PUBLIC UTILITY PRICING AND INDUSTRIAL DECENTRALIZATION IN SOUTH AFRICA

Dissertation

submitted in partial fulfilment of the requirements for the degree of
DOCTOR OF PHILOSOPHY of Rhodes University

by

JOSEPH LYALL WALLIS

December 1983

ACKNOWLEDGEMENTS

I would, firstly, like to express my deep gratitude to my supervisor, Prof. M.L.Truu for the encouragement, assistance and advice he gave me during the writing of this thesis. I am also indebted to Dr P.J.D. Du Toit; and Mr T.L.Kennedy of the National Institute of Transport and Road Research for the helpful discussions on the practical aspects of transportation policy; to Mrs Corona Langhan and Mr Oakley West, Cartographer at the Geography Department of Rhodes University, for the extremely competent way in which they assisted with the graphs and diagrams; and to Mrs Lalita Ranchhod who typed the thesis both efficiently and cheerfully.

The financial assistance of the following bodies is gratefully acknowledged:

(i) Human Needs, Resources and the Environment Section (Regional Development Programming), National Programme for Environmental Sciences, Co-operative Scientific Programmes, Council for Scientific and Industrial Research;

(ii) the National Institute of Transport and Road Research, Council for Scientific and Industrial Research; and

(iii) the Human Sciences Research Council.

To my wife, Kim, I am deeply grateful for the love and support she has given during the time of writing this thesis.

J.L.Wallis
GRAHAMSTOWN
December 1983

CONTENTS

Page

LIST OF TABLES AND FIGURES

Tables		Page

Figures		Page

INTRODUCTION

1. BACKGROUND TO THE THESIS

During the 1950's and 1960's it would appear that the explicit objectives of economic policy in South Africa were full employment and economic growth with some occasional emphasis on the pursuit of relative price stability. Other goals such as efficiency in resource allocation and the pursuit of an "acceptable" income distribution were at best implicit and subordinate to these objectives. This is exemplified by the fact that a number of key prices which were controlled by the authorities such as the exchange rate, interest rates and public utility tariffs were generally set at levels which were either over-or under-priced relative to factor scarcities throughout this period. South Africa was certainly not unique in this respect since, as Balassa has pointed out, the policies followed in many countries "... generally reflected an inadequate appreciation of the sensitivity of economic agents to price signals."[1] Although such price-distorting policy measures applied may have contributed to the expansion of manufacturing industry during this period this was often at a cost to the national economy in the form of inefficiencies in the allocation of resources. Furthermore such costs tended to accumulate as the price distortions caused by public policy persisted over time. The underlying need to reform the system of incentives affecting the allocation of resources therefore grew steadily in importance up till the 1970's, when it became critical because of the international upheavals in energy prices, interest rates, and exchange rates. It was during this period that policymakers in many countries came to recognize that the resource misallocation caused by distorted prices was not just a painful side-effect of economic growth but could actually

---

1. Balassa, B., Policy Reform in Developing Countries, Pergamnon Press, Oxford, 1st ed., 1977, p. 8.

be a factor limiting future growth. For example, in a statistical analysis of the relationship between price distortions and economic growth in 32 developing countries in the 1970's, the World Bank found that:

"... those countries with the worst distortions experienced significantly lower domestic saving and lower output per unit of investment, thus leading to slower growth."[2]

In South Africa there also appears to have occurred a recognition by policymakers of the consequences of resource misallocation for national economic development as indicated by the movement towards more "market-orientated" measures which characterizes recent monetary exchange rate and industrial decentralization policy proposals. This is exemplified by the statement made by the Prime Minister in his "Carlton" address that the South African Government was committed to the "general deregulation of the economy as evidenced by the re-examination of price, rent and exchange control".

It is against the background of this general direction in economic policy that this thesis examines the present pricing policies followed by the suppliers of rail,transport electricity and water services in South Africa and indicates the direction in which they should be changed in order to contribute towards a more efficient allocation of resources in the economy. Furthermore since a high priority has been attached by the Government to a policy of encouraging industrial decentralization, largely by means of a system of price incentives, this thesis will also examine the effect on this policy of both the existing and the recommended pricing policies pertaining to these services.

---

2. The International Bank for Reconstruction and Development, World Development Report 1983, Washington, July 1983, p. 57.

2. THE ECONOMIC CHARACTERISTICS OF PUBLIC UTILITIES

Rail, electricity and water services are, in most cases provided by so-called public utilities. The supply and demand conditions for public utility services are disntinguished by a number of characteristics which affect and reflect the degree of State involvement in these sectors of the economy.

The production of public utility services is usually characterized by the existence of economies of scale over the entire output range of the market, the physical connection of producer and consumer and highly capital-intensive methods of production. As a result public utilities tend to be "natural monopolies" with a high proportion of their total costs consisting of overhead investment.

The services provided by public utilities have a number of distinctive characteristics:

(i) they are used on a continuos or repeated basis in a multiplicity of activities throughout the economy;

(ii) they are usually not consumed directly but are a primary inputs in the direct production of other goods and services;

(iii) they cannot be stored; and

(iv) they are frequently required at levels which fluctuate on a more or less regular daily, seasonal or annual basis.

As a result of characteristics (i) and (ii), the demand for public utility services tends to be both price and income inelastic although the price elasticity of demand may rise over time as an increasing number of users are able to adjust their stocks of durable goods in response to price changes. The typically low price elasticity of demand for public utility services may afford them the opportunity to appropriate the consumers' surplus of their clients through the practice of first and third degree price discrimination.[3] As a result of characteristics

---

3. First degree price discrimination occurs where a monopolist can change different rates to each individual consumer and third degree price discrimination occurs where different rates are charged to different classes of consumer.

(iii) and (iv), public utilities have to provide facilities of sufficient capacity to meet demand during peak periods. The peak capacity requirement per consumer can, however, be reduced as a public utility expands its services to a greater number of consumers and regions, as the total peak demand facing it is likely to be less than the sum of individual peaks since the peak periods of some groups of consumers may coincide with the off-peak periods of others. Thus there often exists an incentive for public utilities to supply their services on a national scale. Both railway and electricity utilities usually supply their services on a national grid whereas it is usually not possible for water to be supplied within a similar national network.

The economic characteristics of public utilities often lead to extensive regulation of their activities by the State. In particular, the pricing and investment policies of public utitilies tend to be closely regulated, since an unregulated public utility can use its monopolistic power to "exploit" its clients either by restricting the volume its services or by appropriating their consumers' surplus through price discrimination. Furthermore, public utilities are often obliged to obtain formal approval from the appropriate Governmental authority to offer a new service, or to extend, change or abandon an existing one. The institutional arrangements for regulating public utilities however, vary from country to country. In the United States public utilities are privately owned enterprises controlled by various utility commissions. In the United Kingdom, however, such control has been traditionally effected through State ownership. In South Africa the British model of a state-owned enterprise has been largely been adopted. The institutional character istics of this type of enterprise now needs to be examined in more detail, since in the discussion of the theoretical principles underlying the pricing policies of public utilities in the first two chapters of this thesis, it will be assumed that public utilities are state-owned enterprises.

3. PUBLIC ENTERPRISES AND PRICE DETERMINATION

A public enterprise is essentially a hybrid type of organization. As an enterprise it must produce and sell goods and services and it is therefore subject to similar market and organizational pressures as those which typically face a large-scale, privately-owned firm. On the other hand, as a public organization it is owned and controlled by the Government or its agent and may therefore be subject to direct and indirect pressures from administrators, politicians and the public at large. Jones has summarized the effect of this dichotomy as follows:

"As hybrids, public enterprises share characteristics of traditional government units and private business enterprises, with some being closer to one pole than to the other".[4]

It has been argued , though, that public enterprises are not essentially different from large scale private enterprises which are subject to governmental and public pressures through legal regulation, taxes and subsidies, moral suasion and the demands of various interest groups. It is certainly true that the study of a large scale business organization can yield fruitful insights into the organizational problems affecting a public enterprise and that the government may be able to attain the same objectives through the regulation of private enterprises as with the control of public enterprises. Public enterprises do, however, differ from large private organizations in the sense that they are subject to direct, or internal, rather than indirect government control. As Jones has put it:

"Governments, of course, can influence private managers indirectly, or externally, by passing laws, imposing taxes, and jawboning but this is quite different from public enterprises where government has the power to hire or fire the manager. For a private manager, on the other hand, the discipline imposed by the capital market is typically the major determinant of his tenure".[5]

---

4. Jones, L.P., Public Enterprise in Less Developed Countries, Cambridge University Press, Cambridge, 1982.

5. Ibid., pp. 2-3.

The distinctive nature of the direct form of control which governments exert over public enterprises can be seen in the way in which their pricing policies are determined. While it is true that the State often intervenes in the determination of prices for a number of goods and services produced by private enterprises in a western-type mixed economy, this intervention is usually of short duration and often concerned with merely setting the maximum and minimum levels for market-determined prices. In the case of state-owned utilities, though, pricing policy is controlled by the Government means of "rules" which determine the manner in which tariffs should be formulated, and/or by requiring the public enterprise to publish its tariffs in official schedules which can usually only be changed with the approval of an appropriate government authority. It can thus be seen that the process of price determination is not an independent function of the public enterprise management. In the case of public utilities, it is considered necessary for the State to directly control price determination in order to prevent these undertakings from using their monopolistic power to pursue potentially "exploitative" pricing practices.

The main criteria which have been applied by governments in regulating the process of price determination by public enterprises have been those of "fairness" (or equity) and "financial viability". They have usually been concerned that the tariff structure for public utility services should be "fair" in the sense that no group of consumers should be discriminated against by being charged more than the cost incurred in supplying them. Furthermore it has been considered important that the tariff structure should generate sufficient revenue to at least cover accounting costs and thereby ensure that the public utility remains financially viable without the need for State subsidization. However, these criteria of "fairness" and "financial viability" are often subordinated to the social objectives of the State where it is considered desirable for the utility to provide unremunerative services.

However, as Crew and Kleindorfer have noted:

"In the operation of regulation and public enterprise economic principles have often been in the background of these activities. The principal consideration behind regulation and public enterprise has probably been the desire to avoid monopolistic exploitation (broadly defined)

rather than promote economic efficiency."[6]
Given that policymakers have recently become increasing by aware of the importance of resource allocation, a major part of this thesis will be concerned with examining how the present pricing policies of a number of public utilities may lead to resource misallocation and indicating the direction in which they should be changed to improve allocative efficiency. To derive a theoretically sound concept of economic efficiency it is necessary to examine certain basic propositions of welfare economics.

4. WELFARE ECONOMICS

According to Boulding the purpose of welfare economics is "... to set up standards of judgement by which events and policies can be judged as 'economically desirable', even though on other grounds (political, national, ethical) they might be judged to be undesirable."[7]
It is thus a field of study which can yield a useful set of propositions to guide economic policies aimed at promoting greater efficiency in resource allocation. It should be noted, though, that while, as Pigou has stated, the aim of public policy should be to "promote welfare"[8] welfare economics usually concerns itself only with one aspect of general welfare, normally, the economic aspect which is said to depend on "those things and services which the individual consumes or enjoys, and which could be exchanged for money, and the work done by each individual."[9]

---

6. Crew, M.A., and Kleindorfer, P.R., Public Utility Economics, Macmillan, London, 1979, p. 122.
7. Boulding, K.E., "Welfare Economics" in Haley, B.F., ed., A Survey of Contemporary Economics, R.D.Irwin, Illinois, 1952, p. 3.
8. Pigou, A.C., The Economics of Welfare, Macmillan, 4th ed., London, 1970, p. 11.
9. Little, I.M.D., A Critique of Welfare Economics, 2nd ed., Oxford, 1957.

There are two basic traditions in welfare economics - the Pigovian and the Paretian.[10] Both approaches are aimed at defining the conditions under which an ideal or optimum allocation of resources may be attained. It can be shown that Pigou's rule that the ideal allocation is a situation in which "... each several sort of resource is allocated in such a way that the last unit of it in any one use yields a physical product of the same money value as the last unit of it in any other use "[11] is equivalent to the marginal conditions for Paretian optimality which will be discussed in the Chapter 1.[12] However, the two approaches differ in the sense that while the Pigovian one is based on the assumptions of "... measureable and interpersonal comparability of utility"[13] so that "... a total social utility function could be derived simply by adding the utilities of all individuals at each configuration of the economic universe."[14] the Paretian approach can on the other hand, even with the limitation of ordinal non-addable utilities define an economic optimum "... as a situation in which nobody can move to a position which he prefers without moving somebody else to a position which is less preferred."[15] It is largely because the Paretian concept of ordinal utility is less

---

10. The forerunner of this approach was Pareto, V., in Manuel d'économie politique, Paris, 1909. It was introduced into English - language economics mainly by Lerner, A.P., in "The Concept of Monopoly and the Measurement of Monopoly Power" and Hicks, J.R. in "Foundations of Welfare Economics", Economic Journal, Dec., 1939.
11. Pigou, A.C., op.cit., p. 33.
12. See for example, Nath, S.K., A Reappraisal of Weflare Economics, Routledge and Kegan Paul, London, 1969, pp. 36-37.
13. Ibid., p. 35.
14. Boulding, K.E., op.cit., p. 11.
15. Ibid., p. 12.

restrictive than the Pigovian concept of cardinal utility that the former approach dominated conventional welfare econimics to the extent that "... virtually the entire edifice of economic theory as we known it today is built on Paretian premises."[16] The evaluation of public utility pricing policies in this thesis will be largely based on the propositions derived from the Paretian tradition in welfare economics.

5. OUTLINE OF THE THESIS

In Chapter One general equilibrium analysis is used to examine whether the adoption of a marginal cost pricing policy by a public enterprise would be consistent with the optimization of economic welfare. It will also be concurred with the question of whether known and significant departures from the optimum conditions completely invalidate the marginal cost pricing rule or whether this rule should be adjusted by means of a piecemeal optimization procedure.

Chapter Two will then use partial equilibrium analysis to establish how the marginal cost pricing rules can be applied to the formulation of an optimal pricing and investment policy for a public enterprise and how this policy should be adjusted to deal with non-optimalities. The extent to which these optimal pricing rules can be used as instruments of decentralized control inducing public enterprises to pursue their objectives without ad hoc State interference in their decisionmaking will then be examined.

Chapter Three examines the pricing policy of the South African Transport Services (S.A.T.S.) which is inter alia, the utility responsible for the provision of rail services in South Africa. The special characteristics of the supply and demand for rail services are initially discussed to provide the theoretical background against

---

16. Winch, D.M., Analytical Welfare Economics, Penguin Modern Economics (1971), pp. 199-200.

which the present pricing policy can be evaluated in terms of its effect on economic efficiency and industrial decentralization. A historical review of transport policy in South Africa will show that many of the problems associated with the pricing policy of S.A.T.S. arise from it being required to pursue conflicting objectives. Finally a framework for deregulation will be proposed and the effect of deregulation on the pricing policy of S.A.T.S. evaluated.

Chapter Four deals with the pricing policy of the Electricity Supply Commission (Escom). In the first part of this chapter the supply and demand characteristics of an electricity utility and the practical implementation of a tariff structure based on marginal cost principles are discussed in general terms. The recent financial performance and present tariff structure of Escom is then examined with this tariff structure being evaluated in terms of criterial of equity, efficiency and regional equality.

In Chapter Five the special characteristics of water supply are described and the present pricing policies with respect to both the water supplied for irrigation and for industrial and domestic purposes are evaluated in terms of the effects on allocative efficiency and industrial decentralization in South Africa. The practical problems associated with the application of optimal pricing principles to the supply of water will then be discussed.

The thesis will be concluded with a brief discussion of the differences and similarities between the existing and proposed pricing policies for rail, electricity and water in South Africa.

CHAPTER ONE

THE APPLICATION OF WELFARE THEORY TO PUBLIC ENTERPRISE PRICING

INTRODUCTION: VALUE JUDGEMENTS AND NORMATIVE WELFARE ECONOMICS

In this chapter the question will be examined whether it is possible to derive from the theory of welfare economics, rules which may be generally applied to public enterprise pricing. Unlike private enterprises which can, at least in principle, adopt pricing policies based on the goal of profit maximization, a public enterprise could be expected to set its prices to maximize net social benefit or welfare. The evaluation of alternative pricing policies and the determination of an optimal public enterprise pricing policy therefore clearly fall within the field of applied welfare economics, since Mishan[1], for example, has defined welfare economics as "that branch of study which tries to formulate propositions by which we may rank on a scale of better or worse alternative situations open to the economy".

Welfare is, however, essentially an ethical term and any attempt to rank different economic states must be based on value judgements. Even the criteria derived from conventional Paretian welfare economics are based on implicit value judgements although it is usually claimed that these are "so widely acceptable", "general" or "minimal" that the welfare propositions based on them may be considered quite general, non-controversial or "more or less objective". Before one can examine whether Paretian welfare criteria may be applied in an evaluation of public utility pricing policies, it is necessary to consider whether the a priori approach

1. Mishan, E.J., "A Reappraisal of the Principles of Resource Allocation", Economica, November 1957, p. 322.

of Faretian welfare economics is justifiable. In other words, should an economist make prescriptions on the basis of norms established by widely accepted value judgements.

In this context, Nath[2], for example, argues that it is a logically invalid procedure to adopt Paretian value judgements as norms on the grounds of their purported wide acceptability. He bases this contention on a statement by Popper that "it is impossible to derive a sentence stating a norm or a decision or, say, a proposal for a policy from a sentence stating a fact"[3]. Furthermore, Nath suggests that the procedure of a priori welfare economics is "based on a confusion between decisions and recommendations or comments".[4] In a democratic society, decisions have to be based on value judgements which are at any moment widely acceptable to the designated political representatives of individual citizens. There is, however, no moral or logical necessity for policy-orientated comments and recommendations to be based on propositions that are widely acceptable. In terms of this view, it is necessary to make a distinction between the positive and normative aspects of applied economics. An economist is said to be qualified to use positive theory to predict the effects of different policies. It is questionable, however, as to whether he is similarly qualified to make a normative assessment of such predictions. According to one point of view the economist should limit his role to that of a "consultant" only. Mishan interprets this viewpoint as follows:

---

2. Nath, S.K., A Reappraisal of Welfare Economics, Routledge and Paul, London, 1969, p. 2.
3. Popper, K.R., The Logic of Scientific Discovery, London, Chapter 5, 1959.
4. Nath, S.K., op.cit., p. 129

> "He is to confine himself to disclosing the implications of the policies being mooted, and perhaps those also of alternative policies, so enabling political representatives of the people to select policies that are consistent with one another and that accord with the wishes of society at large".[5]

Nath, however, considers that, in his role of commentator, an economist can make normative recommendations, but that he must also make his value judgements explicit and clearly state that they are based on his own predilections. In Nath's opinion "there is no logical warrant for economists in a democratic society to choose some value judgements and make them into basic norms in their professional evaluative work <u>because</u> they are widely acceptable. Any proposed welfare criterion and <u>type</u> of welfare functions are no more than personal opinions of the particular economists, which the public and the other economists may accept or reject depending on how much moral appeal those criteria make to them and quite irrespective of how widely these basic value judgements may be accepted".[6]

Mishan has perceived that this sort of argument attacks the very foundations of conventional welfare economics since "... it carries with it the judgement that there is at present no rule or method by which we may judge of the relative efficiency of alternative forms of economic organization; indeed, that any conceivable set of production plans or any conceivable set of prices no matter how arbitrary, are - in our present state of knowledge - to be considered as good as any others."[7] He concedes that there is no role for <u>a priori</u> welfare economics in a society where decisions are

---

5. Mishan, E.J., <u>Welfare Economics: An Assessment</u>, North Holland Publishing Company, Amsterdam, 1969, p. 14.
6. Nath, S.K., <u>op.cit</u>., p. 129.
7. Mishan, E.J., "A Reappraisal of the Principles of Resource Allocation", <u>Economica</u>, November 1957, p. 324.

reached on the basis of a majority rule voting procedure in the face of fixed individual rankings. Where, however, there are political institutions which provide a forum for informed debate "consistency of principle will be one of the guiding principles".[8] The search for consistency will tend to lead to the formulation of a common set of criteria which are applicable to a range of comparable economic issues:

"Once these criteria have been made explicit, economists can question certain economic decisions not merely by juxtaposing their implications with those of other decisions, but more directly by reference to these criteria themselves".[9]

Mishan, therefore, rejects the view that the role of the economist is that of a consultant only. He suggests that it may be impracticable for the economist to avoid making recommendations since where the economist presents the policymaker with a detailed list of all the economic implications of each of the several policy proposals under consideration, he "is almost sure to be asked to 'organize' the raw data, and to provide some method by which the large variety of consequences for each policy may be summarized, or weighted in some way, to enable the politician to compare them and reach a decision on the 'best' policy to adopt".[10] If the advising economist did this on the basis of his personal predilections, this would place rather arbitrary powers in his hands. Mishan suggests that the economist should use in his normative evaluative work, a consistent set of criteria derived from what he terms the "virtual constitution" of the community.

---

8. Mishan, E.J., op.cit., p. 20.
9. Ibid., p. 20.
10. Ibid., p. 16.

This virtual constitution includes "the minimal number of ethical provisions to which the community will unanimously subscribe irrespective of their present exigencies and material interests".[11] Mishan considers that only the following two ethical propositions can be ascribed to a virtual constitution for a western-type democratic society:

> "(i) that social welfare rises if there is a "Pareto improvement", if everyone is made as well off or better off, and
> (ii) that social welfare is increased if the distribution of welfare is better in some sense."[12]

Mishan's propositions fall clearly within the Paretian tradition of welfare economics. It is interesting to compare them with Pigou's definition of a welfare improvement. Pigou contends, firstly, that " provided that the dividend accruing to the poor is not diminished, increases in the aggregate national dividend of the community, unless they result from coercing people to work more than they wish to do carry with them increases in economic welfare".[13] As was explained in the previous section the cardinalist concept of a maximum "national dividend" is comparable in a static sense to the ordinal Paretian concept of an optimum allocation of resources. Secondly Pigou proposes that :

> "Any cause which increases the absolute share of real income in the hands of the poor, provided that it does not lead to a contraction in the size of the national dividend from any point of view, will, in general, increase economic welfare."[14]

---

11. Ibid., p. 18.
12. Ibid., pp. 25-26.
13. Pigou, A.C., The Economics of Welfare , 4th ed., Macmillan, London, 1960, p. x
14. Ibid., p. 89.

This proposition is based upon the "law of diminishing marginal utility" and implies that, since the marginal utility of money income of the rich will be lower than that for the poor, a redistribution of income from the rich to the poor will, ceteris paribus, lead to an improvement in welfare. The problem with this cardinalist approach is that it assumes that all individuals have identical utility functions for money income, so that with an equal income distribution all would have the same marginal utility of money. As Koutsoyiannis has pointed out:

> "This assumption is too strong. Individuals differ in their attitudes towards money. A rich person may have a utility for money function the lies far above the utility (for money) function of poorer individuals. In this case a redistribution of income (towards more equality) might reduce total welfare."[15]

Mishan appears to have recognized this problem since he does not specify the nature of the income redistribution which will lead to an improvement in welfare although he concedes that "in western societies a 'better' distribution invariably suggests a more equalitarian distribution of the material product."[16]

In order to examine the relevance of a priori welfare economics to this study, admittedly on its own terms, it has been decided to accept Mishan's argument that it is justifiable to derive norms from value judgements which are based on the virtual constitution of the community. By examining the necessary conditions for Pareto optimality and how they are affected by introducing some of the complications which characterize the operation of public utilities within a market economy, it is hoped to establish a framework from which criteria can be derived for a piecemeal assessment of different public utility pricing policies.

---

15. Koutsoyiannis, A., Modern Microeconomics, 2nd ed., Macmillan, 1979, p. 525.
16. Mishan, E.J., op.cit., p. 26.

1:1 THE OPTIMALITY OF MARGINAL COST PRICING

The starting point for any comparison between alternative public enterprise pricing policies and the formulation of an optimal public enterprise pricing policy is the concept of marginal cost pricing (MCP). This section will therefore examine whether it is possible to derive from the basic propositions of welfare economics the conclusion that the adoption of a MCP policy by a public enterprise will result in an improvement in welfare compared to a policy which results in prices being set above or below marginal costs.[17]

In order to make this welfare comparison, it will be assumed that the public enterprise operates within what has been termed a "first best" economy.[18] In other words, although the public enterprise has the discretion to choose its own pricing policy, all other markets are assumed perfectly competitive. The firms operating in these other markets aim at maximizing their profits while every consumer is assumed to aim at maximizing his utility. It is also assumed that there is perfect information and resource mobility, consumers tastes are independent, transitive and rational, production factors are homogeneous and divisible while production functions are characterized by perfectly divisibility, unlimited factor substitutability, diminishing

---

17. The marginal cost pricing rule is usually compared with the rule that prices should be set equal to average cost (AC). It is only at the minimum point of the AC function that AC equals MC so that, at this point, the application of either rule will result in the same price being charged. At any other level of output the application of the average cost pricing rule will result in prices being set above or below marginal cost.

18. See, for example Rees, R., Public Enterprise Economics Weidenfeld and Nicolson, London, 1978.

marginal productivities and constant returns to scale. Furthermore it is assumed that in such an economy there are no externalities in either production or consumption.

If one accepts the Paretian value judgements that:

(i) the welfare of society depends only on the welfare of the individuals comprising society;

(ii) any non-economic causes affecting an individual's welfare can be ignored;

(iii) an individual should be considered the best judge of his own welfare; and

(iv) social welfare rises if any one person is made better off provided that no one else is made worse off;

then one can evaluate alternative pricing policies in terms of the extent to which they satisfy the various conditions for Pareto optimality.

Figure 1.1 represents a "first best" economy where it is assumed for simplicity that there are two homogeneous groups of consumers A and B, who each own two production factors, labour (L) and capital (K), which are used to produce two commodities X and Y with X being produced by perfectly competitive firms and Y being produced by a public enterprise. The public enterprise must choose between adopting a MCP policy which would result in the economy moving to point M where OX_1 of good X and OY_1 of good Y are produced or a policy of setting its prices PY above marginal cost, in which case the economy will attain the output mix (OX_2OY_2) at point N.

It can be seen that both output mixes lie on the transformation curve TT'. This is a locus of all output combinations which satisfy the Paretian condition for an efficient allocation of factors between commodities that the marginal rates of technical substitution of L for K in the production of X and Y be equal; or

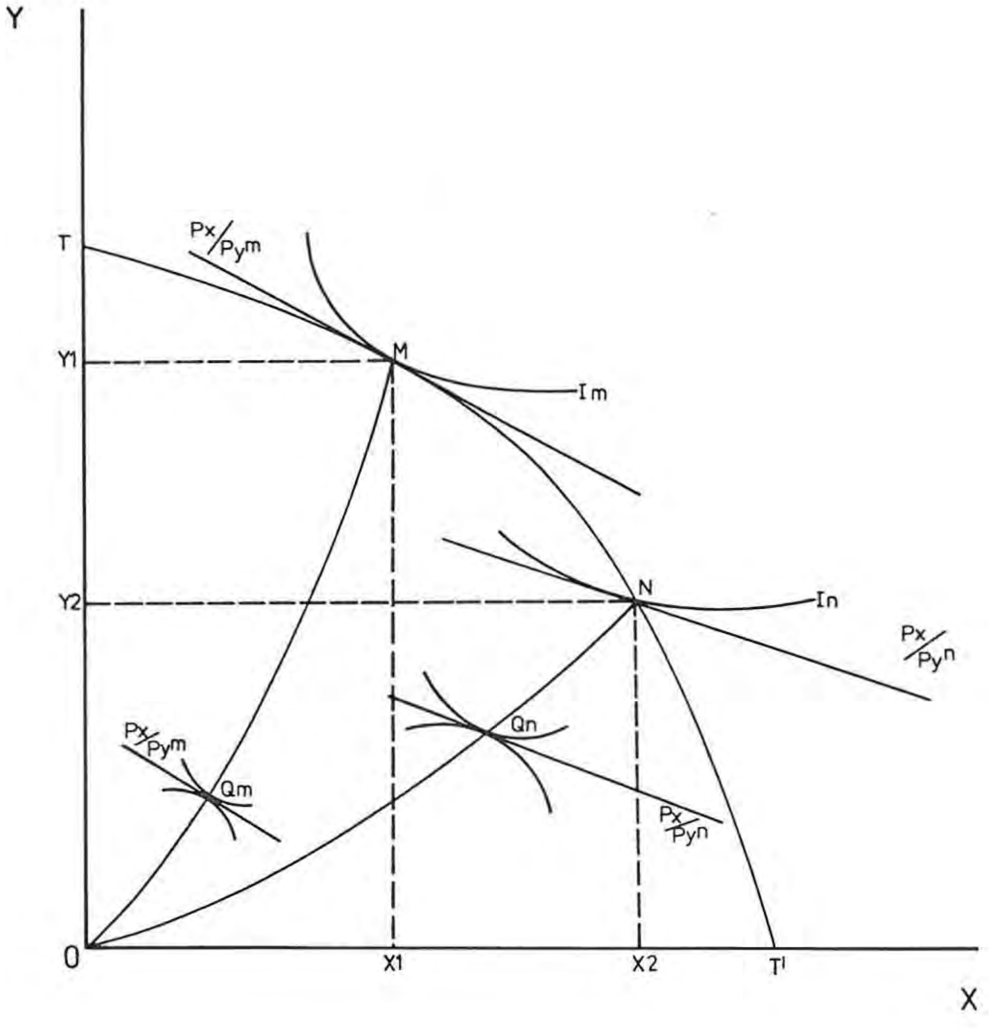


FIGURE 1:1 Pareto Optimality in a First Best Economy

$$\frac{dQx/dLx}{dQx/dKx} = \frac{dQy/dLy}{dQy/dKy} \qquad (1{:}1)$$

or $$\text{MRTS } e^{x}_{,}k = \text{MRTS } e^{y}_{,}k \qquad (1{:}1)^1$$

If this condition is satified then it can be shown that the slope of the transformation curve, which is termed the marginal rate of product transformation of X for Y (MRPTx,y) will be equal to the marginal cost ratio MCx/MCy[19]. The Paretian condition for

---

19. Since each point along TT' fulfills the condition that

$$\frac{dQx/dLx}{dQx/dKx} = \frac{dQy/dLy}{dQy/dKy} \qquad (1{:}1)$$

this can be rearranged as:

$$\frac{dQy/dLy}{dQx/dLx} = \frac{dQy/dKy}{dQx/dKx}$$

Now since the marginal cost of producing X and Y (MCx and MCy) can be expressed as

$$MCx = \frac{w}{dQx/dLx} \quad \text{or} \quad MCx = \frac{r}{dQx/dKx}$$

and

$$MCy = \frac{w}{dQy/dLy} \quad \text{or} \quad MCy = \frac{r}{dQy/dKy}$$

where w and r are the prices of labour and capital respectively it follows that:

$$\frac{MCx}{MCy} = \frac{dQy/dLy}{dQx/dLx} \qquad (1{:}2)$$

technical efficiency in (2:1) will be satisfied if a public enterprise operates in a first best economy and if it is assumed that the public enterprise minimizes costs by choosing that input combination where MRTSI, $^{y}k = w/r$, where w and r are the prices of capital and labour respectively, since in the perfectly competitive factor markets it will face the same factor prices as the other perfectly competitive producers who will only be able to survive if they minimize costs by equating their marginal rates of technical substitution with the factor price ratio, w/r. As a result the marginal technical rate of substitution for the public enterprise will equal that of the perfectly competitive firms and the Paretian condition for technical efficiency will be attained at both M and N.

In figure 1:1 the conventional Edgeworth - Bowley exchange box diagrams OX_1MY_1 and OX_2NY_2 can be drawn to determine the Paretian efficient allocation of the respective output mixes between consumers A and B. In both exchange boxes, point C represents the origin for the indifference maps of consumer A while points M and N represents the respective origins for the indifference maps of consumer B. The exchange contract curves OM and ON each represent loci of output combinations at which A's and B's respective marginal rates of substitution of X and Y are equalized.

$$\frac{dUa/dQ^{a}x}{dUa/dQ^{a}y} = \frac{dUb/dQ^{b}x}{dUb/dQ^{b}y} \qquad (1:3)$$

or $$MRSx^{a}_{,}y = MRSx^{b}_{,}y \qquad (1:3)^{1}$$

These exchange contract curves satisfy the Paretian condition for an efficient distribution of commodities between consumers since each point along these curves indicates the maximum utility attainable for one consumer, given the utility level of the other consumer. Regardless of whether the public enterprise sets its price above or equal to marginal cost, a point on either the exchange contract curve OM or CN will be attained since all consumers maximize their utility by equating their marginal rates of substitution with the price ratio, and, if the public enterprise operates in a first best economy, all consumers will face the same product price ratios, so that their marginal rates of substitution will be equalized, or;

$$MRSx^{a}_{,}y \quad = \quad MRSx^{b}_{,}y \quad = \quad Px/Py \qquad (1:4)$$

It can thus be seen that the attainment of the two lower order conditions for Pareto optimality will be independent of the pricing policy adopted by the public enterprise. The alternative pricing policies can only be compared by considering the extent to which they satisfy the top order condition for a Pareto optimal allocation of factors between commodities which requires that the MRPT between any two commodities be equal to the MRS between the same two goods:

$$MRPTx,y \quad = \quad MRSx^{a}_{,}y \quad = \quad MRSx^{b}_{,}y \qquad (1:5)$$

This top level condition for Pareto-optimality will be attained where the rate at which goods can be transformed into another is equal to the rate at which consumers are willing to exchange a good for another. For example, if MRSx,y is in the ratio of one to one and MRPTx,y is in the ratio one to two, producers can reduce production of x by one unit and make two more of y. Of these, only one y is needed to compensate consumers for the loss of x. The other is available to make everyone better off. Thus, in principle, a rearrangement of production can increase welfare when MRS does not equal MRPT.

In a first best economy, if a perfectly competitive firm is to survive it must set its price equal to marginal cost so that

$$Px = MCx.$$

The survival of a public enterprise is, however, guaranteed and it has the discretion to choose its pricing policy. If it chooses a pricing policy which results in prices being set above marginal cost, or

$$Py > MCy$$

it follows that:

$$\frac{Px}{Py} < \frac{MCx}{MCy}$$

so that the Paretian condition for allocative efficiency will not be attained, that is:

$$MRSx^{a}_{,}y = MRSx^{b}_{,}y = \frac{Px}{Py} < \frac{MCx}{MCy} = MRPTx,y$$

If, however, the public enterprise adopts a marginal cost pricing policy then:

$$Py = MCx$$

and

$$\frac{Px}{Py} = \frac{MCx}{MCy}$$

so that allocative efficiency will be attained:

$$MRSx^{a}_{,}y = MRSx^{b}_{,}y = \frac{Px}{Py} = \frac{MCx}{MCy} = MRPTx,y \qquad (1{:}6)$$

In figure 1:1 the community indifference curve,[20] IM represents a locus of all the output combinations for which the equalized marginal

---

20. The problems associated with the construction of a community indifference curve were first raised by Scitovsky, T., in his article "A Reconsideration of the Theory of Tariffs", Review of Economic Studies, IX, 1942, pp. 89-110, and later elaborated on by Samuelson, P.A., in his article "Social Indifference Curves" in the Quarterly Journal of Economics, February 1958, pp. 1-22.

marginal rates of substitution for the two consumers have the same value as at point QM on the exchange contract curve OM where the marginal rates of substitution are equal to the price ratio Px/Py^m which results when the public enterprise adopts a marginal cost pricing policy, i.e.

$$MRSx_{,y}^{a} = MRSx_{,y}^{b} = \frac{Px}{Py^m}$$

It can be seen that this indifference curve is tangential to the transformation curve, TT', at point M which indicates that the adoption of a marginal cost pricing policy satisfies the condition for Paretian allocative efficiency in (2:6). The community indifference curve IN represents a locus of all output mixes where the equalized MRS's are equal to their value at point Qn where $MRSx_{,y}^{a} = MRSx_{,y}^{b} = Px/Py^n$. The fact that IN is not tangential to the transformation curve indicates that Paretian allocative efficiency will not be attained when the public enterprise adopts a pricing policy which does not result in price being equal to marginal cost.

The tangency between the community indifference curve IM and the transformation curve, TT', at point M is, however, sensitive to changes in the distribution of income if the relative preference for good X and Y differ at the margin for consumers A and B. A redistribution of income will change the value of the equalized marginal rates of substitution and result in a new community indifference curve which will intersect IM at point M. Point M will thus no longer represent an allocatively efficient output combination. The Pareto optimality of any point on the transformation curve can thus only be established in terms of a given distribution of income.

If the point of welfare maximization is to be determined then an interpersonal comparison of utility must be made. It must be assumed that a Bergson welfare function of the form

$$W = W(Ua, Ub) \qquad (1:7)$$

exists.

This encompasses all the Paretian value judgements and presupposes that an ethical valuation of the relative merit or worthiness of the individuals concerned can be made. This Bergsonian value judgement has been described by Baumol[21] as the "social judgement par excellence". The Bergson welfare criterion is more complete than the dual welfare criteria included in a "virtual constitution" since it assumes that society can rank any possible distribution of utility between individuals and can therefore judge whether a particular economic situation has a higher level of welfare than another situation even though it may not be a Pareto or distributional improvement in welfare.

The Bergson welfare function can be represented in commodity space if one accepts the Paretian value judgement that an individual is the best judge of his own welfare which is not affected by non-economic causes so that

$$Ua = Ua(Q^a x, Q^a y) \quad \text{and} \quad Ub = Ub(Q^b x, Q^b y) \qquad (1{:}8)$$

It follows then that the welfare function

$$W = W(Ua, Ub)$$

can be expressed as

$$W = W \;\; Ua(Q^a x, Q^a y), \;\; Ub \;\; (Q^b x, Q^b y) \qquad (1{:}7^1)$$

It follows that using the "function of a function rule"

$$W = V(Qx, Qy) \qquad (1{:}7^{11})$$

and $$Wo = V(Qx, Qy) \qquad (1{:}9)$$

---

21. Baumol, W.J., <u>Welfare Economics and the Theory of the State</u>, London, G.Bell & Sons, 1965.

(1:9) is one of a set of non-interesting community indifference curves determined after a social value judgement has been made as to the optimal distribution of utility between consumers.[22] One such community indifference curve is shown in Figure 1:2. The rate of change of this community indifference curve in output space is termed the community marginal rate of substitution of X for Y (CMRSx,y) that is

$$\frac{dW/dQx}{dW/dQy} = CMRS_{x,y} \qquad (1:10)$$

Combining (1:10) with (1:6) we get

$$CMRS_{x,y} = MRS^{a}_{x,y} = MRS^{b}_{x,y} = MRPT_{x,y} \qquad (1:11)$$

which brings together all the necessary and sufficient conditions for the maximization of social welfare. This "optimum optimorum" is depicted at point E in Figure 1:2 where the transformation curve is tangential to the community indifference curve Wo. At E the social and individual utility levels are all maximized simultaneously.

Now if a public enterprise adopts a MCP policy the economy may attain a point on the transformation curve where the output mix is different from that existing at the position where the

$$CMRS_{x,Y} = MRS^{a}_{x,y} = MRS^{b}_{x,y} = MRPT_{x,y} \qquad (1:11)$$

---

22. Graaff, J., de V., has shown that a Bergsonian community indifference curve represents the inner limit of a set of Scitovsky community indifference curves in Theoretical Welfare Economics , Cambridge University Press, 1957, p. 49.

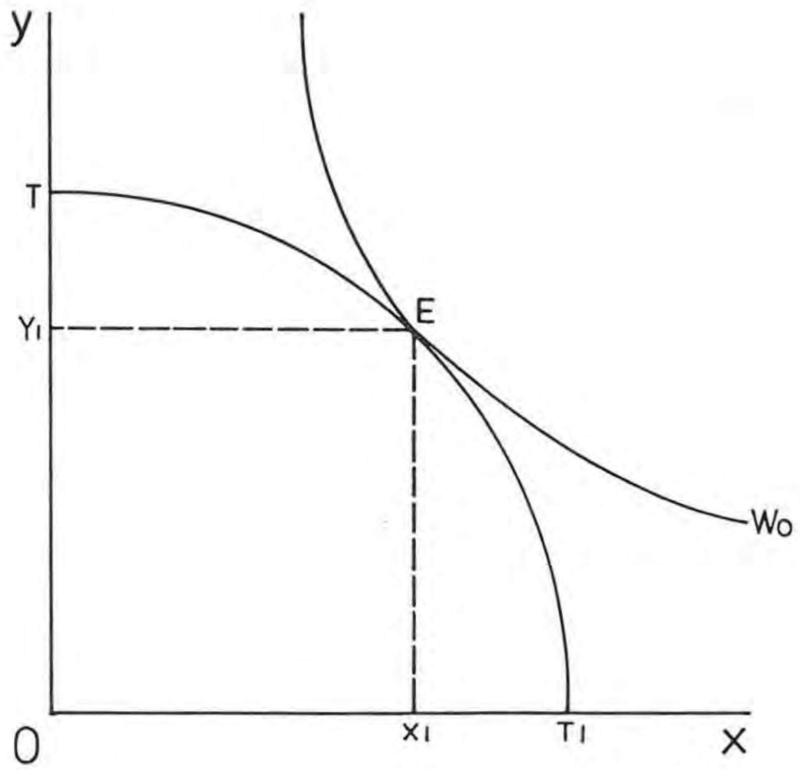


FIGURE 1:2 Community Indifferences Curve

If the MCP pricing policy is to be consistent with welfare maximization in a first best economy, it is necessary to assume that it is possible to redistribute income through lump sum taxation so that the marginal conditions for optimality are not violated. In this case the distribution of income can be adjusted to the level consistent with the optimal output mix at point E. Marginal cost pricing by a public enterprise is therefore a necessary but not a sufficient condition for welfare maximization. Any other pricing policy, however, will not result in the optimal output combination and income distribution being simultaneously obtained.

It can be concluded that marginal cost pricing will be optimal if:

(i) Public enterprises operate within a first best economy;

(ii) A welfare function which contains an explicit interpersonal comparison of utility exists;

(iii) Redistribution of income through lump sum taxation is feasible.

The next section will examine whether it is possible to make a welfare comparison between alternative public enterprise policies without making an interpersonal comparison of welfare. In other words, even if one cannot definitely say that welfare is maximized at an output mix such as point M in Figure 1:1, the question that still remains is whether one can say that an improvement in welfare occurs when the economy moves from point N to M as a result of the public enterprise adopting a marginal cost pricing policy.

1:2 A WELFARE COMPARISON OF ALTERNATIVE PRICING POLICIES WITHOUT AN INTERPERSONAL COMPARISON OF UTILITY

Figure 1:3 indicates how the criteria embodied in Mishan's concept of a "virtual constitution" can be used to make a welfare comparison between a marginal cost pricing policy which results in the economy attaining the output mix at point N and a policy of changing a price above marginal cost which results in the point N on the transformation curve TT' being reached.

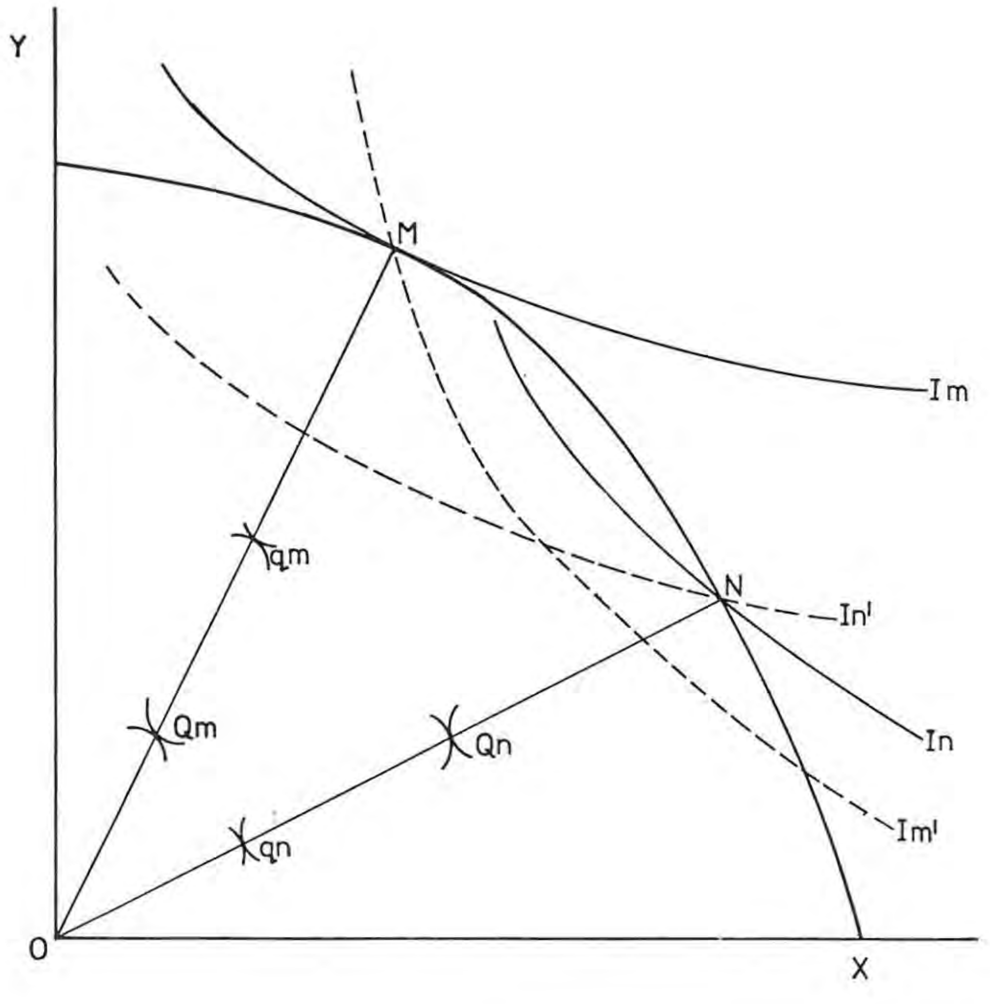


FIGURE 1:3 A Welfare Comparison between Alternative Pricing Policies without an Interpersonal Comparison of Utility

A simple welfare comparison between these two points based on Pareto principles is not possible since the distribution of income at point M will be different from N. This can be explained if it is assumed that the public enterprise producing good Y is generally more capital intensive than the competitive firms producing good X. In order to move from N to M on TT' a reallocation of factors from N to M must take place which will alter both the factor price ratios obtained in factor markets and the pattern of factor rewards in the economy. This redistribution of income will result in some households being better of while others are worse off so that it is not possible to state that M represents a welfare improvement over N without making some judgement about the respective distributions of incсme at these respective points.

The hypothetical compensation technique developed by Kaldor[23] and Hicks[24] can be used to determine whether M represents a higher level of welfare than N without having to make an interpersonal comparison of utility. Point qn on ON represents a hypothetical distribution of the output mix at N between consumers A and B, such that it also represents the same distribution of utility between the two individuals as point Qm on the exchange contract curve OM. A community indifference curve In' can be generated to pass below the point M indicating that, beginning with the distribution of utility associated with Qm and qn, society will be at a lower level of social welfare with the N as opposed to the M output combination. The adoption of a marginal cost pricing policy which would result in a move from N to M can thus be seen to lead to a clear improvement in welfare in terms of the Kaldor-Hicks compensation criterion.

---

23. Kaldor, N., "Welfare Propositions in Economics", E.J., 1939.

24. Hicks, J.R., "The valuation of Social income", Economica, 1940.

However, as Scitovsky[25] demonstrated, the Kaldor-Hicks test is negated once the economy moves to a new position and is then compared with the old position. In Figure 1:3 the point qm on the exchange contract curve OM represents the same distribution of utility between the two consumers as the point Qn on the curve ON. The community indifference curve Im' can be generated tc pass through M and below N indicating that when the Scitovsky reverse test is applied, point N represents a welfare improvement over point M. It should be noted again that M only represents a point of tangency between the transformation curve and the community indifference curve Im, associated with the Qm distribution, and that if the hypothetical qm distribution of M is employed in order to make it Pareto comparable with N the tangency characteristic of M is destroyed.

Little[26] has suggested a possible solution to this problem by proposing a dual welfare criterion similar to that included by Mishan in his "virtual constitution". If the distribution of utility at Qm and qn are considered to be distributionally superior to the distributions at qm and Qn then it can be demonstrated that a movement from N to M will result in an unambiguous improvement in welfare. This follows from the fact that while Qm is Pareto superior to qn which is distributionally superior to Qn (Qm pqn dQn) and although Qn is Pareto superior to qm, qm is considered to be distributionally inferior to Qm (Qnpqmd Qm.) Thus only M is superior according to both the efficiency and distributional components of the dual welfare criterion.

---

25. Scitovsky, T., "A note on Welfare Propositions in Economics", R.E.S., 1941.
26. Little, I.M.D., A Critique of Welfare Economics, 2nd ed., O.U.P., 1950.

An implication of this analysis is that unless the distribution of income associated with the top level optimum, which would be reached if a marginal cost pricing policy is successfully implemented m is favoured above all other distributions, then it cannot be claimed that such a policy would result in an unambiguous improvement in welfare. As Mishan has stated:

> "Far from an optimum allocation of resources representing some kind of ideal output separable from an independent of interpersonal comparisons of welfare, a particular output retains its optimum characteristics only if we commit ourselves to the particular welfare distribution associated with it."[27]

There are many situations, however, where there is a trade-off between equality and efficiency[28], in which case the propositions of welfare economics can offer no clear-cut indication of which situation is unambiguously better or worse. In this thesis it will be examined whether such a trade-off exists between efficient public enterprise pricing policies and the impact of such policies on the regional distribution of income. It would seem that in cases where this trade-off exists, the economist can do no more than indicate the costs in terms of equity of an efficient policy and vice versa, recognizing that the final choice must be based on political value-judgements.

---

27. Mishan, E.J., op.cit., p. 70.
28. Okun, A., has examined the implications of this trade-off in a democratic society within the context of the Rawlsian Theory of Justice in Equality and Efficiency: The Big Tradeoff, The Brookings Institution, Washington, 1975.

1:3 THE EXTENT TO WHICH WELFARE COMPARISONS BETWEEN PRICING POLICIES ARE AFFECTED BY RELAXING RESTRICTIVE ASSUMPTIONS

Having defined the conditions under which the adoption of a marginal cost pricing policy by a public enterprise is likely to result in an unambiguous improvement in welfare, it is now necessary to examine the extent to which the optimality of requiring public enterprises to follow a marginal cost pricing rule will be invalidated if a number of the assumptions which define a first best economy are relaxed. This is not an exhaustive study of all the imperfections which could possibly occur in a market economy, but rather a discussion of some of the departures from the conditions for Pareto optimality which are particularly likely to affect the activities related to the public enterprise sector.

1:3:1 INCREASING RETURNS TO SCALE

Public enterprises often experience increasing returns to scale over the entire output range in the markets where they operate. This is even more likely to be the case in a country like South Africa where the size of the market for their goods or services is relatively small, say, compared to the United States.

Now, the existence of increasing returns to scale is incompatible with perfect competition and may cause a natural monopoly to come into being.[29] Figure 1:4 illustrates why this may occur. The consequence of increasing returns to scale is that the long run average cost curve (LAC) falls as output increases. As a result

---

29. See, for example, Howe, K.M. and Rasmussen, E.F., Public Utility Economics and Finance , Prentice-Hall, New York, 1982, p. 19.

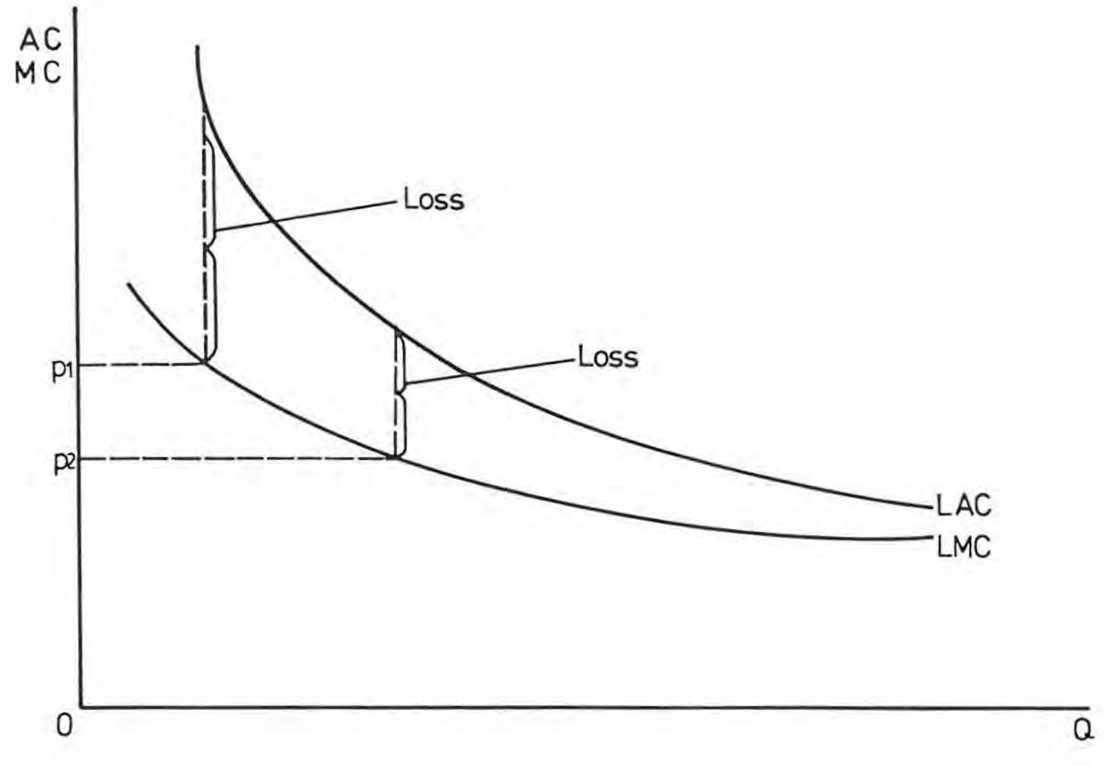


FIGURE 1:4 Increasing Returns to Scale

the long run marginal cost (LRMC) curve lies below it over the entire output range. The condition for profit maximization in a perfectly competitive market is that all firms set their MC equal to the respective market prices. In an industry where there are increasing returns to scale this will lead to individual firms incurring perpetual losses and closing down until eventually a natural monopoly will emerge. Since a monopolist will maximize profits by restricting output to the point where marginal revenue equals marginal cost and charging a price above marginal cost, this will lead to a deviation from the conditions required for Pareto optimality.

