–37, 59
  capital stock in 28–31
  change in capital and change in income 33

deficiencies 31
desired growth of capital 31
modifications 35–6
non-linear accelerator in 32–3
technological progress in 31
Goodwin, R.M. 7, 18, 21, 28, 29, 30, 31, 47, 59, 61, 61–2, 62, 63, 220, 251
Goodwin's Hypothesis 34
government bonds 133
government deficit financing 246
government prices 87
growth, equivalent rate 21

Haavelmo, T. 48, 49, 63
Haberler, G. 1, 4, 17, 18, 19, 222, 226, 237, 251, 253
Hall, R.H. 91, 100
Hansen-Samuelson model 8, 10–12, 22–5, 34
accelerator coefficient in 13–14
assumptions in 8–9
consumption coefficient 12–13
definition of income in 22
floors and ceilings 24
national income in 8
output and capital stock in 13, 14
Robertsonian relation 22
state of the economy in 22–4
time series 11–12
types of behavior in 22–3
Harrod, R.F. 21, 22, 60, 61, 67, 82, 118, 126, 193, 218, 219
Harrod-Domar case 27
Harrrod-Domar models 21–2, 61
Hawtrey 72
Hicks, J.R. 2, 5, 6, 8, 18, 48, 61, 62, 75, 81, 83, 220, 222, 225, 251
Hicks' Trade Cycle Model 2–3, 44, 81, 226
Hicks' type
accelerator model 240–41
linear model 25–7, 47
state of economy in 25–6
Hitch, C.J. 91, 100
horizontal demand curve 92
households, behavior 10

imperfect or monopolistic competition theory 91
income
change in 58, 70, 77
and investment 85
and induced investment 79
investment as a determinant 17
level of 15–16
income generating function 55
income variables, Keynesian analysis of 3
induced disinvestment 68, 210
induced investment 50, 51–2, 81, 82, 102
aggregate 212–13
definition 82
direct 177
effect of market structure on 177–219
lagged and immediate 229
and market rate of interest 214
and monetary behavior 220–58
realized 102
and rises in income 79
schedule 178
industrial structure 86, 93
and the accelerator process 259
and business cycles 88
industries, ranked by 'freedom of entry' 94
industry demand curves 52
inherent stability 222
inner-circle relations 3, 4, 6
innovation, and investment 260
innovational investment 82
installation costs 93, 100
interest rate-investment graph 254
interest rates 150, 230, 234, 235
and demand cure for investment 233
floor 235
interest rate–investment relation 80–81, 213–17
internal instability 169
internal rate of return 114
internal-external rate of return 82
investment
and change in income 85
and cost curves 102–26
as a determinant of income 17
financing of investment and fixed money supply 235–6
and innovation 260

and output 68
see also realized investment
investment behavior, and balance sheets 259
investment behavior models 13
investment decisions 80, 102, 120
investment demand 87
investment demand schedule 178–9
investment goods
demand for 66
industries 220
investment relation 50
'investment threshold' 103, 104
iso-product curve 115, 121
iso-product map 120
iso-profit curves 160, 162
iso-return curves 113

Kalecki, M. 146, 159
'The Principle of Increasing Risk' 147
Keynes 3, 18, 80
general theory 80
Keynesian analysis, of income variables 3
Keynesian consumption function 22
Keynesian Liquidity Preference Theory of Money 232
Keynesian liquidity trap 236–7, 240
Keynesian models 16
Klein, L.R. 76, 84, 96, 101

Lange, O. 91, 100
Leontief, W.W. 4, 18
Lerner, A.P. 113, 118, 126
Lester, R.A. 91, 100, 169, 175
liabilities 134, 147
linear models 22–7
linear process models 5
liquidity preference 16, 225–6, 234, 244–5
liquidity preference relation 233
long run average cost curves (LRAC) 103, 104, 113, 114–15, 119, 144, 146, 149, 154, 155, 157
long run cost conditions 124
long run cost curves 89, 118
long run cost relations 123
long run marginal cost curves (LRMC) 113, 114–15, 119, 125, 144–6, 148, 154, 179
conditional monopoly 202–5
types of 121
long run supply curve 113
long run survival average cost curve 143
Lundberg, O. 50, 63
Lutz, F. 181, 218
Lutz, V. 181, 218

macroeconomic business cycle theory 3–4
macroeconomic model, variables in 7
marginal analysis 90–91
marginal cost curve 142
Marginal Cost equals Marginal Revenue 109
marginal propensity to consume 76
marginal revenue curve 179
jump 168
market constraints 160–76
market demand curves 78, 140–142, 189
market rate of interest 215
and induced investment 214
market share rule 78
market structure 85, 86, 87–8
effect upon induced investment 177–219
varying 105
markets 17
non-competitive 93
Marshall, A. 14, 19, 90, 100, 160, 161, 175
Marshallian industry 14, 16, 50, 78–9, 90
maximization principles 181, 192, 199
maximum price 170, 172
mechanical interrelation model 7
mechanical pendulum business cycle models 75
Metzler, L. 6, 18, 177, 218
Mints, L. 222, 251
Mitchell, W.C. 2, 15, 18, 19, 76, 83, 88, 100
Modigliani, F. 14, 19
monetary behavior, and induced investment 220–58
monetary changes 229–33
monetary systems 231–3
changes in quantity and velocity of money 243–6

quantity of money changes 238–43
quantity of money constant: change in velocity 233–8
money, asset properties 234
money income ceiling 237
money market 87, 134
effect on plant size 146–52
money supply 230, 233, 239, 242
ceiling 242–3
money-interest rate relation 80
monopolistic control of prices 86
monopolistic firm, equilibrium 163–4
monopoly 86, 87, 95
and competition 77–8, 85, 93
conditional 160
monopoly firms 191–8
optimum equity base 194
plant 194
monopoly profit 164, 193
'monopsonistic' supply curve 147
monopsony, absence of 93
multi-plant firms 162