The chief economic justification for the nationalization of natural monopolies is that social welfare will be improved by requiring these public utilities to set their prices according to optimizing welfare criteria. If none of the other assumptions underlying the welfare maximization model have been violated and if distributional consideration are ignored, this will normally involve public utilities setting their prices equal to marginal cost. As can be seen from Figure 1:4, however, this will still leave them with the problem of financing their resulting losses. Hotelling[30] suggested that marginal cost pricing would be optimal even where a public utility incurred perpetual losses, if these losses were financed by raising lump sum taxes. It has, however, generally been accepted that there are no feasible taxes available to the government which are of a truly lump sum character. Any feasible tax is thus likely to distort the conditions for Pareto optimality. This is obviously true of indirect taxes which result in the price ratio facing consumers being different from that facing firms so that:

$$MRSx^{A}_{,y} = MRSx^{B}_{,y} = \frac{Px}{Py} \neq \frac{MCx}{MCy} = MRPTx,y.$$

---

30. H.Hotelling "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," Economica, 1938.

Less obviously, it is also the case with an income tax which distorts the Pareto optimal rate of substitution between work and leisure.[31] It is therefore clear that a marginal cost pricing policy may not be consistent with a Pareto optimal general equilibrium being attained in the economy as a whole. Furthermore a deficit financed by taxation may result in a redistribution of income from the taxpayer to the user of the services of the public enterprise. Such a redistribution is likely to ignore considerations of equity and may thus be regarded as arbitrary in nature. In the previous section it was shown that even where a new pricing policy moved the economy to a position which represented a potential Pareto improvement in efficiency, if it did not also involve a distributional improvement, then it could not be concluded that there is an increase in welfare. It follows that the adoption of a marginal cost pricing rule by a public enterprise does not unambiguously raise the welfare level in terms of criteria pertaining to either efficiency or equality.

1:3:2 EXTERNALITIES

In asserting that Pareto optimality will be attained when public utilities adopt a marginal cost pricing rule and all other markets are perfectly competitive, it has been implicitly assumed that all benefits and costs of producers and consumers are reflected in market prices, and that there is no divergence between private and social costs and benefits. However, for a long time, economists have recognized that externalities arising out of interdependencies between production and utility functions may cause a perfectly competitive economy to depart from the conditions for Pareto optimality.

---

31. Ruggles, N., "Recent developments in the theory of Marginal Cost Pricing", Review of Economic Studies, 1950-1.

During the 1950s and 1960s a great deal of the discussion on externalities focussed on their definition and classification.[32] The definitions proposed concentrated largely on the element of interdependence among utility or production functions. This interdependence may give rise to externalities between different producers, different consumers and between producers and consumers. In particular the provision of public utility services may result in externalities since they affect a broad range of activities and may be interdependent with the activities of a number of other producers and consumers.

Buchanan and Stubblebine[33] have defined an externality to be present in the relationship between two individuals, A and B, if at least one of the utility functions takes the form:

$$U^A = U^A(a1, a2, \ldots, an, bk) \qquad (1:12)$$

where a1, a2,...., an represents the actitivies of individual A and bk represents the single activity of individual B which affects the utility of individual A. A marginal externality exists when,

$$dU^A / dbk = 0. \qquad (1:13)$$

---

32. The main contributions in this area were by Meade, J.E., "External Economies and Diseconomies in a Competitive Situation", Economic Journal, Vol. 62., March 1952, Scitovsky, T., "Two Concepts of External Economies", Journal of Political Economy, Vol. 62., April, 1954, Buchanan, J., and Stubblebine, W., "Externality", Economica, Vol. 29., November, 1962 and Whinston, A., Price Coordination in Decentralized Systems, Carnegie Institute of Technology, June, 1962.

33. Ibid., p. 372.

Buchanan and Stubblebine have defined a marginal externality to be "... Pareto-relevant when the extent of the activity may be modified in such a way that the externally affected party, A can be made better off without the acting party, B being made worse off"[34]. In particular, when B maximizes his utility an externality will be "Pareto-relevant" if the partial derivative of A's utility function with respect to his own activities is not equal to zero since in this situation there can still be some gains from trade between individuals A and B. Figure 1:5 illustrates the nature of the externalities which arise between two producers using congested facilities. It should be noted that essentially the same analysis will apply in the case of consumption externalities.

For the purposes of this example, it is assumed that highway facilities along with other resources are used in the production of X and Y. The production of these goods is thus considered as a process which is only completed when the goods are finally delivered to consumers. In the initial position at point E, highway facilities are used to such and extent by both X and Y, that congestion and transport delays result. It follows that altough the marginal rate of technical substitution between highway facilities and other resources is the same at point E, this may not necessarily be an optimal allocation of resources, since the reduction in the use of highway facilities by X will increase the productivity of highway facilities for Y. The producers of good X move from point E to F along isoquant X1 as they exactly compensate for the loss in output resulting from reducing their use of highways by increasing their use of alternative forms of non-congested output. The increase in the productivity of highway facilities for the producers of Y is depicted by the inward shift of their set of isoquants to the Oy origin. The isoquant for the output of Y1 units of Y is now showned by the dashed line Y1'. At point F, the production of Y will be at Y2', a higher level than before. At the same time there will

---

34. *Ibid.*, p. 374.

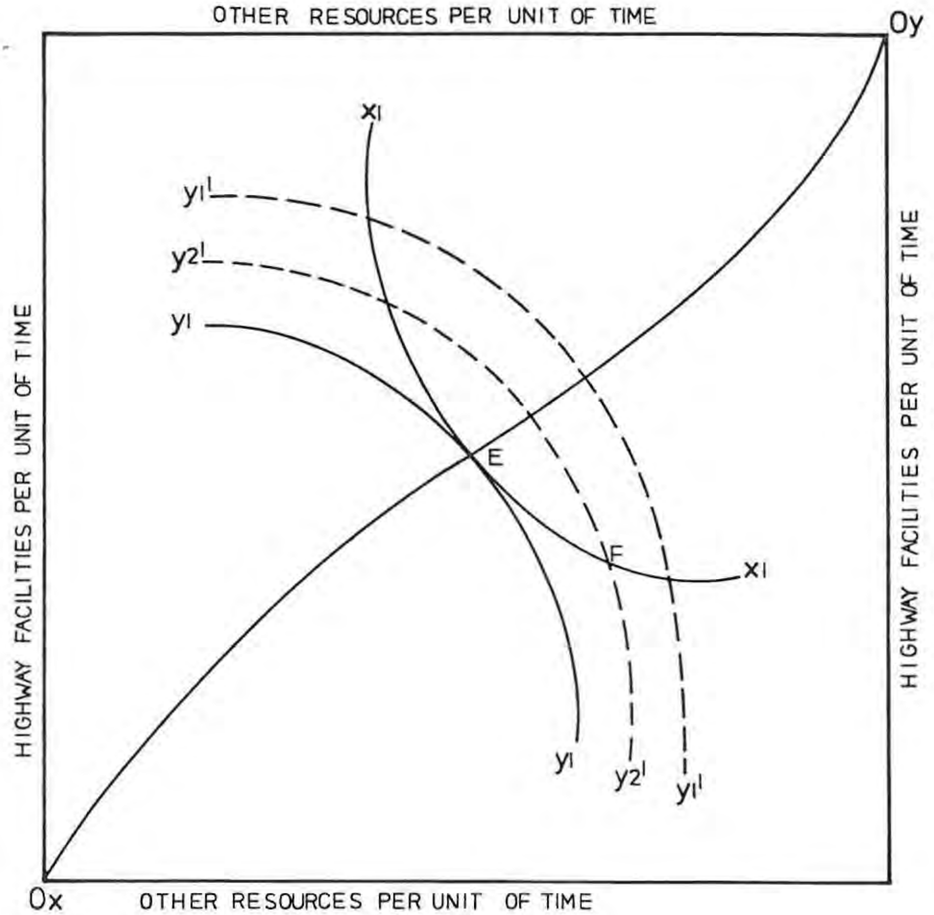


FIGURE 1:5 Externalities in Production

have been no change in X production. The efficiency of production has therefore been increased by the resource exchange.

It should be pointed out that interdependence is a necessary but not a sufficient condition for a marginal externality to affect the conditions for Pareto optimality. As Pigou has explained:

> ".... the essence of the matter is that one person A, in the course of rendering some service, for which payment is made, to a second person B, incidentally also renders services or disservice to other persons (not producers of like services), of such a sort that payment cannot be exacted from the benefited parties or compensation enforced on behalf of the injured parties."[35] (Italics mine).

It follows that a second necessary condition for an externality to exist is that due to institutional and/or technological reasons, A and B cannot enter into trade regarding activity bk.

The main impact of externalities on the pattern of resource allocation is that they may cause a divergence between private and social costs and benefits. Thus if the price of a public utility service is not equal to its marginal social benefit (MSB) or the marginal cost incured by the utility does not fully reflect marginal social cost (MSC) then the strict application of the marginal cost pricing rule by the public utility will no longer be consistent with the conditions for Pareto optimality. In fact the condition for socially optimal production becomes:

$$MRSx_{,y}^{A} = MRSx_{,y}^{B} = \frac{MSBx}{MSBy} = MRPTx,y = \frac{MSCx}{MSCy} \qquad (1:14)$$

---

35. Pigou, A.C., op.cit., p. 84.

Pigou recommended that a system of taxes and subsidies should be imposed by the State to ensure that relative prices satisfied these conditions for Pareto optimality. Buchanan and Stubblebine,[36] however, have argued that ".... full Pareto equilibrium can never be attained via the imposition of unilaterally imposed taxes and subsidies until all marginal externalities are eliminated". Nath[37], on the other hand, has defended the logic of the Pigovian solution by showing that a tax-subsidy system will bring about equilibrium provided that ".... the marginal rate of subsidy (tax) at any point is always equal to the value of the marginal externality for that point". If the validity of Nath's argument is accepted, then the implications for a public enterprise are that, if externalities only arise in its activities, then it should adjust its price above or below marginal cost to reflect the marginal effects of externality.

There are, however, a number of technical problems associated with this Pigovian price adjustment. Coase[38] has suggested that the magnitude of the external effect of a given act depends not only on the decision of the actor but also on the activity patterns of the persons affected. If, for example, the smoke caused by a power station results in local residents having to pay higher laundry costs, the more people who live within the vicinity of the power station, the more serious will be the externality problem. Thus a rule which states that a public enterprise should adjust its prices upwards by an amount equal to the cost of the disadvantages its activities impose on others is likely to be neither unambiguous nor obviously just.

---

36. Buchanan, J., and Stubblebine, W., op.cit., p. 383.

37. Nath, S.K., op.cit., p. 71.

38. Coase, R.H., "The Problem of Social Cost", Journal of Law and Economics, October, 1960.

Davis and Whinston[39] have distinguished between "separable" and "non-separable" externalities and have shown that no determinate equilibrium can be attained when externalities are non-separable. If, for example, the activity level of actor A affects the welfare of B and vice versa, then in the case of separable externalities, the utility functions of the two actors will take the form,

$$UA = f(a) + g(b) \quad ; \quad \text{and} \quad UB = b(b) + j(a) \qquad (1:15)$$

and in the case of non-separable externalities

$$UA = f(a,b) \quad ; \quad \text{and} \quad UB = g(b,a). \qquad (1:16)$$

If externalities are separable then B's decision will have no effect on A's optimal decision as to his own activity since B's activity will affect the total welfare of A but not the marginal yield to A of A's own activity.

On the other hand, if externalities are not separable then A's decisions may affect B's activity level which in turn will change A's former behaviour, and so on, with the result that no determinate equilibrium will be attained since the utility functions of both actors will be continually changing. In this situation it will not be possible to specify the appropriate level of corrective price adjustment for a public enterprise. Baumol, however, has pointed out:

> ".... in defence of the conventional Pigouvian position that his policy recommendations were formulated in terms of a setting of pure competition. Here the external effect of the decisions of any one firm on anyone else is by definition negligible. Just as in a

---

39. Davis, O. and Whinston, A., "Externalities, Welfare, and the Theory of Games", Journal of Political Economy, Vol. 70, June, 1962.

> competitive industry firm A's output decision does not directly influence that of firm B, it will similarly produce no indirect effect on B's decisions through any externalities which it may impose."[40]

In general, therefore, it can be seen that externalities disturb the conditions for Pareto optimality in a perfectly competitive economy. Although, in theory, corrective price adjustments may take account of the effects of externalities, the determination of an optimal general equilibrium solution clearly becomes more difficult as the effect of externalities becomes more widespread throughout the economy.

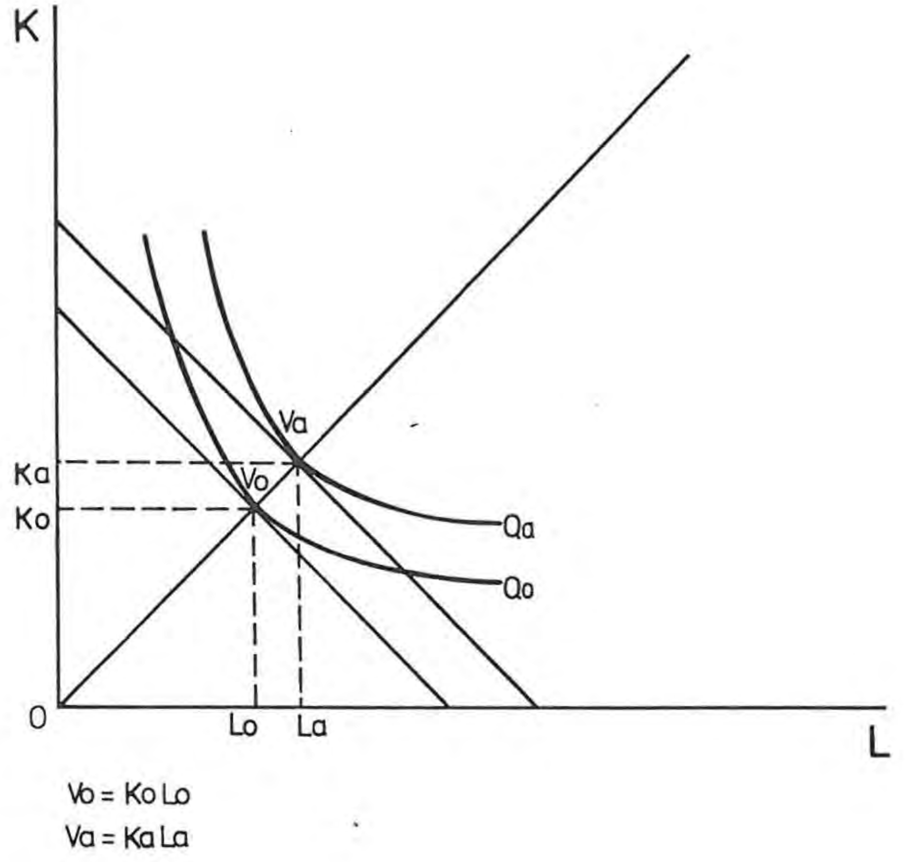
1:3:3 X - INEFFICIENCY

In defining the conditions for the optimality of a marginal cost pricing policy by a public enterprise in a first best economy it is necessary to explicitly include the assumption that public enterprise minimizes costs. This assumption does not need to be made in the case of perfectly competitive firms since if these firms are to survive in a perfectly competitive market they have to minimize costs. In the case of a public enterprise, however, its survival is normally guaranteed by the state and it is possible that inputs can be used with varying degrees of effectiveness within the enterprise. When an input is not used effectively, the difference between the actual output and the maximum output attributable to that input is a measure of the degree , of what Leibenstein[41] terms "X-inefficiency". This is illustrated in Figure 1:6(a). Vo represents the value of the minimum inputs needed to produce a given output Qo while the actual inputs used is designated by Va. The ratio of the difference between the two over the actual expenditure, (Va-Vo/Va) is a measure of X-inefficiency.

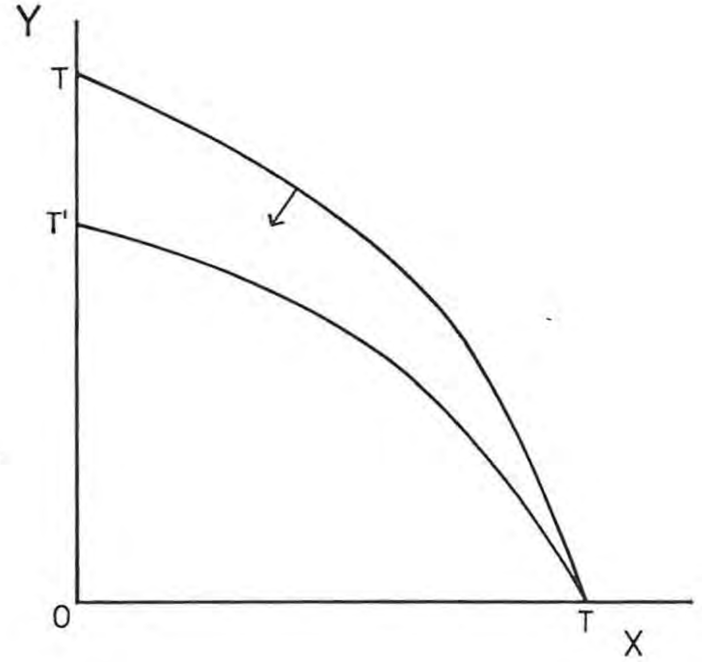
---

40. Baumol, W.J., op.cit., p.32.

41. Leibenstein, H., The General Theory of X-Efficiency and Economic Underdevelopment , Oxford University Press, 1978.



(a)



(b)

FIGURE 1:6 "X-Inefficiency"

The effect of X-inefficiency in the public enterprise on the economy as a whole is illustrated in Figure 1:6(b). The transformation curve TT' is shifted inwards. Now, the marginal cost pricing rule for public enterprise is derived from the Pareto optimality conditions for allocative efficiency. However, allocative efficiency is a top level condition for Pareto optimality and will not be attained when the lower level condition of technical efficiency is not obtained. The theoretical basis for marginal cost pricing as a necessary condition for optimality is thus further undermined by the existence of X-inefficiency since as Leibenstein says "allocative efficiency depends on X-efficiency and not the other way round".[42]

The question still arises, though, as to whether a comparison can be made between marginal cost pricing and an alternative pricing rule if X-inefficiency exists in both cases. It might be possible to do this where neither pricing policy will have any effect on the degree of X-inefficiency within a public enterprise. However, it has been suggested by Rees[43] that a policy of pricing above marginal cost to meet an exogenously determined profit target may place pressure on a public enterprise to contain X-inefficiency, particularly if tight control over the public enterprise prevents it from using its monopolistic power to raise prices to pass on its inefficiency to consumers. In subsequent chapters the problem of X-inefficiency will be addressed as the effects of alternative pricing policies on particular public utilities in South Africa are analysed.

---

42. Ibid., p. 159.

43. R. Rees, Public Enterprise Economics, Weidenfeld and Nicolson, 1978.

1:4 SECOND BEST THEORY AND THE DESIRABILITY OF PIECEMEAL OPTIMIZATION

In section 1:3 it was found that a number of the conditions which define a first best economy are restrictive in the sense that, if they are relaxed, marginal cost pricing may no longer be a necessary condition for attaining the highest possible level of social welfare. For a long time it was considered that despite the fact that public enterprise prices have to be set within an environment in which there are numerous departures from the conditions of a first best economy, they should still be set equal to marginal cost since it was implicitly believed that the greater the number of the various necessary conditions which could be satisfied in an economy, the nearer the system would be to a social optimum. In their seminal article on the theory of the "second best", Lipsey and Lancaster established the proposition that:

> ".... given that one of the Paretian optimum conditions cannot be fulfilled, then an optimum situation can be achieved only by departing from all the other Paretian conditions".[44]

The optimum situation finally attained is termed "second best" because it is achieved subject to a constraint which, by definition, prevents the attainment of a conventional Paretian optimum.

To illustrate the effect of this, suppose the economy contains n commodities, and one wishes to determine the optimal ratio of price to marginal cost for commodity i, $\hat{k}i$, taking the price/marginal cost ratios, kj, of the other n-1 commodities as given. Furthermore, let k represent the least and K the greatest of these fixed ratios.

---

44. Lipsey, R.G., and Lancaster, K., "The General Theory of Second Best", R.E.S., (1956-7), p. 11.

As Green[45] has shown, $\hat{k}i$ is a weighted sum of the fixed ratios with the weight wij attached to kj being positive if commodities i and j are substitutes and negative if they are complements. If it is assumed that k is not less than unity the following conclusions can be derived about the optimal price ratio, $\hat{k}i$, in a second best economy:

(a) $\hat{k}i$ will be greater than k if $\sum_{j=i}$ wij(kj-k) is positive. This will occur where the positive terms arising from indirect taxes or monopolistic pricing on substitutes are greater in magnitude than the negative terms which result from the prices of complements being set above their marginal costs. In this case, the optimal pricing policy of a public enterprise will result in prices being set above marginal cost. It may therefore be possible that the adoption by the public enterprise of a pricing policy other than marginal cost pricing may actually lead to an improvement in welfare.

(b) $\hat{k}i$ will be less than k if $\sum_{j=i}$ wij(kj-k) is negative. This may occur if taxes or monopolistic pricing are heavily concentrated on goods complementary to the public enterprise. In this case, an optimal pricing policy for the public enterprise will actually result in prices being set below marginal cost.

It can thus be seen that, in a second best economy, an a priori evaluation of alternative public enterprise policies in terms of general welfare propositions is not possible. Furthermore optimal pricing policies can only be formulated for a particular public enterprise taking into account the non-optimalities affecting the markets related to that enterprise. It follows that no general rule can be derived for optimal public enterprise pricing in a second best economy. As Graaff has noted:

---

45. Green, H.A.J., "The social optimum in the presence of monopoly and taxation", R.E.S., October, 1961.

> "It seems fairly clear that the condition which have to be met before it is correct (from a welfare viewpoint) to set prices equal to marginal cost are so restrictive that they are unlikely to be satisfied in practice."[46]

The question still remains, however, as to whether it is valid under second best conditions to base public enterprise prices on marginal cost, adjusting them, where necessary, in a piecemeal manner to take account of non-optimalities. This approach has been favoured by Turvey who contends that:

> ".... there is a presumption in favour of marginal cost pricing in most economies with a fairly freely functioning price mechanism ...., the right policy is to pursue marginal cost pricing .... subject to 'corrections' made only for those non-optimalities which are known to have a significant effect on the demand or cost structures of (the public enterprise)".[47]

Turvey distinguishes between three types of non-optimalities:

(i) Deviations from optimal conditions which are known and significant but which should be corrected by policies falling beyond the scope of a public enterprise pricing policy: For example, even if a public enterprise does not consider the distribution of income and wealth to be acceptable it should not attempt to correct this through its policies unless authorized by the government to subsidize a particular group of consumers. Otherwise it should act as if the existing distribution of income and wealth were acceptable. It is interesting to compare

---

46. Graaff, J., de V. op.cit., p. 154.

47. Turvey, R., Optimal Pricing and Investment in Electricity Supply, Allen & Unwin, London, 1968, p. 87.

this approach with the proposal by Graaff "that the only price a public enterprise or nationalized industry can be expected to set is what we may as well call a just price - a price which is set with some regard for its effect on the distribution of wealth as well as for its effect on the allocation of resources".[48]

Graaff argues that the task of the economist ".... is to provide the positive knowledge, not to recommend the level at which price should be set."[49] However, if a public enterprise is not required to take account of the distributional effects of its policies, as Turvey has suggested, then it may be valid, even in terms of Graaff's analysis, for an economist to indicate to a public enterprise the effect on resource allocation of various pricing policies based on marginal cost.

(ii) Non-optimalities which are either unknown or which have a trivial impact on the demand and cost structure of the public enterprise: For example, if the price of shaving cream was substantially above marginal cost with the result that the use of electric razors is larger than it would be in a first best situation, it would be ridiculous to suggest adjusting the price of electricity to compensate for this. On the other hand, although income taxation may have a significant impact on the supply of labour, the implications of this non-optimality for electricity, would be unknowable so that prices could not be adjusted for it.

(iii) Non-optimalities which are known, significant and which can be corrected by adjusting the pricing policy of a public enterprise: For example, it will generally be more optimal to set prices above marginal cost where

---

48. Graaff, J., de V., p. 155.

49. Ibid., p. 155.

(a) important close substitutes sell at significantly above marginal cost or generate large external economies;

(b) products in whose production the service of the public enterprise constitutes a major input sell at significantly below marginal cost or involve large external diseconomies;

(c) important close complements sell at significantly below marginal cost or generate large external diseconomies; and

(d) major inputs of the public enterprise are bought at significantly below marginal cost or involve large external diseconomies.

Where non-optimalities of this nature are unalterable, a piecemeal adjustment to marginal cost pricing in the form of mark-ups or mark-downs on marginal cost is considered to be desirable by Turvey.

The implication of the analyses of the second best problem by economists such as Bohm[50] and McManus[51], however, is that this type of piecemeal approach is invalid since public enterprise prices should only be derived from general equilibrium models which take market distortions into account as additional constraints.

Davis and Whinston[52], in turn, reject the Bohm-McManus approach on the grounds that it will ".... in general, require that the policy maker (or his analyst) know the solution to the grand maximization problem

---

50. Bohm, P., "On the Theory of 'Second Best'", Review of Economic Studies, July, 1967.

51. McManus, M., "Comments on the General Theory of Second Best", Review of Economic Studies, (1958-9), pp. 209-224.

52. Davis, O. and Whinston, A., "Piecemeal Policy in the Theory of the Second Best", Review of Economic Studies, p. 325.

before the values of the policy instruments (taxes or subsidies in the case of Bohm and prices for McManus) can be determined". It is difficult to argue that one can have a priori knowledge of the solution to a general equilibrium problem without having to construct it. Construction, however, implies a knowledge of the utility functions of all individuals and the production functions of all firms in the economy. It must be accepted that this is an unreasonable informational requirement and that "real policy makers cannot reasonably hope to ever have an exact prior knowledge of the solution in advance of the design and selection of measures for policy."[53] The piecemeal approach to policy suggested by Turvey quite clearly does not have such stringent informational requirements.

Davis and Whinston take the position that the central purpose of the theory of the second best is to establish when and under what conditions does the existence of imperfections in other areas of the economy cause it to be undesirable to design policies for the achievements of the Pareto conditions in the area of immediate concern. Although a theory of second best inherently involves the construction of a simplified general equilibrium model, Davis and Whinston argue that the model is not constructed to provide real policies but rather to answer the question of whether a particular piecemeal policy is justified. It is therefore relevant to the evaluation of any public utility pricing policy since such policies must necessarily be piecemeal in nature because they are designed to maximize welfare in a particular sector, of the economy, taking the conditions in all other sectors as given.

In particular, it would seem that the piecemeal approach to the formulation of an optimal pricing policy, proposed by Turvey is justifiable. Such an approach, however, can only take place

---

53. Ibid., p. 326.

within a partial equilibrium framework. The functions of partial and general equilibrium analysis in the application of welfare economics to public enterprise pricing will be examined in the next section.

1:5 CONCLUSION: A COMPARISON BETWEEN THE PARTIAL AND GENERAL EQUILIBRIUM RATIONALE FOR MARGINAL COST PRICING

While the basic propositions of Paretian welfare economics can be incorporated into a general equilibrium analysis to establish the conditions under which marginal cost pricing is optimal and will lead to a welfare improvement if it is adopted in the place of another pricing policy, a partial equilibrium approach is more useful to derive the pricing rules which a public enterprise should apply to deal with the various complications of a second best economy. In other words, general equilibrium analysis is the most appropriate tool for the justification of marginal cost pricing or any modification of this policy. As Rees has said:

> "The basis economists have for advocating marginal cost pricing is that provided by the necessary conditions for a Pareto optimum in a first best economy. These conditions are derived from a general equilibrium analysis, as should be any other sets of conditions which try to deal with the inadequacies of the first best model."[54]

Partial equilibrium analysis, on the other hand, appears to be the most appropriate tool for the application of marginal cost pricing principles to particular public enterprises operating in particular markets. Turvey thus unashamedly adopts a partial equilibrium analysis to extend and modify the marginal cost principle to deal with a number of the problems typically facing public enterprises, justifying his approach as follows:

---

54. R.Rees, op.cit., p. 46.

> "Much of the discussion of optimal pricing by public enterprises which is to be found in the literature is couched in terms of general equilibrium analysis. Despite its intellectual elegance this approach will not do here. We are concerned to derive rules for the behaviour of one, or a group of public enterprises which will maximize the Social Benefit of their activities less their Social Costs, given the environment in which they work, so the rules can only cover things under their control. Furthermore, if the rules are to be of any use they must not require the public enterprise to know things which it is in practice impossible for them to ascertain. The analysis .... therefore has three features: It relates to second best optimization, it is partial rather than general equilibrium and it is rough and ready".[55]

The rationale of marginal cost pricing must now be translated into partial equilibrium terms. To do this it is necessary to accept the three basic postulates proposed by Harberger[56] as the basis of a "conventional" framework for all applied welfare economies. The postulates are:

(a) the competitive demand price for a given unit measures the value of that unit to the buyer;

(b) the competitive supply price for a given unit measures the value of that unit to the seller; and

(c) when evaluating the net benefits or costs of a given action, the costs and benefits accruing to each member of the relevant group should normally be added without regard to the individuals to whom they accrue.

Harberger's plea is thus for the general acceptance by economists working in the field of welfare economics of social welfare

---

55. Turvey, R., op.cit., p. 21.

56. Harberger, A., "Three Basic Postulates for Applied Welfare Economica: An Interpretative Essay", J.E.L., (Sept., 1971), pp. 785-97.

function that maximizes the sum of consumers' and producers' surplus over competive cost, and which is neutral with respect to income distribution. A public enterprise should therefore aim to

Maximize $W = TR + S - (TC-R)$ (1:17)

where W = net economic welfare, TR = Total revenue, S = consumers' surplus, TC = total cost and R = inframarginal rent.

Implicit in this approach is the view that public policy analysis should be concerned principally with issues of efficiency in resource allocation, leaving equity considerations in the hands of fiscal authorities.

Assuming, for illustrative purposes, that all factors are available in completely elastic supply, inframarginal rents will be zero and the net welfare gain is:

$$W = TR + S - TC$$

Now $TR + S = \int_O^Q P(Q')dQ'$ where $P(Q')$ is the demand curve. Differentiating this expression with respect to Q yields

$$(TR + S) = \int_O^Q P(Q')dQ' \qquad (1:18)$$
$$= P(Q)$$

and $\frac{d}{dQ}(TC) = MC$ (marginal cost) (1:19)

Thus the necessary and sufficient conditions for optimization are:

$$\frac{dW}{dQ} = \frac{d}{dQ} (TR + S) - \frac{d}{dQ} TC = 0 \qquad (1:20)$$

whence $P - MC = 0$ (1:21)

and $$\frac{d^2W}{dQ^2} = \frac{dP}{dQ} - \frac{d^2}{dQ^2},(TC) \quad 0 \qquad (1:22)$$

It can thus be seen that according to this partial equilibrium approach, marginal cost pricing is a necessary optimality condition. The intutive rationale for the optimality of marginal cost pricing is that if the price of a good gives a money measure of the subjective value, to each consumer, of the marginal unit of consumption of the good and if marginal cost gives a money value of the value of output sacrificed by supplying the marginal unit of the good, then, only if price equals marginal cost is the value of the marginal unit to consumers just equal to the value of what has to be sacrificed to provide it. As Rees has stated:

> ".... the intention of marginal cost pricing is to impose on the consumer the cost of providing his marginal unit of consumption, so as to cause him to adjust his total consumption to the point at which the value of the marginal unit to him is just equal to its costs."[57]

Although this partial equilibrium rationale of marginal cost pricing provides a useful framework for the formulation of public enterprise pricing rules in the next chapter, it does not, in itself , constitute a particularly rigorous justification of the optimality of the marginal cost pricing rule since it does not make clear the welfare value judgements on which it is based, nor the implicit assumption about compensation for the redistribution of income nor the fact that its validity is totally dependent on the existence of a first best economy.

Thus, in brief, it can be seen that although the analysis in this chapter has been conducted at a fairly abstract level, it is nonetheless necessary to support by logical argument the validity of any normative recommendation concerning the optimality, or otherwise, of any rule for public enterprise pricing.

---

57. Rees, R., op.cit., p. 47.

CHAPTER TWO

OPTIMAL PRICING RULES AS A INSTRUMENT FOR DECENTRALIZED DECISION TAKING

INTRODUCTION

In the previous chapter the justifiability of using a piecemeal optimization procedure to formulate a pricing policy for a public enterprise was established. This procedure usually involves the following steps

(i) The determination of a structure of prices for the goods or services sold by the public enterprise which is based on the marginal cost of producing these goods and services;

(ii) The adjustment of this price structure for what Turvey has classified as "known and significant non-optimalities" which cause a departure of the conditions for optimal pricing from a simple equality between price and marginal cost.

The first part of this chapter will describe how this piecemeal optimization procedure can be used to formulate the conditions for optimal pricing by public enterprises. These conditions will form the normative basis against which the actual pricing policies of particular public enterprises in South Africa will be evaluated.

The second part of this chapter will consist of an evaluation of the extent to which the pricing rules derived from these conditions can be used as instruments of decentralized control inducing public enterprises to pursue their objectives without ad hoc ministerial interference in their decisionmaking.

2:1 THE BASIS FOR A MARGINAL COST PRICING POLICY

2:1:1 THE APPLICATION OF MARGINAL COST PRICING AND INVESTMENT RULES

In the previous chapter it was established that the marginal cost pricing principle is an acceptable starting point in the process of determining an optimal pricing policy for a public enterprise. Although marginal cost may be defined as the addition to total cost resulting from a given increment in output, the actual level of marginal cost will vary according to the time range over which it is planned to produce the increment in output. In this respect a distinction is usually made between short run and long run marginal cost. Short run marginal cost is the cost of an increment in output produced during a period when certain inputs cannot be varied whereas long run marginal cost is determined for a period when all inputs are variable. Investment policy can clearly only be formulated for the long run whereas prices are essentially determined for the short run. Rees[1] has drawn attention to the interrelationship between price and investment decisions and has suggested that a marginal cost based pricing and investment policy should be based on the following rules:

(i) in the short run the public enterprise should set price equal to marginal cost; and

(ii) in the long run the goal of investment policy should be to adjust the capacity of plant to produce the level of output at which price is equal to long run marginal cost.

The application of these rules is illustrated in Figure 2:1. DoDo represents the expected level of demand for period O. The optimal level of output planned for this period is therefore qo, since at this level of output the demand function intersects the long run marginal cost curve (LMC). Now the minimum unit cost of producing this output is indicated by the long run average cost curve (LAC) as being qox. Once the enterprise has installed plant of the scale required to produce the output qo at minimum unit cost qox, the elements of total cost which depend on the scale of plant will be fixed and the short run marginal and average cost curves for period 0 will be SMCo and

---

1. R.Rees, Public Enterprise Economics, Weidenfeld and Nicolson, 1978.

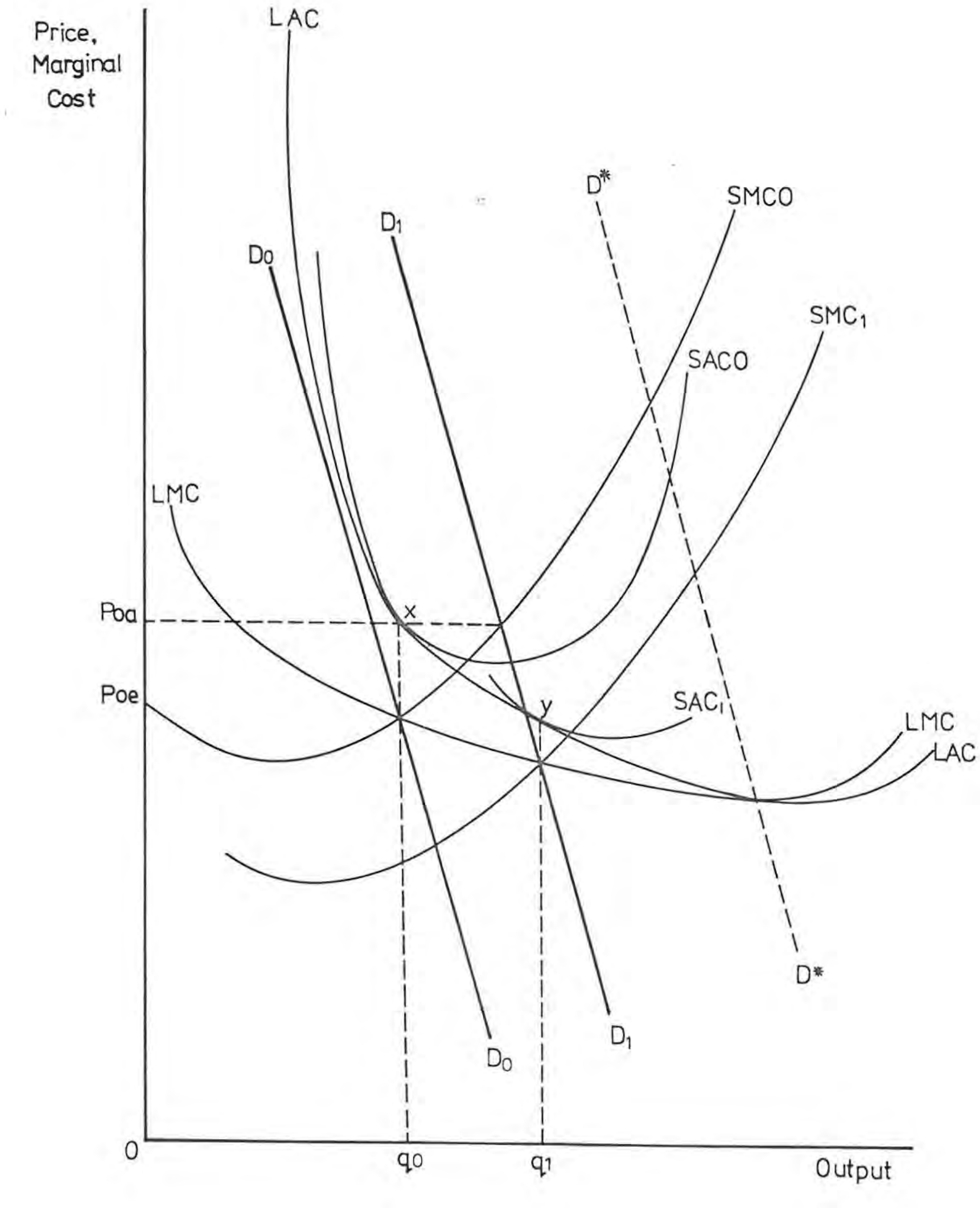


FIGURE 2:1 The Application of the Marginal Cost Pricing and Investment Rules

SACo respectively. If the expected level of demand DoDo is actually realized in period 0 then the application of rule (i) would result in price being set at poe which is equal to short run marginal cost and also long run marginal cost. If, however, the actual demand in period 0 rises to D1D1, then the application of the rule that price should be set equal to short run marginal cost will result in the price poa being charged which is higher than long run marginal cost. It can thus be seen that price should only equal both short and long run marginal cost if the expected level of demand in any period is actually realized.

It is assumed that for a public enterprise formulating an investment policy in period 0, all inputs only become freely variable at the beginning of period 1. If the public enterprise expects the level of demand D1D1 to persist in period 1 then it will implement an investment programme designed to expand the scale of plant to the size necessary to produce the level of output q1 at which the demand price is equal to long run marginal cost.

It is clear that in strictly applying these pricing and investment rules the public enterprise does not take into account their implications for the financial performance of the undertaking. It is, however, generally accepted that although a public enterprise should not earn excess profits, it should generate sufficient revenue to cover its costs and thereby remain independent of Treasury support. In most cases a public enterprise is statutorily obliged in terms of its charter to attain a financial equilibrium over a reasonable period of years. The public enterprise management cannot therefore be indifferent to the surpluses or deficits which arise under a marginal cost based pricing and investment policy. The implications of marginal cost rules for the profitability of the undertaking depend, to a large extent, on the assumptions one makes concerning the shape of the cost functions facing the public enterprise.

In figure 2:1 it can be seen that all the cost functions are "U-shaped". The U-shaped nature of the short run average and marginal cost functions is illustrative of the proposition that the input requirements per unit increment of output will be affected by the law of variable proportions. This law is the converse of the law of eventually diminishing productivity and holds that as an enterprise increases its output, at first the variable input requirements per unit increment of output will fall due to the increasing marginal productivity of the variable inputs but that a point will be reached after which input requirements will rise due to the diminishing marginal productivity of the variable inputs used. If one makes the ceteris paribus assumption that factor prices and technology are constant then marginal and average costs will first fall to a minimum point and then increase in the manner shown in Figure 2:1.[2]

The U-shaped characteristic of the LAC and LMC is illustrative of the laws of returns to scale. According to these laws the unit costs of production decrease as plant size increases, due to the economies of scale which the larger plant size makes possible. The traditional theory of the firm assumes that it is not possible to derive further economies of scale if the scale of plant is increased beyond a certain optimum plant size. Diseconomies of scale arise after this optimum plant size has been exceeded since the managerial function of coordinating and controlling activities within an organization become more complex and the decisionmaking process becomes less efficient. The turning up of the LAC curve is attributable to managerial diseconomies of scale since technical diseconomies can be avoided by dupicating the optimum technical plant size.

It can be seen from Figure 2:1 that only when demand is at the level D*D* will the application of the marginal cost pricing and investment rules be consistent with the attainment of financial equilibrium by the

---

2. SAC will be declining at all points where SMC $<$ SAC and increasing at all point where SMC $>$ SAC. It follows that the SMC curve will intersect the SAC curve at its minimum point.

public enterprise. This is because at this level of demand price equals long and short run marginal cost at the optimum output q* which can only be produced at minimum unit cost using the optimally sized plant associated with minimum point of the LAC curve. If the public enterprise expects demand to be at any level below D*D* then according to the optimal investment rule (ii) it should adjust its plant size to produce a level of output at a minimum unit cost which will exceed its long run marginal cost. It follows that if this expected level of demand is actually realized the application of the rule that price should equal short run marginal cost will result in insufficient revenue being generated to cover total cost. It is possible though that where the actual demand in period exceeds expected demand, the setting of a price equal to short run marginal cost may actually result in a surplus of revenue over cost being earned. This surplus will only exist in the short run if demand is still expected to be below D*D* since if the public enterprise follows the optimal investment rule it will expand plant capacity to produce output at a minimum unit cost which is again greater than long run marginal cost.

It is clear, therefore, that a conflict may arise between the objectives of efficiency and financial viability when a public enterprise formulates its pricing policy according to the optimal pricing and investment rules discussed in this section. In the next section the question of how some of the assumptions underlying the shape of the cost curves should be modified in the light of some of the insights of modern cost theory will be examined.

2:1:2 THE IMPLICATIONS OF MODERN MICROECONOMIC THEORY OF COSTS FOR MARGINAL COST PRICING AND INVESTMENT RULES

The U-shaped cost curves of the traditional theory have been questioned by various writers on theoretical a priori and on empirical grounds. It is considered that a greater realism can be introduced into the representation of cost functions if it is assumed that:

(a) short run average variable costs are constant over a range of output; and that

(b) managerial diseconomies are offset by technical economies of scale for all scales of plant capacity so that the long run average cost curve does not turn up at any feasible level of output.

The implications of these assumptions for the marginal cost pricing and investment rules discussed in the last section will now be examined.

A. CONSTANT SHORT RUN AVERAGE VARIABLE COSTS

It is traditionally assumed that the short run average variable cost (SAVC) and the short run average total cost (SATC) functions are U-shaped. This implies that each plant size is designed a single level of output. As Koutsoyiannis explains this means that:

> "Any departure from (this level of output), no matter how small .... leads to increased costs. The plant is completely inflexible. There is no reserve capacity, not even to meet seasonal variations in demand."[3]

A number of writers including Stigler,[4] Andrews[5] and Wiles[6] have proposed that in choosing the size of plant to produce a planned level of output, an enterprise decisionmaker will not only take into account the minimum unit cost of plant but also its built-in flexibility. In particular a public utility producing a service that cannot be

---

3. Koutsoyiannis, A., Modern Microeconomics, Macmillan Press Ltd., 1979, 2nd ed., p. 113.
4. Stigler, G., "Production and Distribution in the Short Run", in Readings in the Theory of Income Distribution, American Economic Association, Blakiston, 1946.
5. Andrews, P.W.S., Manufacturing Business, Macmillan, 1949.
6. Wiles, P.J.D., Price, Cost and Output, Blackwell, 1961.

stored will allow for unexpected fluctuations in demand by providing for a certain amount of reserve capacity. In Figure 2:2 the flat stretch of the SAVC curve corresponds to the built-in-the plant reserve capacity. It can be seen that from q1 to q2 the SAVC is equal to SRMC, both being constant per unit of output. It follows that if demand fluctuates between D1D1 and D2D2 then the application of the short run marginal cost pricing rule will be equivalent to setting price equal to unit variable cost. If the SAVC function of a public enterprise actually resembles that shown in Figure 3:2 then it is clear that the practical difficulties usually associated with calculating marginal cost for the purpose of setting prices will largely fall away, since most undertakings are able to calculate their unit variable costs of production. In a comprehensive summary and critique of a wide range of statistical cost studies, Johnston[7] shows that the evidence from most statistical studies is that in the short run the AVC is constant over a considerable range. Although these studies have been criticized on the grounds that they do not distinguish between accounting and economic costs and that they fail to deal adequately with changes in technology and in factor prices, Koutsoyiannis makes the valid point that:

> "the fact that so many diverse sources of evidence point in general to the same direction (that is, lead to broadly similar conclusions) regarding the shape of costs in practice, surely suggests that the strictly U-shaped cost curves of traditional theory do not adequately represent reality".[8]

Figure 2:2 shows that if unit variable costs are constant over the range of output which reflects the planned reserve capacity, the

---

7. Johnston, J., Statistical Cost Analysis, McGraw-Hill, 1960.
8. Koutsoyiannis, op.cit., p. 138.

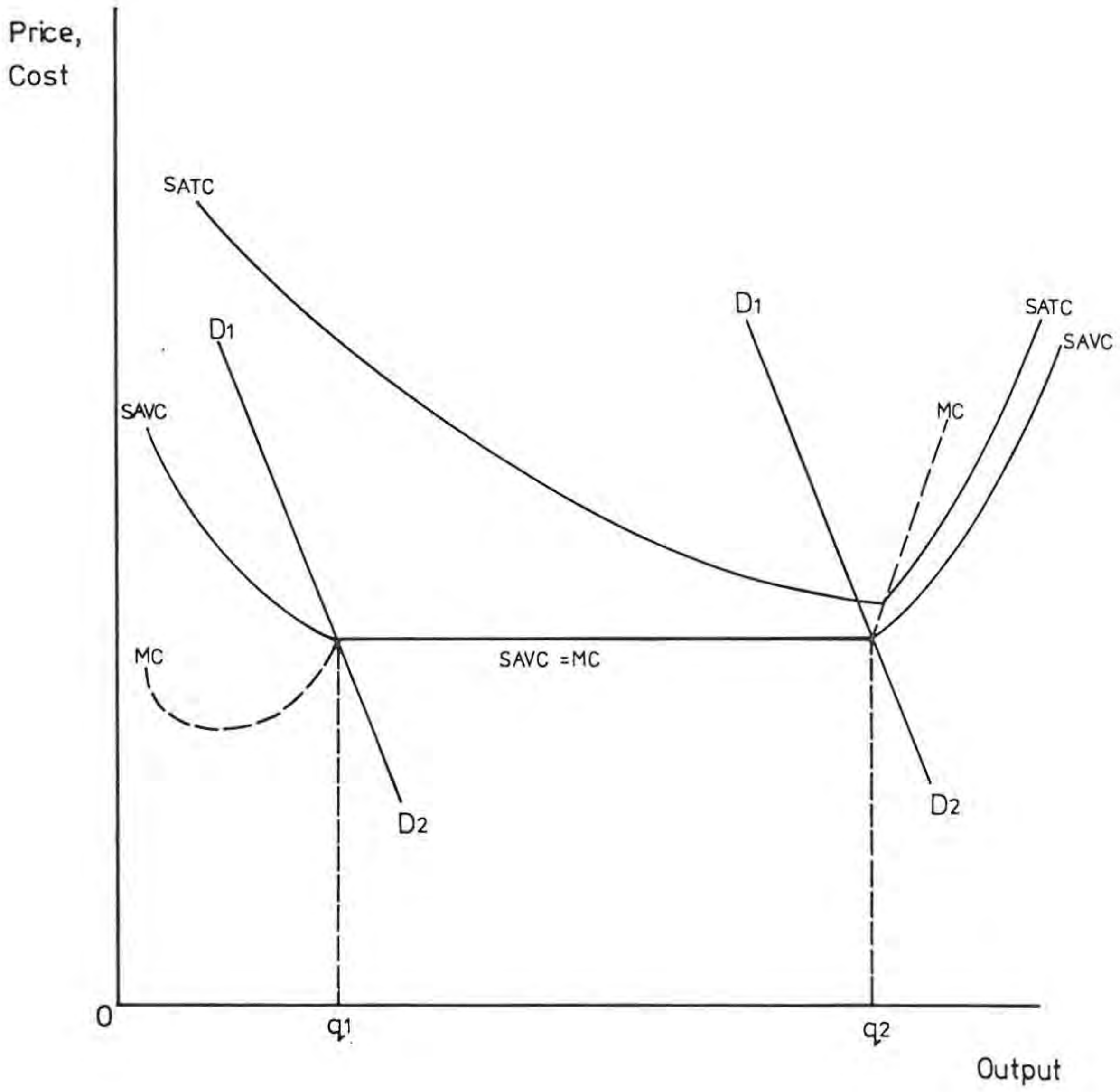


FIGURE 2:2 The Effect of the Provision of Reserve Capacity On Cost Curves

short run average total cost curve (SATC) will fall continuously up to the level of output q2 at which reserve capacity is exhausted. If output exceeds this level, the SATC curve will start rising and the MC curve will intersect it at its minimum point. It can be seen that if demand falls between D1D1 and D2D2, the application of the rule that price should equal short run marginal cost will clearly result in a deficit being incured in the short run. Now, Bain has proposed that a plant

> "will have a certain load factor reflecting the ratio of average rate of use to the capacity or best rate of use, and this load factor will generally be smaller than one."[9]

Koutsoyiannis[10] has suggested that a typical load factor will be between two-thirds and three-quarters of capacity. If this is valid then it follows that not only will a decisionmaker expect the normal level of utilization of the plant to fall within the range q1q2 in Figure 2:2 but also that there will be a "cushion" of reserve capacity to meet any unexpected increase in the level of utilization. There is therefore a greater likelihood that a marginal cost pricing policy will result in a public enterprise incuring a short run deficit where the enterprise has planned to use reserve capacity to meet any unexpected fluctuation in demand. It can be concluded, therefore, that while the assumption of constant short run average variable costs over a range of output may make it easier for a public enterprise to apply the short run marginal cost pricing rule, it also weakens the case for such a pricing rule since there may be a greater probability that the short run financial objectives of the organization will conflict with the pursuit of this efficient pricing policy.

---

9. Bain, J.S., <u>Barriers to New Competition</u>, Harvard University Press, Cambridge, Massachussets, 1956, p. 63

10. Koutsoyiannis, A., <u>op.cit.</u>, p. 121.

B. "L" - SHAPED LAC CURVE

If it is assumed that a decisionmaker will plan to expand the size of the plant where the level of utilization exceeds the typical load factor of, say, two-thirds of full capacity, a long run average cost curve (LAC) can be drawn by joining the points on the SATC curves corresponding to two-thirds of the full capacity of each plant size. Figure 2:3 shows that this will give rise to a "L-shaped" LAC curve which does not turn up at very large scales of output and does not envelop the SATC curves, but rather intersects them at the level of output defined by the typical load factor of each plant. Now it appears to be accepted by modern theorists that production costs continue to fall smoothly even at very large scales, while managerial costs may rise only slowly at very large scales. The LAC function can therefore be represented by a "L-shaped" curve as in Figure 2:3 if the fall in technical costs more than offsets the probable rise in managerial costs. The statistic cost studies referred to by Johnston[11] tend to support the view that there are no diseconomies of scale at large scales of output although this empirical evidence does not establish conclusively whether costs remain constant beyond a certain minimum optimum scale, or fall continuously with scale.

If the LAC curves fall continuously with scale the LMC curve will lie below LAC at all scales of output. It follows that the application of the investment rule that capacity should be adjusted to the level of output where demand equals long run marginal cost will result in the public enterprise incuring a deficit at every level of demand. The size of this deficit is, however, likely to become less significant as demand grows over time.

On the other hand a number of economists such as Bain have suggested that there is a minimal optimal scale of plant at which all possible scale economies are derived. If the scale of plant is expanded beyond this minimum optimal scale then constant returns to scale are likely

---

11. Johnston, J., op.cit.

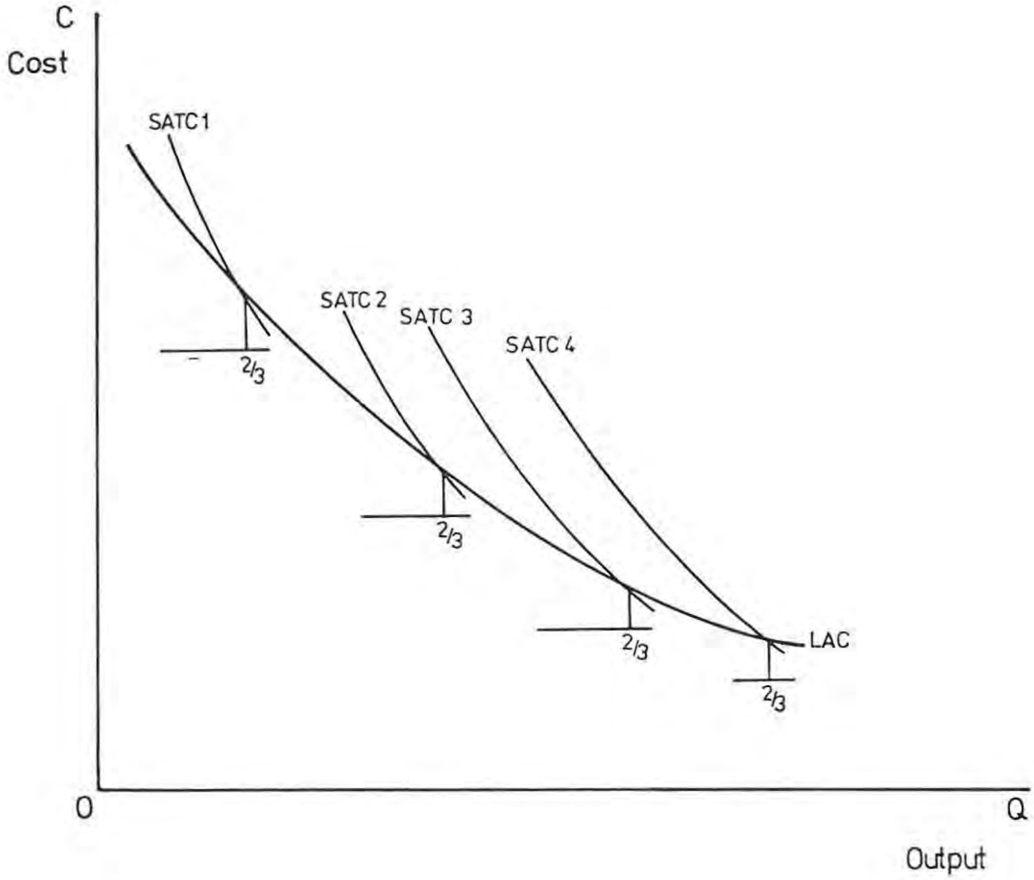


FIGURE 2:3 The Deriviation of an "L" - Shaped Long Run Average Cost Curve

to set in. Over this flat section of the LAC curve, LAC will equal LMC and the application of the optimal investment rule will result in all costs being covered. The case for a marginal cost based pricing and investment policy is therefore strengthened where there are constant returns to scale.

A number of economists such as Rees[12] and Turvey[13] have developed a strong case for marginal cost pricing under conditions of constant returns to scale where it is assumed that price is used as an instrument to ration demand to the capacity of fully utilized plant. The characteristic features of the Rees - Turvey (R-T) model of public enterprise pricing will be examined in the next section.

2:1:3 THE REES - TURVEY MODEL OF PUBLIC ENTERPRISE PRICING

A. APPLICATION OF OPTIMAL PRICING AND INVESTMENT RULES

The R - T model of public enterprise pricing is based on the following assumptions:

(i) in the short run, a maximum output level is defined, which cannot be exceeded at any cost;

(ii) there are constant returns to scale so that LAC is equivalent to LMC;

(iii) unit variable costs are constant so that SRMC is equivalent to SRAVC; and

(iv) finite increments to plant capacity of perfectly divisible.

---

12. R.Rees, "Public Enterprise Economics", Weidenfeld and Nicolson, 1978.

13. Turvey, R., Economic Analysis and Public Enterprises , London, Allen and Unwin Ltd., 1st ed., 1971, pp. 13-79.

Figure 2:4 illustrates how the marginal cost pricing and investment rules will be applied under these conditions. In the figure, $\overline{qo}$ represents the maximum output capacity of the plant in the short run, r denotes unit running cost which is equivalent to both short run marginal and average variable cost up to the level of output $\overline{qo}$, while r + k represent the sum of unit running and capital cost which, given the assumption of constant returns to scale, is equivalent to both long run average and marginal cost.

If demand is at the level D1D1 the public enterprise will follow the short run marginal cost pricing rule by setting price equal to unit running cost. It can be seen from the figure that there will still be excess capacity if the public enterprise applies this rule but that a greater underutilization of capacity would occur if the price was set at any higher rate. By setting the price equal to unit running cost, however, the public enterprise will incur a deficit of $k\overline{qo}$ in the short run. According to the optimal investment rule that capacity should be adjusted to the level where price equals long run marginal cost, the public enterprise should implement a disinvestment programme with the goal of reducing the output capacity of the plant to $\overline{q1}$ if it expects demand to remain at the level D1D1. It can thus be seen that any deficit which arises under a marginal cost based pricing and investment policy will only be temporary in nature and will be eliminated by the adjustment of capacity to optimal levels. The investment policy of the public enterprise will therefore be used as the instrument to reconcile any conflict between the financial and efficiency objectives of pricing policy.

If demand is at the level D2D2 then an excess demand will result if price is set equal to unit running cost. In this situation both Rees and Turvey recommend that this excess demand should be rationed through the normal operation of the market mechanism. In other words price should be allowed to rise to the level p2 at which the quantity demanded is equal to the maximum output capacity of the plant, $\overline{qo}$. It can be seen that at the price p2, a deficit equal to $((r + k) - p2)\overline{qo}$ will be ocurred in the short run and that the purpose of investment policy should be to reduce the output capacity of the plant to $\overline{q2}$ so that a

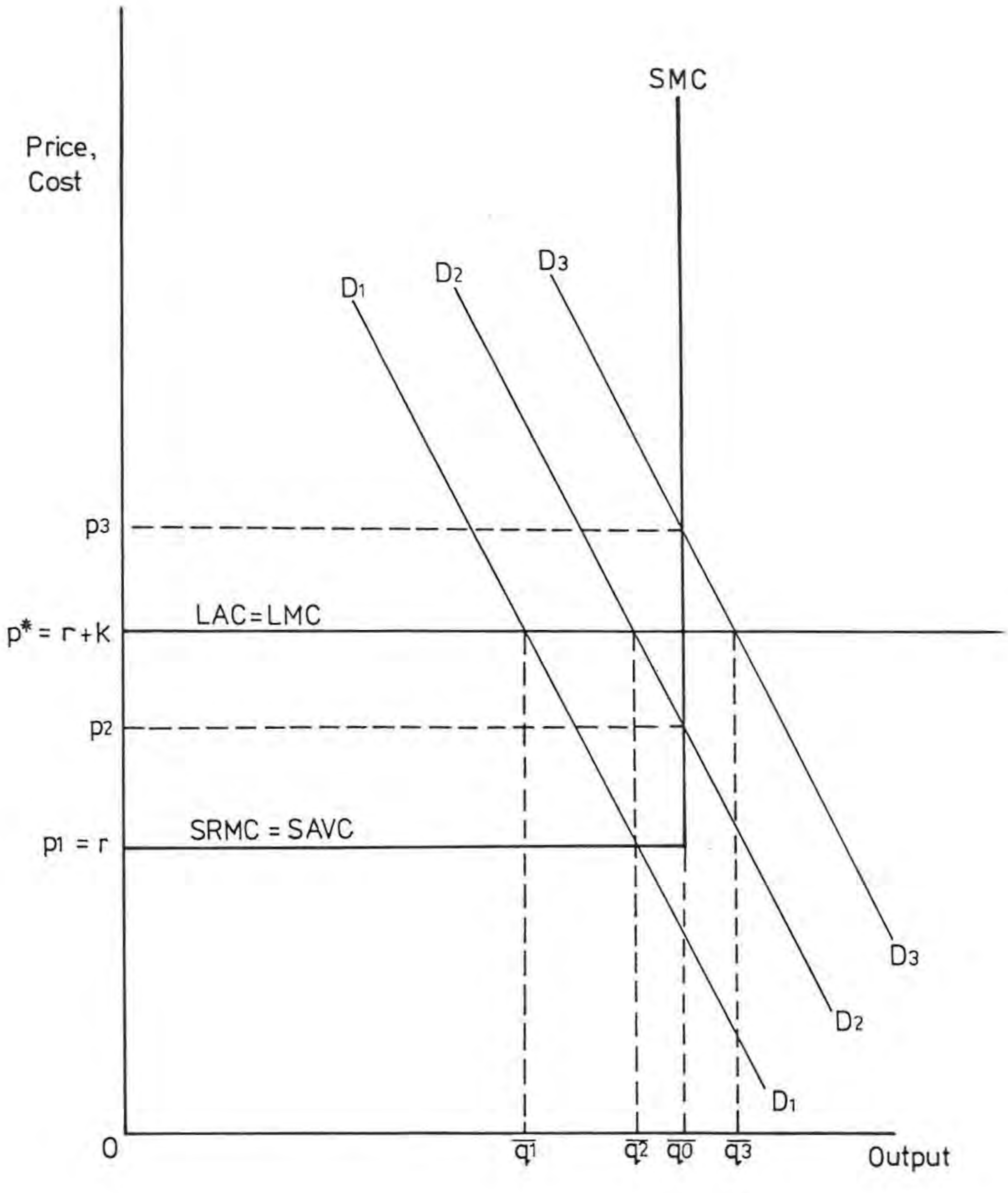


FIGURE 2:4 The "Rees-Turvey" Model

marginal cost pricing policy can generate sufficient revenue to cover total costs.

If demand is at the level D3D3 then the price p3 will ration demand to the maximum output capacity $\overline{qo}$. A surplus of $(p3 - (r + k)\overline{qo}$ will be earned in the short run. This surplus will, however, be eliminated if the public enterprise follows the optimal investment rule of expanding the output capacity of plant to $\bar{q}3$.

The Rees - Turvey model strengthens the case for requiring public enterprises to adopt marginal cost based pricing and investment policies in a number of ways:

(i) it indicates the manner in which pricing policy can be used as a tool to promote the efficient utilization of facilities;

(ii) it indicates the appropriate response which a public enterprise should make to the signals which arise in the production and marketing of its product. For example, excess capacity should provide the signal for the public enterprise to adjust its price downwards provided that price is not allowed to fall below the level of unit running cost while excess demand should be regarded as a signal for the public enterprise to adjust its price upwards. Furthermore, if the public enterprise expects to earn a surplus this should be regarded as a signal for it to expand the scale of its plant while an expected deficit should indicate the need to rationalize its operations by reducing the capacity of its plant;

(iii) it suggests that any conflict between economic efficiency and financial viability should be only temporary and that, in the long run, the pursuit of an efficient pricing and investment policy is fully consistent with the attainment of financial equilibrium by the public enterprise; and

(iv) the model can be extended to deal with the problem of formulating prices where demand fluctuates between peak and off-peak periods. This will be discussed in the next sub-section.

B. THE EXTENSION OF THE MODEL TO DEAL WITH PEAK LOAD PRICING

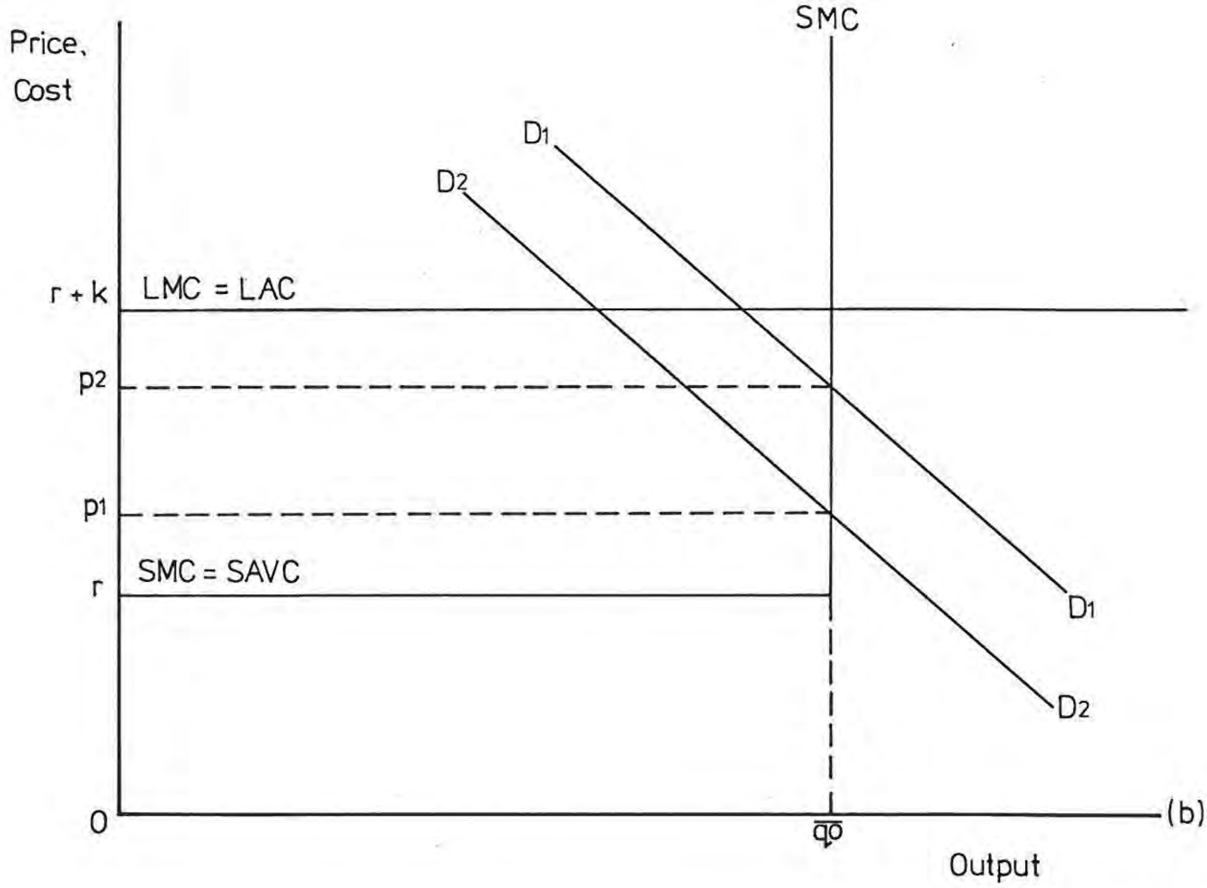
It is often the case that there is a systematic fluctuation in the demand for the services provided by a public utility within a given period, this pattern repeating itself from period to period. These fluctuations are usually not sufficiently prolonged to enable the public utility to adjust the capacity of its plant to match them. Consequently, public utilities often resort to intertemporal price discrimination or peak load pricing policies to deal with this problem of fixed capacity and fluctuating demand.

The theoretical solution to this problem was developed by Boiteux[14], Steiner[15], Hirschleifer[16], Buchanan[17] and Williamson[18]. Figure 2:5 shows how a peak-load pricing policy can be applied within the framework of the R-T model.

Figure 2:5 is drawn on the assumption that demand fluctuates on a regular daily basis with there being a period of peak demand, D1D1, between, say, 6 a.m. and 6 p.m. and a period of off-peak demand, D2D2, between 6 p.m. and 6 a.m. The basic guidelines of the R-T model for pricing policy are that:

---

14. Boiteux, M., "Peak load Pricing" in Nelson, J.R., (ed.), "Marginal Cost Pricing in Practice", Englewood Cliffs, N., 1964, pp. 59-90.
15. Steiner, P.O., "Peak loads and efficient pricing", Quarterly Journal of Economics, November, 1957, pp. 585-610.
16. Hirschleifer, J., "Peak loads and efficient pricing: Comment", Quarterly Journal of Economics, August, 1958, pp. 451-62.
17. Buchanan, J.M., "Peak loads and efficient pricing: Comment", Quarterly Journal of Economics, August, 1966, pp. 463-71.
18. Williamson, O.E., "Peak-load pricing", American Economic Review, Vol. 56 (1966), pp. 810-27.



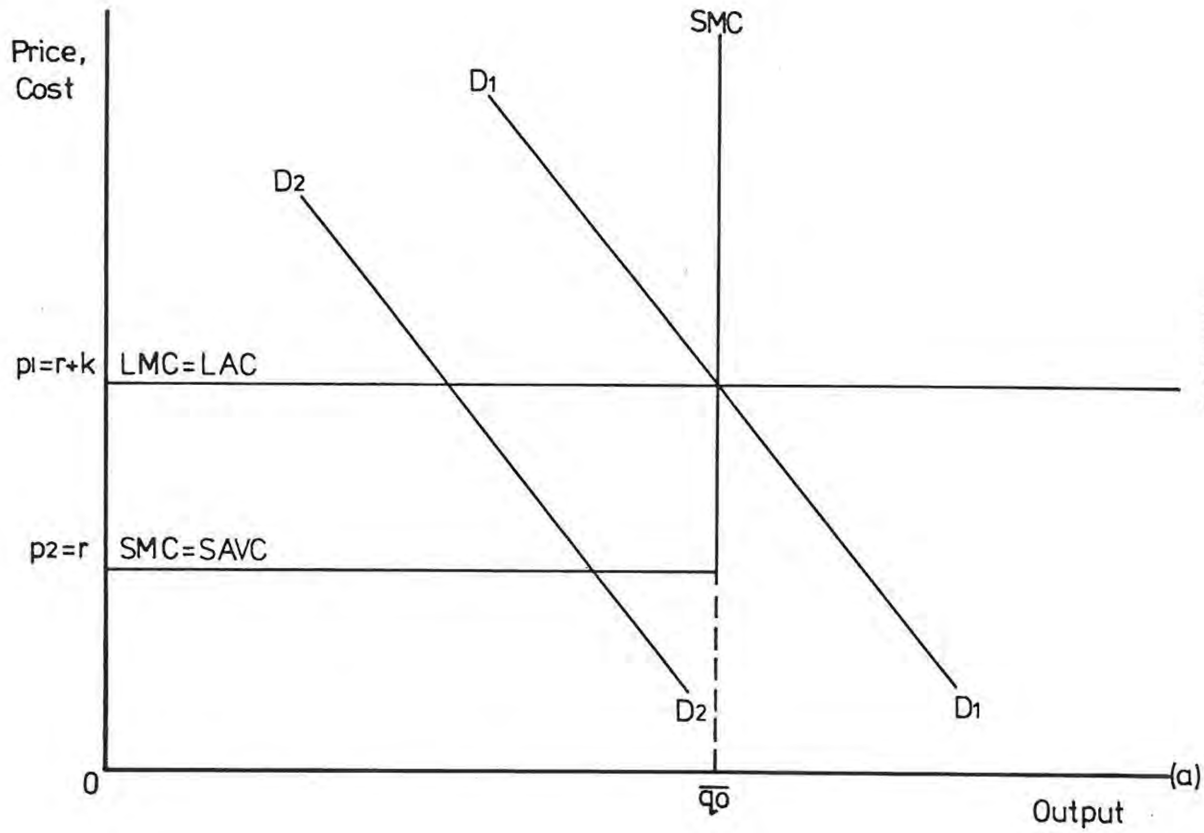


FIGURE 2:5 Peak Load Pricing

(i) where there is excess capacity, price should be dropped as low as possible to induce the maximum utilization of this excess capacity provided that prices still cover unit running costs, and that

(ii) where capacity is fully utilized, prices should rise to restrict demand to capacity and thereby generate revenue which will make a contribution to capital costs.

It can be seen from Figure 2:5 that the application of these guidelines will lead to peak-off-peak price differentiation.

Figure 2:5 (a) illustrates the situation where off-peak demand is at such a low level that excess capacity results even where price is set equal to unit running costs. If price is allowed to ration peak demand to the maximum output capacity of the plant, $\overline{qo}$, then it can be seen that it is only during the peak period that price should make some contribution toward unit capital costs. This prediction has given rise to the practice by some utilities of allocating capital costs only to peak period consumers. It should be pointed out though that it is implicitly assumed in the peak load pricing model represented in Figure 2:5 that the demand functions in the two periods are independent. It may, however, be the case that by charging a much higher price in the peak period the public enterprise may induce such a shift of consumption to the off-peak period that it in fact becomes the new peak period. As a result of being aware of the possibility of this problem of "shifting peaks" arising, public enterprises usually do not in practice allocate all capital costs to the peak period.

Furthermore it is possible that by setting price equal to unit running cost in the off-peak period, excess demand may arise in this period. Figure 2:5 (b) illustrates the conditions in which this situation may occur. In the figure the excess demand which arises in the off-peak period is removed by allowing price to rise to p2 while the higher price, p1, is charged in the peak period. Quite clearly it would be inefficient for the public enterprise to allocate all capital costs to peak period consumers in this situation.

In formulating an optimal investment policy it should be recalled that the net social benefit or marginal value to consumers of an increment in the output capacity of plant is measured by the difference between price and running charges (p-r). Now since an increment in plant capacity generates outputs and therefore benefits in both peak and off-peak periods the relevant measure of the marginal benefit to consumers of an increment in capacity is the sum of the marginal values of capacity peak and off-peak consumers. The planned daily capacity of the plant should therefore be expanded up to the optimal output where the sum of the marginal values of capacity in the two sub periods is equal to marginal capital cost that is:

$$(p1 - r) + (p2 - r) = k, \qquad (2:1)$$

Figure 2:6 indicates how this investment rule should be applied. If the capacity of plant is initially fixed at the output level $\overline{qo}$, then the application of a peak load pricing policy will result in the price p1 being charged in the peak period and the price p2 in the off-peak period. It can be seen from the figure that $(p1 - r) + (p2 - r) > k$. This provides a signal for the public enterprise to implement an investment programme to expand plant capacity up to the optimal output level $\overline{q1}$ at which the contributions generated in peak and off-peak periods are sufficient to cover capital costs.

There are often additional administrative and metering costs associated with differentiating prices over an additional sub-period. Consequently an additional pricing period should only be introduced if the net social benefit of doing so exceeds the additional administrative and metering costs. Figure 2:7 indicates the welfare losses associated with charging a uniform price over an entire period as an alternative to differentiating between the prices charged to peak and off-peak consumers. In Figure 2:7 the public enterprise has to decide between charging a uniform price p* in both periods which would generate an equal contribution per unit to capital costs or charging a price p1 in the peak period and a price p2 in the off-peak period with the contribution to capital costs only being made in the peak period. It can be seen that if p* is charged in the off-peak period, excess capacity equal

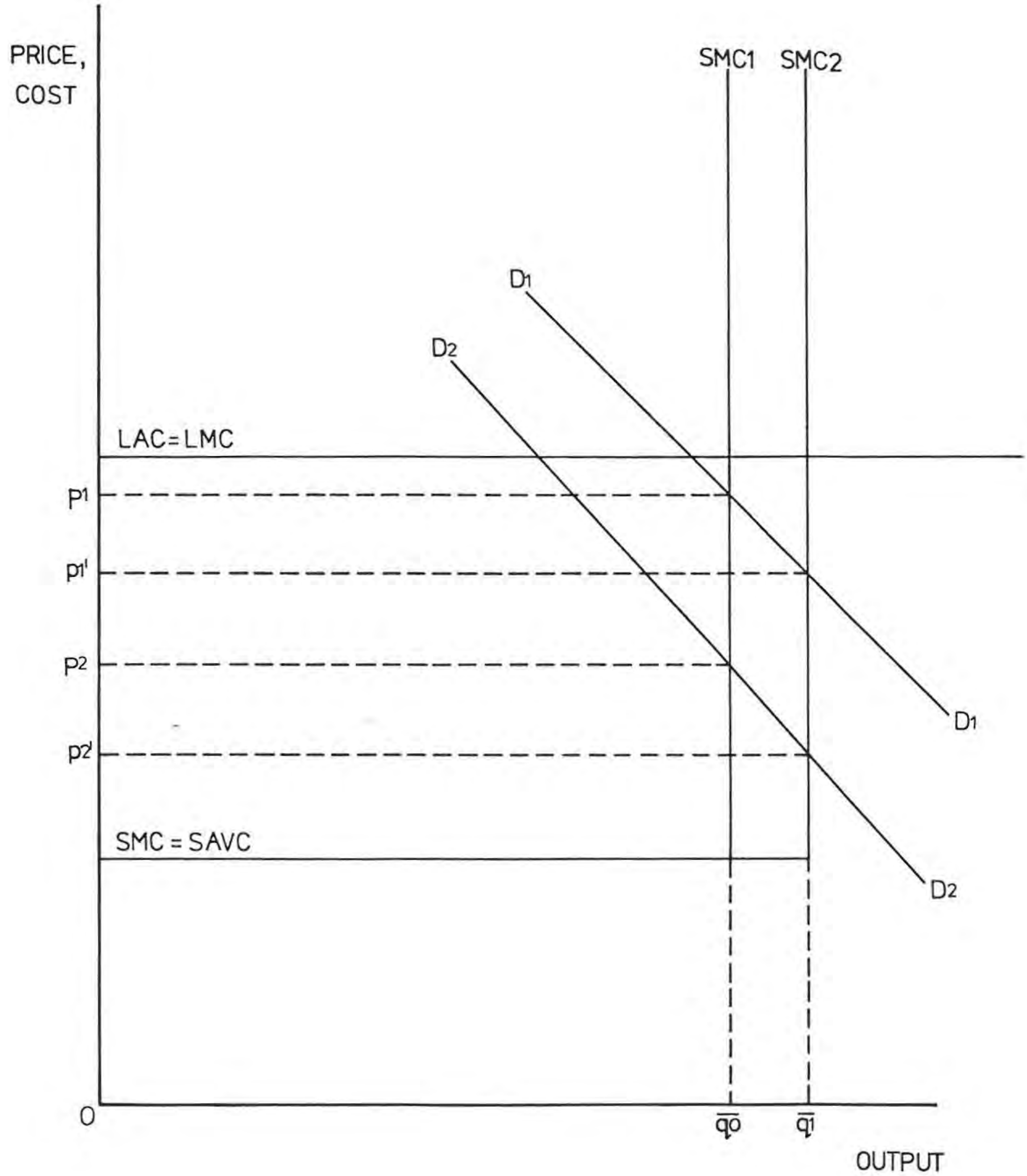


FIGURE 2:6 The Application of the Optimal Investment Rule

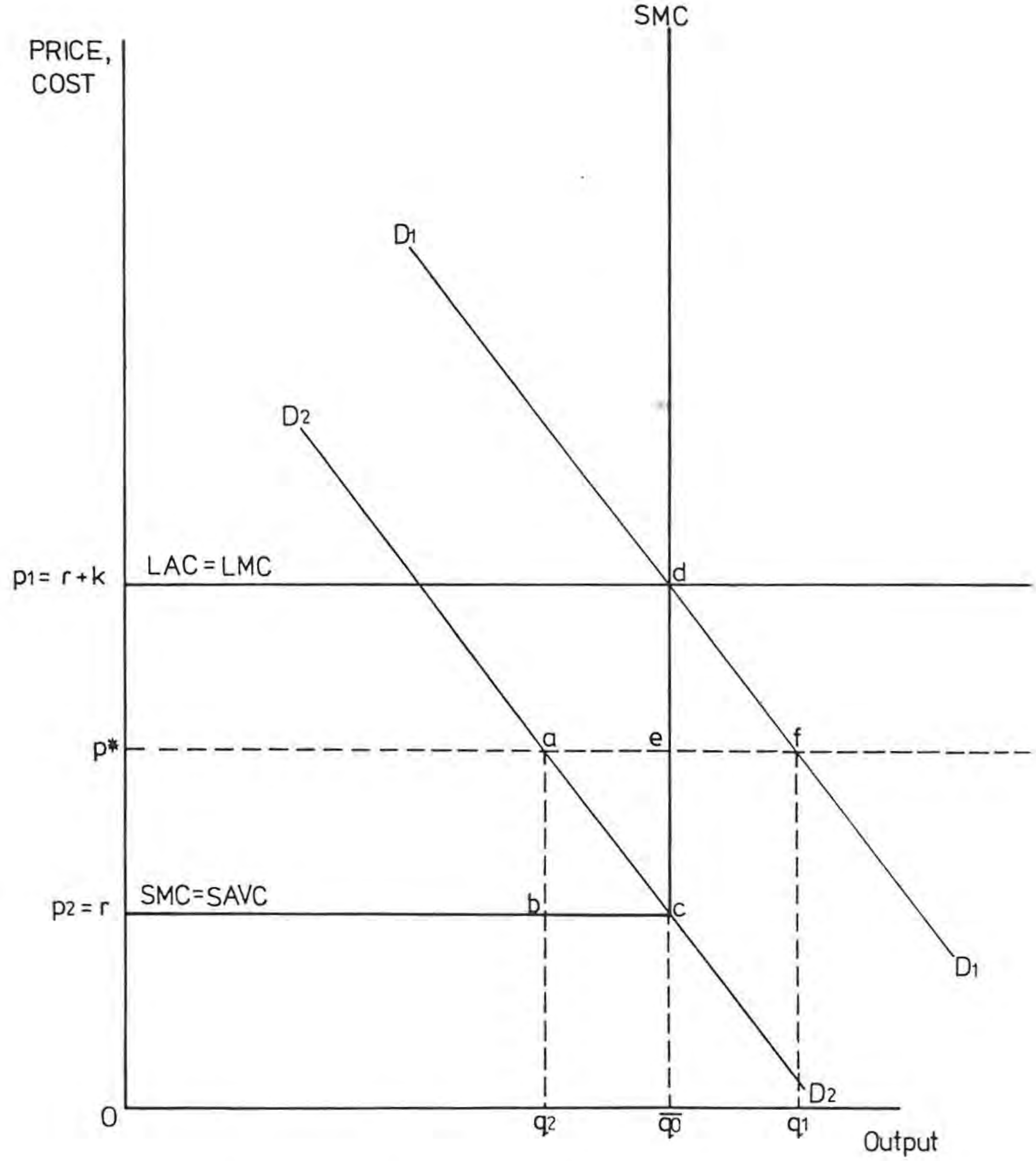


FIGURE 2:7 The Welfare Losses Incurred in Charging a Uniform Price

to ($\overline{qo}$ - q2) will arise. The welfare loss associated with this excess capacity is indicated by the area of the triangle abc which measures the net loss in consumer's surplus resulting from charging a price higher than short run marginal cost. If p* is charged in the peak period, excess demand equal to (q1 - $\overline{qo}$) will arise. This may induce the public enterprise to implement an investment programme designed at expanding the output capacity of plant to q1. The welfare loss associated with charging p* in the peak period can therefore be measured in terms of the additional capital costs of investment less the net gain in consumer's surplus equal to the area def which arises from charging a lower price than p1. It will thus be justifiable for the public enterprise to differentiate between peak and off-peak prices if the welfare losses associated with charging a uniform price in both periods are greater than the administrative and metering costs. The same principle can be applied in deciding whether to introduce any further pricing subperiod.

C. EVALUATION OF THE R - T MODEL

It has been shown that there is no long conflict between the efficiency and financial objectives of a public enterprise if a marginal cost pricing and investment policy is applied under the conditions which characterize the R - T model. The assumptions underlying the shape of the cost curves in the R - T model may be valid in the following circumstances:

(i) where the provision for reserve plant capacity gives rise to a flat stretch of the SAVC curve where unit running costs are constant;

(ii) where the output required can only be produced by scales of plant capacity greater than the minimum optimum scale at which, as Bain has suggested, constant returns to scale set in; and

(iii) where it is not possible to increase the maximum output capacity of plant in the short run by requiring workers to work overtime.

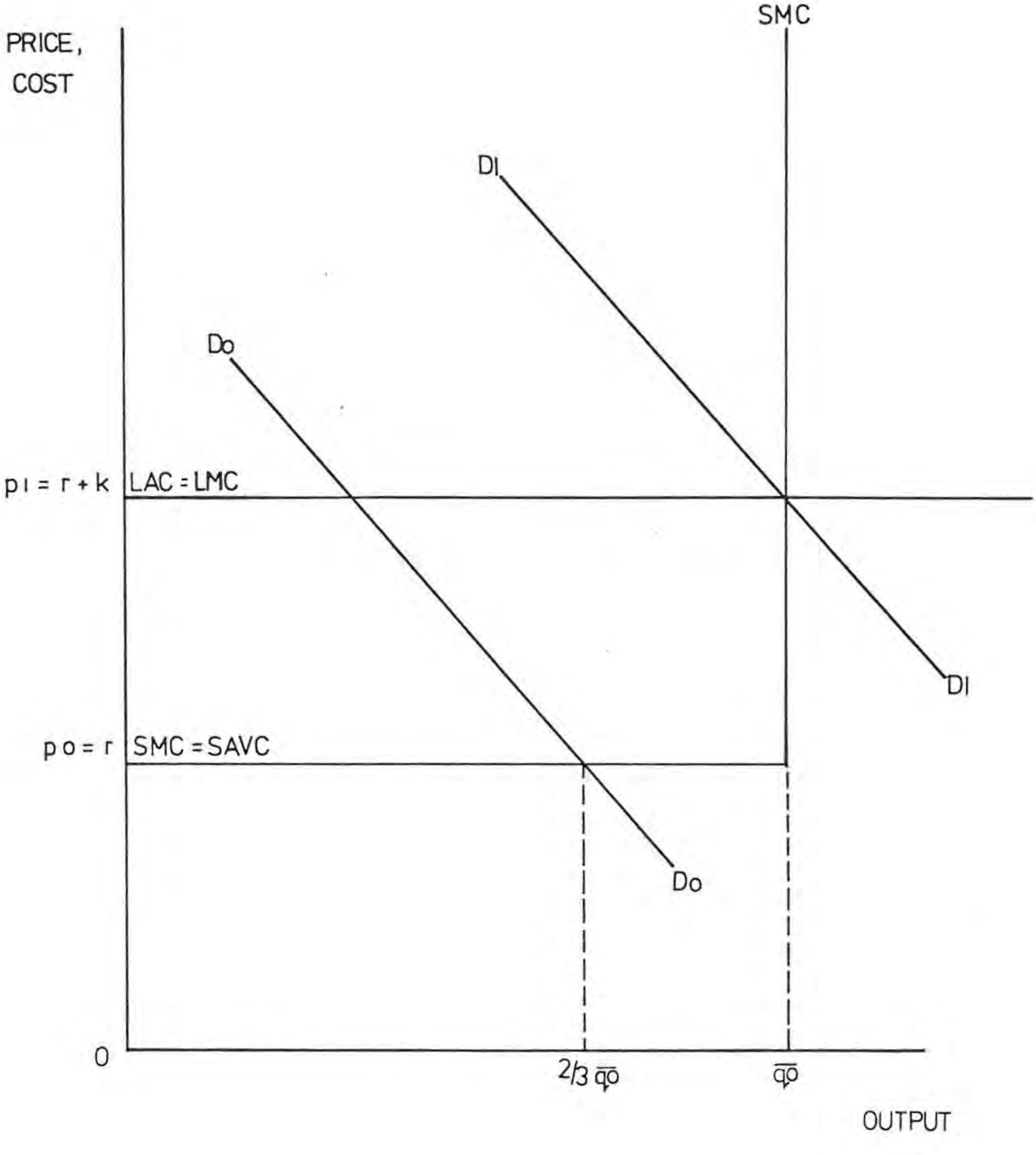


FIGURE 2:8 The Effect of the Provision of Reserve Capacity on the Rees-Turvey Model

However, even if the cost curves facing a public enterprise approximate the shape indicated by the R - T model, there may still be a conflict between efficiency and financial viability if provision is made for reserve plant capacity in excess of that amount associated with the typical load factor of the plant. Figure 2:8 indicates how this provision for excess reserve capacity will affect the central predictions of the R - T model. If demand is at the level DoDo, the setting of price at po which is equal to short run marginal cost, will result in plant being utilized at its typical load factor of 2/3 of full capacity $\overline{qo}$. Now, if demand is expected to rise to D1D1 then according to the R - T model, price should be allowed to increase to p1 to ration demand to existing capacity and no expansion of plant capacity would be required since total costs would be covered at the price p1. However, in an uncertain environment a public enterprise may respond to the expected increase in the utilization of the plant beyond the typical load factor of 2/3 $\overline{qo}$ by implementing an investment programme to expand the capacity of plant so that this typical load factor can be maintained under the new demand conditons. If this investment programme takes place then the application of the marginal cost pricing rule will still result in the public enterprise incuring a deficit. It follows that under conditions of uncertainty where public enterprises make provision for a "reserve plant margin", the short run surpluses or deficits incured by the public enterprise can no longer be regarded as signals for investment or disinvestment.

The predictions of the R - T model will also be affected if one drops the assumption of perfect divisibility of finite increments in plant capacity. This problem of indivisibility will be examined in the next section.

D. THE EFFECT OF INDIVISIBILITIES

In practice it is often found that the plant used by public enterprises is indivisible over wide ranges of output and can only be varied in larger indivisible lumps. In Figure 2:9 the capacity output $\overline{qo}$ is produced by a single unit of plant so that it can only be increased in an indivisible lump from $\overline{qo}$ to $\overline{q1}$ by purchasing and installing

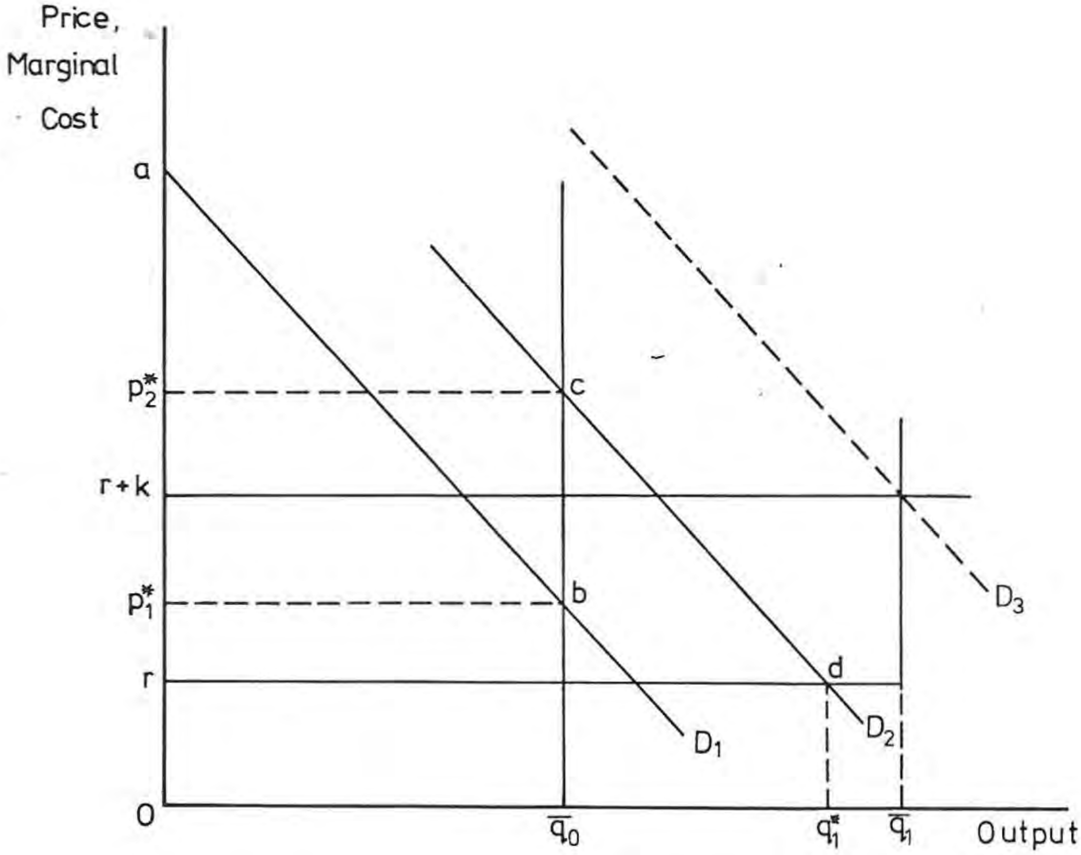


FIGURE 2:9 The Effect of Indivisibilities on the Application of the Optimal Pricing and Investmet Rules

a second unit of plant and decreased by closing down the plant. There will be no change in the application of the marginal cost pricing rule in the short run. However, if demand is expected to remain at D2 so that the public enterprise expects to make a surplus equal to $(p^*2 - (r + k)\overline{qo})$, then, in the long run, the public enterprise must decide whether it will continue to make this surplus or whether it will implement an investment programme to increase the maximum output capacity of plant from $\overline{qo}$ to $\overline{q1}$, charge a price equal to r and incur a loss of $k\overline{q1}$. If the demand price is assumed to measure the marginal social value of an additional unit of consumption, then the enterprise will expand capacity to $\overline{q1}$ if the gain in the total social benefit equal to the area $Cdq^*1\overline{qo}$ is greater than the sum of the increased capital costs $k(\overline{q1} - \overline{qo})$ and increased running costs $r(q^* - qo)$. If the public enterprise does decide to expand capacity by this indivisible amount then it can be seen that a deficit equal to $k\overline{q1}$ will be incurred until such time as demand grows to the level D3 where the optimal price will be sufficiently high to cover all unit costs. It is clear, therefore, that in this case deficits are likely to be for more prolonged in duration than under the implicit condition of the R - T model that plant capacity is perfectly divisible and investment can vary by any amount.

Figure 2:9 also illustrates the problem facing a public enterprise when it makes a loss on a large indivisible unit of plant capacity. If demand is at the level D1, then the application of the short run marginal cost pricing rule will result in the public enterprise making a loss equal to $(r + k - p1^*)\overline{qo}$. If demand is expected to remain at this level, then the public enterprise must decide between continuing to operate these facilities at a loss or closing them down. If a private enterprise was in a similar situation it would face no such alternative and would have to close down the plant since it could not survive in the long run by making a loss. It may, however, be economically justifiable for the public enterprise to continue operations provided that there is a net social benefit to be derived from pursuing this policy. It would thus compare the benefit society derives from the provision of output $\overline{qo}$ at price p^*1 which is measured by the area $ab\overline{qo}$ O against the cost of provision $(r + k)\overline{qo}$.

If $ab\overline{qo}\,O$ is greater than $(r + k)\overline{qo}$ then society would derive a net benefit frcm the continued operation of the plant. On the other hand, the plant should be closed down if $(r + k)\overline{qo}$ is greater than $ab\overline{qo}\,O$. If, however, the continued operation of the plant at a loss is economically justifiable, then it follows that a loss will persist on a more or less permanent basis unless demand increases to a level sufficiently high for a marginal cost pricing policy to generate a surplus or if it is decided to close down the plant.

E. CONCLUSION

To sum up, it appears that the application of the marginal cost pricing and investment rules will result in a public enterprise incurring a deficit in the following circumstances:

(i) in the short run, where excess demand is absorbed by the reserve plant margin and is not rationed by raising price above unit running cost; and

(ii) in the long run, either where there are increasing returns to scale so that the LAC curve lies above the LMC curve or where indivisibilities cause the public enterprise to retain unutilized plant either because demand has not grown to the level where the plant will be fully utilized or because the social benefits obtained from continuing operations at a loss exceed the welfare losses associated with closing down the plant.

It is therefore unlikely that the price structure determined according to the rules discussed in this section will be acceptable to the policy-maker. It is therefore necessary to consider the next stage in the piecemeal optimization procedure whereby prices are adjusted above or below marginal cost to satisfy the conditions for second best optimality in an economy where there are known and significant non-optimalities.

2:2 OPTIMAL ADJUSTMENTS TO THE MARGINAL COST PRICING RULE

2:2:1 THE EFFECT OF A BINDING REVENUE CONSTRAINT

It is necessary for a public enterprise to pursue pricing policies which result in the prices charged for its services being higher than their marginal cost when the application of the marginal cost pricing rule is not expected to generate sufficient revenue to satisfy the revenue target of the enterprise. It will be assumed that the amount of the revenue target is decided by a body independent of the public enterprise's management so that it acts as an exogenous constraint on pricing policy. Although the revenue target is usually not set below a level sufficient to cover total costs (net of State subsidies) it may be raised above this level to enable the public enterprise to generate the internal funds necessary to finance future capital development. Where this revenue target acts as a binding constraint upon the pursuit of a marginal cost pricing policy, the public enterprise must determine the pricing policy which will maximize net social benefit subject to this revenue constraint. This approach clearly differs from that which underlies the pursuit of an average cost pricing policy where the primary purpose of pricing policy is not to optimize economic efficiency but rather to generate sufficient revenue to cover total costs. In formulating a pricing policy which maximizes net social benefit subject to a revenue constraint, a public enterprise must decide between the following approaches:

(i) whether a constant unit rate will be charged for each of its products or whether the price will take the form of a two-part tariff with a fixed charge and a rate per unit of consumption; and

(ii) whether the objective of pricing policy should be to obtain maximum economic efficiency or whether it should serve the objectives of both economic efficiency and distributional equality.

A. THE RECONCILIATION OF FINANCIAL AND EFFICIENCY OBJECTIVES THROUGH PRICE DIFFERENTIATION

Where a multiproduct public enterprise cannot attain its revenue target if it applies a marginal cost pricing policy, it is necessary to allocate the shortfall among the different outputs in such a way as to minimize the loss of allocative efficiency. In the case where the public enterprise produces two outputs with independent demand functions, the optimal outputs of the two goods must satisfy the conditions:

$$\frac{p_1 - MC_1}{MR_1 - MC_1} = \frac{p_2 - MC_2}{MR_2 - MC_2} = Z \qquad (2{:}2)$$ [19]

where Z is some negative number whose value is determined by the revenue constraint. This expresses the condition that the ratio between the marginal welfare losses and marginal profit should be equal for both goods since otherwise it will be possible to reallocate outputs in such a way as to leave profits unchanged but to secure a greater welfare gain in one market than in the other.

Now since the marginal revenue of each good can be written as

$$MRi = pi(1 - 1/ei) \qquad i = 1,2. \qquad (2{:}3)$$

where ei is the price elasticity of demand for good i, it follows that when the optimality condition expressed in (2:2) is attained, then for both outputs:

$$pi(1 - 1/\hat{e}i) = MCi \quad i = 1,2 \qquad (2{:}4)$$

---

19. Rees, R., op.cit., p. 105, derives the optimality of this condition.

where

$$\hat{e}i = -\frac{(1-Z)ei}{Z} \qquad (2:5)$$

In other words, the optimal deviation of price from marginal cost will vary inversely with the elasticity of demand of the good. As will be discussed in Chapter Three, a pricing policy of setting prices on the basis of "what the market will bear" provided that the prices of all goods at least covers marginal cost, may result in a pricing structure which approximates this optimality condition, since, other things being equal, higher prices are likely to be charged on goods with inelastic than those with elastic demand. Baumol and Bradford[20] consider that the intuitive rationale for such a rule is that the damage to welfare resulting from departures from marginal cost pricing will be minimized since the relative quantities of the goods sold will approximate their marginal cost pricing proportions.

If the two goods are interdependent in demand then the optimal prices and quantities must satisfy the relationship:

$$\frac{p_1 - MC_1}{MR_{11} + MR_{12} - MC_1} = \frac{p_2 - MC_2}{MR_{22} + MR_{21} - MC_2} = Z \qquad (2:6)$$

where MRi,J, i,j, = 1,2, is the effect of the ith output on the revenue of the j th. The sum $MR_{11} + MR_{12}$ represents the total effect on the revenue of the enterprise, of a change in output 1, via the effect on both its own demand, and that of good 2. The only difference between the optimality condition in (2:6) and that in (2:2) is that in the former this total effect is taken into account. The intuitive rationale for (2:6) thus remains the same as for (2:2).

---

20. Baumol, W.J. and Bradford, "Optimal Departures from Marginal Cost Pricing", A.E.R., 1970, pp. 267-283. The writers trace the antecedents of this discussion to F.Ramsey's "A Contribution to the Theory of Taxation" in Economic Journal, March 1927. This argument was taken up by Pigou, A.C., in A Study of Public Finance, 2nd ed., 1947, while Boiteux, M., was the first writer to present the analysis in the form of a discussion of public utility pricing in "Le 'revenue distributable' et les pertes économiques," Econometrica, April 1951. He then completed a definitive analysis of the subject in terms of Pareto optimality in "Sur la gestion des Monopoles Publics astreints à L'equilibre budgètaire", Econometrica, Jan., 1956.

An implicit assumption of the analysis in this section is that a public enterprise can only charge a single unit price for any of its services. In the next section the question whether in the second best situation created by the existence of a revenue target, a smaller welfare loss may result from charging a two-part tariff consisting of a fixed charge and then a price per unit of consumption, will be examined.

B. TWO PART TARIFFS

The rationale for a two part tariff is that the fixed charge is in the nature of a lump sum tax since it is paid for out of consumer's surplus and therefore does not affect the value the consumer places on the marginal unit of consumption. If this is accepted then it should be possible to meet the profit target from the fixed charge and set prices equal to marginal cost. It can however, be shown that changes in the fixed charge will affect the demand for a public enterprise good and hence resource, allocation, firstly, because of the income effect on demand and secondly, because some consumers will rather not consume the good at all than pay the fixed charge.

These two effects are illustrated in the consumer choice problem depicted in Figure 2:10. In the absence of any fixed change, the consumer's initial position of equilibrium will be at point d where the budget line yo is tangential to the indifference curve Uo. If a fixed charge F is imposed on good 1 then before the consumer can buy any amount of the good he must first pay F so that his income is reduced to y_1 = yo-F, which defines the new budget constraint shown as y_1, in the figure. If the consumer refuses to pay the fixed charge and buys nothing of good then his income will remain at yo but he must remain at point a in the figure since he must spend all of y_1 on q_2. The consumer will do this if the new consumer equilibrium is at point e on the indifference curve cV1 since in this case point a represents a higher level of utility than point e. On the other hand, if point e lies on the curve bV1 the consumer will pay the fixed charge and consume q_1* of good. It should be noted though that the quantity consumed of good 1 falls from q^o_1 to q_1* due to the income effect.