National Bureau of Economic Research 88
national income
change in 85–6, 102
and disposable income 79
negative accelerator coefficients 81, 82
non-competitive markets 93, 94–5, 109
non-linear models 27–47
alternative 37–47
non-linear and stochastic assumptions, combination 59
non-linear theory 7
non-linearity 7
normal profit 103, 105
normal rate of return 111–12

Ohlin, B. 227, 251
oligopoly 85, 92, 160
operating losses 128
optimum plant 214–15
output
and investment 68
short run and long run change 71
owners' equity 127, 136–8, 144, 147
rate of return on 200
zero rate of return 97

Patinkin, D. 17, 19
Pigou, A.C. 17, 19, 233, 252
Pilvin, H. 21, 60
planning curves 118, 179, 195, 215, 242, 244
planning marginal cost curve 150, 212
planning rate 116, 117
plant 105
average cost curve 139–40
cost conditions 122
cost curves 136
cost schedules 109
definition of 88
flexible vs. inflexible 124
modification 120, 123–4
modification vs. additional 123
monopoly firms 194
optimum 214–15
scale of 89, 106, 119, 132, 150
short run output cost relations 109–10
short term average cost curves 104
plant factors 113
plant size 181, 186
effect of money market on 146–52
Poisson type distributions 63
portfolio money 234
price constraint, and conditional monopolies 216
price flexibility 87, 88
price flexibility theory 87
price leadership 170–71
price level, and savings coefficient 246–51
price policy, regulated monopoly 174
price rigidities 86, 87
price theory 90
price-quantity value, firms 90
prices
governmentally regulated 87
monopolistic control of 86
'The Principle of Increasing Risk' 147
probability distribution 49–50, 55, 56
product demand curves 79
product markets, with few firms 167
product price rise, reactions to 183–4
production conditions, modification of existing plant 120

production function 89, 106
  definitions 89–90
productive capacity 21
profit
  in the investment function 69
  monopoly 164
  normal 103
profit maximization 89, 90, 91–2, 93, 95, 99, 201
  constraints on 92
  long run conditions 117
  Marginal Cost equals Marginal Revenue 109
  and monopolized industries 198
  planning rate 117, 118
  traditional diagram 111
profit rate 162
profitability 80
'Propagation Problems and Impulse Problems' 24

quasi-cartels 78, 79
quasi-rent doctrine 113

random process approach 50
random shocks 48
random variable 53
'random variable' model 54
rate of return 183, 187, 192, 212
  internal 114
  normal 111–12
  on owners' equity 200
  and short run and long run cost curves 116
  and short term marginal cost curve 115
realized investment 66, 76, 227, 230, 231, 237, 238, 249
  and ex-ante savings 259–60
Reder, M.W. 93, 100, 160, 175
regulated monopoly, price policy 174
renting or hiring 130–31
replacement costs 93
resources, employed and unemployed 86
retained earnings 152
Ricardian Rent 112, 126, 185, 193
risk 104, 105, 148, 149, 157
risk premium 98, 131, 149–50
'riskless' competitive world 182–3
Robertsonian relation 22
Robinson, J. 91, 100, 163, 164, 175

'sampling problem' 40
Samuelson, P.A. 8, 18, 22, 61
saving 230, 250
savings 221, 223
  ex-ante and ex-post 226–9
  Robertsonian definition 73–4
savings coefficient, and price level 246–51
Schumpeter, J.A. 222, 223, 229, 231, 238, 251, 252, 253
Schumpeterian hypothesis 47
Schumpeterian innovation hypothesis 24
'secular stagnation' 46
short run marginal cost (SRMC) 136, 143, 188, 197, 198
short term average cost curve 104, 105, 151
short term marginal cost curve, and rate of return 115
standard price theory 108
Stigler, G. 118, 126, 163, 175, 187, 218
'stochastic energy source' 25
stochastic models
  error approach 47–9
  probability distribution 49–50
  stochastic variables 49–60
stochastic and non-linear assumptions, combination 59
supplementary relations 6–7
supply curve, 'monopsonistic' 147
'supplying a market' behavior 202
survival constraints 163
survival cost curve 142
survival of firms 127–59
  long run 139
Sweezy, P.M. 168, 175

tangency solution 164–6
technological progress, in Goodwin model 31
time series 48, 55, 60, 67
Tinbergen, J. 2, 3, 4, 6, 18, 37–8, 62, 67–8, 69, 76, 82, 83
Trade Cycle 25, 81

Tsiang, S.C. 9, 18, 69, 70, 72, 75, 80, 83, 84
turning points 4, 5, 6, 24

unemployment 220–21
United States, cyclical behavior 44

'value of market position' 93
'value of the organization' 93
variables, in macroeconomic model 7
velocity 230, 231
  autonomous changes in 244
  change in 233–8
  of circulation 232
  of money 243–6
Viner, J. 86, 87, 99, 100, 102, 112, 115, 118, 119, 125, 126, 185, 218
  price flexibility theory 87
vulnerability conditions 105
vulnerability of firms 95, 160–76

Wang 118, 126
wealth 16–17
'welfare economics' 160
Wilson, T. 65, 66, 82

zero return
  curve 148
  on owners' equity 97
  on total investment 119