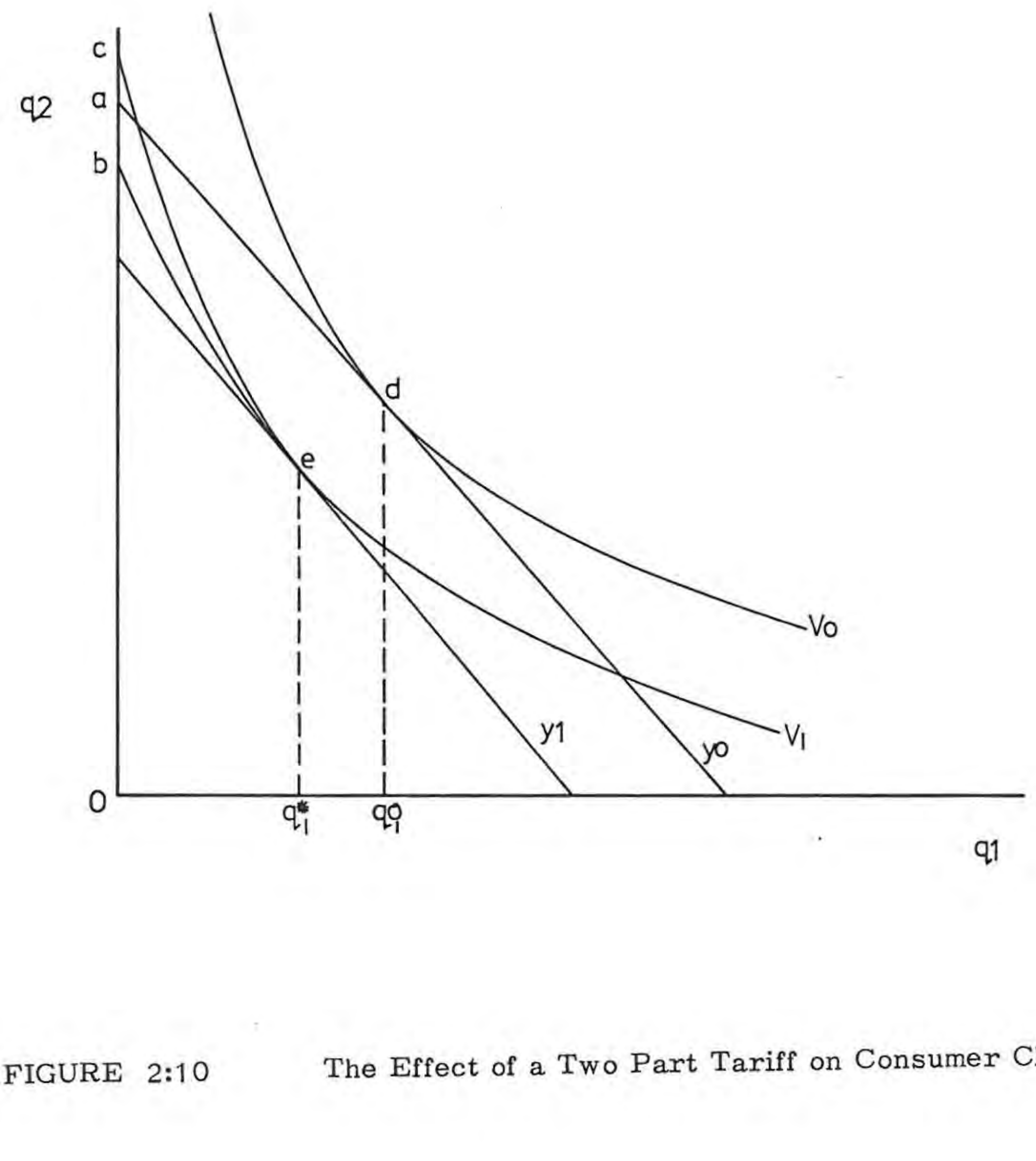


FIGURE 2:10 The Effect of a Two Part Tariff on Consumer Choice

Now if these results are generalized over the whole market for good 1 it can be seen that an increase in the fixed charge will cause both a fall in the number of consumer's n, and in the level of consumption by those who pay the fixed charge and who continue to consume the good. The imposition of a fixed charge to meet a profit target will thus cause welfare losses due to the underconsumption of the good and it is therefore necessary to determine an optimal two part tariff which will generate the required profit target with a minimum loss in efficiency.

The profit constraint, $\bar{S}$, under a two-part tariff can be expressed as:

$$p_1 q_1 + Fn - C(q_1) = \bar{S} \qquad (2:7)$$

where $p_1 q_1$ is the revenue from sales of the good produced by the public enterprise and Fn is revenue from the payment of fixed charges. Given that

$$q_1 = f_1(p_1, F) \quad \frac{dq_1}{dp_1} \quad , \quad \frac{dq_1}{dF} < 0 \qquad (2:8)$$

and

$$n = f_2(p_1, F) \quad \frac{dn}{dp_1} \quad , \quad \frac{dn}{dF} < 0 \qquad (2:9)$$

it can be shown that the optimal price and fixed charge, which through (2:8) and (2:9) will determine total demand q_1 and the number of consumer's n must satisfy the conditions

$$\frac{p_1 - MC_1}{MR_1 - MC_1 + \phi_1} = \frac{(p_1 - MC_1)\,\frac{q_1}{F}}{MRF + (p_1 - MC_1)\,\frac{dq_1}{dF}} = Z \qquad (2:10)$$

where

$$\phi = F \frac{dn}{dp_1} \quad \frac{dp_1}{dq_1} > 0 \qquad (2:11)$$

Equation (2:10) sets out the condition for optimality that the ratio of the marginal welfare loss of output to the marginal profit of output, adjusted for the effect of changes in output on the revenue from the fixed charge, should be equal to the ratio of the marginal welfare loss with respect to the fixed charge to the marginal profitability of the fixed charge. If these ratios are unequal then it would be possible to find variations in price and the fixed charge which would increase welfare while leaving total profit unchanged.

C. THE RECONCILIATION OF FINANCIAL, EFFICIENCY AND EQUALITY OBJECTIVES THROUGH PRICE DIFFERENTIATION

The acceptance, in principle, of Graaff's proposal that a public enterprise pricing policy should be formulated ".... with regard for its effect on the distribution of wealth as well as for its effect on the allocation of resources"[21] implies that both the objectives of economic efficiency and distributional equality should be taken into account in allocating among the outputs produced by a public enterprise, the shortfall from the revenue target which would be expected to result if a marginal cost pricing policy was applied. If consumers have differing preferences and consume goods in different proportions, the change in relative outputs resulting from the adjustment of public enterprise prices above marginal cost will also effect the distribution of welfare among consumers. A policymaker may therefore seek to differentiate the price adjustment applied to the different outputs produced by a public enterprise in a way which improves distributional inequality and contains the loss of allocative efficiency. Rees[22] has defined the conditions

---

21. Graaff, J., deV., Theoretical Welfare Economics, Cambridge University Press, 1st ed., 1957, p. 155.
22. Rees, R., op.cit., p.121.

for optimal price differentiation in a case where a public enterprise produces two goods as:

$$pj\ (1 - \frac{kj}{ej}) = MC \qquad j = 1,2 \qquad (2:12)$$

$$\text{where} \quad kj = 1 + \frac{Dj}{Bj} \qquad (2:13)$$

Bj being the Lagrangean multiplier associated with the revenue constraint and Dj representing the "distributional characteristic" of the respective goods. The size of this distributional characteristic is determined by the relative proportions of the good consumed by different consumer groups weighted by the marginal social utility of income associated with each group.

It can be seen from (2:12) that if the two goods have the same demand elasticities, then the greater relative divergence of price from marginal cost should be allowed for that good with the lower distributional characteristic. In other words, when the shortfall from the revenue target is allocated over goods, relatively smaller welfare losses are imposed on those goods, large proportions of which are bought by consumers with high marginal social utilities of income. It follows that generally the distributional characteristic may offset or reinforce the effects of different demand elasticities among goods. A good with both a high demand elasticity and high distributional characteristic will have a relatively lower price/marginal cost ratio then one with opposite characteristics. It can therefore be concluded that the overall allocation of the shortfall among outputs will depend in part upon the relative losses in allocative efficiency as represented by demand elasticities, and in part upon the relative losses in distributional equality, as measured by the distributional characteristic.

2:2:2 SECOND BEST PROBLEMS

The marginal cost pricing rule for a public enterprise is only strictly applicable to a first best economy in which there is no divergence of price from marginal cost in any other sector of the economy. The theory of second best suggests that where there are deviations of price from marginal cost in sectors which are interdependent with public enterprises, then it is no longer optimal for the latter to set their prices equal to marginal cost since second best optimality may require public enterprises to allow their prices to deviate from marginal cost in such a way as to minimize welfare losses. The key element in this problem is the relationship between the output of the deviant sector, q_1, and the public enterprise price p_2.

$$q_1 = f(p_2) \qquad (2:14)$$

It can be shown that the optimal public enterprise price p_2^* must satisfy the condition

$$p_2^* - MC_2 = -(p_1^* - MC_1)(dq_1/dp_2)(dp_2/dq_2) \qquad (2:15)$$

If q_1 is a substitute for q_2, then dq_1/dp_2 will be positive so that the term on the right hand side will be positive as a whole. It follows that in this case the public enterprise price should exceed marginal cost by an amount which is greater the greater is the price-marginal cost divergence in the deviant sector. Conversely if q_1 and q_2 are complementary goods, then the optimal p_2^* will be below marginal cost.

The marginal net social benefit of the public enterprise good q_2 depends not only on the difference between its own price and marginal cost, but also on the changes it induces in the output of the deviant sector. It can be shown that marginal net social benefit will be zero and total net social benefit will be maximized if the public enterprise sets its price to attain condition (2:15) since this condition can be rewritten as:

$$(p_2^* - MC_2) + (p_1^* - MC_1)\frac{dq_1}{dq_2} = 0 \qquad (2:16)$$

where the first term can be interpreted as the partial marginal net social benefit of q_2 in market 2, and the second as the partial marginal net social benefit of q_2 in market 1.

The relationship between these marginal social net benefits and output is shown in Figure 2:11. $\overline{q_2}$ is defined to be the output at which $p_2 = MC_2$ and curve A shows how the value of the difference $p_2 - MC_2$ increases as q_2 is reduced below $\overline{q_2}$. Curve B is drawn on the assumption that the second partial marginal net social benefit $-(p_1 - MC_1)dq_1/dq_2$ falls as q_2 is reduced. This assumption is essentially a second order or stability condition.

This figure clearly shows why it will no longer be optimal for a public enterprise to adopt a marginal cost pricing policy in a second best situation. At output $\overline{q_2}$, the loss in welfare resulting from a small reduction in q_2 is given by $p_2 - MC_2$ and is therefore zero, while the gain in welfare in market 1 is measured by $- p2 - MC2 - (p_1 - MC_1)\ dq_1/dq_2$ represented by point b in the figure. Since the gain in welfare in market 1 from a reduction in q_2 is greater than the loss in market 2 there is scope for Pareto improvement. Consumers in market 1 would be better off and no-one else worse off if they could pay the consumers of good 2 to reduce their consumption and release resources into production of good 1. The market, however, would fail to organize this mutually beneficial exchange. The public enterprise though could achieve the same result if it reduced its output from q_2 to q_2*, where A and B intersect, by allowing its price to rise above marginal cost by the amount $p_2^* - MC_2$. It can thus be seen that by adopting an appropriate pricing policy, the public enterprise could bring about the required resource reallocation and thereby exhaust all possibilities for mutually beneficial exchange.

A second type of second best problem arises when a public enterprise buys its inputs from a sector which charges a price greater than marginal cost. This is reflected in Figure 2:12 which illustrates a bilateral monopoly situation where the public enterprise is the sole buyer of a good and a private enterprise the sole seller. Although the public enterprise is the sole buyer it chooses not to exercise its monopsonistic power and accepts the supply prices determined according to the private enterprise's marginal cost curve, MC. The demand curve, D, shows at each level of output, q, the marginal value product of the good for the public enterprise which, in an otherwise first best economy, measures its marginal social value. The optimal

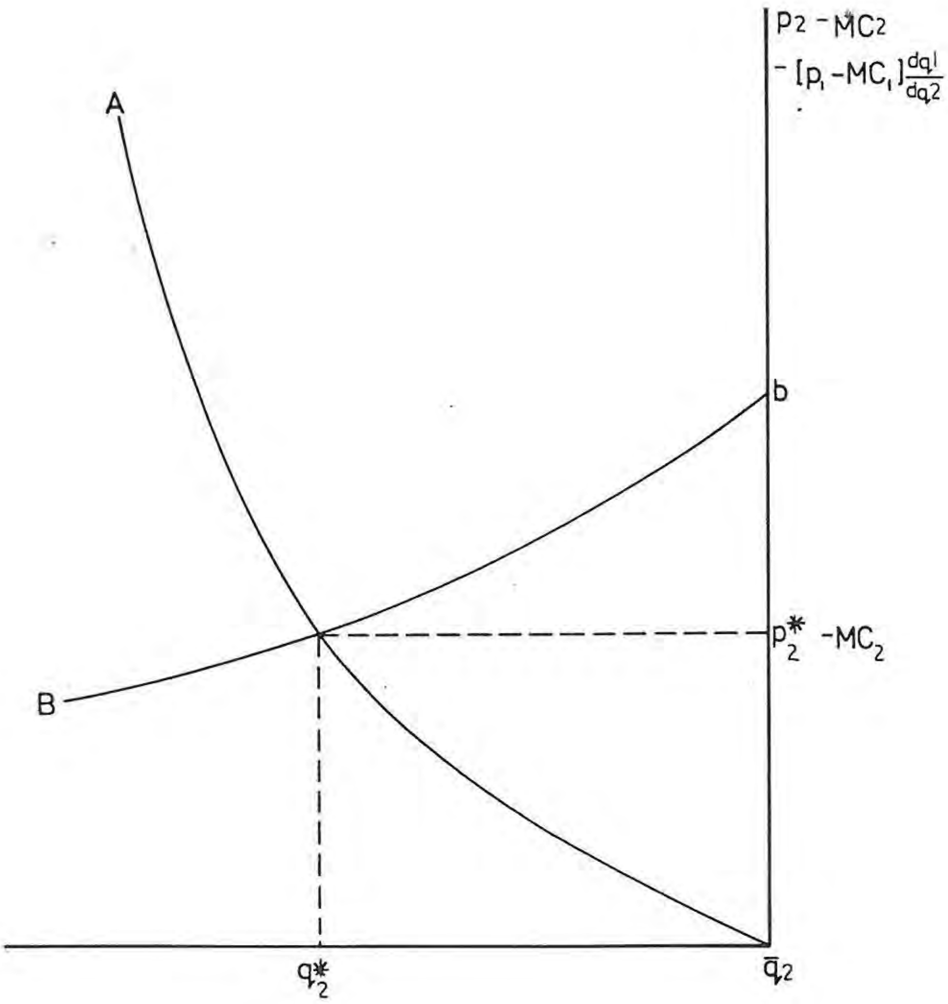


FIGURE 2:11 Second Best Optimality

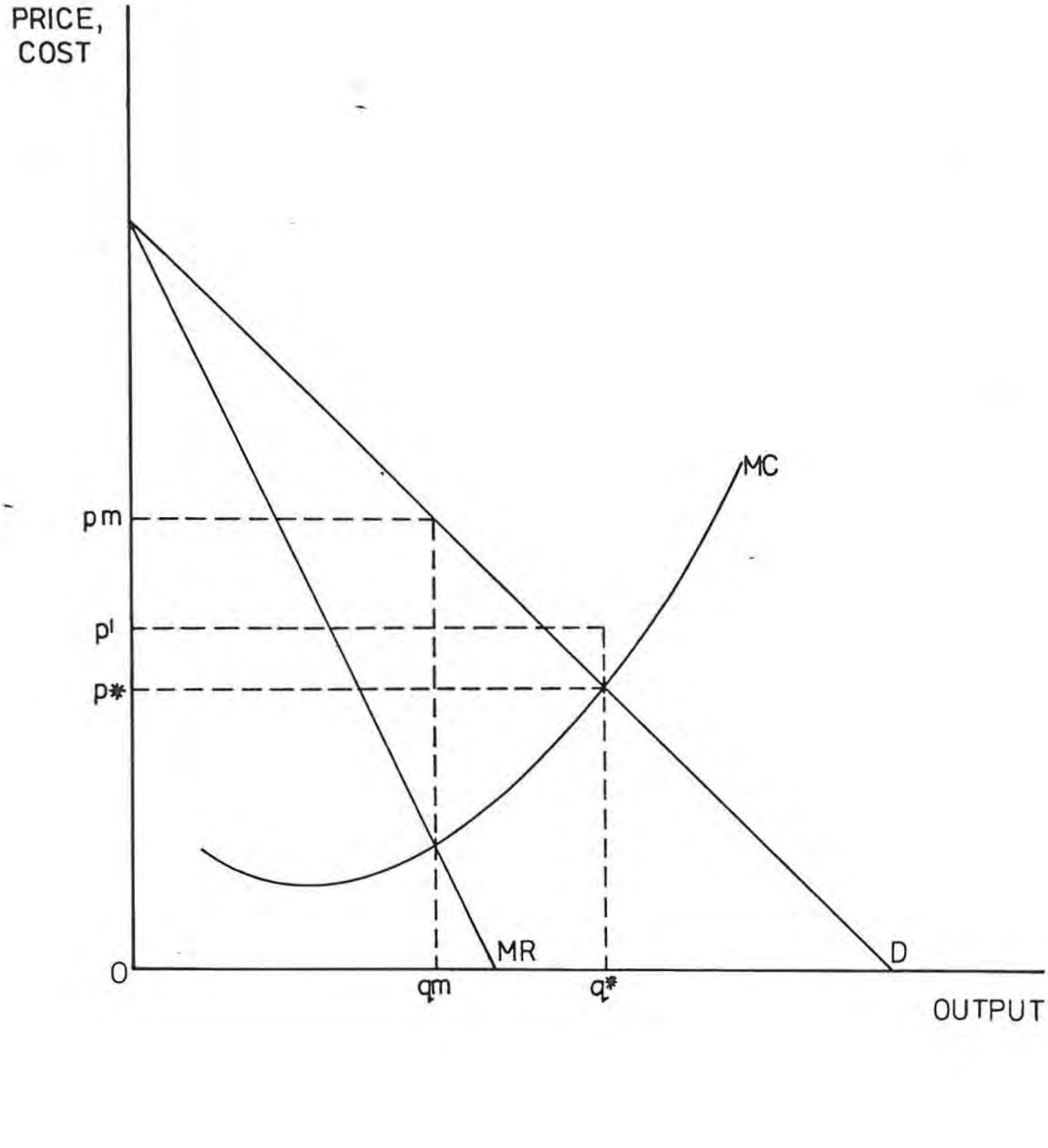


FIGURE 2:12 Bilateral Monopoly

quantity of the good which should be produced by the monopolist and bought by the public enterprise is q* since at any other output, the marginal value product of the good will differ from its supply price. The monopolist, however, is unlikely to accept the profit associated with the optimal price p* since it will maximize its profit at price p^m by restricting its output to q^m where marginal revenue is equivalent to marginal cost. The public enterprise may therefore have to adopt a strategy of announcing that it will buy an amount q* and then negotiate with the monopoly over price. If a price is higher than p* such as p^1 is determined by negotiation, the public enterprise should still use in its determination of input mix, output and final price, the "shadow price" p* rather than the actual price p'. The significance of p' is that it determines the profit or loss of the public enterprise and so the result of the payment of a price p' instead of p* for the input is that there will be a transfer payment from the Treasury, which has to subsidize any public enterprise operating loss, to the monopolist. If it is assumed that there are no welfare effects associated with a public enterprise deficit or surplus it follows that the value of p' will affect income distribution rather than resource allocation.

The next section will deal with the problem of how "shadow prices" may be applied when the market rate of interest does not affect the public enterprise cost of capital.

2:2:3 THE PUBLIC ENTERPRISE COST OF CAPITAL

In formulating rules for optimal pricing in the previous section, the existence of an interest rate, r, which determines the cost of capital to a public enterprise has been presupposed. Although it is not necessary to consider the cost of capital in short run pricing problems involving fixed capacity, it becomes an important parameter in the analysis, once capacity is treated as variable in the long run. This is because it will, in relation to other prices, determine the least cost input combination and thereby, the relation between total costs and output so that it becomes an important component of marginal cost. The cost of capital is also that rate used in discounting net social benefits of future consumption, so as to determine that level of investment which maximizes their net present value. It follows, therefore, that ".... sophisticated pricing rules are unlikely to achieve economic efficiency .... if the value of the cost of capital is non-optimally chosen , given its pervasive influence on the solution."[24]

In a first best economy, a public enterprise will satisfy the necessary conditions for a Pareto optimal allocation of resources to investment if it invests up to the point where its marginal rate of return on investment (mg) is just equal to the market rate of interest.[25] It will thus adopt a similar investment appraisal rule to private enterprises, only undertaking those investment projects for which the time stream of benefits discounted at the market rate of interest are greater than the cost of the project, and doing this until mg = r. It follows that private firms will also maximize their net present value by expanding investment to the point where their marginal rate of return (mp) is equal to r.

Now since mg = r

and mp = r

it follows that mg = mp. (2:17)

This satisfies the condition for the optimal allocation of investment between private and public enterprises. Now, since every household will maximize its utility by equating its marginal rate of time preference (p) with r, a Pareto optimal allocation of present and future consumption among households will occur since:

pi = r for households 1 ...., n

and a Pareto optimal allocation of resources to total investment will

---

24. Rees, R., op.cit., p. 124.

25. A derivation of these conditions is conducted by R. Rees, op.cit., pp. 125-132.

also occur since from

$$mg = r = mp = pi. \qquad (2:18)$$

In sum, it can be seen that, if the public enterprise adopts the market rate of interest as its cost of capital and pursues an investment appraisal rule designed to maximize the net present value of investment, then it will satisfy the necessary conditions for Pareto optimality in investment in a first best economy. The market rate of interest will furthermore provide information about the opportunity cost of resources devoted to capital which can be used to derive the value of marginal cost upon which the pricing policy of the public enterprise will be based.

A seond best problem occurs when the market rate of interest no longer measures both

(i) the time preference rate of every household and thereby fails to provide information on the value of current versus future consumption; and

(ii) the marginal rate of return on investment in the private sector and thus does not provide information on the opportunity cost of resources devoted to public rather than private investment.

An example of such a second best problem occurs when the profits of private enterprises are taxed. In this case private enterprises will maximize their net present value by investing up to the point where their marginal rate of return equals the tax adjusted interest rate, r^*, that is:

$$mp = r^* = \frac{r}{1-t} \qquad (2:19)$$

where t denotes the rate of company tax. It is necessary to examine the effect of this imperfection on the three conditions for Pareto optimality in investment:

(i) the Paretian optimality of the allocation of present and future consumption among consumers will be unaffected since every consumer will still maximize utility by equating his time rate of preference with the market rate of interest;

(ii) the Pareto optimality of the allocation of investment between private and public enterprises will, however, be affected. This is because private enterprises will maximize their net present value by equating their marginal rate of return on investment with a tax adjusted interest rate ($r' = r/1-t$) which is actually higher than the market rate of interest. If the public enterprise uses the market rate of interest as the cost of capital used to discount the time stream of future benefits associated with investment projects, it follows that in the equilibrium position where private and public enterprises are both maximizing their net present value of investment:

$$mp = \frac{r}{1-t} > mg = r \qquad (2:20)$$

This represents a suboptimal allocation of resources to investment, since if investment could be diverted from the public enterprise to private enterprises, the total return on investment in the economy could be increased. Competitive markets and the application by public enterprises of the familiar investment appraisal rule would, however, fail to effect this mutually beneficial exchange. It follows that the cost of capital of the public enterprise needs to be adjusted upwards to correct problem of overinvestment by public enterprises and underinvestment by private enterprises in this second best situation; and

(iii) the necessary condition for a Pareto optimal allocation of resources to total investment will also be violated if private enterprises are taxed since it follows from (2:18) that $mp > pi = mg$.

If consumers reduce their current consumption and invest it in private enterprises, they will be more than compensated for their sacrifice. However, if public enterprises continue to adopt the market rate of interest as their cost of capital, perfectly competitive markets will fail to effect this mutually beneficial exchange.

If the policy of taxing companies is unalterable, then the public enterprise may be able to move the economy to a second best optimum by adjusting upwards its cost of capital and thereby reducing the optimal level of its investment. In pursuing this policy of reducing investment and increasing the cost of capital, the public enterprise planners must realize that this will require a reduced bond issue and hence a fall in the market interest rate, r. This will cause a rise in both current consumption and private enterprises investment to an extent determined by their sensitivity to the interest rate. Taking these two factors into account Rees derived the condition for second best optimal investment as follows:

$$1 + mg = \theta 1(1+r) + \theta 2\left(1+\frac{r}{1-t}\right) \qquad (2:20)$$

where $0 < \theta 1, \ \theta 2 < 1, \ \theta 1 + \theta 2 = 1.$

The equilibrium marginal rate of return for the optimal level of public enterprise investment is thus a simple weighted average of the consumer's time preference rate (which is equivalent to the market rate of interest) on the one hand, and the marginal rate of return to private sector investment, on the other hand, where the latter can be determined by a simple correction to the market interest rate. The weights are:

$$\theta = \frac{\frac{dxo}{dr}}{\frac{dxo}{r} + \frac{dyp}{r}} \quad \text{and} \quad \theta 2 = \frac{\frac{dyp}{dr}}{\frac{dxo}{dr} + \frac{dyp}{dr}} \qquad (2:21)$$

where xo represents the level of current consumption and yp the level of private enterprise investment.

As Rees has stated:

> ".... the weight which the market interest rate receives depends entirely on the strength of the interest effect on current consumption ..... relative to that on investment. The reason is that these determine the extent to which the public enterprise investment displaces, at the

margin, consumption or private investment".[26]

In other words, the above condition expresses the equality between the marginal rate of return on public enterprise investment and its opportunity cost, which can be determined in terms of the displacement of current consumption and private enterprise investment which occurs at the margin.

It follows, therefore that the public enterprise cost of capital will take the form:

$$r^* = \theta r1 \text{ and } \theta 2 \frac{r}{1-t} \qquad (2:22)$$

The public enterprise can thus satisfy the conditions necessary for optimal investment if it maximizes net present value, vg; in other words

$$vg = \frac{xg}{1+r^*} - yg \qquad (2:23)$$

using r^* in discounting.

In conclusion, it can be seen that in a second best situation the adoption by a public enterprise of a "test discount rate" based on estimates of the marginal rate of return on private investment may be suboptimal since generally, we must have

$$r < r^* < \frac{r}{1-t} \qquad (2:24)$$

to compensate for the resource misallocation that results from the tax distortion. The analysis in this section indicates in a general way, the way in which the cost of capital should be

---

26. <u>Ibid</u>., p. 138.

adjusted to correct for unalterable market distortions. It should be noted that, in the same way as in other second best situations, the informational requirements of the public enterprise planners who formulate those rules may substantially increase so that that the optimality conditions may be merely useful to indicate the direction in which price adjustments should occur.

2:2:4 CONCLUSION: THE OPTIMAL TARIFF STRUCTURE

The discussion of optimal pricing principles in this section leads to the conclusion that there are two main considerations which should be borne in mind in formulating an optimal structure of tariffs for a multiproduct public enterprise. Firstly, it is necessary to determine the structure of marginal costs. This will be affected by the technical relationship between inputs and outputs, market or "shadow" input prices, and the degree of capacity utilization. Secondly, the optimal relationship between prices and marginal costs must be determined. The structure of optimal price/marginal cost ratios have been shown to depend on the nature and degree of interdependence with markets in which prices diverge from marginal costs, on the demand elasticities of the goods, and on their distributional characteristics.

In the next section the extent to which these optimal pricing rules go towards accomplishing the objectives of a decentralized system of control of public enterprises will be evaluated.

2:3 EVALUATION OF PRICING RULES

2:3:1 PUBLIC ENTERPRISE DECISION TAKING AND THE ROLE OF PRICING RULES

Public enterprise decision taking takes place in a delegated choice situation since the public enterprise does not make choices in terms of its own interests but is rather obliged to make decisions in the "national interest". The national interest may be explicitly or implicitly defined by a minister in terms of an objective function which requires the public enterprise to pursue policies

(i) which are consistent with an efficient allocation of resources within the economy and productive efficiency within the public enterprise;

(ii) which ensure that the public enterprise remains financially viable and able to break even without recourse to Treasury financing; and

(iii) which further the aims of government distribution and macroeconomic stabilization policy.

It is often the case that conflicts exist between these different objectives and that the pursuit of any particular objective involves trade-offs in terms of performance with respect to other objectives. There are thus two requirements for public enterprise decisions to conform to the national interest:

(a) there must be a satisfactory statement of objectives which resolves the conflicts between different objectives by attaching relative weights to these objectives. If a consistent, stable and comprehensive statement of objectives cannot be obtained then the minister may seek to resolve the conflict by sequential attention to the interests of various interest groups. This may involve _ad hoc_ intervention in the decisions of a public enterprise to further the aims of a particularly pressing policy objective. However as Rees put it:

> ".... such intervention blurs the lines of authority and responsibility, duplicates decision-taking, absorbs scarce resources, leads to delay in implementing decisions, and breeds dissatisfaction among public enterprise executives who feel their discretionary power sharply constrained."[27]

---

27. Rees, R., _op.cit._, p. 11.

As the demands upon public enterprises increase, so to does the necessity to grant these enterprises special privileges relative to private enterprises to enable them to fulfill these demands in the 'national interest'; and

(b) once a statement of objectives has been formulated by a public enterprise it is necessary to devise a system of control to ensure that decisions actually taken are consistent with these objectives. In designing a system of control the following two kinds of costs will have to be traded off against each other:

(i) the costs of operating the system which include the costs of resources absorbed in monitoring decisions, collecting and communicating information, and examining alternatives as well as the delays in the implementation of decisions which the control system might cause. These costs will be minimized in a decentralized control system where the public enterprise is given the discretion to take any decisions which will lead to the attainment of a clearly specified set of objectives. On the other hand these operating costs will be maximized if a centralized system of control is adapted where the decision-taker has to justify every decision before it is implemented since in these circumstances there will be a duplication of decision-taking with the decision-taker acting essentially as preliminary organizer of information; and

(ii) the costs which occur when decisions do not correspond with objectives. These costs will be higher for a decentralized then a centralized system since in the former case there will be a tendency for the decision-taker to develop his own preference ordering over the outcomes of decisions which differs from that he is supposed to adopt.

To facilitate decentralized decision taking a minister should derive ground rules from the objective function which can be generally applied by the public enterprise in its day to day management. Examples of such ground rules are marginal cost pricing, target rates of return, and the use of a test rate of discount in applying discounted cash flow techniques to investment appraisal. The various optimality conditions for pricing and investment by a public enterprise which were discussed in sections 2:1 and 2:2 should clearly form the basis for these ground rules. There are, however, two major problems which face a public enterprise in formulating ground rules from these optimality conditions.

Firstly the borader the range of objectives which the public enterprises are required to pursue, the less effective is likely to be their performance of any particular objective. This is because the public enterprise may be constrained in its pursuit of any single objective by a requirement to meet target or minimal levels of performance on other objectives. This can be illustrated by reference to the conditions for optimal pricing set out in the previous section. Where a public enterprise is required to meet a profit target to satisfy the objective of financial viability or where it is required to take into account the objective of distributional equality, these additional objectives act as further constraints on optimal pricing, with the result that the optimal pricing policy is that which attains these objectives with the minimal loss of allocative efficiency. The problem is thus that although public enterprises have an impact on a wide range of economic variables, there are diminishing returns to extending the number of policy objectives which use public enterprise pricing policies as instruments to secure their attainment. In fact if one accepts the Tinbergen[28] rule that to achieve any given number of policy objectives requires at least an equal number of instruments, then public enterprises should be used as instruments to obtain a strictly limited number of objectives. For example, public enterprises may be required to employ the most efficient pricing and investment policies

---

28. J.Tinbergen, On the Theory of Economic Policy, North-Holland Publishing Co., Amsterdam, 1952.

subject to meeting a net cash flow constraint while other instruments should be used to attain regional, distributional and macroeconomic objectives. In this case however, the other instruments may have to offset the distributional effects of "efficient" public enterprise pricing policies.

Secondly, a related problem is that the information required to implement an optimal pricing rule tends to increase as the number of practical constraints placed on decentralized decision taking by the public enterprise increases. In the simplest case where optimality conditions are fulfilled by a policy the public enterprise is required to know the following information:

(a) the demand function for the current period (Do) when inputs cannot be varied;

(b) the demand function (D1) for the future period in which inputs can be varied;

(c) how input requirements vary with output
  (i) when one input is fixed and
  (ii) when all inputs are variable; and

(d) the price of variable inputs, the cost of capital and the annual rate of plant depreciation. Where there is a systematic pattern of demand fluctuations within a period then it is also necessary to know the peak and off-peak period demand functions.

This information is a basic requirement for the application of any optimal pricing rule since all such rules require that price should be related to marginal cost in such a way as to minimize welfare losses. Additional information will be required in each of the following cases:

(i) where there is a departure from marginal cost pricing in a market related to the one served by the public enterprise then, as indicated in equation 2:15, the public enterprise must know how output in the related market varies with changes in its output, $\frac{dq_1}{dq_2}$;

(ii) where the public enterprise purchases inputs at prices which differ from their opportunity costs then it must know their "shadow prices";

(iii) where a multiproduct public enterprise is required to meet a profit target by allocating it over the different outputs then it must have information about both the own price elasticities of demand and cross elasticities of demand of the various outputs;

(iv) where the public enterprise meets a profit target by charging a two-part tariff then it is required to know the functional relationship between

(iv.i) output and the price and fixed charge $q_1 = f_1\ (p_1, F)$ and

(iv.ii) the number of consumers and the price and the fixed charge, $n = f_2\ (p_1, F)$; and

(v) where a multiproduct public enterprise is to take distributional considerations into account then it should have information about the marginal social utilities of the consumers consuming its different products, and the proportion of each output which they consume, as indicated in equation (2:12).

It can be seen that while the additional information requirements (i) and (ii) result from a realistic adjustment by the public enterprise to the environment in which it operates, the last three requirements (iii), (iv) and (v) result from the integration of additional public enterprise objectives with the primary pricing policy objective of securing allocative efficiency. It is clear that it becomes increasingly difficult to apply pricing rules in a process of decentralized decision taking where they are required to further additional policy objectives. It can be concluded that the cost in terms of foregone efficiency and additional information requirements of integrating additional policy objectives into public enterprise pricing policies should be compared with the increased burden placed on other instruments if public enterprises ignore these objectives.

In a perfectly competitive environment where there are no externalities, market determined prices provide the information which coordinates the decisions of producers and consumers in a decentralized manner. The basic rationale behind the use of pricing rules as instruments for the decentralized control of public enterprises is that they should constrain these enterprises to make decisions which result in the highest level of welfare for the economy as a whole being attained. In evaluating pricing rules in the next section the following factors will be considered:

(i) The extent to which they substitute for the functions normally performed by a competitive market process, and

(ii) The effectiveness with which they internalize into public enterprise decisions, factors which would be external if it were to act as a private enterprise.

2:3:2 PUBLIC ENTERPRISE DECISION RULES AS A SUBSTITUTE FOR COMPETITION

It is a common feature of public enterprises that they operate in an environment in which competition has been limited to a greater or lesser extent. Any policy recommendation for a public enterprise should therefore suggest rules which substitute for the various functions of competition. This approach obviously requires that there should be a clear understanding of the nature of competition. Unfortunately economic theory has been historically characterized by considerable ambiguity concerning the meaning of competition. McNulty[29] considers that the failure of distinguish between market structure and competition lies at the root of this ambiguity.

Whether it was seen as price undercutting by sellers, the bidding up of prices by buyers or the entry of new firms into a profitable industry, competition was introduced into economics by classical

29. McNulty, P.J., "Economic Theory and the Meaning of Competition", Quarterly Journal of Economics, Volume 82, 1968.

writers as a concept which was intended to have empirical relevance and operational meaning in terms of contemporary business behaviour. Adam Smith viewed competition as an ordering force which, in the long run, drove price to its lowest sustainable level and thereby ensured allocative efficiency. Following from the introduction of the concept of perfect competition into economics by Cournot, neoclassical economists focussed on competition as a description of an ideal market structure which could lead to a general equilibrium which satisfied the Pareto optimality conditions. It is paradoxical that in this state of perfectly competitive general equilibrium no price-cutting or any other form of competitive rivalry between firms would be possible. As Littlechild[30] has put it:

> "The situation depicted in the perfect competition model represents the situation when the competitive process has run its course".

In evaluating pricing rules as a substitute for competition, competition will be viewed as a description of the process of rivalry between firms seeking to gain a competitve advantage over each other by, - say, cutting prices, improving quality, cutting costs or introducing new products, techniques or methods of organization, in which no firm is able to sustain its advantage by preventing other firms from imitating it.

In a dynamically changing economy with imperfect dissemination of information, the incentives and pressures generated by a competitive market process induce private enterprises to meet consumers' wants at the lowest possible cost. If they fail to do this then they

---

30. Littlechild, S.J., "The Fallacy of the Mixed Economy" Hobart Paper No. 40, Institute of Economic Affairs, London, June 1978, p. 30.

are unlikely to survive for any length of time. A key role in this process is played by the entrepreneur who is characterized as being alert to profit opportunities not so far grasped by other market participants. By the very act of exploiting these opportunities the entrepreneur will bring them to the attention of other market participants and they in turn will respond with new alertness so that profits which exist due to unnoticed opportunities are competed away. In a competitive environment profits and losses therefore serve two functions. Firstly, they provide a feedback on performance since entrepreneurs who are successful in discovering and meeting the wishes of consumers are rewarded by profits; those who are unsuccessful are penalized by losses. Secondly, they signal which activities ought to be encouraged and which industries expanded in order to move the economy in the direction of greater allocative efficiency.

Although optimal pricing rules replicate the conditions for an allocatively efficient equilibrium they clearly need to be supplemented by mechanisms which provide the incentives and pressures necessary to induce the public enterprise to take advantage of the opportunities to increase net social benefit which continually arise in a dynamically changing environment.

A competitive market process will also generate strong pressures in private enterprises to contain costs by reducing X-inefficiency. A firm which tolerates X-inefficiency will find it difficult to survive in a competitive environment since it cannot shift the responsibility for this inefficiency on to consumers by exercising market power to charge higher prices. In a survey of a number of studies dealing with the relationship between X-inefficiency and market structure, Scherer found that firms generally respond to an increase in competitive pressures by making vigorous attempts to reduce costs and concluded that "incentives for cost reduction under a pressure theory of motivation had more powerful behavioral effects than incentives under a reward theory alone."[31]

The optimal pricing rules of a public enterprise are generally formulated on the assumption that costs are minimized and therefore take no account of X-inefficiency. Where however, a profit target is incorporated into the pricing rule this can place pressure on the public enterprise to contain costs provided that sufficient control is exercised over prices to ensure that the enterprise does not use its market power to raise those above their optimal levels to cover up internal inefficiency. The centralized monitoring of prices increases can, however, result in the loss of many of the advantages of a decentralized system of control. Rees has suggested instead "that the pursuit of technological and managerial efficiency should be divorced from the question of profitability and should be undertaken directly, rather than indirectly through pricing policy."[32] The appropriate instrument would be an "efficiency audit" in terms of which an independent outside body would be exclusively concerned with producing analyses of the efficiency of public enterprises, choosing its own area of investigation, and having full rights to information. The gains in decentralization would come about because this outside body would not be involved in the process of formulating decisions, but rather in appraising and monitoring the outcomes of decisions and procedures. In this way the attempt to stimulate greater managerial and technological efficiency could be made consistent with greater decentralization of pricing and investment decisions.

---

31. Scherer, F.M., Industrial market structure and economic performance, Rand McNally Publishing College, Chicago, 2nd ed., 1980, p. 466.
32. R.Rees, op.cit., p. 111.

A competitive environment is likely to create a strong incentive among firms to introduce new products and new techniques provided that the advantages gained from such innovation are not competed away too rapidly. Quite clearly optimal pricing rules are not directed at the innovative ability of the public enterprise so that it is necessary to consider whether there will be sufficiently strong pressures and incentives, to stimulate it to introduce the new products and techniques which would have been introduced in a competitive environment. Leibenstein[33] has suggested that a public enterprise is "likely to be consistent with organizational structures under which it is difficult to introduce improvements and there is little to prevent such organizational structures from surviving for long periods of time". There are a number of reasons why this is so:

(a) there is generally no penalty to the management of a public enterprise for not introducing an innovation even when it is profitable to do so;

(b) there is insufficient reward in terms of promotion for introducing a profitable innovation when profit targets are already being met;

(c) the losses attributable to innovation are conspicuous and may create promotion difficulties for the management involved; and

(d) there is frequently a division in the organizational structure of the public enterprise between these capable of introducing changes as against those interested in doing so.

It follows that a system of rewards and penalties needs to be imposed on a public enterprise which stimulates the management to adopt organizational structures which are more responsive to innovation opportunities. In the next subsection the question of how the scheme of bonuses and promotion may accomplish this purpose will be discussed.

---

33. Leibenstein, <u>op.cit.</u>, p. 176.

In a competitive market process firms choose their role by reference to profitability. Their chosen role involves them in a complex of actions which benefit consumers of goods and services. If a firm decides to cease to provide a good or service others take its place. Furthermore some activities may be undertaken or services provided only because they foster customer goodwill and thereby improve the firm's standing vis-à-vis its rivals. The duties or obligations of a firm operating in a competitive market do not need to be defined because they are explicitly defined by the market. A public enterprise, on the other hand, is given a function to perform and is usually protected from competition in order to perform it. In the absence of competitive pressures, the imposition of pricing rules does not ensure that the public enterprise will perform services which a private firm would perform in the interests of fostering customer goodwill, so that its social obligations and duties need to be explicitly defined. Nove[34] has suggested that minimum service obligations may be regarded as "a necessary substitute for competition" and that the formulation of optimal pricing and investment rules should take these obligations into account as a practical constraint upon policy.

2:3:3 INTERNALIZATION OF EXTERNALITIES

The pricing and investment rules for public enterprises can also be evaluated according to the extent to which they internalize to the public enterprise what would normally be externalities. In formulating these rules in the previous section, externalities were ignored since it was assumed that marginal net social benefit could be measured by the difference between price and marginal cost. There are two broad approaches to dealing with externalities through the use of the price system:

---

34. Nove, A., Efficiency Criteria for Nationalised Industries, Allen & Unwin, London, 1973, p. 53.

(a) a system of state taxes and subsidies can be imposed which cause demand prices to reflect marginal social benefits and supply prices to reflect marginal social costs; and

(b) externalities can be incorporated into the objective function of a public enterprise so that pricing rules are adjusted to take account of them. For example in section 2:2:1 the optimal tariff structure of a multiproduct public enterprise was adjusted to reflect the distributional characteristics of the goods it produces.

Both these methods of internalizing externalities are optimal in a static world where demand, cost and social welfare functions are known with perfect certainty. However, in a world of dynamic change and uncertainty there are likely to be competing means of efficiently internalizing externalities in a decentralized way without ad hoc interference by the government in the decisions of the public enterprise.

Granick[35] has studied the systems of incentives established to induce decentralized decision takers to internalize externalities in both Soviet enterprises and in the subunits of large American firms.

In Soviet enterprises managers are assumed to maximize their discounted future income subject to the constraint of avoiding actions that are likely to lead to dismissal. Managerial income is largely derived from bonuses which depend on the enterprises fulfilment of a small number of specified plan indicators such as output or sales, cost reduction, profit and labour productivity. Annual plan indicators are set at levels which are quite ambitious

---

35. D.Granick, "The Internalizing of Externalities in Socialist Enterprises and in Subunits of Large American Firms", in Shepherd W., op.cit., pp. 77-95.

in relation to the potentialities of a high proportion of enterprises so that they usually cannot be fulfilled except by violating other plan instructors to which bonuses are not specifically attached. Furthermore managers will often be reluctant to overfulfill a plan indicator in one year since this may result in higher plans being set for following years. Given that the set of centrally fixed prices relevant to the enterprise do not reflect supply and demand conditions, Granick concludes that "enterprise managers fail to internalize a surprisingly important group of performance criteria which are internalized in a capitalist firm".[36] For example, the product mix produced is often not that requested by users,quality of product tends to be low and the introduction of new products is often slow.

The Russian failure to handle externalities through non-parametric mechanisms may be contrasted which the apparent success which Granick found in a study of the reward system for senior managers in a small number of large decentralized companies in the United States. Bonuses do not play the same role of discouraging externalities that they do in Russia since they are based not so much upon a set of well-defined and weighted criteria of performance as upon subjective evaluation of the manager's performance. Furthermore the pattern of career development is more important than bonuses in the American company reward system. Given that managers tend to rotate rapidly among major job functions and among major subunits within the company, it follows that the executive who maximizes performance of his own subunit at the expense of other subunits within the company will be jeopardizing his future career advancement since this usually depends on his acceptance within these other organizations. Externalities will thus be internalized to the subunit managers through the criteria for advancement which are used in the company.

---

36. <u>Ibid</u>., p. 83.

The results of Granick's study can be considered relevant to public enterprises operating in a market economy in that it indicates the need to develop direct incentives in terms of career advancement, salary and bonuses for the managers of these enterprises to pursue their delegated objectives. There are however, a number of factors which may limit the transferability of the incentive system operating in large decentralized private enterprises to a public enterprise:

(a) the pattern of career advancement followed by top managers in a public enterprise is unlikely to lead them into jobs in the departments responsible for controlling the enterprise;

(b) the subjective evaluation of managerial performance is complicated by the multiplicity of conflicting objectives public enterprises are required to pursue; and

(c) it is very difficult to reward successful risk-taking activity by top managers in the substantial fashion that is possible in privately owned companies.

A particular case of this problem of internalizing externalities will be examined with next section, that is, whether public enterprise pricing policies can be used as instruments of regional policy.

2:3:4 PUBLIC ENTERPRISE PRICING AND INVESTMENT POLICY AS AN INSTRUMENT OF REGIONAL POLICY

Regional economic policy is directed against those cases of regional inequality which constitutes problems and which cannot be corrected by the automatic operation of market forces. Stillwell[37] has identified three types of problem regions:

(a) underdeveloped regions where there are deficient employment, trading and consumption opportunities and a limited provision

37. F.J.B.Stillwell, Regional Economic Policy, MacMillan Studies in Economics, MacMillan, London, 1972, pp. 10-25.

of infrastructural facilities and public services. Oettle[38] considers that such regions may exhibit any of the following characteristics:

(i) low wages, a narrow spectrum of job choice and limited promotion opportunities for workers;

(ii) limited local or regional marketing and procurement opportunities for firms;

(iii) high price levels and a narrow spectrum of product choice for consumers; and

(iv) unfavourable conditions of use, incomplete availability and low quantitative and qualitative capacities of infrastructure facilities;

(b) regions which are depressed due to their above-average dependence on industries which are static or declining in demand at the national level or which suffer from some endemic disadvantage such as their peripheral location as regards the major market centres; and

(c) regions suffering from congestion since increases in concentration add more to total diseconomies than to total economies.

In South Africa the basic strategy of regional policy has been and remains to encourage private investment in selected industrial development points in order to propagate a self-generating process of growth at these poles which would stimulate the development of underdeveloped regions and counterbalance the forces of concentration in congested areas.[39] There are two broad groups of incentives directed

---

38. K. Oettle, "Obligations on transport for underdeveloped regions in W. Blonk, ed., op.cit., pp. 22,23.

39. See, for example, the information newsletter of the Department of Foreign Affairs and Information <u>The Promotion of Industrial Development. An element of a Coordinated Development Strategy for Southern Africa</u>, Pretoria, April, 1982.

at stimulating private investment at these growth poles:

(a) cash grants offered to firms during the first years of investment to overcome the liquidity problems often faced during this period; and

(b) a system of taxes, subsidies, and infrastructural investment designed to offset the long term cost disadvantages associated with locating investment at these growth poles rather than in areas of high industrial concentration.

There are a number of ways in public enterprise policy can be used to further regional policy objectives:

(a) the public enterprise can be obliged to provide a minimum level of service to underdeveloped regions regardless of whether a perpetual loss is incurred on these services. If the losses on these unremunerative services are financed by state subsidy this will clearly have implications for allocative efficiency and resource allocation within the economy but it will not affect the pricing policy adopted by the public enterprise. On the other hand if the public enterprise is required to finance the unremunerative services from the revenue earned on its other services, this will have the effect of raising both the profit constraint and tariff structure for these services. To apply the optimality rules developed in this chapter, the public enterprise will thus have to adjust its tariff structure to meet the new profit target with a minimum loss of allocative efficiency. A system of control will also have to be established to ensure that X-inefficiency is contained in the provision of the unremunerative services;

(b) the public enterprise may be required to offer subsidized rates to consumers in the underdeveloped regions. As explained above the implications of these subsidized rates may depend on whether they are internally or externally subsidized; and

(c) the public enterprise may be influenced to locate its new investment in the underdeveloped regions. This may necessitate a diversion of resources from areas of high industrial concentration where there is rapid growth in the demand for public services to underdeveloped regions where demand may initially lay behind the increased provision of these services. It follows that this policy may initially worsen excess capacity in underdeveloped regions and congestion in concentrated areas. If prices are related to marginal costs, then there is likely to be a price differential in favour of the former region which will act as a subsidy offsetting the cost disadvantages of private investment in this region.

The relationship between public enterprise and regional policy could only be considered in a general way in this subsection. It is clear however, that there are limits to such generalization and in the following chapters this relationship will be considered with reference to specific public enterprises in South Africa.

CHAPTER THREE

TRANSPORT

3:1 THE FRAMEWORK FOR PUBLIC TRANSPORTATION POLICY

3:1:1 THE STRUCTURAL CHARACTERISTICS OF THE TRANSPORT SECTOR

This chapter will apply the analytical tools discussed in Chapter Two to evaluate the pricing policies of a public transport undertaking in general and South African Transport Services (SATS) in particular. Before this can be done it is necessary to examine the special characteristics of the transportation sector within which a typical public transport undertaking would operate.

The products of transportation are the services supplied in moving goods or people from one place to another at the time the movement is desired. Transportation thus produces place and time utilities. The demand for transport therefore arises due to the advantages that people see in doing different things in different places. An individual's decision to make a trip is dependent on the attraction and accessibility of the destination. Klaassen[1] expresses the potential of a destination in terms of the perceived benefits and costs of making the trip. The benefits are composed of the number and quality of the opportunities at location while the costs include the sacrifices of money and time and the risk of making the trip. The transport user's perception of cost is influenced by his level of income, (since high-income users will place a relatively higher valuation on time and risk) and the extent to which he is able to choose between different modes of transport for the trip.

---

1. Klaassen, L.H., "Optimum use of transport networks and its possible contribution to welfare are maximization," in Polak, J.B. and van der Kamp, J.B., ed., op.cit., pp. 98-119.

This modal choice will not only be determined by the relative prices of the different modes of transport but also by such quality dimensions as their relative speed, safety and reliability. Each mode of transport provides a wide range of heterogeneous services which may be either specific to a particular mode, or substitutable for, or complementary to, the services provided by other modes. It is clear that any evaluation of the pricing policies of a public transport undertaking must take into account these intermodal substitution and complementary relationships within the transportation sector.

The economic structure of the transportation sector is profoundly affected by the distinction between fixed and mobile plant which characterizes the capital equipment used in this sector. Fixed plant or infrastructure is extremely costly and long-lasting and has little alternative use so that capital invested in transport infrastructure is irretrievable for other purposes. Its operating costs are low relative to capital costs and there are usually economies of scale up to a certain plant size while there are also big indivisibilities between practical levels of capacity. Finally, it is costly to expand infrastructure in stages since it may stimulate development around it and raise land values, and hence the cost of acquiring and/or expropriating private property. These characteristics of fixed plant explain why it is usually provided out of public funds or by a public enterprise or natural monopoly. In the case of railways, where the owner of track facilities is able to become the sole supplier of operating services on the track since these services are consumed in direct conjunction with the track facilities, a natural monopoly is likely to arise because:

(a) the initial capital investment is large and thus represents a significant barrier to entry; and

(b) when the plant is fully utilized, double tracking will require a large additional investment, which cannot profitably be made unless a large increase in traffic is expected.

If, instead of double tracking a new firm entered the market, an almost complete duplication of facilities would be necessary and the immediately available traffic would presumably be inadequate to allow either firm a profit. Competition between two firms could not survive because of the continued and unavoidable presence of excess capacity and a monopoly would thus emerge. In most countries railways are operated as public enterprises since a private firm would have the potential to exercise its monopoly power to restrict output and raise prices and would be unlikely to provide socially desired services if it was unprofitable to do so.

In contrast to fixed plant, the mobile units of transport equipment are relatively cheap and short-lived, highly flexible since their mobility enables them to have a multitude of alternative uses, and limited in their optimal size since the range of sizes over which economies of scale are possible is generally not large relative to demand. Furthermore they are likely to have high operating costs relative to fixed costs.

The effect of these characteristics on market structure can be seen in the road transport industry where there is a public provision of road infrastructure so that private firms are only involved in the provision of operating services. Effective competition between a large number of road operators is likely to result from the easy conditions of entry into this industry which are caused by:

(a) the small size of the initial investment required;

(b) the ease with which mobile equipment can be shifted between different locations and uses; and

(c) the fact that investment can increase by small amounts in response to growth in traffic.

The effective intramodal competition within the road transport industry is likely to induce private road hauliers to seek ways to improve the quality and reduce the costs of their services since in a competitive environment prices cannot be raised above opportunity costs for any length of time. However private road hauliers are unlikely to take account of externalities, in particular congestion. Congestion occurs when a marginal increase in traffic causes a decline in the average speed of traffic. The costs of congestion include the value of time losses and the increase in fuel consumption, wear and tear, labour and overhead costs which result from traffic delays.

3:1:2 INTERMODAL COMPETITION

The economic structure of the transport industry is likely to be characterized by considerable intermodal competition between public and private enterprise. This is exemplified by the competition in freight transport services between railways and road hauliers which exists in most countries.

Intermodal competition may improve the efficiency of the transport system as a whole if it is unrestricted since it will enable the user to choose that mode which offers the most preferred service in terms of cost, speed, safety, reliability and traffic utility. There will thus be a strong inducement for the different modes to improve their competitiveness in terms of the cost and quality of their services.

The main argument against unrestricted intermodal competition is that there is considerable inequality in the structural, institutional and legal conditions affecting each mode. For example, railway interests often argue that railways have to meet their infrastructural costs whereas road hauliers can make use of publicity provided road infrastructure. It therefore appears that a precondition for deregulating intermodal competition in freight transport services is that a system of road user charges should be imposed which reflect the social opportunity costs of using road infrastructure including the cost of congestion. A discussion of the practical problems associated with road user charging and of the various types of road user charges falls beyond the scope of this thesis. It is important to note, though, that if road user charges do not reflect opportunity costs then this will create a "second best" problem in formulating the pricing policies for a public transport enterprise since there are likely to be complementary and substitutionary relationships between its services and those provided by private road operators.

It can thus be concluded that in formulating the pricing and investment policies of a public railway undertaking, it is necessary to take account of the economic environment in which it operates and the nature and extent of intermodal competition in the transport sector. It is also necessary to consider the special characteristics of the Railway production process. This will be examined in the next section.

3:1:3 THE PRODUCTION PROCESS ON THE RAILWAYS

Railways technology consists of the simultaneous operation of three distinct processes which are physically connected by the flow of empty and loaded railway trucks. The three processes are loading of trucks, switching of trucks and line-haul operation.

(a) LOADING PROCESS

The loading process generally consists of the assembly of the freight, the loading of railway trucks and the forwarding of trucks. It is an operation which is carried on at the terminal and involves the employment of the following input services: labour, fuel, equipment and fixed plant including terminal building space. The output is the number of loaded railway trucks.

(b) SWITCHING PROCESS

Yard switching involves the employment of switch engines in picking up, assembling, and sorting railway trucks into trains. In the switching process the following input services are employed: labour, fuel, equipment and fixed plant. The output is the number of transferred railway trucks. The productivity of the switching process is affected by a number of factors which vary between different switching yards:

(i) the distance between the assembly yard and the loading docks and sidings which contain the trucks awaiting pickups. As this distance increases, the distance covered by a switch engine as it transfers a given number of trucks also increases; and

(ii) the degree of circuity in the routing of switch engines. Circuity of routing is necessitated by the scatter of railway trucks. If there are fifty trucks awaiting pickup, they may be at fifty different locations, at the same location, or distributed in some intermediate fashion. Circuity implies a greater expenditure of engine time to pick up a given number of trucks over a route of a given length.

(c) LINE-HAUL OPERATION

The line-haul process involves the transfer of loaded and empty trucks between terminals. Labour, fuel, equipment and fixed plant are employed as input services. The outputs produced are loaded truck/kilometres loaded, transported railway trucks; and empty trucks/kilometres. The output of the line-haul process may thus be varied by changing either the number of trucks to be hauled or the distance over which they are to be transported. It is also necessary to take account of the flow of empty trucks which are returned from consignees to points where they are to receive new consignments of freight. An imbalance in the directional flow of trucks may cause empty trucks to accumulate at "surplus" terminals which receive a greater number of loaded trucks than are sent out. These trucks are then transported, without return loads, back to terminals in "deficit" areas.

The special nature of the Railway production process must clearly be taken into account in evaluating the allocation of traffic between the different modes. This will be discussed in the next section.

3:1:4 INTERMODAL COORDINATION

One of the main objectives of transport policy is to ensure that there is an efficient allocation of traffic between the different modes. This means that each mode should specialize in the carriage of those traffic flows in which it has a comparative cost advantage. In this section the factors which should be taken into account in any evaluation of the allocation of traffic between modes will be considered.

The optimal distribution of traffic to different modes of transport depends on a large number of variables relating to the cost and service characteristics of the modes and the nature of the goods being shipped. The main intermodal service difference is speed, with trucks being faster than trains on any given arc since trains often have to collect traffic at intermediate points and may be slowed down by the periodic stops of other trains in the system, whereas trucks can carry smaller loads over the entire arc without any lengthy stops. Furthermore, besides having faster arc speeds truck service will usually be faster by virtue of having fewer transloadings. This is illustrated in Figure 3:1. If goods are to be transported from factory A to factory B by rail, then they will first have to be loaded on to trucks at factory A and then sent by road to station A where they will be offloaded. They must then be loaded on to a train at station A and railed to station B where they will be offloaded and then loaded onto a truck which will take them by road to their final destination at Factory B. If the traffic is carried by trucks, then they will be loaded on to the trucks at factory A and carried directly by road to factory B. Rail transport in this case, will involve three loading and unloading operations compared to the single operation that would occur if the goods were transported by road.

It also can be seen from the figure that the Railways is likely to incur far higher terminal costs than road hauliers. At the same time rail costs are likely to increase at a slower rate as the length of haul increases compared to road transport costs. The effect of this is shown in Figure 3:2 where it can be seen that at a particular length of haul, OD*, rail transport will become the cheaper mode. Obviously if the length of haul is shorter than OD* then the goods should be carried by road.

The allocation of freight traffic between the different modes can be examined in terms of a "decision tree" proposed by Abouchar[2] which is shown in Figure 3:3. According to Abouchar the main factors affecting the allocation of traffic between road and rail are the value of the commodity, the length of haul and the density of traffic along the route. As can be seen in Figure 3:3 high-value and medium-value freight are assigned to road, except when a rail

---

2. Abouchar, A., Transportation Economics and Public Policy with Urban Extensions, Wiley & Son, Toronto, 1977.

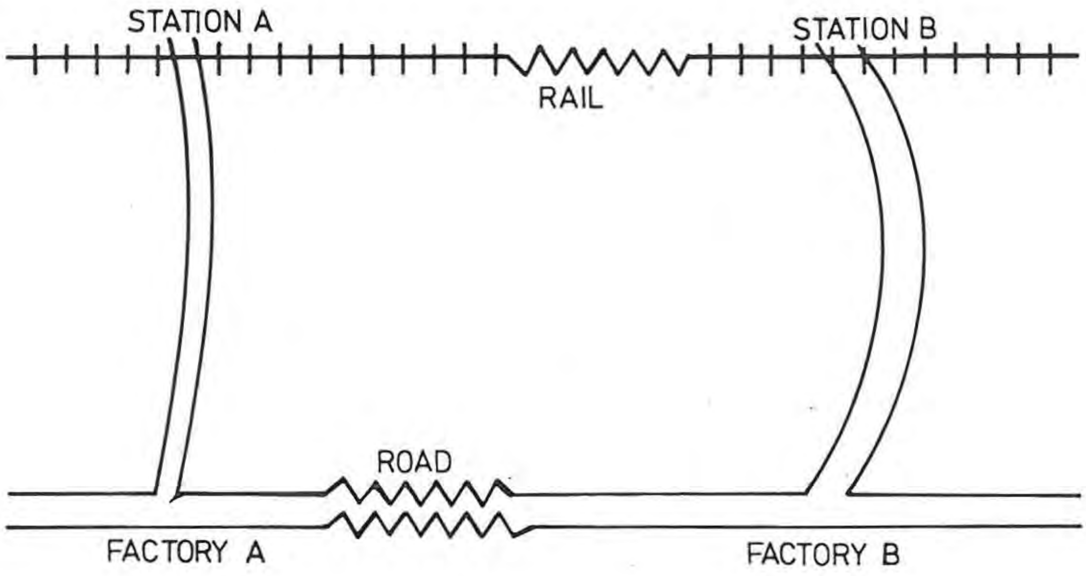


FIGURE 3:1 Intermodal Co-ordination

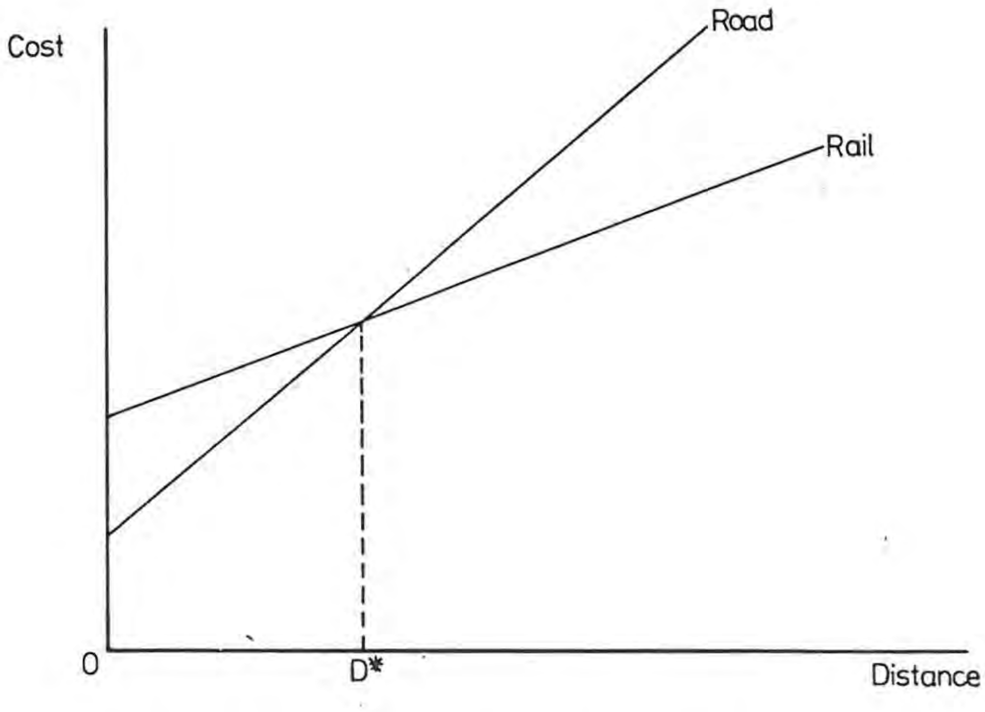


FIGURE 3:2 A Comparison of the Cost - Distance Relationship of Rail and Road Transport

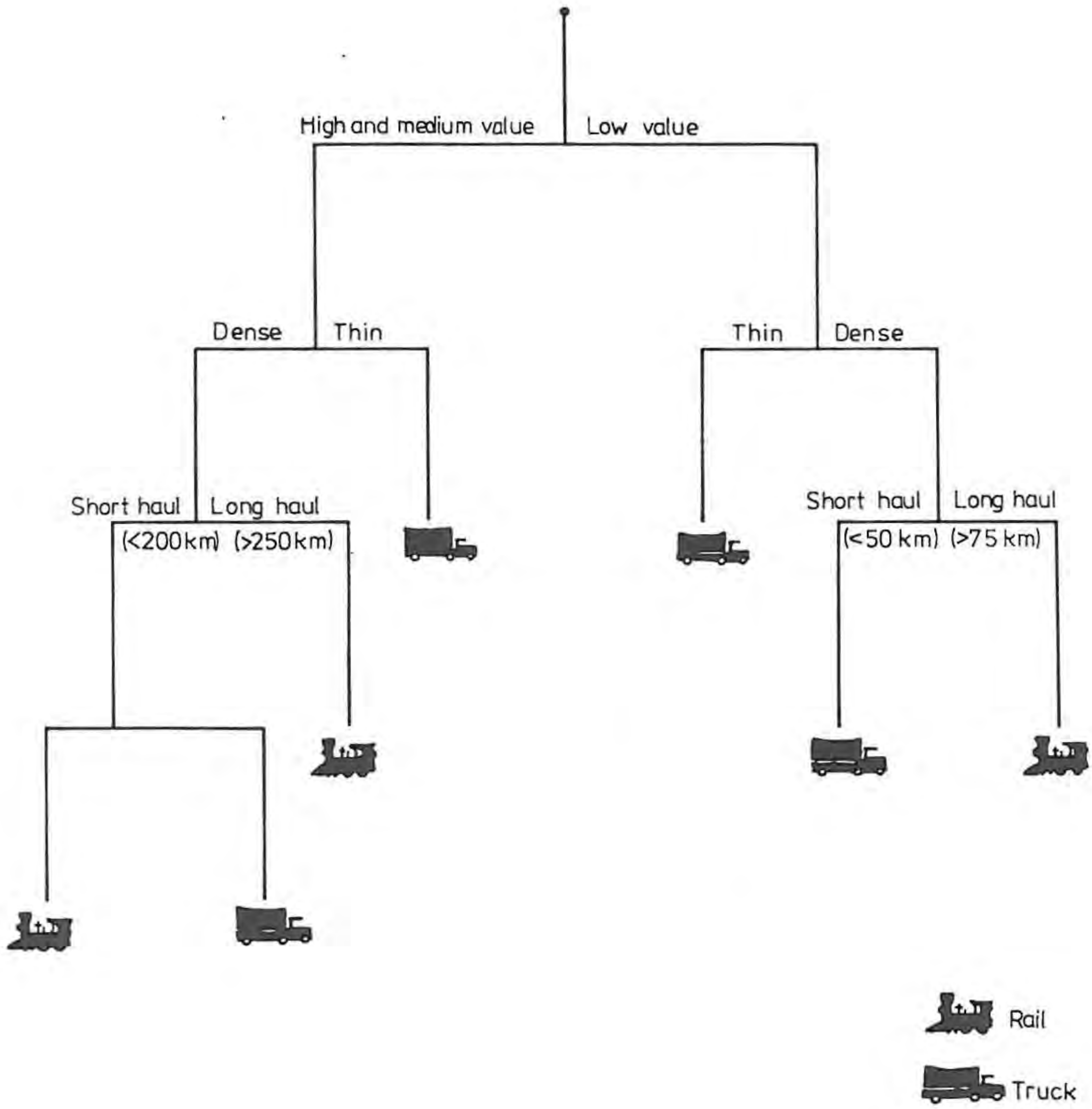


FIGURE 3:3 A Decision Tree for the Intermodal Allocation of Freight Traffic

connection exists on the arc and the individual commodity flows are dense. In shipping high-and-medium-value commodities shippers usually choose the mode on the basis of speed. They will thus only choose rail if they can reduce their shipping times to times attainable by road. This usually requires the shipper to have his own siding and to ship to buyers with sidings.

The decision tree also indicates that all commodities which are shipped over a long haul should be carried by rail. However, the minimum length of the haul over which a commodity should be shipped by rail will vary according to the nature of the commodity and the circumstances in which it is being shipped. For example, the owners may own a small fleet of vehicles which are normally unwanted at the end of the day and may thus be available for a short haul to the station. The cost of doing this may be low especially if the goods can be moved direct from lorry to railway wagon. The minimum length of haul at which rail becomes less costly will be at a much shorter distance than in the case where goods can only be shipped at peak periods to the station and when there is a lower probability that they can be loaded directly on to the appropriate railway wagon. Thus although, the comparative advantage of rail generally increases with length of haul it is not possible "to lay down that costs per ton forwarded by road or rail for any given mileage."[3]

Figure 3:3 also indicates that low-value, short-haul dense flows should be shipped by road. An example of such a traffic flow would be the shipping of building materials which are quarried in or close to cities. These would not typically originate at a railhead. Therefore, if they are to be shipped by rail, at least one truck-rail transfer would be required. In this case, road transport is normally cheaper than rail transport.

---

3. Ponsonby, G.J., Transport Policy: Coordination through Competition , I.E.A., Hobart Paper, No. 49, 1969.

Another low-value assignment which would normally be shipped by road would be a traffic flow with such a low density that railroad construction along its arc would not be justified in the first place.

The considerations reflected in this decision tree are usually taken into account by Governments in regulating intermodal competition in the transport sector. A problem with this approach, however, is that it is difficult to prescribe general rules to assist in the rational allocation of traffic since the circumstances in which a modal choice decision is made will only vary between shippers but also over time. In section 3:9 it will be argued that efficient intermodal coordination is best achieved when the costs of the different modes are reflected in their prices so that the individual shipper be able to choose the mode which transports his goods at the lowest cost. Co-ordination is thus generally more effectively achieved through competition than by regulation although it is admitted that regulation must occur when the competitive process breaks down.

The provision of transport services by the different modes clearly takes place in a spatial context. In the next section the question of how the supply of transport both affects and is affected by the spatial distribution of economic activities will be considered.

3:2 THE IMPACT OF PUBLIC TRANSPORT ENTERPRISE POLICIES ON REGIONAL DEVELOPMENT

The reciprocal relationship between space and transport must be borne in mind in evaluating the impact of the pricing and investment policies of a public transport enterprise on regional development. The distribution of economic activities in space should not only be viewed as a static determinant of the shape of the transport network existing in a country but should also be seen as a dynamically changing structure which can be influenced by many factors including the availability, cost and quality of transport services.

The pricing and investment policies of a public transport enterprise may affect the availability, cost and quality of transport services and hence the spatial structure in the following ways:

(a) AVAILABILITY

If the priorities for the investment projects to be undertaken by a public transport enterprise are determined according to net present value criteria which ignore regional development objectives, then there may be a tendency for investment resources to be diverted from regions with declining demand toward growing regions. At the same time the public transport enterprise which operates on these "commercial criteria" is likely to adopt a policy of closing down unremunerative lines in declining regions. The availability of transport services to users and shippers in these regions is therefore likely to diminish and this will be a factor contributing to their decline. The government may influence the availability of transport services in these regions by:

(i) insisting that public transport enterprises incorporate regional factors into their investment criteria; and

(ii) obliging these enterprises to provide a minimum level of service in these regions as a public duty.

(b) COST

The pricing policies adopted by a public transport enterprise will have a regional impact on the cost of transport. The regional implications of particular pricing policies will be examined in section 3:5.

(c) QUALITY

The quality of a transport service offered in terms of its ability to handle a large volume of goods, speed, ability to form networks, regularity, frequency of service, safety and comfort will vary according to the mode and route of transport. The affinity of the users of transport services for these different dimensions of quality will vary according to the objective properties of the goods to be transported and the subjective preferences of the transport users. It follows that the wider the modal choice the greater the range of users whose quality requirements will be satisfied. If a public transport enterprise adopts certain

> pricing policies it may only be able to remain financially viable if it is protected from intermodal competition. Modal choice can therefore be expanded and the overall quality of service improved if the public enterprise adopts pricing policies which do not require the restriction of intermodal competition.

For the purposes of this section an improvement in either availability, cost or quality is defined as an improvement in transport services. In predicting the regional impact of an improvement in transport it is important to consider whether it affects the transport system linking two regions or the transport system within a region. The effect of an interregional transport improvement on the development of different regions of a country can be examined with the use of a simple model based on the following assumptions:

(a) The same commodity is produced in the two regions of a country, regions 1 and 2;

(b) In region 1 the cost per unit of output, at the factory is C1 while in region 2 the cost is C2;

(c) The distance between the production centres in each region is M;

(d) The commodity is not produced at any site between the regions; and

(e) The transport costs are a constant t money units per kilometer in both directions.

The market area boundaries measured along the direct route between the production centres in each region, as illustrated in Figure 3:4 may be represented by M1 and M2 where

$$M1 + M2 = M \qquad (3:1)$$

The equilibrium values of M1 and M2 will be the distances from the region 1 and region 2 production centres at which production costs plus transport costs will be equal, given by:

$$C1 + tM1 = C2 + M2 \qquad (3:2)$$

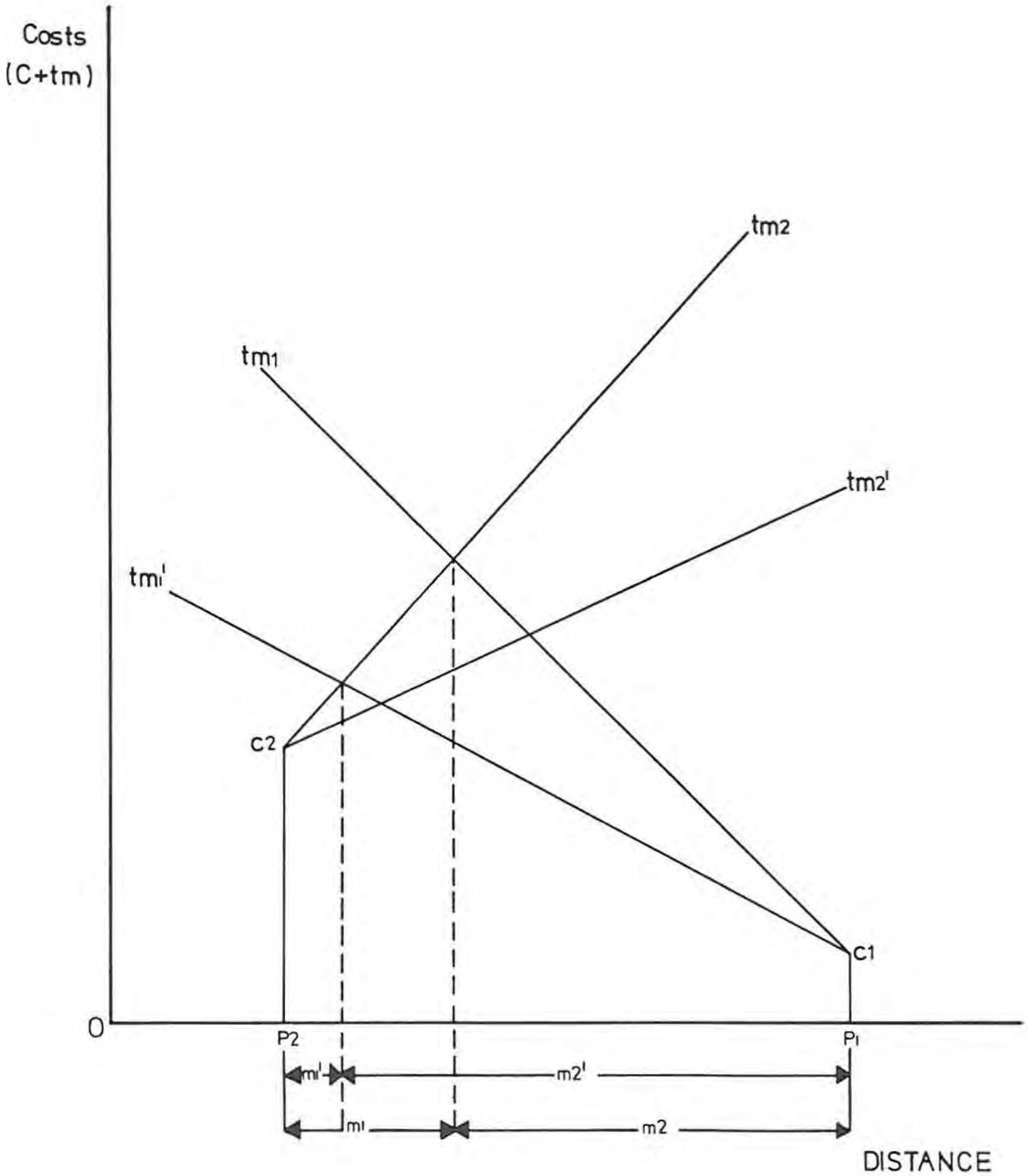


FIGURE 3:4 The Effect of an Interregional Improvement in Transport

If (M - M1) is substituted for M2, then the equation which can be derived for M1 is

$$M1 = \frac{1}{2} \left(M + \frac{C2 - C1}{t}\right) \qquad (3:3)$$

It follows that if production costs are lower in region 1 than in region 2 (C1 C2), then the market area M1 will increase when transport cost fall. This is shown in Figure 3:4 where the initial equilibrium market boundary is at point a where

$$C2 + tM2 = C1 + tM1$$

If the transport cost per kilometer falls by the same amount in both areas from t to t' then the new equilibrium market boundary will occur at b where C2 + t'M2 = c1 + t'M1. It can be seen that the market area for the product produced in region 1 expands from M1 to M'1 while that for region 2 contracts from M2 to M'2.

Generally, if there are two producing regions and if one region had relatively low manufacturing costs, then improving interregional transport would enable the low cost region to expand its market and in some cases, this could involve the complete elimination of production in the high cost region. This conclusion would lend support to Bonnafous'argument that an improvement in interregional transport between developed regions and underdeveloped regions which are " not in a positiion to bear being put in close communication with more advanced regions" could actually "aggravate inequalities of development" by exposing the underdeveloped region to more intense interregional competition and the draining of its resources.[4]

---

4. Bonnafous, A., "Underdeveloped regions and structural aspects of transport infrastructure", in Blonk, W., ed., Transport and regional development, Farnborough Saxon House, 1979, p. 49.

However, it is unlikely that the cost differences between two regions would be of the same order, or even in the same direction, for all forms of output so that transport improvements would normally encourage specialisation and increase interregional trade flows. It follows that in certain underdeveloped regions where there is an insufficient transport supply given its activities and prospects then an improvement in interregional transport will favour its development prospects by providing greater accessibility for its firms to the markets for products and inputs in more developed regions. An example of this would be a region with considerable tourism potential which has not been fully tapped due to poor transport service.

An improvement in intra-regional transport will usually assist in the development of a region by expanding the local marketing, procurement, employment and consumption opportunities of firms and households located at the designated growth poles of the underdeveloped region.

The analysis in this section has so far focussed upon the structure forming effects of an improvement in transport which results from changes in the pricing and investment policies of a public transport enterprise. However, as stated at the start of this section, there is a reciprocal relationship between transport and the spatial distribution of activities so that it is also necessary to consider the effects of this type of transport improvement where it is necessitated by the location of new investment at growth poles in underdeveloped areas by private firms, in response to regional policy incentives. If these firms are not to suffer from a long-term cost disadvantage relative to firms located in more developed regions with larger local markets and hence lower transport costs, then an improvement in transport is necessary and the government may therefore influence the pricing and investment policies of public transport services to improve the transport services in underdeveloped regions. Although an improvement in transport may not remove the private cost advantage of the region with the larger market, they would make it less important in relation to total costs and thus reduce the extra cost associated with production in the underdeveloped region.

It can be concluded that while an improvement in transport may not itself cause economic development to take place, when government action results in new industrial growth in an outlying region, then improved transport facilities must be provided and this consideration should be taken into account in formulating public enterprise pricing and investment policies.

The considerations discussed in this chapter would have to be taken into account by a public transport enterprise formulating its price and investment policy in any country. In the rest of this chapter, attention will be focussed upon the policies of the public transport enterprise in South Africa, applying the general framework of analysis developed so far in this and the previous chapter.

3:3 ADMINISTRATION OF TRANSPORT POLICY IN SOUTH AFRICA

The administration of transport policy in South Africa is vested at government level in the Minister of Transport. His portfolio comprises two separate and independent departments of state, namely the South African Transport Services (S.A.T.S.), which is essentially a state owned public utility transport undertaking, and the Department of Transport, whose primary functions are the regulation and coordination of various aspects relating to transport and the rendering of certain related services.

Amongst the principal functions of the Department of Transport are the provision and maintenance of an efficient system of national roads, the control of certain forms of motor transport defined by laws, including the regulation of commercial motor carrier road transport; motor vehicle insurance in so far as this concerns the protection of third parties against losses from motor vehicle accidents; the promotion and coordination of commercial civil aviation and air safety; the administration of the South African Merchant Shipping Act and the relevant regulations for ensuring marine safety and welfare; the provision of meteorological services; and the control of sea and beach pollution by oil. The Department of Transport also owns all the national and certain other major airports and provides the attendant air control services.

S.A.T.S. is a multimodal organisation and operates and provides a wide range of transport services on a closely integrated and coordinated basis. These services include a network of rail service covering the length and breadth of the country; its commercial harbours; an extensive road transport service; pipelines for the conveyance of crude oil, petroleum and related products; and the domestic overborder and international airline known as South African Airways.

The organisation is entirely state-owned, ownership being vested in the State President while its control and administration is exercised by the Minister of Transport who is assisted and advised by a board of three commissioners. The day to day management of SATS is entrusted to a director general who is assisted by two deputy general managers and eight assistant general managers. On account of the size of South African and by reason of the extensive nature of its rail service, the network is divided for local control purposes into ten separate geographic regions or systems, each under the control of a system manager.

The rail transport policy of South Africa is, by virtue of the economically strategic value of transportation, embodied in various Acts of Parliament, and in this respect Parliament plays an important role in determining that policy. The Republic of South Africa Constitution Act provides, inter alia, that "the railways, ports and harbours of the Republic shall be administered on business principles, due regard being had to agricultural and industrial development within the Republic and the promotion by means of cheap transport, of the settlement of an agricultural and industrial population in the inland portions of all provinces". In addition, the Act provides that, in so far as revenue is concerned, the organization's total earnings shall not be more than sufficient to meet the necessary outlays for working, maintenance, betterment, depreciation and contributions to its sinking fund and the payment of interest on capital. In other words, while it must operate on business principles it cannot do so with a profit incentive.

Other important provisions of the Constitution Act in regard to the transportation policy of S.A.T.S. are that no harbour or railway line for the conveyance of public traffic shall be built without the specific approval of Parliament, and furthermore that the estimates of expenditure on both capital projects and expenditure from the revenue derived from the operation of the various transport modes and related services shall be submitted annually to Parliament for scruting and approval.

The funds required for capital expenditure are, after appropriation by Parliament, in the main obtained from the South African Treasury by means of loans. These loans are, generally speaking, not redeemed and interest on such loan drawing is paid in perpetuity to the Treasury. To the extent that available Treasury funds may be insufficient, and also to afford greater flexibility in the financing of the capital development programme, direct access to the capital market is available. This access is subject to the approval to the Minister of Transport in consultation with the Minister of Finance. The repayment of such direct loans, as well as the interest thereon, is guaranteed by the Minister of Finance.

3:4 THE TARIFF POLICY OF S.A.T.S.

The tariff policy of S.A.T.S. is based upon the following principles:

(i) S.A.T.S. MUST BE MANAGED AND OPERATED ON BUSINESS PRINCIPLES

Since S.A.T.S. is administered as a Government business undertaking independent of the Central Government in so far as it concerns financial matters, the provision that it should be run according to business principles implies that it should balance its budget without recourse to the State Revenue Fund. Although it is usually not possible for S.A.T.S. to balance its budget exactly each year, it is accepted that an equilibrium between revenue and expenditure must be attained in the long term. To facilitate this, annual provision is made for either surpluses to be transferred into a Rates Equalisation Fund or for deficits to be financed from this Fund. Table 3:1 sets out the operating results of S.A.T.S. for the period 1977/8 to 1983/4. It can be seen that while the organization has been able transfer its excess earnings to the Rates Equalization Fund throughout this period, the financial position appears to have deteriorated and that it is estimated that this Fund might have to be used to finance substantial deficits during 1982/3 and 1983/4. Although the requirement that S.A.T.S. be managed on business principles implies that the organization as a whole should attain financial equilibrium, it does not preclude the cross-subsidization of losses incurred on certain services from the excess earnings generated by other services.

TABLE 3:1 OPERATING RESULTS OF THE S.A. TRANSPORT SERVICES 1977/78 TO 1983/84

Year	Surplus/(Deficit) R Million
1977/78	185,8
1978/79	274,1
1979/80	111,0
1980/81	196,9
1981/82	95,0
1982/83*	(370,2)
1983/84*	(631,7)

*Estimate

Source: South African Transport Services, Memorandum by the Minister of Transport Affairs setting out the estimated results of working for the financial year 1982/83 and anticipated revenue and expenditure for the year 1983/84 together with the latest traffic and other statistics, W.P.A - 83, Pretoria, Government Printer.

It will be shown that cross-subsidization takes place between the different-modes, services, routes and traffic flows served by S.A.T.S. Table 3:2 illustrates the cross-subsidization which has taken place between the different modes of transport provided by S.A.T.S.

TABLE 3:2 OPERATING RESULTS OF THE DIFFERENT TRANSPORT MODES PROVIDED BY THE S.A. TRANSPORT SERVICES, 1977/78 TO 1983/84

Mode	Operating Results: Surplus or (Deficit) R million						
	1977/78	1978/79	1979/80	1980/81	1981/82	1982/83	1983/84
Railways	26,5	45,0	(27,9)	(1,4)	(150,2)	(654,3)	(854,7)
Harbours	46,1	61,0	84,4	95,9	139,9	239,5	189,0
Airways	6,1	42,4	(54,2)	(39,6)	(57,9)	(99,2)	(120,0)
Pipelines	107,1	125,7	106,7	142,0	163,2	143,8	154,0
All Services	185,8	274,1	111,0	196,9	95,0	(370,2)	(631,7)

*Estimate

Source: South African Transport Services 1983, Memorandum by the Minister of Transport Affairs setting out the estimated results of working for the financial year 1982/3 and anticipated revenue and expenditure for the year 1983/84 together with the latest traffic and other Statistics , W.P.A - '83, Pretoria, Government Printer.

It can be seen that over the past five years the deficits incurred by the Railways and Airlines have been cross-subsidized out of the surpluses earned by the Harbours and Pipelines.

It fact this cross-subsidization is necessary if S.A.T.S. is to be independent of Treasury support and at the same time fulfill its socio-economic obligations in terms of the following principle.

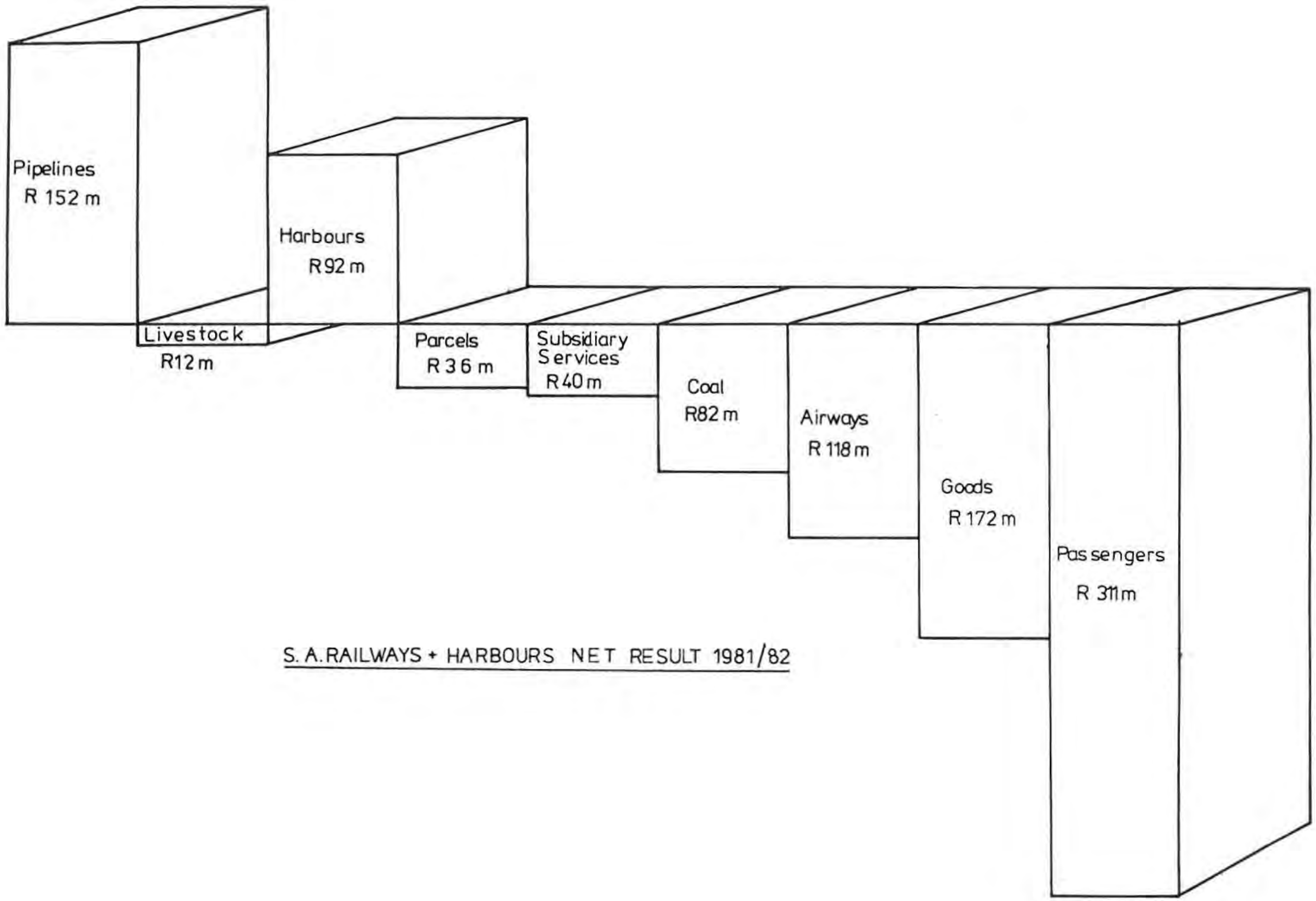

Pipelines
R 152 m
Harbours
R 92 m
Livestock
R12 m
Parcels
R 3 6 m
Subsidiary
Services
R 40 m
Coal
R82 m
Airways
R 118 m
Goods
R 172 m
Passengers
R 311 m
S. A. RAILWAYS + HARBOURS NET RESULT 1981/82

FIGURE 3:5

(ii) THE TARIFF POLICY OF S.A.T.S. HAS TO TAKE INTO CONSIDERATION THE ECONOMIC NEEDS AND TRANSPORT REQUIREMENTS OF THE REPUBLIC

In applying its tariff policy S.A.T.S. must not only seek to protect its financial viability but it must also ensure that its tariffs are determined in such a manner that they promote the economic development of the country. In certain instances this leads S.A.T.S. to set uneconomic tariffs for socio-economic services. These have been defined as services:
"for the promotion of the community and its needs;
for the development of the country;
for promotion of the community's standard of living;
that are politically orientated, and
that provide vital transportation links as cheaply as possible."[5]
In other words socio-economic services are provided for the benefit of the community at rates which do not cover the costs attributable to these services. The passenger services provided by S.A.T.S. probably bear the greatest burden of uneconomic social services. As a result an annual loss is usually incurred by the passenger division which is partly subsidized out of State Revenue and partly cross-subsidized out of the earnings of other services. Figure 3:5 shows that in 1981/82 a loss of R311 million was incurred on passenger services after taking into account the state subsidy of R287 million received in respect of these services. Although Figure 3:5 indicates that a loss was incurred by freight traffic, this was probably the result of the particularly poor performance in 1981/2 and it appears that in previous years surpluses on freight traffic were used, together with those

---

5. S.A.Railways and Harbours Board Report, R.P. 79/80, Pretoria, Government Printer, p. 25.

earned by the Harbours and Pipelines, to cross-subsidize the passenger losses.[6] The provision of uneconomic social services therefore tends to make cross-subsidization between services necessary.

(iii) THE "COLLECTIVE PRINCIPLE"

According to the "collective principle" the same rate per tonne/kilometre or passenger must be charged on traffic of the same class over the same distance and under the same conditions and circumstances notwithstanding the fact that different rates may be justifiable on similar traffic over different sections of lines due to say, differences in the density of traffic, empty running of trucks, standard of track, gradients and curves. This leads to cross-subsidization between the different routes served by the railway network. It is difficult to quantify the extent of this cross-subsidization since S.A.T.S. does not disclose costs and revenues by routes but it is clear that the failure by S.A.T.S. to reflect route cost differentials in its tariffs does cause such regional cross-subsidization to exist.

(iv) TARIFF DIFFERENTIATION

Tariff differentiation is an accepted commercial principle underlying the tariff policy of S.A.T.S. By differentiating its tariffs, S.A.T.S. seeks to improve the utilization of its

---

6. S.A.T.S. does not disclose in its annual reports the surpluses or deficits incurred by the separate divisions but independent studies tend to indicate that freight services were, as a whole, profitable in recent years.

facilities and at the same time generate revenue to cover total costs and the losses incurred by uneconomic social services. The Transport Services Act of 1981 restricted the application of tariff differentiation by prescribing discrimination in tariffs between persons using the same service under like conditions but allowed preferential rates to be granted subject to certain conditions to achieve certain goals.

Rate differentiation has been applied by S.A.T.S. according to the value of service principle or "charging what the traffic can bear". This principle has been interpreted in the following way:
".... the Railway Administration has been careful to avoid charging a higher rate on low grade traffic than such traffic could afford to pay. In other words, this principle has been a reductive factor-not a factor for increasing tariffs. The principle can thus be more accurately described as 'not charging more than the traffic will bear' which is really a benevolent means of 'tempering the wind to the shorn lambs'."[7]
The Schumann Committee found that this principle was complementary to the principle that rates should be set according to the cost of service since the maximum rate which may be levied on a particular commodity should be based on "what the traffic can bear" while the minimum rate should at least cover the costs directly attributable to the particular service. Railway traffic mainly consists of freight and passengers. It is now necessary to consider how these tariff making principles are applied to freight and passengers respectively.

3:4:1 FREIGHT TRAFFIC

(a) TRAFFIC VOLUME AND COMPOSITION

The dominant importance of freight traffic in the activities of S.A.T.S. is indicated in Table 3:3 which sets out the

---

7. R.S.A., Report of the Committee on Railway Rating Policy and Location in South Africa, Johannesburg, April 1964. In the text this will be referred as the Schumann Report.

TABLE 3:3 RAILWAYS REVENUE 1977/78 TO 1981/82

	1977/78		1978/79		1979/80		1980/81		1981/82	
Transportation Services	Rm	%	Rm	%	Rm	%	Rm	%	Rm	%
Passengers	181,3	9,4	195,5	8,9	334,1	13,3	437,5	14,5	530,6	14,7
Parcels	47,8	2,5	49,2	2,2	49,2	2,0	63,1	2,2	79,0	2,2
Mails	9,0	0,5	10,1	0,5	10,2	0,4	13,1	0,4	17,1	0,5
Goods	1379,0	71,7	1576,1	71,8	1673,9	66,4	1939,8	64,3	2299,3	63,7
Coal	199,1	10,3	241,3	1,0	295,1	11,7	361,1	12,1	450,6	12,5
Livestock	12,0	0,6	11,6	0,5	12,6	0,5	13,0	0,4	14,3	0,4
Rents & Storage	21,3	1,1	22,2	1,0	24,7	1,0	28,2	0,9	34,8	1,0
Miscellaneous	75,1	3,9	88,0	4,0	1188,8	4,7	156,7	5,2	123,6	3,4
TOTAL	1924,6	100,0	2194,0	100,0	2519,3	100,0	3017,5	100,0	3612,2	100,0

Source: Annual Reports S.A. Railway and Harbours 1977/78 to 1979/80. S.A.T.S. 1980/81 to 1981/82 contributions to revenue by the various services provided by the Railway division of S.A.T.S. during the period 1977/78 to 1981/82.

It can be seen from Table 3:3 that the contribution of goods traffic to total revenue ranged from 71,8 per cent in 1978/79 to 63,7 per cent in 1981/82. Although it may appear that the contribution of goods traffic has declired significantly since 1978/79 this can largely be attributed to the fact that substantial increases in State compensation to passenger services have been included in total revenue and that if one excludes these amounts then the contribution of freight traffic during this period can actually be represented by the percentages shown in Table 3:4

TABLE 3:4 CONTRIBUTION OF FREIGHT TRAFFIC TO REVENUE COLLECTED FROM TARIFFS: 1977/78 TO 1981/82

Year	Goods %	Goods, Livestock and Coal %
1977/78	73,3	84,5
1978/79	73,7	85,5
1979/80	71,3	84,4
1980/81	69,9	83,6
1981/82	69,2	83,1

The average contribution of goods traffic during this period has thus been 71,5 per cent while that of goods, coal and livestock has been 84,2 per cent. This relatively dominating position of goods traffic is of particular significance in the financial structure of S.A.T.S.

The total volume of freight conveyed by rail has exhibited a fairly rapid rate of growth. The Schumann Committee reported that between 1910 and 1963 the tonnage of freight transported by the Railways had grown at an average annual

TABLE 3:5 S.A.T.S. GOODS TRAFFIC 1977/78 TO 1981/82

	1977/78		1978/79		1979/80		1980/81		1981/82	
	Ton Millions	%	Ton Millions	%	Ton Millions	%	Ton Millions	%	Ton Millions	%
Live Stock	0,4	0,3	0,4	0,2	0,4	0,2	0,4	0,2	0,3	0,2
Live Animals & Animal Products	0,6	0,4	0,6	0,4	0,6	0,3	0,5	0,5	0,4	0,2
Vegetable Products	18,5	11,9	19,5	11,9	17,7	9,7	18,2	9,9	19,1	10,1
Animals & Vegetable Fats & Oils	0,	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
Timber and Paper Products	7,2	4,6	7,1	4,4	7,7	4,2	8,0	4,3	7,3	3,9
Total Agriculture	26,9	17,3	27,7	17,0	26,5	14,5	27,2	14,7	27,2	14,5
Mineral Products	56,7	36,5	57,6	35,3	66,7	36,4	61,6	33,4	62,0	32,9
Base Metals	9,0	5,8	10,5	6,4	12,2	6,7	11,7	6,3	11,2	6,0
Coal traffic : revenue earning	39,5	25,4	43,5	26,7	52,7	28,8	57,4	31,1	60,8	32,3
Locomotive coal	2,2	1,4	2,0	1,2	1,7	1,0	1,7	0,9	1,5	0,8
Total Mining	107,4	69,1	113,6	69,7	133,3	72,7	132,4	71,7	135,5	72,0
Prepared Foodstuffs:										
Chemical Products										
Articles of store, cement and asbestas Ceramic products, glass & glassware	0,6	0,4	0,6	0,4	0,7	0,4	0,9	0,5	0,8	0,4
Machinery & mechanical appliances, electrical equipment & Components	0,6	0,4	0,4	0,2	0,4	0,2	0,5	0,3	0,5	0,3
Vehicles, aircraft and spares, transport equipment	0,5	0,3	0,5	0,3	0,5	0,3	0,5	0,3	0,5	0,3
Textile and footwear	0,8	0,5	0,8	0,5	0,8	0,4	0,8	0,4	0,8	0,4
Container traffic	0,3	0,2	1,4	0,9	1,9	1,0	2,2	1,2	2,5	1,3
S.A.T.S. Stores	6,0	3,9	6,4	3,9	6,7	3,6	6,7	3,6	6,7	3,6
Miscellaneous	0,5	0,3	0,5	0,3	0,6	0,3	0,7	0,4	0,5	0,3
Grand Total	155,3	100,0	163,0	100,0	183,3	100,0	184,5	100,0	188,2	100,0

rate of increase of 4,2 per cent from 10 million tonnes to 93,5 million tonnes. During the period 1971/2 to 1980/1 the total freight conveyed by rail increased at average annual rate of 6 per cent. Figure 3:6 indicates this upward trend in freight tonnage.

The average length of haul of goods traffic has exhibited an upward trend in the long term. During the period 1911 to 1928 the average length of haul increased relatively rapidly from 286 kilometres to 403 kilometres. Thereafter there was no upward or downward tendency for about twenty years, but after 1948 the average length of haul again increased reaching 456 kilometres in 1962. The Schumann Committee attributed this upward trend to the following factors:

"(i) private road transport and the introduction and expansion of exempted and cartage areas, as a result of which the Railways has been deprived of a large quantity of short -distance traffic;

(ii) electrification of lines near coal fields, which has resulted in falling away of short distance coal traffic;

(iii) the mining of export minerals in the interior and consequent long distance transport thereof to ports;

(iv) the increase in the export of agricultural products, especially maize, resulting in more long-distance traffic to the ports"[8]

Since 1962-63 the average length of haul has continued to increase reaching 552 kilometres in 1981/82. Although all four of the factors mentioned by the Schumann Committee have probably continued to underly this upward trend, the most important has probably been (i) since intermodal competition for short-distance traffic has become more intense. It is interesting to note that since the expansion of the exempted areas in which roadhauliers

---

8. Ibid., p. 37.

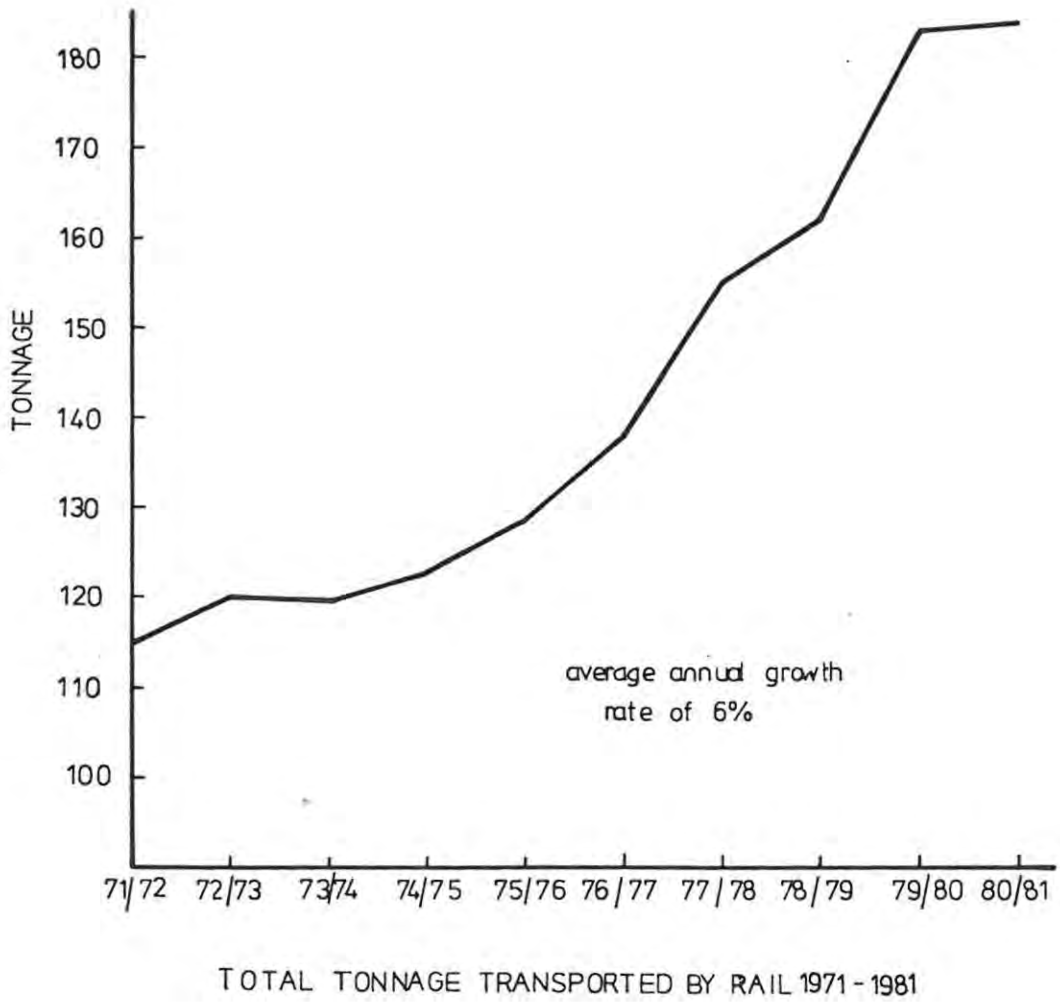


FIGURE 3:6 S.A.T.S. Total Tonnage Transported by Rail, 1971/72 to 1980/81

can freely compete with S.A.T.S. in 1977, the average length of haul has increased from 535 kilometres to 552 kilometres. Although the long term upward trend in the average length of haul has probably resulted in an increase in revenue per ton, it has also probably had a detrimental effect on the financial performance of the Railways since the cost of conveyance per ton usually rises at a greater rate with increasing distance than revenue per ton.

It can be seen from Table 3:5 that the bulk of goods traffic consists of goods produced in the primary industries of agriculture and mining. During the period 1977/78 to 1981/82 the share of total tonnage contributed by agriculture and mining averaged 86,6 per cent. The importance of mining appears to have grown relative to agriculture during this period with the share of agricultural goods falling from 17,3 to 14,5 per cent and the share of mining goods rising from 69,1 to 72,0 per cent. This can be attributed to the development of the Richards Bay Line and the increase in coal export traffic by 128 per cent from 13,4 million tons in 1977/78 to 30,6 million tons in 1981/82. Since 1911 the following trends in traffic composition can be discerned:

(i) the importance of the agricultural sector grew with the application of more modern farming techniques, with its contribution to total tonnage rising from 15,4 per cent in 1911 to 25,9 per cent in 1962-63 but appears to have declined in recent years as indicated by the fall in its contribution to 14,3 per cent in 1981/82;

(ii) the transport of mineral products grew from 1 ½ per cent in 1911 to 16 per cent in 1962-63 to 32,9 per cent in 1981/82 due to the rapid growth in the mining and export of base minerals;

(iii) the transport of coal declined in importance from 59,2 per cent in 1911 to 33,4 per cent in 1962-63 while in 1981-82, local shipments of coal contributed only 16,8 per cent of the total tonnage conveyed by the Railways. This downward trend can be attributed to the increasing utilization of alternative sources of fuel in the domestic economy of which the electrification of railway lines is but an example. The substantial increases in export shipments of coal has to some extent contributed to a reversal of this downward trend; and

(iv) the contribution of merchandise to rail traffic has not increased over the last seventy years despite the rapid expansion of domestic trade and industrial production. This is probably due to the concentration of industrial activity in areas where road transport plays an important role.

The physical tonnage of goods offered for shipment mainly affects the operating requirements of the Railways while the financial implications of traffic volume and composition can only be ascertained after the structure of rates for freight traffic have been considered.

(b) TARIFF POLICY WITH RESPECT TO FREIGHT TRAFFIC

In the case of freight traffic, the application of a policy of rate differentiation involves three steps:

(i) the specification of a schedule of tariff scales setting out the relationship between the rates charged on fifteen different tariff classes;

(ii) the classification of commodities into the tariff scale; and

(iii) the determination of special rates to take account of the special circumstances of particular commodities which cannot be classified into the tariff scale or of particular groups of users who require special treatment in terms of national development policy.

The basic structure of rates is embodied in the tariff scale which comprises fifteen tariff classes consisting of two groups; that is tariffs 1 to 10 for high-rated traffic, and tariffs 11 to 15 for low rated traffic. The relationship between these tariff classes is shown by means of the index in Table 3:6. When compiling the basic tariff scales, considerations regarding particular commodities or regions have no bearing at all because such factors are taken into account during the classification of commodities or the determination of special rates. The basic tariff scales, therefore, are not designed to suit a particular commodity, but to serve as a basic framework for the tariff structure. The railways generally accept the view that the interrelationship between tariff scales must be maintained when rate adaptations are necessary, so that the basic structure will not in the course of time become distorted.

TABLE 3:6 INDEX OF TARIFF SCALES AT 225 KM IN 1980-81

TARIFF	INDEX
1	100
2	86
3	64
4	57
5	50
6	43
7	32
8	32
9	25
10	21
11	14
12	14
13	11
14	11
15	11

Table 3:6 indicates that, while the gaps between tariffs are not very large, there is an appreciable difference between the highest and lowest tariffs. It has long been contended that the degree of rate differentiation embodied in the tariff structure is excessive in magnitude. In 1960, Verburgh[9] noted that "... the wide margin between rates in the highest and rates in the lowest class of the South African railway tariff scales is clearly illustrated .... where a comparison is made with a number of European countries. In South Africa, a number of the lowest railway rates, such as the rates on lime and sand, are only one tenth of the highest rates, like those on furniture and several other high-rated manufactured articles. By contrast the lowest rates on the railways of ... European countries ... are about one third of the highest rates". The Schumann Report also drew attention to the fact that rate differentiation by the South African Railways was excessive and recommended that the rates on high tariff classes should be lowered while the rates on low tariff classes should be raised. If the rate differentiation applied by the railways is considered excessive, this implies that an element of cross-subsidization has become entrenched in the tariff structure in the sense that the rates charged on high-rated goods are kept artificially high to generate a surplus necessary to subsidize the uneconomic rates on low-rated goods which fail to cover their direct costs. S.A.T.S. appears to have accepted the contention that its rate discrimination is excessive since it has adopted a policy of increasing the rates on low-rated goods by a greater percentage than on

---

9. Verburgh, C., South African Transportation Policy, Bureau for Economic Research, University of Stellenbosch, 1960, p. 59.

high-rated goods on every occasion when there is a general upward adjustment of rates. This is reflected in Table 3:7 which compares the maximum rate increases for high-rated and low-rated goods, for each year in which there was an upward adjustment of rates during the period 1970-1980.

TABLE 3:7 COMPARISON OF RATE INCREASES FOR HIGH - AND LOW - RATED GOODS: 1970 - 1980

	HIGH - RATED	LOW - RATED
1970 - 1	-	-
1971 - 2	-	-
1972 - 3	16,8%	26,1%
1973 - 4	-	-
1974 - 5	5,3%	21,2%
1975 - 6	-	-
1976 - 7	17,2%	19,8%
1977 - 8	-	-
1978 - 9	11,4%	12%
1979 - 80	-	-

Source: Annual Reports, S.A. Railways and Harbours, 1979-80.

The tariff scales are also characterized by a tapering of rates since the average charge per kilometre becomes less as distance increases. The degree of taper is not based solely and exclusively on the taper in unit cost but also to a considerable extent on the principle of "what the traffic can bear". There is thus a greater degree of taper in respect of the lower tariffs as compared with the higher tariffs, as well as in respect of longer distances compared with shorter distance.

The actual freight rate which is levied on a particular commodity is finally determined by the tariff class in which it is classified. The factors which are considered important by the railways in framing the goods classification can be grouped into those demand factors which determine the highest class in which a particular commodity should be classified and into cost factors which determine the lowest class into which it should be classified:

(i) DEMAND FACTORS

There are a number of criteria of the ability of a particular article to bear a certain rate, but according to the Schumann Report the Railways generally considers the following two to be the most important:

(a) VALUE

The value of a commodity in relation to its weight clearly affects its capacity to bear a particular railway rate. This is because the transport cost per unit of weight as a percentage of value becomes less as value increases. For example, the transport cost of clothing, even at a high rate, is insignificant in relation to the value of the clothing, whereas the cost of a long distance haul of coal even at a low rate may as much as the selling price at the colliery. It follows that commodities with a high value can usually bear a higher rate than those with a low value.

(b) SIMILARITY

In practice, commodities which are similar in nature, and with respect to their use and relationship with other commodities, are usually grouped in a common tariff class. This is considered to be fair by the Railways and obviates the problem of having to deal with continuous representations by consignors who feel they are being unfairly discriminated against since their commodity is classified in a different tariff class to a closely related article.

(ii) COST FACTORS

There are a number of factors which affect the direct costs which must be covered by the rate charged in the particular tariff class into which a commodity is classified. The following factors are usually considered important by the Railways in determining the lowest class into which a commodity should be classified:

(a) RISK

Where an article is liable to break or become damaged or lost easily, there will be additional expenses involved in its transport such as packing and insurance which should be taken into account by the Railways in determining its minimum tariff class.

(b) LOADABILITY

The actual weight of a commodity which can be loaded into a truck will be affected by factors such as its weight in relation to its size, its ease of handling and its shape. If an article is so voluminous or awkwardly shaped that, say, less than 2 tonnes of it occupies all the space of a 10 tonne railway truck then it will clearly have a higher transport cost per unit than an article which utilizes the space as well as the weight capacity of the truck. It follows that the more loadable commodity should be classified on a higher tariff scale than the less loadable commodity.

(c) HANDLING

In determining the correct tariff class for a commodity, account should be taken of the extent of the physical handling of the commodity by the railways since this will clearly increase the cost involved in its despatch.

(d) TRAFFIC VOLUME

The transport cost of a particular commodity will clearly be affected by the volume, frequency and regularity of its despatch and these factors should be taken into account by the Railways in classifying the commodity.

Apart from the compilation of different tariff scales and the grouping of all commodities into different tariff classes there are also special rating arrangements which differ from the usual pattern and which have arisen in special circumstances. The rates charged in these special instances appear to be based on cost reflecting the gradual trend in railway tariff policy from value of service to cost principles.

The granting of cost based export rates is an accepted policy of S.A.T.S. Although the Railways has generally charged lower tariffs on exported commodities since the time of the Union, the application of the cost principle has been followed since 1971 when the Reynders Commission recommended that this principle should be the dominant factor in determining rail tariffs for exports. Based on this principle, it is generally possible to bring about a significant reduction in rates for high-valued commodities railed for export. However, it has also resulted in the increase of certain uneconomic rates for low-valued commodities. S.A.T.S. will only grant a cost based export rate to shippers if they can prove the normal rail rate from the point of dispatch to the nearest port is the factor militating against the successful marketing of the product at a competitive price abroad.

S.A.T.S. has introduced contract rates on a cost basis in instances where the traffic to be conveyed is offered in such quantities that special traffic arrangements have to be introduced to use optimally the mobile plant. The reduction in cost which is attained through these special traffic arrangements is then passed on to the shipper by a rates reduction which could not have been made if the traffic were conveyed under normal conditions of carriage. Since it departs from these normal conditions of carriage S.A.T.S. has to enter into a contract with the shipper whereby the shipper can only be charged at contract rates if he offers for conveyance a tonnage greater then a specified minimum tonnage per annum.

S.A.T.S. has also introduced block-load rating arrangements to pass on to shippers the cost-savings it incurs when traffic is conveyed in block-train loads comprising 39 vacuum-braked or 50 air-braked bogie trucks from one point to another without S.A.T.S. being involved in the marshalling of the train.

With the introduction of large scale containerization on 1 July 1977 S.A.T.S. deemed it necessary to provide for unit container rates based on the size of the container irrespective of its contents or mass since only the size of the container would affect the cost of conveyance. Although these rates were initially only applied to traffic shipped between City Deep and Durban they have since been extended to other major harbours of South Africa.

There also exists a system of rebates in respect of outgoing traffic from industrial development points which is designed to compensate for the longer term disadvantages in terms of transport costs which are associated with locating at these points. In recent proposals for a regional development strategy it was proposed that:

> "The existing system of rail rebates in respect of outgoing traffic will be retained and expanded in the case of industrial development points. All existing industries in identified industrial development points

will be eligible for these rebates."[10]

It should be noted, though, that these rebates are paid by the State to shippers so that the Railways is not required to depart from commercial principles by being obliged to apply these rebates in its own tariffs.

(c) THE FINANCIAL PERFORMANCE OF RAILWAY FREIGHT SERVICES

The practical result of the application of a differential rating system is that the contributions which high and low-rated traffic respectively make to total freight revenue will not be in the same proportion as the contribution which they make to total freight tonnage. This is indicated in Figure 3:7 where it can be seen that high-rated traffic makes a far greater contribution to revenue than to traffic volume. The reason for this is that agricultural and mining products which make up the bulk of railway freight traffic are classified at low tariffs so that they do not make the same impact on the revenue generated by the railways.

Figure 3:7 also shows that the share of high-rated traffic has declined significantly during the last ten years both in terms of traffic revenue and volume. This can be attributed to the increasing intensity of intermodal competition which mainly affects high-rated traffic since the Railways tends to

---

10. Department of Foreign Affairs and Information, The Promotion of Industrial Development: An Element of a Coordinated Regional Development Strategy for Southern Africa S.A. Digest, April 2 1982.

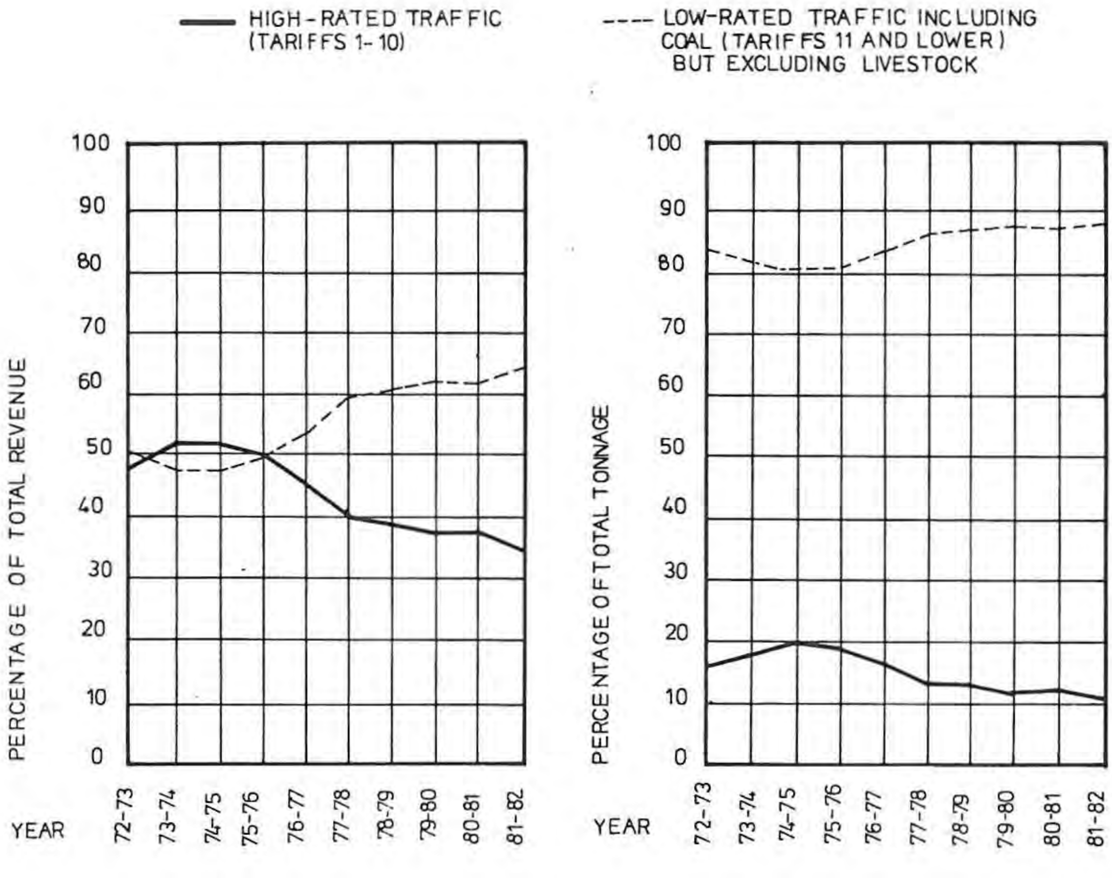


FIGURE 3:7 Share of Total Freight Transport Market Supplied by Different Modes

have a comparative advantage in the long distance bulk transport of the products which are classified at low tariffs. The declining market share of rail transport in the face of competition from own and road transport for freight is reflected in Figure 3:8. It can be seen that during the ten years between 1971/72 and 1980/81 the share of rail transport fell from 61 per cent to 51 per cent and that it is expected to fall to 40 per cent in 1987/88. This decline may be intensified if the process of deregulation continues.

In the annual reports of S.A.T.S. no distinction is made among the costs related to the various classes of goods traffic. It is thus not possible to obtain from the information published by S.A.T.S. any indication of profitability by tariff class. In an independent study Kennedy[11] used the internal costing information of the Railways to calculate the profitability of each tariff class. He found that in 1981/82 surpluses on high-rated commodities, container traffic, and traffic carried at contract and block-load rates cross-subsidized the losses incurred on low-rated commodities, ore and livestock, so that only a small overall surplus was earned by the freight division of S.A.T.S. It is also interesting to note that Kennedy, found that for commodity classes 1 to 10, a large deficit was actually incurred for commodity class number 1 which was offset by the surpluses earned by other high rated traffic. It can thus be seen that

---

11. Kennedy, T.L., S.A.T.S.: Profitability by Tariff Class, unpublished memorandum, National Institute of Transport and Road Research, C.S.I.R., Pretoria, 1982. The actual amounts of the deficits and surpluses for each tariff class calculated by Kennedy are confidential.

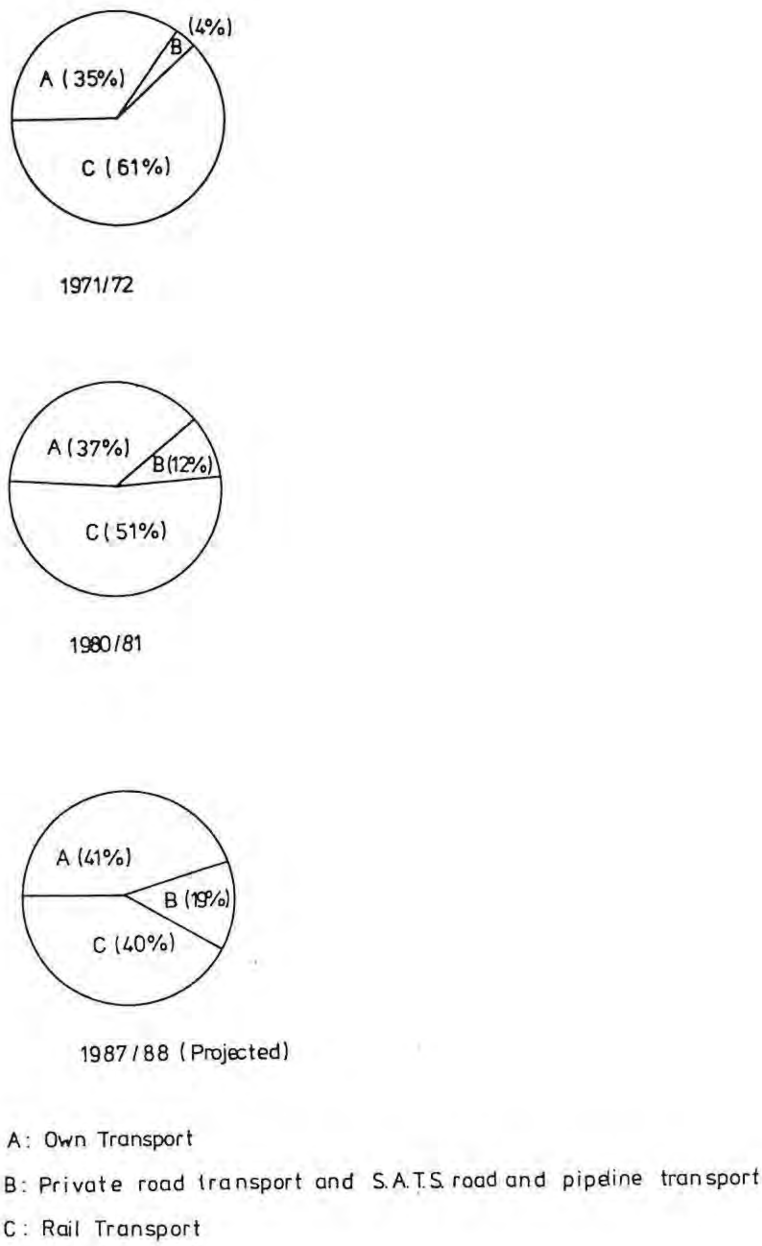


FIGURE 3:8 Market Share of Rail Transport, 1972/73 to 1981/82

the differential rating system applied by the Railways actually results in a significant amount of cross-subsidization between commodities.

3:4:2 PASSENGER TRAFFIC

(a) TRENDS IN TRAFFIC VOLUME AND COMPOSITION

Passenger traffic consists mainly of passengers, parcels and mail, the former being of predominant importance. Table 3:8 indicates the percentage of the total revenue of the railway portion of the undertaking which was generated from these sources during the period 1977/78 to 1981/82.

TABLE 3:8 PERCENTAGE OF TOTAL RAILWAY REVENUE CONTRIBUTED BY PASSENGER TRAFFIC

Year	(1) Percentage	(2) Percentage
1977/78	10,1	12,3
1978/79	9,1	11,6
1979/80	8,8	15,6
1980/81	9,0	17,0
1981/82	9,4	17,3

Source: S.A.T.S. Annual Report

Column (1) in Table 3:8 represents the percentage contribution of passenger traffic revenue before taking into account the compensation received by S.A.T.S. from the State for operating uneconomic social passenger services while column (2) represents the percentage contribution after this amount has been included in passenger traffic revenue. The Schumann Committee[12]

12. Republic of South Africa, op.cit., p. 43.

reported that in 1962-63 the percentage contribution of passenger traffic to total revenue was about 14 per cent compared to more than 25 per cent at the time of the Union and the peak figure of 31 per cent in 1919-20. Although column (2) of Table 3:8 might indicate that this declining trend in the contribution of passenger revenue has been reversed in recent years, a comparison between column (1) and column (2) shows that this can be mainly attributed to increases in the compensation received from the State and that the actual tariff revenue received from passengers, parcels and mail has actually averaged 9,3 per cent of total revenue during the period 1977/78 to 1981/82. The long term declining trend in the relative importance of passenger traffic is fundamentally due to the growing intensity of competition from other modes of transport. Nevertheless the total number of passenger journeys has increased from 28,2 million in 1910 to 339,7 million in 1962-63 and 753 million in 1981/82.

At the time of Union, first and second class traffic accounted for approximately two thirds of the total number of passenger journeys. Since then the scene has changed considerably with third class traffic constituting 83 per cent of total passenger traffic in 1981/82. This change can be ascribed to the increasing urbanization of Blacks which has led them to rely increasingly on rail transport whereas Whites have tended to rely more and more on road transport.

Passenger traffic is categorized into two main types for tariff making purposes: suburban traffic and main-line or long-distance traffic. Suburban passenger traffic has shown a rising trend with the total number of journeys increasing from just under 42 million in 1918-1919 to 318 million in 1962-63 and 707 million in 1981-82. The Schumann Committee reported that while all classes of suburban traffic increased during the years prior to the Second World War, White suburban traffic remained virtually static in volume in the first twenty years after the war in spite of population growth and low passenger fares. This static condition of White suburban traffic appears to have persisted during the last twenty years with about 110 million journeys being undertaken in 1962-63 and 123 million in 1981/82. This trend can be almost entirely attributed to the increased utilization of motor transport. The relatively rapid increase in suburban third class traffic resulted in an increase of its contribution to the total number of suburban journeys from just under 30 per cent in 1918-19 to 66 per cent in 1962-3 to 78 per cent in 1981-1982. The growth in black suburban traffic can be attributed to the rapid rate of growth in the urban black population and their location in townships outside the large urban areas which has compelled them to rely predominantly on train services in travelling to and from work.

The share of total passenger traffic constituted by main line traffic has shown a long term declining trend falling from 18,7 per cent in 1918-19 to 6,0 per cent in 1981-82. After reaching a peak of 17,2 million in 1945 the number of first and second class main-line journeys has actually decreased to 3,2 million in 1981-82. On the other hand, there has been an upward trend in the number of third class main-line journeys with about 3,7 million journeys being undertaken in 1918-19, 17 million in 1948-49 and 42,3 million in 1981-82.

(b) THE STRUCTURE OF PASSENGER FARES

A different structure of fares is set for main-line and suburban traffic. There are three different journeys in respect of main-line journeys, the fares being in inverse proportion to the number of passengers per compartment. As first, second and third class compartments provide accommodation for four, six and eight passengers respectively, the second class fare is equal to 4/6 or 2/3 of the first class and third class 4/8 or 1/2 of first class. There is also a lower scale applicable to journeys by Blacks on special trains between Black residential areas and, especially, the mining areas. In 1981/82 only R0,4 million was generated from the fares charged on those special trains. The same rate schedule is charged for single and return journeys, although it is based on a tapering scale so that for long distances the return fare is considerably less than double the single fare. The revenue generated from the fares charged to main-line passengers is significantly reduced by the fact that railway staff, government officials, public officers and others are entitled to free or privileged tickets.

Suburban fares are set at rates which are significantly lower than those charged for main-line traffic. This arises from the density of suburban traffic and the use of special cheaper types of passenger coaches which provide more accomodation per coach than those used for main-line traffic. The difference between single and return fares is however, far less accentuated for suburban passengers than for main-line passengers. Only two classes of accomodation are now provided for suburban traffic. The standard rate differential between these two classes is less than that which exists for main-line passengers. However, a substantial part of third class traffic consists of black passengers being transported on the lines constructed between resettlement schemes and the main urban areas. The authorities have taken the view that these black passengers could not afford to pay rail fares at a rate which would produce sufficient revenue to enable the Railways to operate these passenger services on remunerative basis. Consequently very low fares are charged and the State undertakes to compensate S.A.T.S. for operating these unremunerative services. The amount of compensation received by S.A.T.S. has risen dramatically by 571 per cent from R42,8 million in 1977/78 to R287,3 million in 1981/82. In addition, S.A.T.S. offers reduced rates on monthly season tickets for the benefit of passengers of both classes who use the services on a regular basis.

The charges for parcels and excess luggage conveyed by passenger train are based on the principle of degressive rates, that is, the freight charge per kilogram decreases as the mass of the article increases. As may be expected the rates are considerably greater than is the case of traffic carried by goods train, especially where small parcels are concerned.

(c) FINANCIAL PERFORMANCE OF THE PASSENGER SECTION

Table 3:9 indicates the amount and composition of the revenue generated by passenger traffic during the period 1977/78 to 1981/82.

TABLE 3:9 REVENUE GENERATED BY PASSENGER TRAFFIC

	1977/78		1978/79		1979/80		1980/81		1981/82	
Main Line	Rm	%	Rm	%	Rm	%	Rm	%	Rm	%
First Class	4,7	2,6	4,6	2,4	6,0	1,8	7,5	1,7	9,6	1,8
Second Class	14,5	8,0	14,6	7,5	17,3	5,2	20,6	4,7	25,1	4,7
Third Class	61,5	33,9	60,3	30,8	69,6	20,8	84,4	19,3	106,2	20,0
Special Trains	0,2	0,001	0,1	0,0	-	-	0,3	0,0	0,2	0,0
TOTAL	80,9	44,6	79,5	40,7	92,9	27,8	112,8	25,8	141,3	26,6
Suburban										
First Class	19,4	10,7	19,4	9,9	23,2	6,9	28,1	6,4	33,1	6,2
Third Class	38,2	21,1	41,6	21,3	46,2	13,8	53,8	12,3	68,9	13,0
Compensation for socio-economic	42,8	23,6	55,4	28,3	171,9	51,4	242,8	55,5	287,3	54,1
TOTAL	100,4	55,4	116,4	59,2	241,3	72,2	324,7	74,2	389,3	73,4
GRAND TOTAL	181,3	100,0	195,9	100,0	334,2	100,0	437,5	100,0	530,6	100,0

Source: S.A.T.S. Annual Reports 1977/78 to 1981/82

It can be seen from Table 3:9 that the share of total passenger traffic revenue contributed by suburban traffic has increased significantly from 55,4 per cent in 1977/78 to 73,4 per cent in 1981/82. This can be mainly attributed to the substantial increases in State compensation for socio-economic services during this period. If the amount of this compensation is deducted from total passenger revenue, the percentage contribution of main-line and suburban traffic can be calculated as the percentages shown in Table 3:10

TABLE 3:10 PERCENTAGE CONTRIBUTION OF MAIN-LINE AND SUBURBAN REVENUE FROM FARES

	1977/78	1978/79	1979/80	1980/81	1981/82
Main-Line %	58,4	56,5	69,4	58,0	58,0
Suburban %	41,6	43,5	30,6	42,0	42,0
Total %	100,0	100,0	100,0	100,0	

Source: S.A.T.S. Annual Report 1977/78 to 1981/82

It can be seen that in 1981/82, although main-line traffic constituted but 5,8 per cent of the number of journeys it contributed 58 per cent of the revenue generated from passenger fares. This can be attributed to both the greater distances covered by main-line journeys and to the higher level of main-line fares relative to suburban fares.

The contribution to revenue made by third class traffic appears to have grown significantly over the years. The Schumann Committee reported that in 1918-19 this traffic contributed 28 per cent to total revenue from passenger train traffic. By 1962-63 this contribution has risen to almost 50 per cent while in 1981-82 third class main-line and suburban traffic contributed about 72 per cent of the revenue generated from passenger fare. If one considers that State compensation was mainly received in respect of third class traffic then the total contribution to total passenger traffic revenue from this traffic amounted to 87,3 per cent. Conversely, it can be seen that the percentage contribution of first and second class traffic has declined considerably, with these classes of traffic only contributions 12,7 per cent of total passenger revenue in 1981-82.

S.A.T.S. does not publish information concerning the allocation of expenditure to its different services so that it is not possible to obtain from its annual reports any indication of the loss incurred by its passenger services. The Schumann Committee reported that in 1962-63 a total loss of R44,3 million was incurred by passenger services. This amount appears to have grown substantially in the last twenty years. The Franzsen Committee[13] reported that in 1977/78 a total loss of R320 million was incurred and estimated that losses of R368 and R407 million would be incurred in 1978/79 and 1979/80. Heydenrych[14] has indicated that in 1981/82 the total loss of rail passenger services had grown to R598 million. The total loss on passenger services has thus grown by 1250 per cent in the last twenty years and by 87 per cent in the last five years. This loss is partly financed by State compensation and partly cross-subsidized out of the revenue derived from other services. The portion of the loss which has been subsidized out of State funds appears to have grown substantially in recent years. In 1977-78 State compensation amounted to R42,8 million which was 13,3 per cent of the total loss of R320 million. In 1981-82 the State contributed R287 million which was 48 per cent of the total loss. The burden of cross-subsidization passenger services places on other services has

---

13. Republic of South Africa, Report of the Committee of Enquiry into Socio-Economic Rail Passenger Services, Pretoria, May 1979.

14. Heydenrych, J.A. and Kennedy, T.L., Economic Efficiency of Rail Transport, Paper delivered at the Annual Transport Convention, Johannesburg, July 1983.

therefore not increased substantially rising from R277 million in 1977-78 to R311 million in 1981-82, an increase of only about 12 per cent in 5 years.

The Schumann Committee reported that in 1962-63 losses on passenger traffic were incurred by first and second class main-line and all classes of suburban traffic, with third class main-line traffic apparently breaking even. It is likely that third class main-line traffic has remained the least unremunerative class of passenger traffic since the share of main-line passenger revenue generated by this traffic has risen from 43 per cent in 1962-63 to 75 per cent in 1981-82. The bulk of the passenger services loss is therefore still likely to arise from the first and second class main-line and suburban traffic.

3:5 EVALUATION OF S.A.T.S. TARIFF POLICY

In this section the tariff policy of S.A.T.S. for both freight and passenger traffic will be evaluated in terms of both its effect on economic efficiency and the spatial distribution of economic activity.

3:5:1 FREIGHT TARIFF POLICY

3:5:1:1 THE SYSTEM OF RATE DIFFERENTIATION

The Schumann Committee levelled the criticism against the tariff structure for freight traffic that the range of rate differentiation was excessive. To examine the validity of this criticism, it is necessary to compare the existing tariff structure of S.A.T.S. with that which would result if tariffs were based on the optimal pricing principles discussed in the previous chapter. According to these principles the rates charged for different commodities should vary according to the marginal cost of transporting them. As will be explained in a following section, a marginal cost based tariff structure can be approximated if the rates for different commodities are set so that they at least cover the costs which could be avoided if the Railways refused to ship a particular commodity. If rates are based on this principle, then there will be some differentiation of rates since the avoidable cost of shipping different commodities will vary according to factors such as risk, loadability, handling and traffic volume. As was mentioned in the previous section, these factors are taken into account by the Railways in classifying commodities into the tariff scale so that to some extend the present tariff structure reflects "cost of service" principles. However, the extent of rate differentiation which would be reflected in a "cost-orientated" tariff strucutre would probably be much smaller than that which characterizes the existing tariff structure of the Railways. For example, Verburgh[15] pointed out that the ratio of the highest to lowest charge in a tariff structure where commodity rates are differentiated primarily according to loadability was 3 to 1 while the current ratio for the Railways is about 9 to 1. Furthermore the tapering of rates also appears to be applied to a greater degree than is necessary to reflect the decline in cost per unit with distance.

To some extent, further rate differentiation beyond that which would characterize a cost-based tariff structure is justifiable, where tariffs based on marginal cost do not generate sufficient revenue to cover the total costs of the railway undertaking. In section 2:2:1 of the previous chapter it was shown that price discrimination will be optimal under a binding financial constraint where the deviation of price from marginal cost is in inverse proportion to the elasticity of demand. The principle is to some extent, followed by the Railways since it takes the volume per unit of weight of a good into account in determining the highest tariff into which it should be classified. The elasticity of demand with respect to changes in railway rates of a commodity with a high value per unit of weight such as clothing is likely to be low since its transport cost per unit of weight is low and a change in railway rates will have little impact on the price at which clothing is sold. The Railways is thus justified in classifying this commodity into a high tariff scale.

---

15. Verburgh, *op.cit.*, p. 59

It is also justified in classifying in a lower tariff scale a commodity such as coal which has a high elasticity of demand with respect to changes in railway rates due to its low value per unit of weight.

The application of the value of service principle by the Railways becomes untenable however, for those classes of traffic for which the Railways is subjected to intermodal competition. For these commodities the limit on the rates which the Railways can charge is set by the cost and quality of the alternative transportation mode. The Railways does not appear to take this factor into account since the tariffs charged for high-rated commodities appear to be uncompetitive in relation to the cost and quality of the transportation services provided by road hauliers. The effect of this can be seen in the serious decline in the share of high-rated traffic carried by the Railways as shown in Figure 3:7. The extent of rate differentiation applied by the Railways is therefore probably excessive and there is a clear need to narrow the gap between the tariffs charged for high-and-low-rated commodities. Although the Railways appears to have recognized this problem and has made some progress towards narrowing the gap by raising the tariffs on low-rated commodities by a greater percentage than those on high-rated commodities every time the tariff structure is adjusted upwards, this had led to a particularly slow process of adjustment. The problem of resource misallocation resulting from excessive price discrimination has thus persisted for a considerable period of time so that the distorted railway rates structure has been incorporated into long term investment and location decisions.

As was explained in the previous section, the necessity for this excessive rate differential arises from the fact that the tariff revenue derived from conveying high-rated commodities is used to cross-subsidize the losses incurred on low-rated traffic and passenger services. In fact the Franzsen Committee estimated that in 1977/78 "the general tariff level on goods services is about

18 per cent higher because of cross-subsidization".[16] If the Railways accepted the principle that the tariffs set for each service should at least generate sufficient revenue to cover its avoidable cost and that if this does not occur, the quantitative and qualitative level of the service should be reduced, unless State compensation is received then the burden of cross-subsidization on its other services would be eased. It would then be possible to reduce the tariffs on high-rated commodities to levels which would make rail services more competitive with those provided by road operators. Furthermore it is also likely that this policy would result in an upward adjustment of the tariffs charged on low-rated commodities. To the extent that these tariffs have been artificially low, they have probably encouraged the excessive growth of this traffic and consequently caused an excessive investment in rail capacity to meet the requirements of this traffic. An upward adjustment of tariffs on low-rated commodities might thus result in some savings on the Railways investment programme. It would be unlikely that this tariff adjustment would meet with significant market resistance since the Railways has a comparative advantage in the bulk transport of the agricultural and mineral products which are classified into low tariff scales.

It can be concluded that the system of rate differentiation applied by the Railways is excessive but that the cause of this distortion is probably rooted in the conflicting demands on Railway tariff policy which underly the pattern of cross-subsidization. These conflicting demands will be examined in more detail in a section 3:6 . It is now necessary to consider in more detail the failure of the Railways to reflect route cost differentials in its rates.

---

16. Republic of South Africa <u>Report of the Committee of Enquiry into Socio-Economic Rail Passenger Services</u> , Government Printer, Pretoria, May 1979.

3:5:1:2 THE PROBLEM OF UNREMUNERATIVE ROUTES

According to the "collective principle" applied by the Railways the same rate per tone per kilometre is charged for a particular class of traffic regardless of the route on which it is carried. This ignores the fact that costs may vary considerably between different routes. As Verburgh has noted:

> "In flat country the cost of construction and the cost of exploitation is much lower than in mountainous country, while on routes with dense traffic and a high degree of utilization the unit costs are lower than on routes where a low degree of utilization is a normal long-run phenomenon."[17]

The effect of applying the "collective principle" is that the traffic on certain routes will fail to cover the costs directly attributable to the routes since either

(i) they handle a disproportionately high percentage of low-rated traffic and/or

(ii) the high unit costs associated with these routes are not reflected in the rates charged for the traffic carried along them.

It follows that the traffic on these routes is effectively being cross-subsidized by traffic on more remunerative routes. If an average cost pricing policy is adopted this will mean that the rates on unremunerative routes should be raised to reflect the higher costs incurred on these routes. Such a policy may, however, be suboptimal since where there is excess capacity on a line the opportunity cost of inducing extra traffic to use the line may be limited to the running costs of providing the rail service since capital costs will have already been sunk into the line and will be unaffected by any increase in its utilization.

---

17. Verburgh, C., op.cit., p. 73.

Hotelling[18] has argued that in this situation of excess capacity, the government will be able to secure an improvement in welfare by requiring the Railway to set its rates equal to the short run marginal cost incurred on the line and financing the resulting deficit through a state subsidy. However, as was shown in the last chapter, the short run marginal cost pricing rule should be accompanied by an investment policy directed at adjusting capacity to the optimal level where the price which restricts demand to capacity output generates sufficient revenue to cover total costs. Subsidization would thus be a temporary measure directed at enabling the public undertaking to pursue an efficient pricing policy while it takes steps to streamline its operations and reduce surplus capacity on unremunerative lines. A railway undertaking such as S.A.T.S., however, is often faced with the problem of indivisible facilities so that it may not able to simply reduce the capacity of its investment in railway track and structures, switching equipment and station buildings on its unremunerative routes to optimal levels but must instead choose between closing down the route or continuing to operate it at a loss. Such a choice should be based on a cost-benefit analysis based on a consideration of all the quantifiable and unquantifiable social costs and benefits including the effect on regional development of the alternative decisions. It can be concluded, though, that regardless of indivisibilities, the pursuit of an efficient pricing and investment policy by the Railways can only take place if it has the discretion to close down unremunerative routes where there is a net social benefit to be gained by doing so. It follows that the application of the "collective principle" by the Railways together with its public obligation to connect all parts of the country with each other by an integrated railway system will result in a cross-subsidization

---

18. Hotelling, H., "The General Welfare in Relation to Problems of Taxation and of Railway and of Utility Rates" in <u>Readings in Welfare Economics</u>, AEA Series, Vol., XII, Allen & Unwin, London.

of unremunerative routes which will not only result in the retention of a large amount of surplus capacity on unprofitable routes but also the restricted development of railway capacity on more profitable lines.

A further problem with this cross-subsidization of unremunerative routes is that the responsibility for inefficiency on these routes may be shifted from its source so that the pressure to contain X-inefficiency within the organization may be less than where each route is required to cover its direct costs.

Nove[19] has however, argued that public enterprises should be required to meet public obligations since they are not subject to the constraint often imposed by the competitive process on private enterprises whereby they are required to pursue activities which may appear unprofitable in themselves in order to foster consumer goodwill. This argument largely applies to circumstances where the Railways is in the position of a natural monopoly. It may be the case that routes which are unremunerative for the Railways can be served by road transport at the same or slightly higher tariffs but at a cost which is only a fraction of rail transport. In such cases it is clear that the Railways should be relieved of its public obligation to provide rail services on these routes.

3:5:1:3 THE EFFECT OF FREIGHT TARIFF POLICY ON INDUSTRIAL DECENTRALISATION

It is now necessary to consider the effect of the Railway's tariff policy upon industrial decentralization in South Africa. It is generally accepted that an adequate railway system exists in this country in the sense that most producers have reasonable access to a railhead. The pattern of regional development is thus in

---

19. Nove, A., Efficiency Criteria for Nationalised Industries, Allen & Unwin, London, 1973.

most cases not affected by the availability of rail services but may, in some cases, be affected by the structure of railway rates. Although it has been argued that the high rates charged on manufactured goods and the low rates charged on raw materials may have contributed to industrial concentration in the four main urban areas of South Africa since transport costs would have been minimized by location near the market rather than near the raw material source, it is now generally conceded that railway rates structure was not a basic factor in this regard. As the Schumann Committee commented:

> ".... the existing location pattern of secondary industries has grown out of the initial development of the harbour-towns and the mining of gold in the Witwatersrand area ...."[20].

In a study of the effect of railway rates upon industrial location in South Africa, the Schumann Committee reached the conclusion that ".... in the case of less than 5% of the number of all industrial undertakings in the Republic of South Africa, railway rates play an important role in respect of the place of location, that in the case of an even smaller percentage of the instances these rates play a deciding role, and that it is confined to the classes of industry producing the following goods: non-metallic mineral products, wood products, chemical products and the basic metal industries, and to a lesser extent also to foodstuffs and metal products."[21] The Railway's rating policy will thus be an importance influence on location decisions of manufacturers of raw materials which satisfy the inter-industry demand. However, even within this group of industries there will be industries where the value of the raw materials and manufactured products is very high in relation to transport charges so that they can be located almost anywhere. These have been termed "footloose" industries.

---

20. R.S.A., op.cit., p. 152.
21. Ibid., p. 80.

Examples of industries whose location may be affected by the Railway's policy of rate discrimination would be the wood-processing and ferro-alloy industries. With respect to both these industries the Schumann Committee found evidence that the wide gap between the tariff charged on manufactured goods and the tariff charged on raw materials has tended to discourage location near the raw material source. Thus the fact that the rate on manufactured timber products such as flooring blocks and paper is considerably higher than the rate on unprocessed timber may have been a factor in preventing large scale wood processing at the forest. Similarly the low rates charged on ores and limestone and the high rates on manufactured ferro-alloy products has been a factor in discouraging the decentralization of this industry to its raw material source. Any change in Railway tariff policy which results in a narrowing of the range of rate discrimination may thus, in certain cases, be compatible with the objective of industrial decentralization.

The effect on industrial decentralization of the high degree of taper which characterises the Railway rate structure depends on the particular circumstances of the industries affected. On the one hand a high degree of taper may expose a manufacturer in an underdeveloped region to competition from manufacturers in large industrial areas. Since the latter can generally take advantage of large scale economies it will be difficult for local manufacturers to compete with them and in this case the high degree of taper may actually discourage decentralization. On the other hand a high degree of taper may promote decentralization by expanding the marketing and procurement opportunities of firms established within the area since they can sell their products and procure their raw materials at competitive prices on markets at greater distances from their location than in the case where there is a lower degree of taper.

As was shown in section 3:2 it appears that in general, a reduction in the cost of transport will benefit both local firms established within the underdeveloped region and firms located in industrial areas who sell their products and obtain their raw materials in the region. The net effect of the reduction in the cost of transport on industrial decentralization is thus difficult to determine although it can generally be said that it will encourage greater regional specialization and foster the development of those industries in which the region has a comparative advantage.

Despite the fact that the Schumann Committee considered that the Railway tariff policy of allowing a larger differential to exist between high and low-rated commodities and permitting a high degree of tapering of rates with distance could, in certain cases, discourage decentralization, it tended to ignore the fact that this rating policy also supports the continued operation of loss-making lines by allowing cross-subsidization to take place. If the Railways were to change to a more competitive rating policy it would have to take route cost differentials into account and if it could not obtain a state subsidy to continue the operation of unremunerative lines, it would have to close them down. This could affect the availability of transport services in remote areas and might thus discourage industrial decentralization to these areas. It follows that although the general principles underlying the current rating policy of the Railways do not always favour industrial decentralization, they do enable the Railways to fulfil its "common carrier" obligation to provide rail services wherever they are needed and may thus work more in the interests of regional policy than any other pricing principles which can only be implemented if the Railways were freed of these obligations.

It seems to be a principle of the current industrial decentralization policy of the South African Government that an underdeveloped region cannot be developed to its full potential unless firms are encouraged to locate in this region by being offered special concessions which counteract the long term cost disadvantages of locating there. One such concession is the rail rebate on all outgoing traffic which is only granted to manufacturers located at the specified industrial development points. This rail rebate thus only benefits industries located or seeking to locate in underdeveloped regions and reduces their comparative cost disadvantage relative to industries located in developed areas. It seems that the objectives of industrial decentralization policy can be better served by a system of special rail rebates than by an adoption of the general principles underlying Railway rating policy.

3:5:2 PASSENGER TRAFFIC

(a) MAIN LINE TRAFFIC

While third class passenger services have remained fairly viable, first and second class services have been characterized by escalating losses, a falling number of journeys and a decreasing share of total passenger revenue. These problems have largely been caused by the increased competition which first and second class rail passenger services have faced from motor and air transport. A situation of excess capacity has therefore arisen for these services which would probably be exacerbated if the Railways adopted the average cost pricing approach of simply raising fares to cover unit costs since this would probably cause a further loss of market share to other modes of transport. According to the optimal pricing and investment rules discussed in the last chapter the Railways should actually rationalize its operations by cutting back both the quantitative and qualitative level of first and second class passenger services and it might also encourage a greater utilization of the capacity of these services by offering discount rates particularly in off-peak, out-of season periods. This could be regarded as a temporary measure until the capacity of the passenger system has been reduced to a more economic level.

In this respect, the existing structure of passenger fares in terms of which second class are set at 2/3 and third class at 1/2 of first class fares may be regarded as a bit inflexible. It is based on the average cost pricing principle that if the cost of running a third class passenger coach is the same as that of a first class coach then the charge per passenger on the third class coach should be half that charged to first class passengers since the former coach can accomodate twice as many passengers as the latter. However, if fares are to reflect marginal cost principles then rates should vary according to the utilization of a particular service since rates should be increased to ration the number of passengers to the capacity of a fully utilized service and lowered to encourage the greater utilization of an underutilized service. In the latter case rates should not be allowed to fall below the unit operating cost incurred for this service.

The Railway practice of issuing free or privileged tickets to employees and public officials should be discontinued since it conflicts with commercial principles and probably contributes to the loss incurred by the passenger service.

(b) SUBURBAN TRAFFIC

It is apparent that the fares for all classes of suburban passenger services are set below the rates which would be charged under a pricing policy based on marginal cost principles. The Franzsen Committee [22] found that in 1977/78 an increase in fares of an average of 162 per cent was needed to cover avoidable costs. In the following five years, the costs of suburban passenger services continued to increase more rapidly than the revenue they produced so that the level of passenger fares continued to fall even further below avoidable cost. There are, however, a number of valid reasons why suburban passenger services should be underpriced:

(i) The Group Areas Act compels Blacks to live in townships outside white urban and industrial areas. As a result Black workers in the main metropolitan centres have come to rely to a great extent or rail passenger services in commuting to and from their workplaces. Third class passenger fares have been set at uneconomic

---

22. Republic of South Africa, op.cit., p. 2.

levels to alleviate the burden imposed on both workers and employers by the locational requirements of the Group Areas Act.

(ii) It is generally accepted by urban policy makers that an increase in the level of suburban passenger services will yield significant external benefits since rail transport makes a far lower contribution to traffic congestion than road transport so that any diversion of traffic from road to rail will not only reduce the social costs of road congestion but also ease the continuos pressure on the authorities to expand the road system. It may therefore be acceptable, in terms of optimal pricing principles, for the Railways to set the fares for its suburban passenger services below their avoidable cost since these fares do not fully reflect the marginal social benefits associated with suburban rail transport.

Despite the fact that it may economically justifiable for passenger fares to be set below avoidable cost, it is questionable whether the Railways should be required to finance the resulting lossses on its passenger services through cross-subsidization from its other services. If the Railways is to be obliged to perform certain social services then it is equitable that it should be compensated by the State for any losses which arise from these social obligations. This principle appears to have been endorsed by the Franzsen Committee which recommended that "S.A.T.S. must be reimbursed from external sources for supplying uneconomical but essential passenger services that are in the country's interests, but for which the the costs cannot be recovered through the market mechanism."[23]

---

23. <u>Ibid</u>., p. 3.

The Committee indicated that the available external sources were a levy on employers, a levy on fuel and a levy on road hauliers but suggested that the most equitable means of financing rail passenger losses would be through an employer levy based on the wages of workers and applied in selected areas on a differential basis. The selective application of such a levy is considered equitable "because all sectors, particularly the agricultural sector, are not affected to the same extent, and because all geographic areas do not experience transport problems to the same extent....."[24] The Committee recognized the fact that such a measure would affect the cost of labour and recommended that "necessary steps must be taken to control the relationship between labour and capital."[25] This confirms the conclusion reached in Chapter One that, unless a subsidy is financed through a lump sum tax, it will not only affect the distribution of income but also the allocation of resources in an economy. It is apparent therefore, that both external and internal subsidization of loss-making services are costly. The main advantage of a State subsidy over a cross-subsidy, however, is that the distorting impact of the former is likely to be more diffuse, spread over the entire economy, whereas the distortionary impact of the latter will be mainly confined to the public enterprise and the sector in which it operates. It can, perhaps, be argued, from the point of fairness, that where a public enterprise such as S.A.T.S. is required to continue the operation of a loss-making service as a social obligation, then the economy as a whole should bear this burden in the form of external subsidies, rather than the particular public enterprise and the groups interested in its operations. Although the share of the losses on rail passenger

---

24. Ibid., p. 3.
25. Ibid., p. 5.

services borne by the State has risen from 13 per cent in 1977/78 to 48 per cent in 1981/82, there is clearly scope for a further increase in State subsidization since the actual amount of the losses which are financed through cross-subsidization has increased from R278 million in 1977/78 to R311 million in 1981/82. The Franzsen Committee, however, indicated that a major concern of the Treasury is that "the present method of subsidization is creating problems for the State because it shows an escalating tendency, and Treasury has no control over the rate and escalation". Since there is a continuous pressure to expand the capacity, particularly of third class suburban passenger services, there is also a need to control the establishment of new services to prevent an excessive escalation of the losses associated with these services, It was in this respect that the Franzsen Committee recommended that the National Transport Commission should monitor the introduction of new services. Furthermore it was considered that "the best method for the Treasury of contributing to transport subsidies" was by "making interest-free capital available for passenger services. This would clearly obviate the problem of rising interest charges contributing to the losses incurred on these services.

A factor which should be taken into account in formulating the tariff structure for suburban passenger services is that the demand for these services fluctuates on a regular daily basis between peak and off-peak periods. Since the capacity of the suburban railway system is determined by peak period transportation requirements, it is a typical problem of such a system that capacity is underutilized in off-peak periods. By failing to incorporate a peak/off-peak price differential into its passenger fare structure, the Railways misses the opportunity to both improve the utilization of its system in off-peak periods when a lower price should be charged and to reduce the pressure to expand its capacity to meet growing peak period transportation requirements and to thereby achieve savings on its capital investment programme. It should be borne in mind though that the external benefits associated with rail passenger services probably reach their greatest level during a peak period so that the optimal peak/off-peak price differential may be significantly less than where these externalities are not taken into account. If, however, the Railways is induced to raise its passenger fares in the future to increase the revenue derived from these services and to thereby ease the burden of subsidization, it may improve the efficiency of the utilization of its facilities by limiting these fare increases to peak period travellers.

The Railway tariff policy for passenger services cannot be regarded as a possible instrument to use in improving the spatial distribution of economic activity, since the passenger rail system is designed to meet the transportation requirements of a particular spatial structure. The current structure and level of passenger fares is therefore directed at maintaining the existing spatial pattern of economic activity

3:5:3 CONCLUSION

In general, it can be seen that many of the problems associated with the Railway's tariff policy with respect to freight and passenger services have arisen from its failure to reconcile the conflict that exists between the demand that, on the one hand, it should be managed according to business principles and, on the other hand, that it should, in certain instances, provide socio-economic services in the interests of broad national economic needs and transport requirements. If there is to be greater consistency of principle in the application of Railway tariff policy there is clearly a need for a consistent definition of the Railway's objectives. It is difficult to do this when there are conflicting demands made by different interest groups on the Railways. In the following section the nature of these conflicting demands will be examined and a way in which they may be reconciled will be suggested.

3:6 CONFLICTING DEMANDS UPON S.A.T.S. POLICY

In Chapter Two it was explained how a public enterprise operated in a delegated choice situation. The "managerial" theories of the firm propounded by writers such as Williamson[26] may therefore provide fruitful insights into the policy problems of a public enterprise such as S.A.T.S. since, like a large private organization, it is characterized by a division between ownership and control, although, in the case of S.A.T.S. ownership is vested in the State and not private share-holders. Management may therefore tend to pursue policies which maximize their own utility subject to safeguarding their job security by meeting the minimum requirements placed on them by the owners.

A public enterprise such as S.A.T.S. differs, however, from a large private organization in that its management is not the final arbiter in the bargaining process[27] between the groups interested in goal formation. This role is performed, in the case of S.A.T.S. by the Minister of Transport who attempts to ensure that the objectives of S.A.T.S. are consistent with the following broad objectives of transport policy:

(a) The optimal allocation of resources to the transport sector.

(b) The efficient utilization of resources in the provision of transport services.

(c) The development of transport services to narrow the gap between transport demand and supply.

---

26. Williamson, O., "Managerial Discretion and Business Behaviour", American Economic Review, 1963.

27. The process of reconciling multiple objectives in a large organization described in this section applies the behavioural model of the firm developed by Cyert, R.M., and March, J.G., A Behavioural Theory of the Firm, Prentice-Hall, 1963.

However, in the process of formulating goals for S.A.T.S., the Minister would have to take account of the demands of the following interest groups:

(a) THE MANAGEMENT

The management of a public transport-enterprise such as S.A.T.S. are likely to be interested in the widest possible degree of autonomy in its decision making to enable them to manage the organization according to commercial principles. To the extent that public policy limits their managerial discretion by regulating ratesetting and imposing uneconomic social obligations, the management of a public transport enterprise is likely to demand compensating privileges such as state aid in the form of subsidies and grants and protection from competitors who are not subjected to these obligations.

According to Williamson[28], the utility of managers is likely to be a function of salaries, power, prestige and job security. These factors are dependent on the growth and performance of the organization so that management is likely to favour policies which improve its growth and performance potential regardless of whether they are in the broad national interest.

(b) THE USERS OF PUBLIC TRANSPORT SERVICES

The consumers of transport services are not a homogeneous group with uniform requirements but they are likely to demand the widest possible choice of transport services so that they can choose the particular combination of price and quality which most closely suited to their needs.

---

28. O.Williamson, op.cit.

(c) THE TREASURY

To an extent that varies with political social and economic conditions and fiscal policy, the Treasury will be faced with a scarcity of funds relative to the public projects that require Treasury financing. Consequently it will favour any policy which ensures that S.A.T.S. at least breaks even since this will obviate the need to finance a deficit on its operations and will free funds for other projects.

(d) OTHER INTEREST GROUPS

From time to time there may be demands on S.A.T.S. to pursue external macro-economic, social or political objectives. For example, S.A.T.S. may be required to promote industrial decentralization or provide services to black passengers at subeconomic rates as a necessary concomitant of the Group Areas Act.

There is a clear need for the transport policy maker to recognize and reconcile these conflicting demands. In South Africa the reconciliation of conflicts has typically been pursued by the sequential attention to demands and ad hoc policy commitments. In other words, priority seems to have been attached to the most pressing demands and policy commitments made to the interest groups involved, while attention to the demands of other interest groups is postponed until future periods. This is illustrated in the following review of the history of the transport policy with respect to the transport of freight by road and rail in South Africa.

3:7 HISTORICAL REVIEW OF FREIGHT TRANSPORT POLICY IN SOUTH AFRICA

The evolution of the transport policy of S.A.T.S. can be traced through a number of stages each of which arose from the policy commitments made in the previous stage.

(a) MONOPOLY MANAGEMENT AND THE PROMOTION OF REGIONAL DEVELOPMENT

The interpretation often placed on the Constitution Act, in so far as it refers to transport, is that S.A.T.S. should not only fulfil the role of a common carrier on a non-profit basis, but that it should also act, unilaterally as it were, to promote development in various regions by constructing development railway lines. This approach was clearly evident in the earlier days when the development of the country outside urban areas was entirely dependent on railway construction since, in the absence of road transport and suitable inland waterways, there was no alternative fast and efficient surface transportation mode. This resulted in a fairly extensive network of new railways being constructed from the ports to various parts of the interior. Special endeavours were made to meet rural transport needs, so that it was necessary to tolerate for long time a relatively low volume of traffic for the initial years of operation in anticipation of adequate expansion with the passage of time. This policy was assisted by the relatively low railway construction costs prevailing at the time.

Furthermore since the State railway undertaking held a monopoly over the provision of overland transport services, it was considered necessary to regulate tariffs by generally obliging the railways to levy the same rate per ton per mile or otherwise on traffic of the same class carried over the same distance and under the same conditions and circumstances. These standardized tariffs were furthermore set according to the principle of charging "what the traffic can bear" rather than according to the cost of providing the service.

The expected financial performance of many of the railway lines constructed during this earlier period proved to be disappointing once the lines were open to traffic mainly because the potentialities of the regions served had been very much overestimated. Since the railways were required to break even and had adopted a tariff policy of charging "what the traffic can bear", the losses incurred by the traffic conveyed on these unremunerative lines was made good by a system of cross-subsidization whereby monopoly gains were made on more remunerative traffic which was conveyed at rates well above the cost of conveyance.

(b) THE MOTOR TRANSPORT ACT OF 1930 AND THE REGULATION OF INTERMODAL COMPETITION

With the development of the motor vehicle during the 1920s as an economically viable and efficient transport mode, a multitude of private transport operators entered into competition with the railways, and towards the end of the decade motor transport developed and flourished virtually uncontrolled. Due to the intense competition that prevailed, not only between road and rail but also between the road operators themselves, it was considered necessary to have some form of regulation and control in order

(a) to coordinate rail and road transport in such a manner that these two modes of transport would be complementary rather than unnecessarily competitive.[29]

(b) to restrict destructive competition between the road hauliers themselves.

---

29. It should be noted that although S.A.T.S. does operate its own extensive road transport services, these are mostly operated as ancillary and feeder services to the railways and are regarded as an integral part of the rail network.

Consequently a Motor Carrier Transportation Act was promulgated in 1930. Its main object was to permit the private carrier to convey short distance traffic under prescribed circumstances in the larger urban areas and to reserve long distance traffic for the railways. The machinery set up by the legislation for the accomplishment of this purpose was the permit system which affected persons providing road motor transportation for hire and reward. In 1932 the permit requirements were extended to those firms wishing to convey their own goods Road hauliers did not have to acquire permits in the following circumstances:

(a) if they were deemed to be outside the definition of "road transportation" in the Act.

(b) if they carried exempted goods, a complete list of which is published in the Government Gazette, and updated periodically.

(c) if they carried goods within the limits of the zone from their place of business as defined in the Act; and

(d) if they carried goods within the limits of "exempted areas" as defined in the Act.

The effect of these exemptions was to permit the development of private and public road haulage within urban areas. However, in all other cases private and public road hauliers were required to submit permit applications to Road Transport Boards which allow road transportation for goods to take place only under the following conditions:

(a) for goods that are needed urgently;

(b) for goods that are impossible to ship by rail;

(c) for goods that are unreasonable to ship by rail, or other existing transport.

Kennedy[30] has undertaken a thorough review of numercus permit applications at several road transport boards and a number of appeals against refusals to grant permits made to the National Transport Commission and, although he found that the interpretation of the Act does not appear to be uniform throughout the country, he was able to make the following observations about how the boards decide whether to reject a permit application on the grounds that reasonable transportation facilities exist:

(a) permit applications generally not granted if all the applicant can prove is that road haulage is more cost efficient than rail transportation;

(b) only in extremely rare instances are permits issued where the primary justification is extreme harm to the applicant;

(c) permits are granted more readily if convincing support is demonstrated by present or prospective shippers and/or users of the proposed service, rather than claims by the prospective haulier himself; and

(d) some boards view a joint rail/road movement utilizing S.A.T.S. facilities as a "reasonable service" while others do not.

The imposition of controls over intermodal competition through the permit system illustrates the typical pattern exhibited by the growth of government regulation in South Africa in that the need for for this new regulation arose from the fact that the existing regulations had become outmoded as a result of unforeseen developments in the

---

30. Kennedy, T.L., Costs of Road and Rail Transport, Technical Note 11/23/82, National Institute for Transport and Road Research, CSIR, South Africa, July 1982.

transport sector. Thus the high rates charged on certain traffic carried by rail, which had been imposed due to the need to cross-subsidize unremunerative lines had made these classes of traffic particularly vulnerable to intermodal competition from road transport. The railways therefore required protection from intermodal competition if they were to continue this policy of cross-subsidization and this led to the institution of the permit system. The effect of the permit system was to confer on the railways what was tantamount to a quasi-monopoly in respect of certain defined traffics, enabling it to pursue its then established policy of charging "what the traffic can bear".

The institution of this policy of railway protection in 1930 was justified on the grounds that at that time the road system was relatively underdeveloped and would be inadequate to handle large increases in freight traffic. Furthermore, since there was an adequate railway infrastructure already in place, it was considered that it would be uneconomical from the country's point of view to allow large amounts of traffic to be diverted from rail to road.

(c) ECONOMIC GROWTH AND THE NEED FOR DEREGULATION, 1930-1977

During the years which followed the promulgation of the 1930 Motor Transportation Act, the conditions which had initially favoured the institution of the permit system gradually declined in importance. The highway network was greatly improved, several routes of the South African Railways became congested, delays to goods shipments were becoming more frequent, and there were growing doubts among industrialists and some government officials as to whether the existing regulatory environment was in the best interests of the country.

This changing attitude prompted the Marais Commission[31] to be appointed in 1965 to investigate the appropriate roles which the various forms of transport must fulfil in order to promote the development of the economy in the most efficient manner. The 1960s was a period of rapid economic growth and there was a widespread belief in the industrial and commercial sectors that the capacity of the railways might inhibit further growth. This seemed to be confirmed by the increasing incidence of delays on the railways.

The essence of the Marais Commission's findings was that it may not be in the country's best interests for all commodities to move by rail, considering the inherent advantages of road transport. Whereas transport policy had previously been aimed at securing the optimum utilization of the railways, it was considered that the allocation of transport resources should be reassessed against the relative cost comparisons of road and rail transport, from the perspective of the user. The recommendations of the Marais Commission were however, never implemented.

In 1977 the van Breda Commission was appointed to investigate the revision of the 1930 Motor Carrier Transportation Act. As a result of its recommendations there was some piecemeal deregulation in the transport sector:

(i) the Act expanded the zones within which public and private road hauliers may convey goods without a permit. For public hauliers, this zone is a 40

---

31. R.S.A., Report of the Commission of Inquiry Into The Coordination of Transport in South Africa, Pretoria, 1969.

kilometre radius from the place of business, and for private hauliers conveying their own goods, the radius is 80 kilometres; and

(ii) "exempted areas" were defined, within which no permits were required. These exempted areas were basically the metropolitan areas of South Africa.

(d) CURRENT SITUATION

The revision of the Motor Carrier Transportation Act in 1977 indicates that the pendulum of government thinking with respect to deregulation of the transport sector has swung away from the strong support for a tight regulatory environment evidenced in 1930. The amendments to this Act are seen by some as evidence that South Africa is following a world-wide trend toward deregulation. However, the deregulation that has taken place has been essentially piecemeal in character and has largely arisen as an ad hoc reaction to the fact that the system of regulatory controls had become outmoded due to the industrial and economic development which has taken place since their introduction.

The problem with the sequential approach to conflict resolution is that it tends to postpone and not satisfactorily reconcile the conflict between opposing interest groups with different objectives. It seems that there is a widening polarisation between the management of S.A.T.S. and its users and competitors. Furthermore none of these groups appear to be satisfied with the current state of the deregulation process:

(i) road hauliers are calling for "more direct road and rail competition within a wider range of commodities and claim that the existing restrictions are prohibiting an efficient road haulage system"[32] due to underutilization of trucks and the costs and delays of obtaining permits.

(ii) shippers are generally dissatisfied with the fact that road transport boards are the authority to determine urgency of goods or the adequacy of existing transport, and not the industry or firm itself. They argue that "transportation is one of the many elements required in the process of production, and the decision of what mode of transport is most suitable should be in the hands of the firm, just as any other production decision."[33] They recognize the need to utilize S.A.T.S. for the majority of their transport needs, and are generally satisfied with the service received. However, they generally express "a great need to have the flexibility to use road motor transport for a small portion of their business which occasionally require quick urgent deliveries". They believe that as shippers they are in the best position to determine these urgent situations; and

(iii) in recent years S.A.T.S. have incurred substantial deficits on their operations, and they tend to be wary of any deregulatory initiatives which would result in their high-rated freight traffic being diverted to road transport and therefore exacerbate their deficit position. They are likely to oppose any further erosion of their protected position which is not accompanied by their being freed from the obligations and restrictions which necessitated their protection in the first place.

---

32. Ibid., p. 15.
33. Ibid., p. 21.

From S.A.T.S. point of view there are basically three prerequisities which must be met before they support further deregulation:

(i) S.A.T.S. should be relieved of the financial burden of meeting certain uneconomic social burdens, the most important being the subsidization from its own resources of rail passenger services;

(ii) it is considered necessary that the competitive potential of the various transport modes be placed on an equal footing as regards infrastructural costs; and

(iii) the tariff structure of the Railways should be such as to enable it to be more competitive and far more flexible than at present.

It appears that there is a need for comprehensive structural change within the transport sector, and a clearly defined delineation of the objectives and responsibilities of S.A.T.S. These issues will be discussed in the next section, but before this can be done it is necessary to examine the changes which have already ocurred in the Railway's policy with respect to regional development.

It is considered by S.A.T.S. that the railway network is sufficiently developed to serve most areas in the country even in rural areas where most farms are within close proximity to a railway station or siding and the movement of crops to the railway can be handled economically by feeder road transport without any need for further railway construction. The stage of development in South Africa has, therefore, been reached where further railway construction may be needed mainly in connection with industrial expansion and the exploitation of mineral resources. Accordingly the policy has been adopted by S.A.T.S. of constructing new railways solely where:

1. "Such lines are required for departmental purposes, for example to link already busy lines with others to provide alternative routes over which to divert existing traffic instead of providing increased traffic on existing lines.

2. It has been definitely established that road transport services cannot adequately cope with the transport requirements of the region to be served and that the quantity and nature of the available traffic is such that a railway line will be an economic proposition.
3. The railway administration is indemnified by interested parties of financial standing against any losses sustained in the exploitation of the line."[34]

The investment policy of S.A.T.S. has thereby come to be determined more by economic criteria than regional development objectives since, "in determining the economic justification for new railway lines, due regard is had to the available traffic offered for transport, as well as the future potential development in the region concerned."[35]

The policy of S.A.T.S. with respect of guaranteed lines indicates the direction in which the reconciliation of the conflict between transport and regional policy may proceed in the future. S.A.T.S. is prepared to construct lines on behalf of interests in either the public or private sector who can furnish an unqualified and unlimited guarantee against working losses. This represents a historical extension of the following principles:

(i) where S.A.T.S. is required by the government sector to provide services at a tariff insufficient to cover costs, the Constitution Act makes provision for the payment of losses out of state funds; and
(ii) where S.A.T.S. has operated privately owned railways on behalf of the owners, it has usually been agreed that S.A.T.S. should be indemnified against operating losses.

---

34. Loubser, J.G.H., "Transport and regional policy in South Africa" in W.Blonk, ed., op.cit., p. 245.
35. <u>Ibid</u>., p. 245.

A guaranteed line can be financed or constructed either by S.A.T.S. or by the interest concerned. The recent tendency, however has been for the government department or private undertaking which applies for a railway line to finance, and in the case of a private undertaking, to construct the line since S.A.T.S. tends to have most of its funds tied up in its existing capital investment programme.

Guarantee agreements are concluded for a period to be mutually agreed upon between the guarantor and S.A.T.S. The guarantor is compensated against any capital expenditure he incurs in the following manner:

(i) they are exempted from paying the capital cost component of the tariff charged for operating services; and

(ii) the capital cost portion of the tariff payable by all other users of the line as well as any operating surpluses are paid over to the guarantor to redeem the capital invested in the line.

S.A.T.S. therefore operates the line at cost until such time as the capital invested by the guarantor has been fully recovered. Although there is no advantage for S.A.T.S. in operating the guaranteed line during the guarantee period, it also suffers no loss, and will benefit from the additional traffic generated for the rest of the railway network and from operating the line after the expiry of the guarantee period. This approach would therefore appear to reconcile the conflict between regional policy objectives and the objective of operating an economically efficient and financially vaiable railway network, since the burden of bearing the risk of losses on a railway line constructed in an underdeveloped region is transferred from the railways to interests directly concerned with the development of the region.

3:8 THE ROLE OF X-INEFFICIENCY AND A FRAMEWORK FOR DEREGULATION

It is apparent from section 3:7 that strong pressures have built up in the transport sector for its deregulation. The major advantage of deregulation is that modal choice decisions are taken directly by shippers. Shippers are therefore likely to choose that transportation mode which meets their quality of service requirements at the lowest cost. Deregulation of intermodal competition may thus induce a public transport enterprise to behave like a private interprise in seeking ways to:

(a) cut costs and reduce X-inefficiency,

(b) meet the quality requirements of shippers in a more satisfactory manner and

(c) introduce new services, production techniques and organi zational methods.

It has already been observed that a large public monopoly with its survival guaranteed by the State will face far less external pressure to adopt measures which improve operating efficiency and service quality than a firm operating in a competitive environment which would threaten its survival if it tolerated any "X-inefficiency". X-inefficiency may thus arise in a large public monopoly such as S.A.T.S. Since individual agents within its organization can shift the responsibility for failing to contain X-inefficiency on to either other agents within the organization or consumers, through higher prices, or the State through higher subsidies. In the case of S.A.T.S. there is evidence that many of the recent measures which have been introduced to improve operating efficiency and service quality arose as a result of the decline of its share of the transport market for high-rated commodities. This is indicated in the following statement:

"During recent years, excessively rising costs of labour, equipment and material coinciding with variations in the conveyance of profitable traffic, have brought about such a transformation in the earning and spending patterns of the past that an in-depth examination of the working methods in general and a critical analysis and re-evaluation of the measure of productivity were called for. Consequently enhanced efficiency, together with intensified efforts, both technological and administrative, towards increasing productivity was applied in an effort to restore the equilibrium between revenue and expenditure."[36]

---

36. S.A. Railways and Harbours, Annual Report 1971-72 .

In this respect van der Veer sees the containerization programme implemented by S.A.T.S. as "an effort to consolidate its portion of high-rated traffic". He goes on to state that:

> ".... this method of transport largely enabled rail transport to overcome two of the disadvantages associated with this mode, namely relatively long transit times and pilferage and damage to freight. Secondly, it could satisfy the growing demand for a door-to-door service. Apart from contributing to a more efficient transport system overall, particularly in respect of the cost benefits from unitisation and standardisation, containerization did succeed in preventing a further decline in the proportion of high-rated traffic".[37]

Other measures introduced during the last decade to improve the operating efficiency of S.A.T.S. included the following:

(i) the elimination of steam traction and the on-going programme of electrification of railway lines;

(ii) the conversion from vacuum and air brakes to improve train and line capacity;

(iii) the development of a high stability bogie to increase the axle-loading capacity of wagons;

(iv) the construction of a new central marshalling yard for the Reef area which will assist in reducing delays to freight trains; and

(v) the introduction of a zoning scheme which "entails the division of the rail system into various zones, each of which is served from a concentration depot, and the running of block load trains from the point of origin to the destination". This has resulted in improved transit times for trains and turnaround times for wagons.

---

37. van der Veer, G.D., The Role of Rail Transport in the Total Transport Scene, Paper delivered at Annual Transport Convention, Pretoria, August 1982, p. 9.

The effect of these measures is reflected in Figure 3:9 which indicates that the operating efficiency of rail freight services measured in terms of gross ton kilometres per train hour has increased by an average of 7,95 per cent annually between 1973/74 and 1980/81. It can be seen from the abovementioned statements by Railway officials that a major factor motivating the Railways to introduce these measures to improve operating efficiency and service quality has been the decline in the Railway's share of high-rated traffic as a result of the increased intensity of intermodal competition from road hauliers.

This tends to confirm Scherer's thesis that "incentives for cost reduction under a pressure theory of motivation have more powerful behavioural effects than under a reward theory alone".[38] Now it is clear that the deregulation of the transport sector would result in the Railways being exposed to a greater intensity of intermodal competition and therefore a greater pressure to implement measures which increase operating efficiency and service quality, than it presently has to face. There is therefore likely to be an even greater incentive for Railway management to seek ways of containing X-inefficiency and reducing slack within the organization.

Furthermore deregulation may enable the railways to specialize in those services for which it has inherent competitive advantages. For example, the movement of bulk commodities, in train loads or multiple wagon loads can be performed very efficiently and effectively by the railways. The fullest possible development of these services may be hindered if the railways is also required to handle other types of traffic for which it is a relatively inefficient mode such as single wagon loads, and small packages.

---

38. Scherer, F.M., Industrial Market Structure and Economic Performance, Rand McNally Publishing Company, 2nd ed., 1980, p. 466.

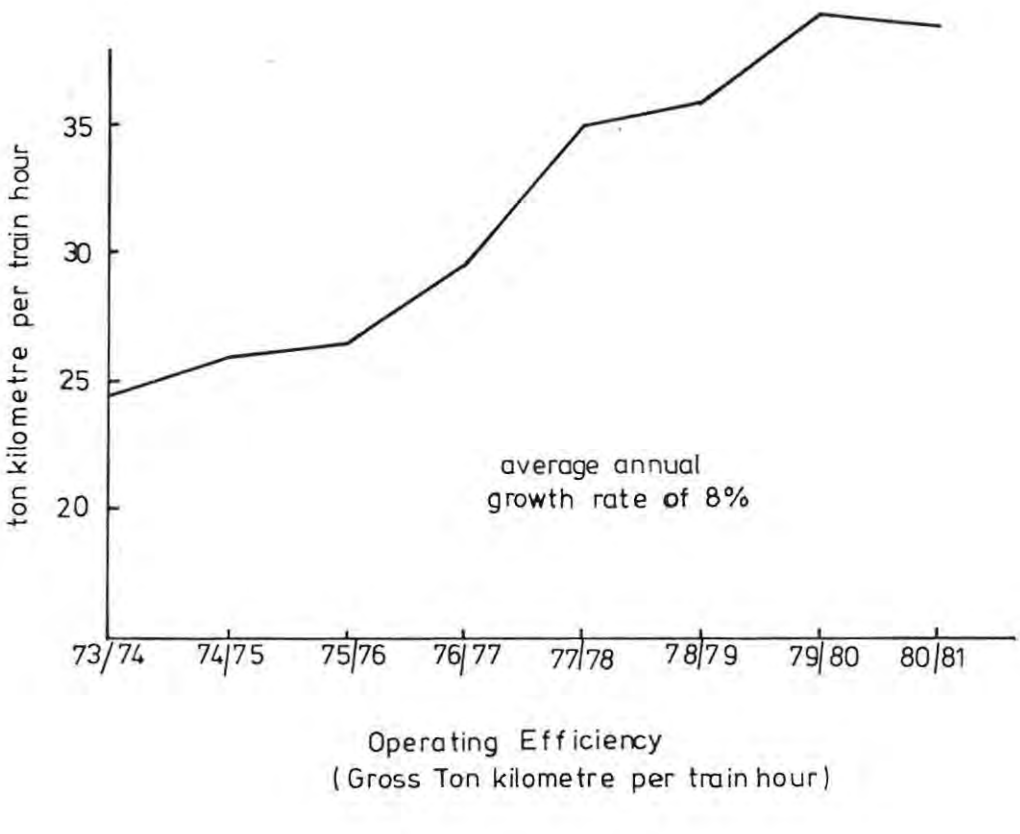


FIGURE 3:9 Operating Efficiency of Rail Freight Services, 1973/74 to 1980/81

A further advantage of the deregulation of intermodal competition is that it would introduce a greater measure of flexibility into the transport sector. As the histrorical review of transport policy in South Africa illustrated, regulations tend to make sense only in the context of the time and circumstances in which they are introduced and tend to become outmoded with new developments in the transport sector.

It is equally apparent, however, that a precondition of unrestricted intermodal competition in the transport sector is that each mode is able to compete with all the others on equal terms. In the area of inland freight transport intermodal competition largely takes place between S.A.T.S. and privately owned road hauliers. It is of crucial importance that the function, responsibilities and compensating privileges of S.A.T.S. be clearly defined in a manner which is equitable to all interest groups involved. The traditional approach in South Africa is to compensate the railways for its burden to perform various social obligations by protecting it from intermodal competition in certain areas. The permit system, however, clearly operated to restrict intermodal competition and it seems that if deregulation is to take place, and, if the railways are still to be required to meet certain obligations, then alternative methods of compensation will have to be introduced. An obvious possibility is that state subsidies and grants should be paid to the railways where it is required to perform uneconomic social obligations. These railway services should thus be regarded as merit goods and the amount of funds made available for their provision would be determined according to the same criteria as any other form of government expenditure.

The main problem with this type of deregulation programme is that it may result in considerable structural readjustment in the transport sector which may be initially painful to railway interests as resources are diverted from rail to road transport. It may also be difficult to ensure separate accountability by the railways for its social and purely economic functions. Furthermore if state grants and subsidies rather than a system of protection are to be relied on to compensate the railways for performing social obligations, this will place further pressure on scarce Treasury funds and will run counter to the current trend to cut government expenditure particularly on social programmes.

An important consequence of this approach to deregulation is that S.A.T.S. should be freed of the responsibility to promote regional development since this often places on this public undertaking a burden to continue to operate unremunerative lines and thereby conflicts with the objective of providing economically efficient and financially viable transport services, as was explained in section 3:2, if this policy leads to a deterioration in the availability, cost and quality of transport services in underdeveloped regions then it may run counter to the current government policy of promoting industrial decentralization by encouraging private investment at development points in these regions.

A solution to this problem of reconciling conflicting regional and efficiency objectives may be indicated by the current policy with respect to guaranteed lines whereby the risk of losses is borne by the parties interested in their construction and not the railways. This suggests that it may be fruitful to investigate the policy of transferring the responsibility for ensuring that regional transport needs are met from S.A.T.S. to a government department interested in regional development. This approach to regional transport policy was proposed for France by L'Huillier and Domenach[39]. They suggested that a regional transport authority (RTA) should be set up in each region to deal with "every mobility problem having to do with the good management of regional space".[40] The RTA should be certified to negotiate from a strong position with all operators of the regional transport system and should be able to contract with some of them for short or long periods, granting subsidies or not, in order to plan for an effective and socially efficient transport service. The RTA would fulfil the role of a consultant, arbitrator and auditor so as to permanently

---

39. D.L'Huillier and O.Domenach, "Transport and regional policy in France", in W.Blonk. ed., op.cit., pp. 121-140.

40. Ibid., p. 137.

survey the quality of transport service and its adequacy in regard to the general objectives of the region.

In situations where the transport needs of a region could be best served by a monopoly then the RTA could simulate the pressures and incentives of a competitive market by calling for tenders from all transport operators and awarding the contract to provide transport services to the lowest bidder. There would thus be competition for the market rather than competition in the market. This type of bidding model was proposed as an alternative institutional solution to the natural monopoly problem by Demsetz[41] who argued that it might eliminate all discretionary power in production forcing a competitive solution. This approach would seem to be entirely consistent with a deregulated environment in the transport sector.

Rowley and Peacock[42], however, have pointed to a number of problems with this competitive bidding solution:

(i) the inputs required for a successful bidder to enter into production must be available to the many potential bidders at prices determined in open markets;

(ii) the costs of collusion on the part of rural bidders must be prohibitively high; and

(iii) the bidder who is initially successful will have substantial advantages at subsequent biddings when contract renewals fall due in terms of market information, property rights to specialist factor inputs and the goodwill built up over the contract period. Further the costs of transferring

---

41. Demsetz,H., "Why regulate utilities?", <u>Journal of Law and Economics</u>, 1964.

42. Peacock,A.T., and Rowley,C.K., "Welfare Economics and the Public Regulation of Natural Monopoly", <u>Journal of Public Economics</u>, 1972, pp. 227-244.

transportation rights to a new operator would tell in favour of the existing operator.

Although these problems may be neutralized through countervailing regulatory measures, this would run counter to the policy of deregulation. Apart from being a vehicle for stimulating intermodal competition within a region, it can be argued that RTAs will also be able to integrate transport and regional policy objectives in a more effective way since they will be a position to treat transport as only one factor to be taken into account in an overall regional development strategy. In view of the mutual interaction between transport and regional development it would seem essential that this integrated approach be taken. This would be in line with the trend towards integrated transport policy which Klaassen observed in Europe in the seventies.[43] Furthermore this approach would ensure that S.A.T.S. was treated on equal terms with all other transport operators and would enable S.A.T.S. to pursue the objective of providing an efficiency and financially viable transport service,

3:9 A RECOMMENDED PRICING AND INVESTMENT POLICY FOR A PUBLIC TRANSPORT ENTERPRISE OPERATING IN A DEREGULATED TRANSPORT SECTOR

3:9:1 THE SPECIAL POSITION OF A PUBLIC TRANSPORT ENTERPRISE OPERATING IN A DEREGULATED TRANSPORT SECTOR

It is important to realize in recommending a pricing and investment policy for an organization such as S.A.T.S., that if the transport

---

43. Klaassen, L.H., and Bourdrez, J.A., "Integrated transport planning" in W. Blonk, ed., op.cit., pp. 63-77.

sector is deregulated, then S.A.T.S. will be in a different position compared both to other public enterprises and to a large private transport undertaking.

A typical public enterprise is normally required to distribute its products and services on a national scale. Since the availability of its products and services is imposed externally, such a public enterprise is required to set its prices to "reflect the marginal cost of that availability in all situations, so that the level of demand will reflect those marginal costs".[44] If the South African transport sector is deregulated in the manner financially viable transport service longer be under a constraint as to the availability of its services. It could provide or not provide them at its discretion while the responsibility for ensuring that adequate transport services even in underdeveloped areas would be vested in the regional transport authority. In markets where there are available good substitutes for rail services, the existence of these substitutes will place a limit on the price which can be obtained for rail services. Now there is no necessary relationship between the maximum price the market is prepared to pay which will depend on intermodal competition and the marginal cost of handling the traffic concerned. This problem will be explored in the next section where it will be shown that "the role of long run marginal cost is, therefore, as an investment indicator and not as a pricing constraint"[45] with the public transport undertaking operating in a deregulated transport environment "relating its long-run marginal costs to its prices and not the other way round."[46]

---

44. Joy, S., "Pricing and Investment in Railway Freight Services", Journal of Transport Economics and Policy, September 1971, p. 233.

45. Ibid., p. 232.

46. Ibid., p. 232.

On the other hand, if S.A.T.S. operated in a deregulated situation it could not be compared to a private transport undertaking which is likely to take any long run opportunity to exploit a monopoly situation and retain and protect any monopoly profits to the extent necessary to satisfy the enterprise's objectives. A public transport undertaking, however, should have as its primary goal, the maximization of rail service subject to a break even constraint. It is therefore likely to invest in any situation where long run profits are being earned with the object of expanding service either in volume or quality to a point where this surplus is eliminated. In other words, as a public undertaking, S.A.T.S. should not become profit-orientated in the same way as a private enterprise.

3:9:2 ADAPTATION OF OPTIMAL PRICING AND INVESTMENT RULES

It will be recalled that in Chapter Two it was recommended that a public enterprise should proceed with the following steps in formulating its pricing and investment policy:

(a) firstly, the following rules should be applied in determining prices and planning the level of investment:

   (i) if there is excess capacity in the short run, price should be set equal to short run marginal cost;

   (ii) if capacity is fully utilized in the short run, price should be allowed to rise to the rate sufficient to restrict demand to capacity; and

   (iii) in the long run the goal of investment policy should be to adjust the capacity of plant to produce the level of output at which price is equal to long run marginal cost; and

(b) secondly, the resultant tariff schedule and investment programme should be adjusted according to the principles of "piecemeal optimization" to take into account "non-optimalities" resulting from, say, departures from marginal cost pricing in industries closely related to the public enterprise or a binding revenue constraint.

Now, if the Railways is to adapt these pricing, and investment rules to its particular circumstances, the following measures are going to have to be implemented;

(i) firstly, the Railways must be freed from its common carrier obligations and permitted to negotiate contracts on mutually acceptable terms with its customers; and

(ii) secondly, the Railways will have to adopt a system of costing which enables the costs attributable to particular routes and particular traffic flows to be measured.

The issues involved in the practical implementation of these measures with respect to Railway freight services will now be examined.

3:9:2:1 NEGOTIATION OF RATES

Once a Railway undetaking is freed of its common carrier obligations, it is able to accept traffic on terms negotiated with its customers. As a result it is no longer constrained to set its rates according to published national scales but can, instead, charge its services at negotiated rates. This considerably improves the Railway's competitive position in a deregulated transport market since its rates need no longer function as a signal to road hauliers of the maximum rate beyond which they cannot increase their own charges. For most long term contracts it will usually be necessary to insert a price variation clause into the agreement in terms of which rates are adjusted upwards either in step with the Railway's own cost index or a national index of wholesale prices. For short term contracts rates would be renegotiated on termination of the contract and it would probably not be necessary to insert a price variation clause.

An advantage of allowing the Railways to negotiate its rates on an individual basis with its customers rather than requiring them to be set according to a published national scale is that a greater flexibility can be introduced into Railway rate schedule. It will thus be possible for the Railways to charge its services on a "per wagon or "per train" rather than a "per tonne/kilometre basis" to take account of the fact that it is cheaper to transport a whole wagon or train load rather than a part load. Furthermore, it will also be possible for the Railways to stipulate consignment train size or regularity minima for traffic carried according to a conventional "per tonne/kilometre" rate to ensure that it is not burdened with uneconomic traffic on an uneconomic route.

In the case of determining tariffs for large consignments of traffic, an operational analysis would have to be carried out to determine whether the required quality of service could be provided. This would draw attention to any unusual costs associated with the traffic consignment. For smaller consignments, however, such operational analysis would not be justified and it would be necessary to exercise control over price minima by an initial screening which could use a table of critical distances showing (i) by traffic type, (ii) by terminal conditions and (iii) by load per wagon, what would be the minimum length of haul at which a tariff competitive with road transport would be profitable for rail traffic. It might also be necessary to also have a review procedure by which any proposed rate could be compared with the market. The screening criteria would have to be reviewed in the light of changes in road haulage prices or in rail costs and volumes.

Now, to apply the rules discussed in Chapter Two, Railway freight tariffs should be allowed to fluctuate between an upper limit at which price rations demand to capacity and a lower limit at which short run marginal costs are being covered. In practical terms, the upper limit for freight tariffs would be determined by the cost and quality of road transport.[47] while the lower limit would be determined by the Railway's unit operating costs.[48] These optimal pricing rules are, however,

---

47. These are likely to be important factors determining the position of the demand function.

48. In Chapter Two it was shown that where plant is designed with a certain built-in flexibility, there will be a range of reserve capacity over which short run average variable costs will be equal to short run marginal costs.

derived in circumstances where the determination of prices is essentially a short term decision. In cases where tariffs have to be negotiated for long term contracts the short run pricing rules are not strictly relevant and tariffs should be based on the long run marginal costs of the particular traffic flows. Quite clearly, a long term contract should not be entered into unless the Railways expects to cover long run marginal cost.

3:9:2:2 THE CONCEPT OF ATTRIBUTABLE COST

Under a system of negotiated rates, costing information would no longer be used to set prices, which would be market determined, but would rather form the basis of two types of decisions:

(i) whether or not to accept a particular traffic or

(ii) whether or not to replace a particular asset being used by a whole variety of traffics of varying levels of long run profitability.

The cost concept strictly applicable to this type of policy problem is long run marginal cost defined by Leftwich as "the change in total cost resulting from a one-unit change in output when the firm has ample time to meet the change in output by changing the scale of the plant".[49] In the British transport sector where there has been substantial deregulation, "British Rail's investment strategy is to replace assets only to the capacity which is justified by traffic which are able to bear their long run marginal costs."[50]

There are, however, considerable practical difficulties involved in obtaining a measurement of long run marginal cost in an organization

---

49. Leftwich, R.H., The Price System and Resource Allocation Holt, Rhinehart and Whinston, New York, 1960.

50. Joy, S., op.cit., p. 233.

such as the railways where facilities are used to jointly provide for a variety of traffic flows and which are indivisible over a large range of activity.

Shillinglaw has proposed the concept of "attributable cost" as "an attempt to bring the long run marginal cost concept into a practical context and make it operational." He defines attributable cost as "the cost per unit that could be avoided, on the average, if a product or function were discontinued entirely without charging the supporting organizational structure".[51] Shillinglaw believes that the concept of attributable cost is applicable to the type of quantitative policy decision which "differs from the short-run decision in the time interval during which the decision is expected to remain in force (since) it does not deal with unique situations, but is expected to provide a continuing answer to a recurring question."[52] Furthermore, Shillinglaw explains that such decisions "do not hinge on the trade-off between present outlays and future benefits but on the expected annual profit contribution of a given segment of the business over an extended period of time."[53] Attributable cost would thus seem to provide very much the sort of information which is required by a transport undertaking operating a system of negotiated rates in a deregulated environment.

In recommending this concept for this type of quantitative policy decision, Shillinglaw notes that it is "less comprehensive but more readily measureable than long run marginal cost."[54] It differs from long run marginal in:

---

51. Shillinglaw, G., "The Concept of Attributable Cost", Journal of Accounting Research, Spring 1963, pp. 73-85.
52. Ibid., p. 74.
53. Ibid., p. 84.
54. Ibid., p. 82.

(a) its assumption of linearity in the long run cost function since "... the objective is to find a unit cost figure that can be used as an estimate of how costs will be affected by a decision to expand or contract a particular segment or aspect of the business assuming that enough time is allowed to permit a permanent change in capacity."[55] This unitization of cost is desirable from the point of analytical flexibility since estimates of attributable cost per unit for each major product function performed provide the cost coefficients necessary for studying any given segment of an undertaking's activities while they may also be used in trend analysis as a means of detecting symptoms of trouble ahead, Shillinglaw does not consider this "unitization to be harmful "as long as the application of the concept is restricted to the class of problems for which it is designed"; and

(b) attributable cost assumes that expansion or contraction will be accomplished by adding to or subtracting from existing facilities whereas long run marginal cost is based on the assumption that all resources are freely flexible in form as well as in amount.

In measuring attributable cost it is clear that short run variable cost should be included but it is also generally accepted that the unitization of short run variable costs only would lead to the exclusion of fixed costs which could and should be allocated to those segments of the undertaking which occasion them. In classifying fixed costs, a distinction can be made between fixed costs which are on the one hand divisible or indivisible and those which are, on the other hand either traceable to a product or common with respect to both products and functions. A fixed cost is divisible if significant shifts in the volume of activity require increases or permit decreases in the total amount

---

55. Ibid., p. 82.

of that cost. Divisibility thus requires either the existence of a substantial number of similar input units or the availability of input units of various capacities. The main problems with attributing fixed costs to a particular segment of an undertaking are thus:

(a) the lack of traceability of a cost to a particular segment

(b) and substantial indivisibilities in this cost.

In discussing which types of fixed cost should be included in attributable cost Shillinglaw proposed that both divisible and indivisible fixed costs which are traceable to the particular line of activity being considered should be part of attributable cost since these costs would be escapable through abandonment of the line of activity involved. Joy[56], however, has also considered what portion of common costs to include in attributable cost.

This is particularly important from the viewpoint of railway pricing since a distinctive feature of the railway production process is that it is characterized by the joint supply of different services to different types of traffic. Although railway services can be broadly divided between those provided for freight and passengers,the railways also transports different kinds of freight while the transport of both freight and passengers over different stretches of track and at different times of the day might all be regarded as separate commodities "jointly supplied by the railway system. Now, Joy finds it is necessary to distinguish between specific, joint and common costs.

Specific costs are those costs incurred in the production of a particular product which, with due notice can be avoided by ceasing to make the product. Common costs are incurred in the production of two or more products where it is cheaper to produce them together than separately but where the level of common cost would decline if the production of one of the products ceased. Joint costs are those costs incurred in the

---

56. Joy,S., <u>op.cit</u>., pp. 237-240.

production of two or more products, the level of which will not change with the abandonment of any of the separable outputs.

Figure 3:10 illustrates the distinction between these costs in the case of a railway route providing for passenger or freight service. Block (a) represents the cost of maintaining the earthworks and the minimum costs of providing and maintaining one line of permanent way and it is fixed for as long as long as the route is to remain open at all and is joint to all the traffics that use the route. The costs of the extra facilities required to provide given passenger and freight frequencies are represented by the blocks (b) and (c) respectively. Block (b+c) represents the total cost of providing these facilities if passenger and freight routes are provided separately. The common cost of providing these facilities for both passenger and freight services on the one route is represented by block (b,c). This is smaller than block (b+c) since the requirements of these two types of traffic are to some extent complementary. The specific costs on the route are the costs of actual wear and tear caused by the passage of trains, and any signalling operation costs which would be avoided if the particular train or trains did not run. Part of the common costs (b,c) can however, also be specifically attributed to either the passenger or freight services as is shown in Figure 3:11.

The common costs <u>caused</u> by freight traffic are represented by the dotted portion of the rectangle (b,c). This is the difference between (b,c) and (b) and is the cost which would be saved if passenger traffic was removed from the route. The dashed portion of the rectangle represents the common costs caused by passenger traffic and is the difference between (b,c) and the freight traffic cost rectangle. Generally, then the common cost caused by a particular traffic can be calculated by subtracting the value of the common cost incurred in transporting all traffics except the particular traffic in question from the common cost of all traffics including this particular traffic.

These attributable common costs can be added to the specific costs for the particular traffic to determine the attributable cost of this traffic. In cases where there is congestion on a route the withdrawal

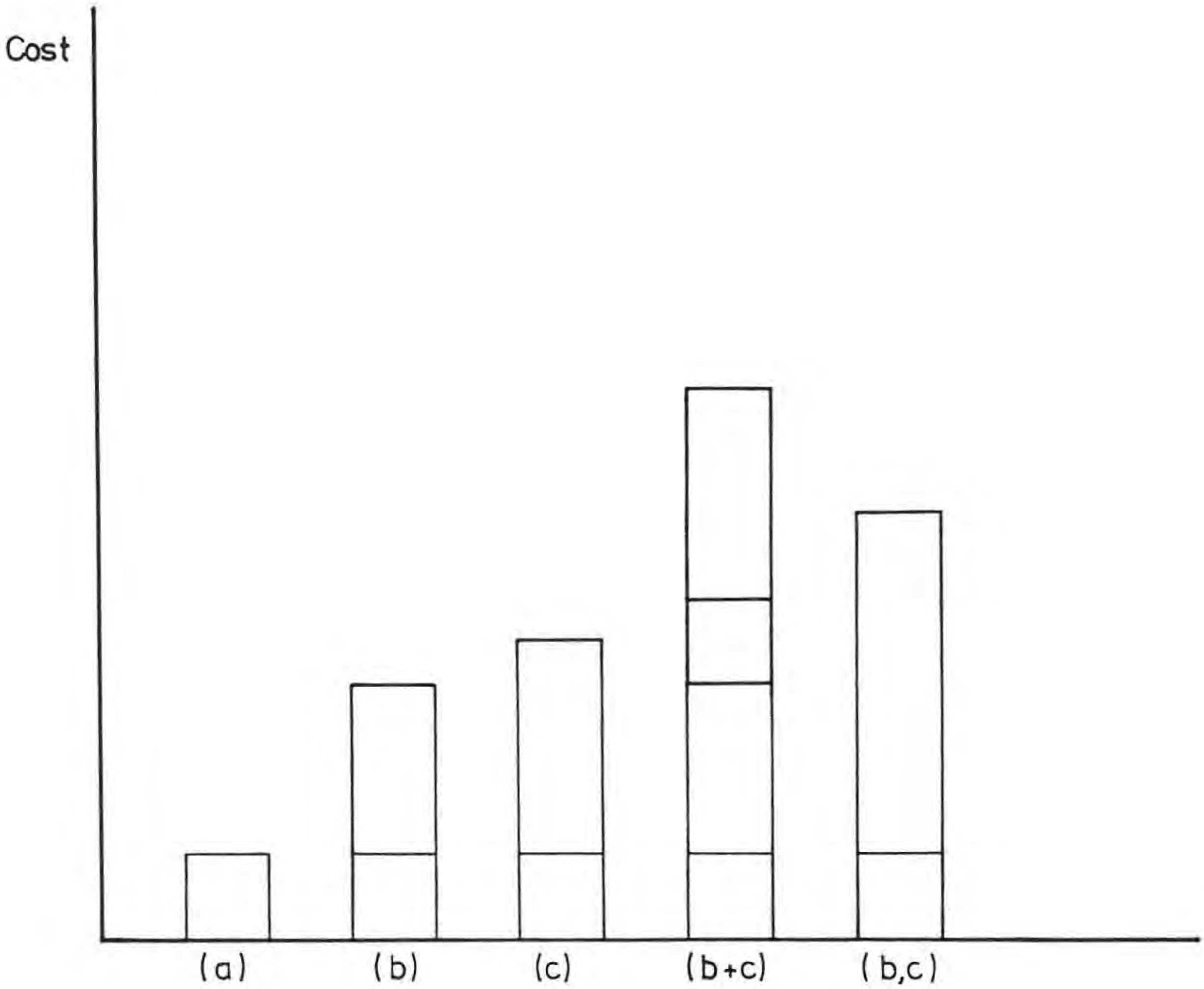


FIGURE 3:10 Specific, Joint and Common, Costs of Providing Rail Passenger and Freight Services

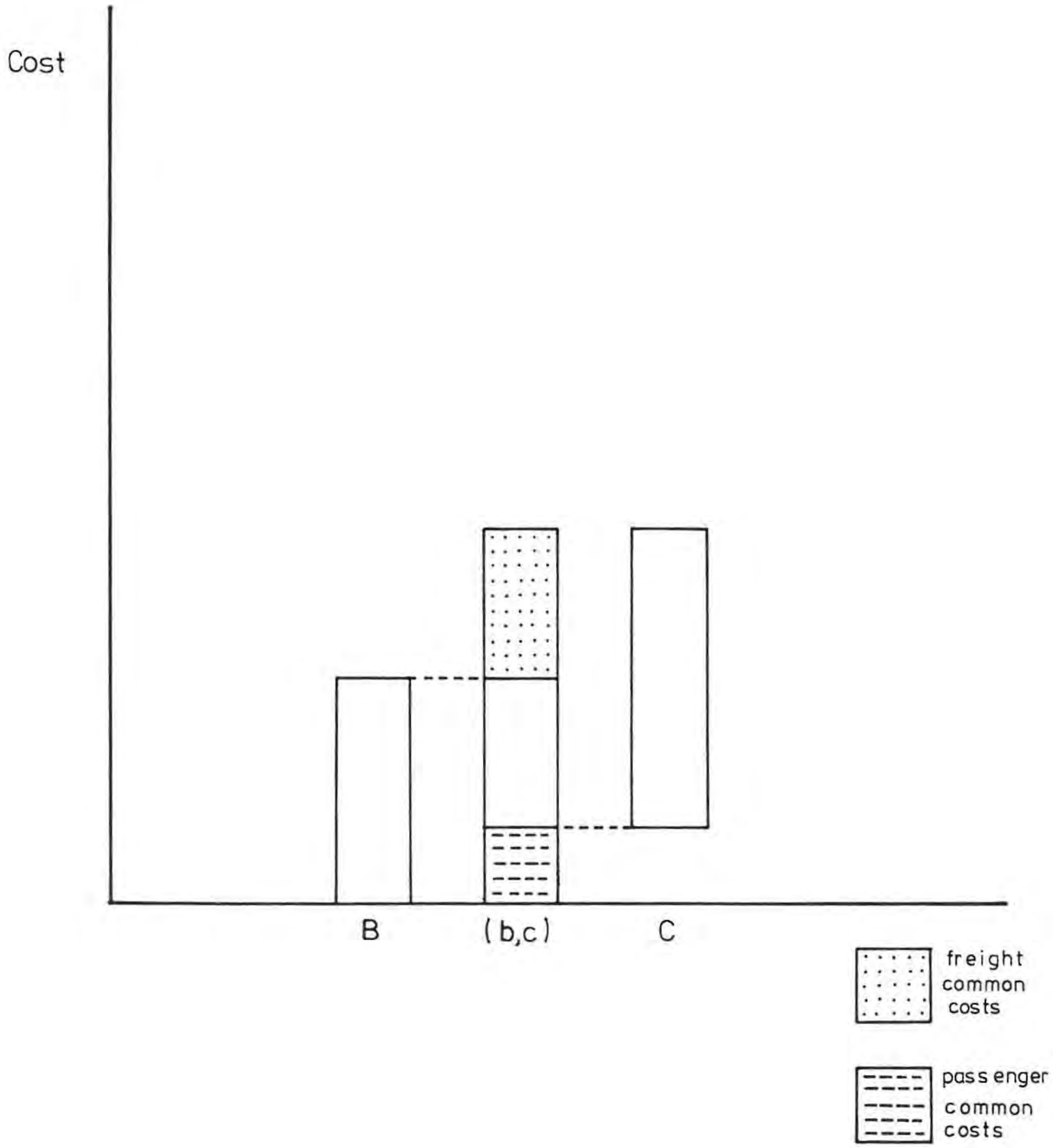


FIGURE 3:11 Calculation of Common Costs of Rail Passenger and Freight Services

of a particular traffic will result in cost savings for the remaining traffic flows which should be added to the traceable costs of this traffic.

Finally, it should be noted that cost causation is best established by finding the peak period and studying the mix of trains during this period since it is the facilities used during this period which determine the capacity of the route.

In conclusion it can be seen that S.A.T.S. is to adopt a system of negotiated rates in a deregulated environment then it will have to know much more than hitherto about:

(i) the variability of all costs with changes in traffic volumes, and

(ii) the revenue and "attributable cost" characteristics of traffics at risk in a reduction or increase in the volume of activity.

It follows that the new role of cost analysis would be to indicate how much railway and how much traffic S.A.T.S. should have and not how much to charge for the traffic. The price to charge will be determined by references to the market through negotiation with users. It is in this sense that Joy suggests that British Rail which operates in similar circumstances "will be relating its long-run marginal costs to its prices, instead of the other way round."[57]

3:9:2:3 DECISION RULES FOR INVESTMENT POLICY

The objective of the railway costing system described in the previous section would be to determine a schedule of the attributable costs of the different types of traffics and routes. Together with a schedule of negotiated rates this cost schedule would provide the information

---

57. Joy, S., op.cit., p. 234.

necessary to apply the following decisions rules:

(a) Steps should be taken to avoid costs if at negotiated competitive rates:

(i) all the traffic on a particular route cannot generate sufficient revenue to cover the costs attributable to the route; and

(ii) an individual traffic flow cannot cover its attributable costs.

In both cases the ultimate type of cost avoidance would be to abandon either the traffic or route concerned. In the case of a route however, cost avoidance may, for example, take the form of a refusal to replace an asset, or the replacement by an asset of lower capacity, or limiting the route to freight traffic only. In the case of an individual traffic, an intermediate step in cost avoidance may be to qualify the acceptance of traffic to periods when there is spare capacity or to a nominated class of wagons available. A problem with cost avoidance is that more time elapses before a major cost can be avoided than is needed to incur a new cost. The railways would however, have to redirect its analytical efforts towards shortening the lag between the loss of traffic and reduction in costs since in a deregulated environment "a transport organization's life is one of acquiring and shedding traffics and acquiring and shedding traffics and acquiring and shedding associated costs."[58] The effect of this decision rule is that no traffic will be carried unless it can at least cover its attributable costs so that most traffics will also be generating sufficient revenue to not only do this but also to make a contribution to non-attributable costs. It is important to note though that this does not imply cross-subsidization since even if

---

58. Ibid., p. 234.

a particular traffic flow can generate enough revenue to cover its attributable costs and still contribute to the attributable costs of other traffic on the route, then the railways should still cease to offer its services to the latter traffic if it cannot cover its own attributable costs.

(b) Given that the goal of a public transport enterprise should be to maximize its service subject to a break even constraint, it follows that if the Railways earns a surplus on any route then it should consider expanding the quantitative and qualitative capacity on this route. The Railways will thus institute a programme of investment designed to expand capacity on those routes where there is the greatest contribution to unattributable costs. The investment decision rules (a) and (b) would thus function to enable the Railways to adjust the capacity and cost of each facility to reflect the profitability of the traffic handled.

In the next section the pricing and investment rules will be evaluated from the viewpoint of the theoretical notion of optimality.

3:10 EVALUATION OF "COMPETITIVE" PRICING AND INVESTMENT POLICY

3:10:1 INVESTMENT POLICY

Essentially the investment decision rules described in the last section represents a practical application of the optimal investment policy discussed in Chapter Two. It will be recalled that the aim of investment is to adjust capacity to the level at which demand equals long run marginal cost. If this capacity was at its optimal level then the price set equal to long run marginal cost would generate sufficient revenue to cover total costs so that in the long run the Railways would be able to attain its long run objective of maximizing service subject to a break even constraint.

However, as was also shown in Chapter Three the optimal application of this investment rule depends on railway facilities being completely divisible and variable in the long run. It is obvious though that in practice railway facilities are indivisible over wide ranges of output. It may thus be the case that where a route is not covering its attributable costs, the only type of cost avoidance available to the Railways may be to close down the route. It will be recalled from section 2:1:3 that where there are indivisibilities it is possible for the social benefits of continuing the operation of a loss-making facility to exceed the costs that would be saved if it was closed down even if one accepts the Harberger postulates that:

(i) the competitive demand price for a given unit measures the value of that unit to the demander;

(ii) the competitive supply price for a given unit measures the value of that unit to the supplier; and

(iii) when evaluating the net benefits or costs of a given action; the costs and benefits accruing to each member of the relevant groups should normally be added without regard to the individuals to whom they accrue.

This was shown in Figure 2:9 where the net social benefit of continuing to operate the route is measured by the difference between the area under the demand curve at capacity output and the total route capital and running costs. This measure of net social benefit may be under-estimated since it does not take account of the following factors:

(a) the revenue accruing to the traffic flows on other parts of the railway network which are generated by the traffic carried along unremunerative routes. In other words, each route should not be viewed independently but one should also consider its contribution to the interdependent system of railway lines;

(b) the external economies accruing to firms and households located in the region served by the unremunerative railway line. Where such external economies exist the demand price of a unit the marginal net benefit of the unit; and the area under the demand curve will no longer measure all the social benefits resulting from the use of the line; and

(c) the marginal social utility of income of the users of the unremunerative railway line may be greater than the national average. McGuire and Garn[59] have suggested that the net social benefit resulting from public projects in particular regions should be weighted by an "area need indicator". They suggested that the "area need indicator" for region i, Ri, could be formulated as follows

$$Ri = a \frac{\bar{E}}{Ei}^{c} + b \frac{\bar{Y}}{Yi}^{d}$$

where $\frac{\bar{E}}{Ei}$ represents the ratio between the national average employment rate and the area employment rate, $\bar{Y}/Yi$, the ratic between the national median family income and the area median family income, while the parameters a, b, c, d, could be set to reflect the "welfare implications of area emplcyment and income statistics".

The employment of such an area need indicator would indicate that the government is not neutral about the regional distribution. The social benefits resulting from the use of a railway line passing through a region with low level of employment at median family income would receive a weighting significantly greater then one.

It can thus be seen that where indivisibilities, externalities, interdependencies and regional income distribution considerations are taken into account, it may be economically justifiable for the Railways to continue the operation of certain loss-making lines. However, if the Railways is to compete on equal terms with private road transport operators it must receive a State subsidy to enable it to continue operating these lines at a more or less perpetual loss. Subsidization, however, is not a costless process although the costs of subsidization are difficult to quantify. If no additional revenue is raised by the

---

59. McGuire and Garn, "The Integration of Equity and Efficiency Criteria in Public Project Selection", Economic Journal, December 1969.

Treasury to enable it to finance the losses on these lines, then scarce Treasury funds will be diverted from alternative uses so that the cost of subsidization can be measured in terms of the effects of reducing the public funds available for other uses. If it is decided to finance the subsidy by raising taxes then there are likely to be net welfare losses since most feasible methods of taxation cause firms to make their production and employment decisions according to a pattern of relative prices which is different to that faced by households. On the other hand, if the subsidy is financed from funds raised by the issue of government bonds, this will have the effect of displacing private sector borrowing on the capital market and will cause an upward adjustment in the interest rate and consequently the cost of funds raised in this market. Regardless of the way in which the subsidy is financed there is always the danger that where the State undertakes to subsidise the loss-making operations of a public enterprise such as the Railways this will lead to a relaxation of the pressure to contain costs and reduce "X-inefficiency" within these operations. In sum, although the costs of a state subsidy in terms of its effect on allocative efficiency within the economy and productive efficiency within a public enterprise can not be properly quantified, they should nevertheless be taken into account in any decision concerning whether or not to close down an unremunerative railway route.

3:10:2 RATING POLICY

The two main implications of the deregulation of the transport sector for Railway rating policy are that:

(a) the Railways would be exposed to intermodal competition from road transport so that it will have to negotiate "competitive" rates with its users; and

(b) the Railways would be relieved of its common carrier obligations and could therefore withdraw its uneconomic services. Furthermore there would no longer be a need to cross-subsidize these uneconomic services by raising the rates charged on its other services. It follows that, as a general rule, the Railways would only continue to perform its uneconomic social obligations if it received financial compensation from the State.

As a result of these changes in the basic constraints on rating policy, a process of adjustment in rates until they reflect the costs attributable to the various traffic flows and routes should take place. For instance, if the rate charged for a particular traffic has been kept artificially above its attributable costs in order to cross-subsidize uneconomic traffic flows, the Railways will have to reduce this rate to forestall the loss of this traffic to road transport operators if the statutory protection of the Railways against intermodal competition is removed. The monopolistic profits earned by any traffic flow are thus likely to be reduced by increased intermodal competition and the rates charged on these traffic flows may tend towards the level where they cover long run marginal costs. If marginal cost pricing principles are adopted by the Railway it is possible that a surplus will be earned on routes with dense traffic flows where rates are raised to reduce congestion of the railway facilities. If the optimal investment policy outlined in this section is adopted these surpluses will only be temporary since the capacity of the Railways will be expanded by investment so that rates will have to be reduced towards the level of attributable costs to ensure that this expanded capacity is fully utilized. As has been seen, where a deficit is incurred on a particular route, the Railways will have to take various measures to reduce the level of service provided for this traffic until the deficit is eliminated. It can thus be seen that the combined effects of intermodal competition and a flexible investment policy should ensure that Railway rates tend toward the level where they reflect attributable costs. In the case of rates negotiated for long term contracts, the Railways would be required to ensure that these cover long run marginal cost. The Railway rate structure is thus likely to bear a closer relation to cost under a "competitive" pricing and investment policy than it does at present.

This may be a factor promoting more efficient intermodal coordination in the transport sector. If the prices charged by the different modes of transport reflect the opportunity costs of providing them, then it is likely that consumers will choose the mode which can provide a particular transport service at the lowest cost. As a result the different modes of transport are likely to specialize in the provision of those services in which they have a competitive advantage. As was shown in section 3:1:4 the transport of goods by rail involves heavy terminal costs so that the comparative cost advantage of rail transport will improve as the length of haul increases. It does not follow that a particular distance should be laid down up to which direct delivery by road is cheapest and best, and above which it would be in the national interest to go by rail since, as Ponsonby has noted," in almost every item of cost the amount will vary according to the circumstances".[60] If the prices charged by the different modes reflect cost then the user will be able to take account of the particular circumstances facing him in deciding which mode to use. For example, even if a user normally ships his goods by rail, an emergency order may arise which cannot be met in sufficient time if the goods are shipped by rail but which can be met if the shipment is made by truck. It can thus be concluded that the transport sector should be more efficiently co-ordinated if the process of intermodal competition and adjustment of rail capacity to demand causes railway rates to reflect cost to a greater extent than is now the case.

The economic impact of the Railways introducing a cost-based tariff policy in a deregulated transport environment has been evaluated by Kennedy.[61] The initial effect of this policy would be that some freight traffic would be diverted from rail to road transport since certain shippers, who under the current permit system are required to use rail transport even where it is less costly for them to use road transport, would be allowed to exercise free modal choice. Shipper's modal choice would, however, not only be determined by the differences

---

60. Ponsonby, G.J., Transport Policy Coordination through Competition, I.E.A., Hobart Paper No. 49, 1969.

61. Kennedy, T.L., Economic Allocation of Transport Resources, Paper delivered at Annual Transport Convention, Johannesburg, July 1983.

between road and rail tariffs but also by the differences in the quality of service provided by each mode. For this reason Kennedy quantified the cost to shippers of using a mode with a higher average transit time in terms of the cost of capital tied up in inventory and the additional warehousing required for higher inventory levels. On the basis of his calculation of the cost to shippers of using either rail or road transport, Kennedy estimated that in 1980/81 approximately 1,18 billion ton kilometres of railway traffic would be diverted to road hauliers.

Kennedy suggested that the overall impact of this diversion could be estimated in terms of:

(i) the benefits to the private sector which would result from reduced transport costs, productivity improvements arising from reduced empty running of trucks and enhanced equipment utilization and the additional traffic growth which would not have resulted if it were not for the increased availability of road motor transport;

(ii) the reduction in net income of the Railways resulting from the loss of revenue earning traffic and

(iii) the increased damage to the roads and the increase of traffic congestion resulting from the greater volume of road haulier traffic.

Although Kennedy's study is at the time of writing, incomplete, his preliminary results indicate that the private sector benefits resulting from this traffic diversion will be significantly greater than the net income loss of the Railways. Kennedy's analysis however, fails to take into account the response of the Railways to the pressures of a competitive environment. It is probable that in a competitive environment the Railways would have a strong incentive to adopt measures which improve its operating efficiency and service quality and thereby help it to recover some of the traffic it loses to road hauliers. Furthermore, if the Railways pursued a flexible pricing strategy of lowering the rates on diverted traffic to near the variable cost level and raising the rates on retained traffic, it could substantially reduce the impact of deregulation on its net income. Kennedy's method of estimating the economic impact of traffic diversion may therefore tend to understate the net economic benefits which could be derived from the Railway's applying the efficient pricing principles discussed in this Chapter within the framework of an equitably deregulated transport environment.

There are, however, two problems which may arise in a deregulated transport sector where the Railway's adopts a 'competitive' rating policy.

(a) Firstly, Railways typically face the problem that there are often considerable lags in adjusting the level of capacity and cost to reflect the profitability of the traffic handled. In particular, at any point in time there may be excess capacity because either;

(i) the Railways has not yet completed a programme of reducing the capacity on uneconomic lines or

(ii) traffic has not yet grown sufficiently to fully utilize capacity which has recently been expanded by a large indivisible amount.

In cases where there is excess capacity, the only costs which the Railways would incur in accepting additional traffic would be the marginal running costs associated with normal railway operations. According to marginal cost pricing principles, the Railways should in these circumstances, lower rates toward the level of unit running costs to attract additional traffic in order to improve the utilization of unutilized capacity. If this policy results in a deficit being incurred then the Railways will need some short term financial compensation to enable it to pursue efficient pricing policies while capacity and demand are adjusting to a long run equilibrium. The application of this type of marginal cost pricing and investment policy may, however, mean that the cost and availability of rail services will vary to a much greater extent then is currently the case. This may lead to an increase in the risk and uncertainty associated with long term location and investment decisions in

South Africa since it would involve a departure from the stable tariff structure and service availability conditions which have historically characterised the provision of railway services in this country.

(b) If the transport sector is deregulated the Railways will be under pressure to lower its rates to reflect costs only if it faces effective intermodal competition. The Railways may, however, have a monopoly over the transport of certain commodities with the result that a measure of external control over the tariffs set for these commodities will need to be exercised to ensure that they are based on the long run marginal cost of conveying them. Furthermore it is conceivable that even where road hauliers do provide substitute services there may still be collusion between the Railways and road hauliers with respect to the fixing of rates. There are however, a number of factors which would make it difficult for the Railways and road hauliers to agree on fixing substantially above competitively determined levels:

(i) there are a large number of road hauliers and entry into this industry is relatively easy since capital costs are low in relation to operating costs. Furthermore many firms transport their goods in their own vehicles. It would thus be difficult for the Railways to establish generally acceptable terms of collusion with such a numerous and diverse group of rivals.

(ii) as has been explained, the cost structures of road hauliers are different from those of the Railways, mainly due to the fact that road operators do not have to incur the capital cost of providing for their infrastructure. The Railways are thus likely to be faced to a greater extent with the problem of adjusting their rates to improve the utilization of capacity and are therefore unlikely to agree with road hauliers on any common level of rates;

(iii) the transport services provided by the different modes are heterogeneous in nature with the cost, speed, frequency and regularity to the user of the services provided by the different modes varying considerably according to the circumstances. The terms of rivalry are thus multidimensional and it would be extremely difficult to reach any collusive agreement to restrict intermodel rivalry.

It can therefore be concluded that in a deregulated transport sector intermodal rivalry is likely to exist on some services between road hauliers and the Railways which will ensure that pressure will be placed on the different modes to reduce rates to competitive levels and that transport users will be able to choose the mode providing a service at the lowest cost. This rivalry is also likely to place greater pressure on the Railways to contain "X-inefficiency" within the organization as described in section 3:8 so that both greater productive and allocative efficiency are likely to result from this policy.

3:10:3 INDUSTRIAL DECENTRALISATION

If the transport sector is deregulated and the Railways is allowed to negotiate rates and withdraw its unremunerative services then, as was shown in the previous section, railway rates are likely to reflect cost more closely than where they are formulated according to 'value of service' principles. As a result the rates charged on manufactured goods are likely to decrease relative to those charged on raw materials. Consequently there might be a certain amount of decentralization toward the raw material source in those industries which process raw materials for intermediate or interindustry demand. Furthermore the reduction in the degree of taper as it tends to reflect cost more than value of service may afford some protection to local industries in underdeveloped regions from competition from industries in the developed areas. However, the impact on regional development of a change in the basic tariff structure is not likely to be significant since the effect of changes in railway rates will vary both in direction and in magnitude between the different sectors.

More important would be the effect of relieving the Railways of its responsibility to promote regional development where this conflicts with commercial principles. This is necessary if the Railways is to compete on equal terms with road operators but it means that the Railways may have to close down unremunerative lines in underdeveloped regions unless it receives adequate financial compensation from the State or Regional Transport Authority. The closure of railway lines in underdeveloped areas and the expansion of the railway network in developed areas may actually result in greater inequalities in regional development since firms are unlikely to locate in areas where there is an inadequate provision of transport services. It follows that any decisions by the Railways to reduce the level of service provided in underdeveloped areas must be closely monitored by Regional Transport Authorities. If it is considered that regional transport needs cannot be adequately met by the existing road services, then such bodies must have the funds to subsidize the continued operation of loss-making lines. In considering the amount of subsidization which should be paid to the Railways, the government should also take account of the considerations discussed in section 3:2 since the effect of a line closure is likely to have national as well as regional implications.

If the Railways is to be freed of its responsibility to promote regional development, it also follows that it will not have to grant special rebates to outgoing traffic from industrial development points where the revenue earned on this traffic falls short of the costs attributable to it. As has been mentioned these rebates are useful components of present South African decentralization policy since they compensate firms located at industrial development points for the long term cost disadvantage of locating at a greater distance from the markets in the densely populated urban areas of South Africa. It follows that the Regional Transport Authority will have to offer financial compensation to any transport operator which carries traffic at special rates from underdeveloped areas.

It can be concluded that if the Railways is to adopt competitive pricing and investment policies in a deregulated transport environment then this is likely to conflict with the objectives of regional policy. There will thus be the type of efficiency/equality trade-off discussed in Chapter Two. As has been suggested the conflict between equality and efficiency should be recognized and dealt with by an adequate division of responsibility between the Railways and the State. The Railways should thus be required to pursue efficient pricing and investment policies and the resulting regional inequalities should be dealt with by public regional policy.

3:11 CONCLUSION: SUMMARY OF FINDINGS AND RECOMMENDATIONS

In conclusion to this chapter the main findings and recommendations will be summarized.

(a) In formulating goals for S.A.T.S. the State has had to recognize and reconcile conflicts between the demands made by the different groups interested in the activities of this organization. Historically these conflicts have been reconciled by sequential attention to the most pressing demands and *ad hoc* policy commitments to the interest groups involved. This approach has tended to postpone and not satisfactorily reconcile the conflict between opposing interest groups with the result that the current situation is characterized by a widening polarization between the interests of S.A.T.S. management and its users and competitors. To break this *impasse* it is suggested that there should be a clearer definition of S.A.T.S. objectives and a delineation of its duties and responsibilities according to the principle that a railway utility should maximize its service subject to a break-even constraint. To facilitate this approach, the recommendation by the Franszen Commission that S.A.T.S. should not be required to provide uneconomic social services unless they are financed through a State subsidy should be implemented.

(b) It appears that the protection from intermodal competition which is afforded to S.A.T.S. through the current permit system has tended to promote inefficiency in the freight transport sector by restricting consumer's freedom of modal choice. Deregulation of the transport sector would reduce the transport and distribution costs incurred by shippers by enabling them to choose the most efficient transport mode and would place pressure on S.A.T.S. to seek ways to retain the share of the freight market by cutting costs and reducing X-inefficiency, improving the quality of service and introducing new services, production techniques and organizational methods. Deregulation should not, however, be undertaken in a piecemeal manner by simply relaxing the permit system since this could seriously undermine the financial performance of the Railways. The State should therefore attempt to ensure that intermodal competition takes place on equal terms by introducing an equitable system of road user charging, subsidizing uneconomic rail services and allowing S.A.T.S. greater flexibility in determining its tariff structure and level of service.

(c) The present freight tariff policy of S.A.T.S. appears to be be characterized by an excessive rate differential between high-rated and low-rated commodities largely because the revenue derived from transporting high-rated commodities is needed to cross-subsidize the losses incurred on other rail services. If the Railways is relieved of the burden of cross-subsidization and it it is allowed to negotiate rates with its customers below those published in the official tariff scales, then this differential is likely to be narrowed to the extent that it largely reflects the differences in the costs of transporting different commodities. If this is the case, rail tariffs will signal to consumers the approximate opportunity cost of choosing this mode.

(d) Route cost differentials are presently not reflected in Railway freight tariffs with the result that unremunerative routes tend to be cross-subsidized from the revenue earned by more profitable routes. If the Railways is to become more competitive in a deregulated environment, it is suggested that it should design a costing system which attributes costs to different routes and different traffic flows on routes. This would provide the Railways with the information necessary to adjust the capacity of services to reflect long run marginal cost. The substantial losses incurred by Railway passenger services should be dealt with through a policy of rationalization rather than by raising tariffs. Although it is desirable for unremunerative social services to be financed by the State, there is a need for the State to control the establishment of new services to prevent the burden on the Treasury from escalating. More efficient utilization of passenger services could also be obtained through the introduction of peak period price differentiation.

(f) It has been suggested that the large differential between the rates charged for manufactured goods and raw materials may have encouraged some industries to locate near their markets rather than their raw material source, but it is unlikely that the Railway tariff structure has been a significant factor influencing the spatial, distribution of economic activities in South Africa. The current system of rebates offered to firms transporting commodities from industrial development points has in any case tended to compensate manufacturers at these points for the high railage costs incurred as a result of locating at relatively great distance from the main urban areas. If the Railways is forced to close down unremunerative lines to remain competitive this may affect the availability of transport services in underdeveloped areas. It is therefore necessary for Regional Transport Authorities to be established to ensure that regional transport needs are being provided.

CHAPTER FOUR

ELECTRICITY

INTRODUCTION

In this chapter the tariff policy of the Electricity Supply Commission (Escom) in South Africa will be investigated and evaluated according to the theoretical framework discussed in Chapter Two. Before this can be done it is necessary to first examine the characteristics of electricity supply and demand which are likely to affect the tariff policy of an electricity utility, and then to consider how a tariff structure can be practically formulated which reflects marginal cost pricing principles since this can be used as a standard against which to compare the existing tariff policy of Escom.

4:1 THE CHARACTERISTICS OF ELECTRICITY SUPPLY

4:1:1 THE SYSTEM OF GENERATION AND TRANSMISSION

Turvey has described the characteristic features of an electrical power system as follows:

"At any moment in time, an electrical power system can be regarded as consisting of a number of nodes, all of which are linked to some of the others by transmission lines. Each node contains generating capacity and/or a load to be met ..... The load to be met at any node is the load of a distribution network so that such a node is a point where transformers reduce voltage to distribution voltage and the current is fed into the distribution network supplied by the node."[1]

---

1. Turvey, R., Optimal Pricing and Investment in Electricity Supply - An Essay in Applied Welfare Economics, Allen & Unwin Ltd., London, 1968, p. 1.

The total load of a power system is therefore the sum of the loads of the consumers served by the network plus its distribution losses. In much of the literature dealing with the application of the theory of marginal cost pricing to the problem of formulating an optimal tariff structure for an electricity utility, it is usually assumed that the power system is operated in an optimal manner. In this section the conditions for the optimal operation of a power system in both the short and long run will be considered.

A. SHORT RUN OPTIMIZATION

To illustrate the nature of the short run optimal solution to the problem of operating a power system, the special case of a system of two nodes, with each node being a point at which a thermal generating plant is operated and a given load requirement is supplied, will be examined. If the two nodes were not connected by transmission lines then the load requirement at each node would have to be entirely by its generating plant. However, in the case where the two nodes are connected by transmission lines and the marginal cost of generation at node 1, m1, is greater than that at node 2, the electricity utility would reduce its fuel and transport costs by generating electricity is excess of the load requirement at node 2 and transmitting it to node 1 where the electricity generated should be less than the load. Since capacity-related costs will be fixed in the short run, it follows that short run optimization involves meeting the given load requirement by operating the generating plant at each node at the level where fuel, transport and other running costs are minimized. In this example, the minimization of these costs will occur when the transmission of electricity from node 2 to node 1 has reached the level where the difference between the marginal generating cost at each node is proportional to the incremental transmission loss, l. The condition for short run optimization is therefore that:

$$m1 = \frac{m2}{1-l} \qquad (4.1)$$

This implies that the incremental cost of increasing the load by

1 kilowatt-hour (kWh) at node 1 must be the same whether the extra unit is generated at node 1 or is provided by generating and transmitting more from node 2. It follows from this optimality condition that once the marginal cost of generation has been calculated at any one node, the marginal cost at every other node can be calculated by making adjustments for incremental transmission losses along the network. As will be shown in section 4:3 this considerably reduces the computational difficulty involved in formulating a marginal cost based tariff structure.

If a network is being operated in a manner which satisfied the optimality condition 4.1, then the marginal generating cost at each node will be given and it is possible to determine which stations should operate and at what rate, how by hour. Turvey explains hourly optimization as follows:

"At any given hour the system has a certain number of generating plants. Some of these will not be available for generation because they are undergoing repair or overhaul. Those which are available can be arranged in merit order, i.e. according to their marginal generating costs. The load required during the hour in question can be most cheaply met by running the highest merit (lowest marginal cost) plant, then the next and so on until total generation equals the given load."[2]

It is clear that once account is taken of transmission losses and constraints, this problem becomes more complicated although the general condition for optimum stated above remains valid. Turvey, however notes that even in the absence of transmission constraints the plant available during any hour may not always be run in full merit order. He lists the following reasons for this:

---

2. Ibid., p. 8.

"(a) at certain times of day, demand rises faster than the rate at which a conventional thermal plant can be brought up to full load;

(b) there is a minimum time of several hours for which it is economic to shut down conventional thermal plant; if demand is expected to rise again within a shorter period it is cheaper to run at minimum load and correspondingly reduce the load on other plants with slightly lower marginal cost than to shut down and then use heat in bringing the plant up to speed;

(c) the unused capacity of the plant with the highest marginal cost among those running may be less than is desired to meet the risk of sudden increase in demand ...... and the risk of a sudden outage of generating plant. It will then be necessary to have several plants running with spare capacity to provide the required reserve."[3]

Although these complications do have an important effect on the day-to-day operation of the system, they play a much less significant role in broad system planning and in the costing of load increments for the purposes of formulating a tariff structure. It will be shown in section 4:3 that the general principle of hourly optimization gives rise to the useful proposition that the short run marginal cost of generation is equivalent to the marginal generating cost of the plant which is used last in merit order to meet any load increment.

B. LONG-RUN OPTIMIZATION

In the long run an electricity utility will be concerned with planning the installation of generating and transmission capacity to meet

---

3. *Ibid.*, pp. 9-10.

forecasted load requirements. The benefits derived from installing a new generating plant largely consist of the fuel savings which arise as a result of using plant with a lower marginal generating cost, mi, than that of the "marginal" plant, $\bar{m}$. The present value of such fuel savings, PV(S), can be represented as:

$$PV(S) = PV(\bar{m} - mi)a \qquad (4.2)$$

where a is unity when the new plant is operating and zero when it is not. Now an increment to one year's plant programme of 1 kW of generating capacity will result in a net change in the present worth of all system's cost of:

$$k - PV(\bar{m} - mi)a + PV(R) \qquad (4.3)$$

where k represents the unit capital cost of the plant and R denotes the annual maintenance and manning costs of keeping the plant in operation. The expression (4.3) is equivalent to the long run marginal cost of generation.

In deciding on what type of plant to install, the electricity utility should aim to adjust the existing plant mix towards an optimal mix at which it would not be possible to reduce the present worth of all system costs by substituting between the capacity of any plants in the system. If A denotes "the annuitized value of", then the marginal condition for optimal plant mix is that the value of

$$A\ (k - PV(\bar{m} - mi)a) + R \qquad (4.4)$$

be the same for all plant used in the system. It is possible that the optimal plant mix may contain a number of different types of plant such as:

(i) gas turbines which have a low capital cost per kW, and high marginal cost per kWh;

(ii) thermal plant which has a high capital cost per kW and low marginal cost per kWh; or

(iii) hydroelectric or nuclear plant which has a very high capital cost per kW and an extremely low marginal cost per kWh.

It follows that in choosing the type of new plant to be included in its plant programme, the electricity utility will usually face a trade-off between savings in capital and in fuel costs. It is usually the case that an electricity utility uses hydroelectric and thermal plant to meet the base load requirements which are likely to occur in every period while gas turbines are used to meet the electricity requirements which are expected to arise only during periods of peak demand. Furthermore it is likely that a number of different vintages of each type of plant may be operated. If it is assumed that technical progress is embodied in newer vintage of plant in the manner suggested by Salter[4], then it follows that each vintage of plant will shift down the merit order of operation as newer vintages with greater efficiency of fuel use are installed as part of an ongoing investment programme. This programme should be conducted in such a way as to minimize the present worth of system costs by ensuring that the marginal condition that the value of $A(k - PV(\bar{m} - mi)a) + R$ be the same for all generating plant, is satisfied in the long run.

In planning a programme of investment in transmission the electricity utility should balance the costs per kW against the benefits of expanding the transmission capacity. Two types of benefits can be derived from expanding transmission capacity:

(i) firstly, it may be the case that it is not possible to operate the system in the optimal manner indicated by condition 4.1 since transmission lines may be loaded to full capacity. Thus if, for example, $m1 > m2/1-l$ due to a transmission capacity constraint then the expansion of the capacity of the lines linking the two nodes will enable the electricity utility to save fuel and transport

---

4. Salter, W.E.G., Productivity and Technical Change, 2nd ed., Cambridge, 1966.

costs by substituting the electricty generated at node 2 for that generated at a higher cost at node 1;

(ii) secondly, the expansion of the capacity of the transmission lines may actually reduce transmission losses along these lines.

Transmission capacity should therefore be expanded to the optimal level, where the present worth of these benefits is equivalent to the capital cost of an increment in transmission capacity.[5] No account has yet been taken of the effect of risk and uncertainty on planning. This will be considered in the next section.

4:1:2 UNCERTAINTY AND RESERVE PLANT MARGINS

In planning the expansion of a system of generation and transmission, an electricity utility will face the following types of risk:

(i) future load requirements may differ from levels forecasted;

(ii) actual weather conditions may deviate from standard weather conditions;

(iii) delays in the construction of new plant and unplanned outages may affect the planned availability of plant; and

(iv) there may be unexpected development in technology and relative input prices.

Turvey assesses these different types of risk as follows:
"..... weather and availability risk can be dealt with in terms of frequency distributions, they can be the subject of probability calculations of more or less sophistication according to the availability of data. The pure forecasting risk can be treated in a similar manner by the arbitrary but useful expedient of attributing a certain variance to forecasting errors in the light of experience and hope. The

---

5. Turvey, R., op.cit., derives this optimality condition mathematically on p. 23.

uncertainty about prices and technology, however, can be expressed in probability, terms only if engineering planners are prepared to express their guesses as subjective probability distributions, something which seems unlikely."[6]

If uncertainty about prices and technology is ignored and it is assumed that forecasting, weather and availability risks can be expressed in probability terms, then the optimal operation of a system requires that the mean expectation of the present worth of system costs be minimized.

The problem of risk and uncertainty mainly impinges on the planning of electricity generation and transmission with respect to determining the appropriate size of the margin of capacity over and above that required to meet the most probable demand. Figure 4.1 shows the probability distribution of peak demands in standard weather conditions in, say, year $t + 6$. $\bar{q}$ is the most probable demand but, if it is assumed that the distribution is normal, than there is a 50 per cent probability that peak demand will actually exceed $\bar{q}$. The consequence of excess demand would be "load-shedding" - some consumers have to be completely deprived of the excess demand so that output is restrained to capacity. If planned capacity output were to exceed $\bar{q}$, however, the risk of excess demand would be less since the area of the distribution above it would be smaller. Given the precise parameters of the probability distribution $f(q)$, the probability that demand will exceed capacity can be found for each possible level of capacity output. The planners can then choose a "critical risk level" or probability that demand will exceed capacity, shown by the area Π^* in Figure 4.1 and found the capacity output corresponding to this, $\hat{q}^*$. There will thus be a margin of capacity output over most likely output.

6. Ibid., p. 61.

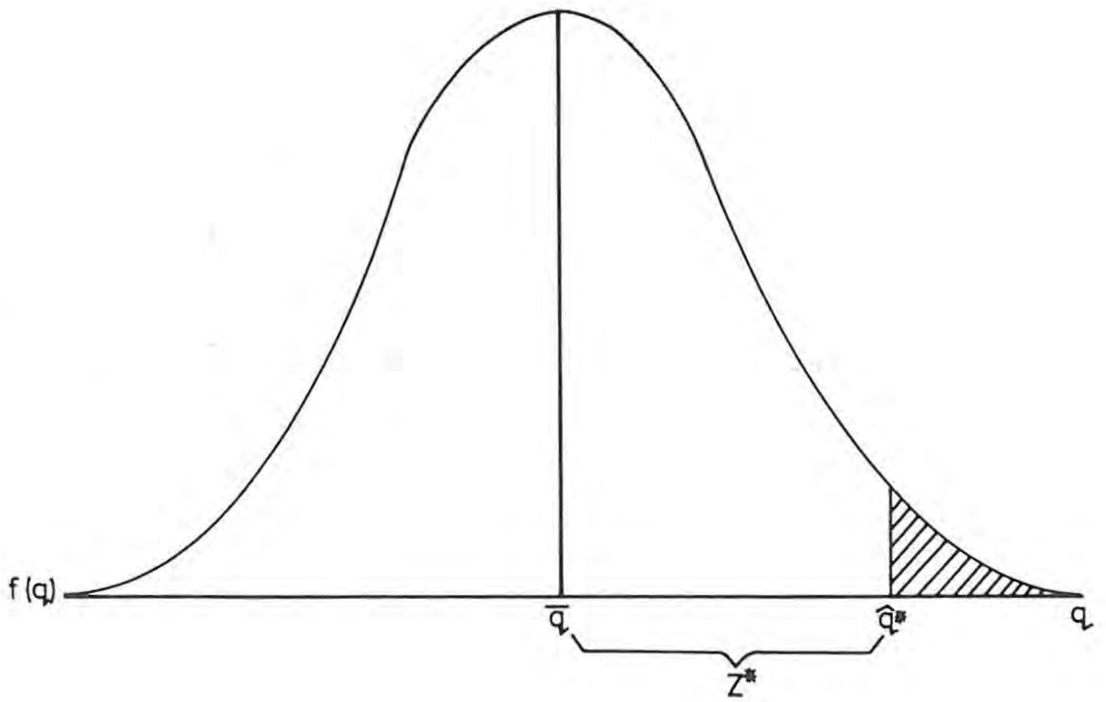


FIGURE 4:1 Probability Distribution of Peak Demands in Standard Weather Conditions

$z^* = \hat{q}^* - \bar{q}$. Consequently there is a margin of installed capacity over the capacity most likely to be required, $K^* - \bar{K}$.

Berrie states that, in practice, the choice of the risk level is based on a "risk standard which enables the good image of electricity at large to be maintained but without committing those consumers to too heavy a capital burden."[7]

There is therefore a trade-off between cost and risk. If the objective of the investment in the capacity margin is to maximize net social benefit then it follows that capacity should increase up to the point where the marginal social benefit of the reduced risk of supply interruption just equals the marginal cost of adding to capacity.

In practice it is thought to be impossible to measure this marginal social benefit, so that the capacity margin decision must be made to maintain "the good image of electricity". In effect this means that the capacity margin is determined by planner's preferences. If adequate control is to be exercised over the public enterprise then this decision should be made by a planning body independent of the daily management of the electricity utility. There are two kinds of effects of power cuts which may concern planners. The immediate effects are public irritation, criticism and political pressure. The longer run effect may be a decline in the growth of electricity demand, as consumers switch to substitutes for electricity whose supply seems to be more reliable. These must obviously be weighed against the costs of spare capacity determining the margin.

---

7. Berrie, T.W., "The Economics of System Planning in Bulk Electricity Supply" in Public Enterprise, R.Turvey ed., Penguin, Harmondsworth, 1968.

4:1:3 THE SYSTEM OF DISTRIBUTION

The system of distribution is much less capable of analytical treatment than generation and transmission, because the concept of overall system optimization has no meaning in the case of distribution. In contrast to generating capacity which can be substituted for capacity in another part of the country, provided that there is sufficient transmission capacity, "each part of the distribution system is specific to the consumers served by it."[8] Consequently there is a significant difference in investment planning in that investment in generation and transmission capacity is related to total load growth and the characteristics of the whole generating system whereas investment in distribution is usually "nothing but the sum of a very large number of individual schemes, each determined by the prospective growth of load in relation to distribution capacity in a particular locality or by a need to replace a particular unsafe or absolete piece of equipment."[9] For example, the capital expenditure required to meet the electricity requirements of a new factory will vary considerably according to the location of the factory in relation to existing supply lines, their voltage level and their present loading. It follows that in contrast with generation and transmission, information about costs cannot be obtained as a by-product of a general optimization procedure so that it is necessary to adopt an inductive approach to costs in the case of distribution.

The function of a particular distribution network is to forward the energy available at each node of the transmission system to final consumers. Figure 4.2 illustrates the distribution network from node B to a number of consumers C1, C2 ... Cn. The capacity to

---

8. Turvey, R., op.cit., p. 68.
9. Ibid., p. 68.

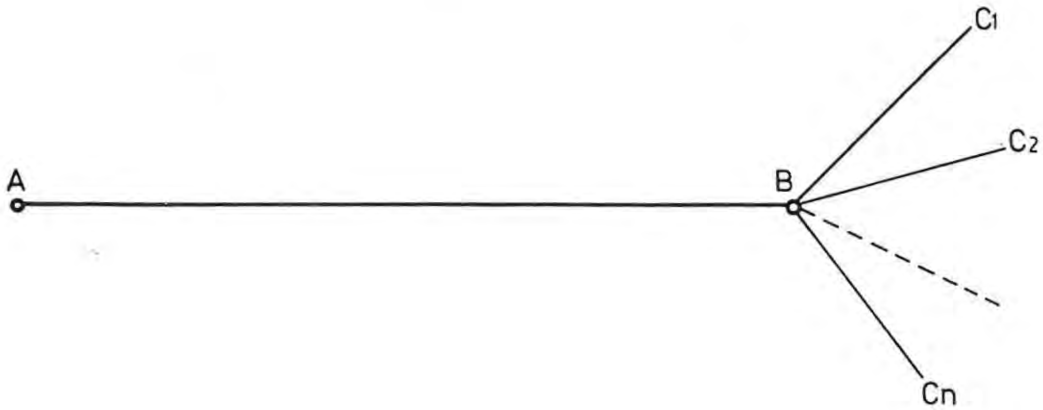


FIGURE 4:2 Distribution Network

be provided for the connection of each client is directly determined by the maximum demand of the client. It is, however, unnecessary to give the transmission line AB a capacity equal to the maximum demands of each consumer. The peak usages of each may be systematically staggered with respect to the others, and, even supposing that they occur at the same time, the uncertain nature of each of the instantaneous demands creates a very slight probability that all customers will use their maximum capacity at the same time.

As was explained in the previous section, the capacity to be installed qo should include capacity z in excess of the most probable demand $\bar{q}$, at the time of the peak, i.e.

$$qc = \bar{q} + z \qquad (4.5)$$

where $$z = Lg \qquad (4.6)$$

Here g designates a parameter characteristic of the random irregularity of the collective demand at the time of the peak[10] and L denotes the Lagrangean multiplier. The capacity margin should become larger with increases in the irregularity g of the demand.

Individual demands are also uncertain and therefore require the distributor to allow for a margin over their average value at peak, $\bar{q}i$. However, this margin is affected not only by an irregularity parameter gOi but also by a correlation parameter Ci measuring the tendency for the accidental peaks of the consumer to coincide with the irregular behaviour of other customers. A low value for Ci thus describes a tendency for the uncertainties of customer i to be partially offset by the set of uncertainties of other consumers,

---

10. The time of the peak will be made up of all hours during which a peak may appear with a certain probability.

while a high value for Ci indicates a tendency for these individual uncertainties to be added to the uncertainties of the collective demand. Consequently, a given irregularity will be responsible for higher irregularity margin to the extent that it is accompanied by a higher value of Ci.[11]

The responsibility of customer i for the collective capacity margin z = Lg will be characterized, up to a factor, by the product of the two parameters Ci and gi i.e.:

$$z = Lg = L \sum_i Cigi \qquad (4.7)$$

The necessary capacity for the line segment AB may therefore be expressed as a function of the characteristics of each individual consumer, in the form:

$$qo = \sum_i (\bar{q}i + LCi.Oi) \qquad (4.8)$$

Now since the collective safety margin is

$$z = Lg$$

and the coefficients Ci which enter into the summation

$$g = \sum_i Ci \; gi$$

are less than 1, it follows that the collective safety margin z is smaller than the sum of the individual safety margins

$$z < \sum zi \qquad (4.9)$$

11. The mathematical equivalents of the parameters $\bar{q}i$, Oi Ci are the mean, the standard deviation and the coefficient of correlation, respectively.

This implies that the capacity necessary for the line segment feeding a node is always less then the sum of the capacities being fed. This phenomenon of reduction due to "diversity" becomes more important for a distribution network serving a large number of individual demands.

4:2 THE DEMAND FOR ELECTRICITY

4:2:3 DETERMINANTS OF ELECTRICITY DEMAND

The demand for electricity is a derived demand. Electricity is not demanded for itself but for how it can be used by suitable equipment to convert electrical energy into other forms of energy such as heat, light, sound and motion. The use of electricity is therefore usually not possible without expenditure by the consumer on electrical appliances and complementary goods. Thus as Hermann has noted "..... (the) demand for electricity and the elasticity of (this) demand with regard to both price and income, is consequently determined not only by the absolute level of the price of electricity as such but also by the cost of appliances, the uniqueness of the function for which the appliance is used, the existence of substitute appliances using other forms of energy and the price of competitive forms of energy."[12] Furthermore, it is also often the case that equipment which uses electricity as its energy source is electricity specific and cannot use any other alternative source of energy. This means that in considering the individual consumption behaviour which underlies the domestic consumer's demand for electricity one must distinguish between the individual consumer's decision about whether or not to purchase an electrical appliance and his decision about the amount of

---

12. Hermann, G.F.K., Electricity Tariffs, Unpublished D. Comm. Thesis, University of Stellenbosch, 1974, p. 8-1.

electricity to use once the appliance has been acquired. A similar distinction can be made with respect to the individual decisions which underly the demand for electricity for industrial, commercial, mining and traction uses. Here electricity is regarded as an input into the production process and the decision maker will often be concerned with the price of electricity insofar as it affects, firstly, the choice between a finite number of production processes with different electricity requirements and, secondly, the amount of electricity to be used by a particular production process once it has been adopted.

The influence of the price of electricity on the decision to purchase an appliance or to adopt a particular production process will depend on the cost of electricity in relation to the cost of equipment and the cost of alternative energy used by substitute types of equipment. The cost of electricity is related to both the price of electricity and the amount of electricity required to use the equipment, and it is generally the case that the demand for electrical equipment is inelastic with respect to the price of electricity where the equipment has low energy requirements, so the cost of electricity is low in relation to the cost of the equipment. Examples of appliances used by domestic consumers where the decision to acquire the appliance is almost completely uninfluenced by the price of electricity are T.V., radio, sound reproduction equipment and various small appliances used in the home. However with appliances used in cooking, food preparation, water and space heating and air conditioning, the cost of electricity is high in relation to the cost of equipment and the same functions can be performed by substitute appliances using other forms of energy so that the price of electricity may become an important factor in the decision to acquire such equipment.

Once an electricity appliance has been purchased or the electrical equipment used in a production process has been installed the price of electricity can usually only affect the amount of electricity used by the particular electrical equipment. In cases where a minimum amount of electricity is required for the normal operation of the equipment so that a reduction in the electrical power used below this minimum would prevent the equipment from adequately performing the function for which it was purchased, the amount of electricity used will be highly inelastic with respect to changes in the price of electricity. A refrigerator is an example of an appliance for which electricity demand becomes highly price inelastic once it has been purchased since a fixed amount of electricity is necessary for its regular use. Equipment used for water-heating or air conditioning provide examples of electrical equipment for which there is a much less definite limit to the amount of energy that may be demanded by the consumer so that changes in the price of electricity influence its consumption by these types of equipment even after they have been purchased.

An individual decision maker's demand for electricity is thus determined by the energy requirements of the electrical equipment he uses or plans to use. It is very difficult to graphically represent on individual's electricity demand since much of this demand consists of energy blocks of fixed width within which the quantity demanded is almost completely inelastic. As Hermann has commented:

> "The blocks of demand do not form any continuous pattern and the number of demand blocks will depend on the appliances available to the consumer."[13]

The discontinuities which characterize individual electricity demand will however tend to disappear when one considers groups of consumers; this is due to the factors such as the variation in consumer habits, consumption patterns relative demand intensities and the law of large numbers, and one could expect a group demand curve to be a conventional continuous downward sloping function. The pattern of individual demand should, however, be reflected in the tariff structure since electricity tariffs affect each consumer individually, which is one of the reasons

---

13. *Ibid*., p. 8-4.

why many electricity utilities adopt a two-part tariff structure is because it causes the average price of electricity paid by each consumer to fall as he increases his electricity consumption and therefore discriminates against his price inelastic demand in favour of his price elastic demand for electricity. In other words, each consumer will consume a minimum amount of electricity in uses for which the cost of electricity is low in relation to equipment cost and there are no alternative energy sources and will be unlikely to reduce his electricity consumption below this minimum amount even when the imposition of a two-part tariff results in him paying a high average price for this first block of consumption. Furthermore this two-part tariff will encourage the consumer to expand his consumption of electricity in uses where there are available substitutes since this will result in a reduction in the average price he pays for electricity.

4:2:2 SECTORAL CONSUMPTION PATTERNS

Since electricity cannot be stored, any unexpected fluctuation in its demand will directly affect the utilization of the system. It follows that in formulating its tariff and capital development policy an electricity utility needs to ascertain whether there are any regularities in the pattern of demand and consumption of the main consumer groups. The consumption and demand pattern of industrial consumers is normally dependent on the nature of production processes, the degree of mechanisation, the type of machinery used and the number of shifts worked. In the case of commercial consumers, the incidence of peak and off-peak periods is determined by shop and office hours, lighting conditions, seasons, displays and the like, while the peak energy demand of electricity used for traction will, for example, depend on the rail traffic flow. It is usually not difficult to ascertain the expected consumption pattern for such consumer categories. However, in the case of domestic consumers, it is usually more difficult to predict the pattern of consumption since, as Hermann has suggested, this "is likely to vary under the influence of social changes, types of entertainment available fashion, eating habits, climate etc." It is therefore necessary for an electricity utility to frequently investigate changes in these consumption patterns to determine whether its estimates of peak and off-peak, daily and seasonal electricity usage are reasonable.

4:2:3 GROWTH IN THE DEMAND FOR ELECTRICITY

Growth in the demand for electricity may be caused by either an increased demand from existing consumers or due to the connection of new consumers. In a dual economy such as South Africa, there is still scope for growth in electricity demand as electricity supply facilities are made available to the economically less developed groups of the population. An example of this would be the proposed electrification of Soweto. In the developed sector of the economy, however, demand growth is likely to almost entirely arise from increased consumption by existing consumers, which may be due either to increased use of existing appliances or the introduction of new uses. It is now necessary to bring this analysis of the demand and supply of electricity together to consider how an electricity utility can practically implement a marginal cost pricing policy.

4:3 THE FORMULATION OF A MARGINAL COST-BASED TARIFF STRUCTURE

It will be recalled from Chapter Two that the adoption of a marginal cost pricing policy by a public enterprise would basically involve applying the following rules:

(i) setting price equal to short run marginal cost (SRMC) when there is excess capacity; and

(ii) allowing price to ration demand to the level of fixed capacity when capacity is fully utilized.

When the system is optimally planned and operated in the manner described in section 4:1, then the price set according to these rules will just cover long run marginal cost (LRMC). In this section the following stages in the formulation of a marginal cost based tariff structure will be described:

(a) the identification and calculation of the various components of long run marginal cost;

(b) the application of rules (i) and (ii) above to derive a tariff structure designed to cover LRMC given the typical pattern of electricity demand and subject to metering costs; and the adjustments which should be made to a marginal cost based tariff structure if it is to reflect financial and second best criteria.

4.3.2 THE CALCULATION OF LONG RUN MARGINAL COST

In the context of electricity supply, Munasinghe defined long run marginal cost (LRMC) as "the incremental cost of optimum adjustments in the system expansion plan and system operations attributable to an incremental demand increase which is sustained into the future."[14] Three broad categories of marginal costs may be indentified for the purposes of calculating LRMC: capacity costs, energy costs and consumer costs. Marginal capacity costs are the investment costs of generation, transmission and distribution facilities associated with providing additional kW of capacity. Marginal energy costs are the fuel and operating costs of supplying additional kWh of electricity consumption. Marginal consumer costs are the incremental costs directly attributable to individual consumers including costs of connection, metering and invoicing. In the first three parts of this section the nature and method of calculation of these components of LRMC will be described while the final part will deal with the use of shadow pricing to ensure that LRMC reflects economic resource costs.

---

14. Munasinghe, M., Electric Power Pricing, Staff Working Paper, World Bank, Washington, July 1979, p. 25.

A. ENERGY COSTS

Boiteux and Stasi[15] have shown that where a system of electricity generation and transmission is optimally planned and operated, it is sufficient to know the marginal energy costs at one node of the network and the optimum configuration of energy movements to deduce, step by step, the marginal cost of energy at all nodes of the network by addition or subtraction of marginal transmission costs. The procedure involved in calculating the marginal energy costs at each node might therefore in practice involve the following steps:

(i) given the load curves anticipated for each node, and their forecasted development through time, it would be necessary to determine the facilities to be installed and the corresponding configurations of energy movements;

(ii) the corresponding marginal energy costs at different hours and seasons of the year would have to be calculated for a selected node; and

(iii) these marginal costs would have to be prepagated from node to node by the addition or subtraction of tramsission losses.

Since the first part of this calculation would be determined independently of the tariff making process as part of the capital development programme of the electricity utility, the calculation of the long run marginal energy costs is reduced to the calculations mentioned in (ii) and (iii) above. These calculations would, however be subject to various cross checks since, as has already been noted,

---

15. Boiteux, M., and Stasi, P., "The Determination of Costs of Expansion of an Interconnected System of Production and Distribution of Electricity" in Nelson, J. R., ed., Marginal Cost Pricing in Practice, Prentice-Hall, New York, 1964.

they rely on the properties of a theoretical optimum which may not exist in practice. The principles of these cross checks will be explained after the main features of the calculation are indicated.

(i) CALCULATION OF MARGINAL ENERGY COSTS AT A NODE OF THE TRANSMISSION NETWORK

The first step in the calculation of marginal energy costs is to choose a base node. It appears that the results obtained will gain significance in proportion to the importance of the node selected and to the extent that generation for the transmission network is predominant at this mode relative to generation for local use or relative to imports.

Let us first consider, for example, the centre of thermal production for South Africa at the Transvaal coalfields. At a particular time on a particular day a given maximum demand must be supplied by the thermal power stations of this region. To satisfy this demand for power at least cost, the load dispatching service for the region is arranged so as to require the most efficient stations to work at full capacity. Only the least efficient power station which must be maintained in operation to meet the total demand for power from thermal stations does not operate at full capacity, since there is no reason for the power demanded to coincide exactly with the productive capacity of an integral number of stations. This last power station will be denoted the "marginal stations". If at this time, a call for extra power is made, this extra amount will be produced by the marginal stations since the more economical stations are already working at full capacity while the other stations are less economical. The cost of an additional kWh is therefore equal at this time to the cost of coal required for the production of kWh by the marginal station. Generally, it follows that at every instant, the marginal cost of production of a kWh in the Transvaal region is equal to the proportional cost of the station which is marginal at that instant.

It follows that, in principle, it is sufficient to ask the dispatching centre which station is marginal for every hour of the year, given the power demanded from the set of thermal plants in the region, in order to deduce the schedule of the 8760 marginal generating costs.

Such a schedule would, however, be of little use to tariff makers since it would be highly impractical to offer a tariff whose price varies from hour to hour. It is therefore necessary to calculate the average marginal energy cost of generation over appropriate rating periods which are selected by examining the system load duration curves to determine the periods during which demand presses on capacity. Studies in France, by Boiteux,[16] and in Britain, by Turvey[17] and Crew[18], indicate that this schedule of the marginal energy costs of generation should be differentiated in both winter and summer between "peak" "full use" and "slack" hours. For the purposes of simplifying the analysis of long run marginal cost it will be assumed that the electricity system does not exhibit marked seasonal variations of demand and that there are only two rating periods by time of day, that is, peak and off-peak periods.

---

16. Boiteux, M., "The Tarif Vert of Electricitè de France" in Nelson, J.R., ed., op.cit.
17. Turvey, R., op.cit.
18. Crew, "Electricity Tariffs" in Turvey, R., op.cit., pp. 258-283.

(ii) THE PROPAGATION OF MARGINAL COSTS

Figure 4.3, illustrates how the marginal cost of energy in peak hours at node P of the network may be calculated, given the marginal energy cost at node A. Since the energy movements in this peak period are known, the directions in which the energy moves between the various nodes A, B, C, ... P and the position of the neutral points (N) are known. By virtue of the property of optimum configurations of energy movements, which has already been discussed the marginal energy cost at B is this marginal energy cost at A, increased by the marginal transmission cost AB. In a like manner, the marginal cost at N can be calculated as the sum of the marginal cost of B and the marginal transmission cost along BN. From N to C the movement is in the direction opposite to the flow of energy so that the marginal cost at C is the marginal cost at N reduced by the marginal transmission cost CN and so on.

The marginal cost of transmitting an additonal kWh along a segment such as AB is the sum of the transmission losses along the line, which are proportional to the square of the current and the deterioration in customer service resulting from the drop in voltage. An additional kWh will, however, only cause a voltage drop when the line is operating at full load. The marginal cost of transmission must therefore be increased at times of full capacity by a term equal to the distribution of the fixed charges of expansion over the kilo watt-hours transmitted.

(iii) CROSS CHECKS AND TESTS

The method of calculation described above lends itself to a number of checks:

(a) the marginal price calculated at, say, point P from point A in figure 4.3 should coincide with the proportional cost of the marginal station at P.

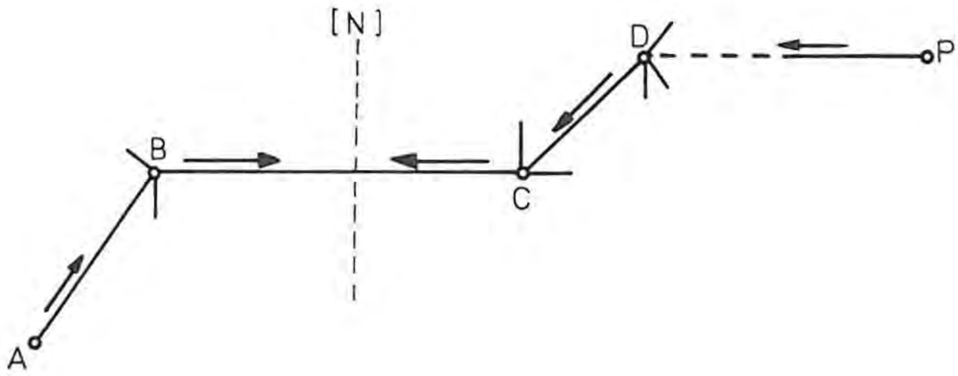


FIGURE 4:3 Energy Movements Along a Transmission Network

The same will apply at every point on the transmission network;

(b) whatever the route followed from one node to another, after adding or subtracting transmission costs as dictated by the energy movements encountered on each route, the delivered price should be the same; and

(c) the schedule of marginal prices calculated for a given region should coincide with the operating cost of the new facilities which are being installed in the region.

If there is any lag in new installations in a region, it will appear as an excessive importer since its costs will rise relative to other regions. By charging for the production of the new plants in a region according to its schedule of marginal costs, on average, a satisfactory equilibrium should be obtained between the annual value of energy produced by each plant, and the annual cost of this production.

Finally, it can be seen that the LRMC of energy during the peak period will be equal to the running cost of the plant used last in merit order to meet an increment in peak period electricity consumption adjusted by the appropriate peak transmission loss factors at each voltage level along the transmission network. Similarly, the LRMC of energy resulting from a load increment during the off-peak period is usually equal to the running costs of the least efficient base load plant used during this period also adjusted for the transmission losses along the network. It should be noted that the transmission loss factors for adjusting off-peak costs will be smaller than the peak period loss factors, since these losses are a function of the square of the current-and current flows are greatest during the peak period.

B. MARGINAL CAPACITY COST

The LRMC of capacity can be ascertained by determining the change in system capacity costs resulting from a sustained increment in long run peak demand. The effect of a given increment in demand is illustrated in Figure 4.4. It can be seen that it would be necessary to change the expansion programme of the system to meet the new incremental load. In calculating the LRMC of generating capcity (LRMC gc) it is important to ascertain whether this change in the expansion programme is effected by either advancing the commissioning date of generating units designed to meet the base load or by installing units which are designed to meet only the future peak load. In many systems gas turbines which have a relatively low capital cost and high operating costs are used to meet peak demand. If this is the case, then the LRMC gc can be estimated in terms of the cost per kW installed of gas turbines annuitized over their expected lifetime, (AC). This figure should be adjusted for the reserve margin (RM%) and losses due to station use (Lsu%) to derive the following expression for the LRMC of generating capacity:

$$\text{LRMC gc} = \text{AC}. \ (1 + \text{RM}/100)(1 - \text{Lsu}/100) \qquad (4.10)$$

If, however, the incremental load is to be met by advancing the installation of base load plant then it should be noted that this plant has a relatively high capital cost and low operating cost, so that it is important to deduct from the adjusted annuitized cost per kW installed of this plant, the potential fuel savings which are expected to result from the displacement of less efficient plant by these new base load units.

The LRMC of the transmission and distribution network should be allocated to incremental capacity, since the designs of these facilities are determined primarily by the peak kW that they carry rather than by kWh electricity requirements. The capacity costs of transmission and distribution facilities must be identified since the consumers at each voltage level should be charged only upstream cost. Thus if, for example, there are three supply voltages categories: high, medium and low, the LRMC of capacity at the high voltage (HV) level would be.

$$\text{LRMC HV Cap} = \text{LRMC Gen. Cap.} / (1 - \text{LHV}/100) + \Delta\,\text{LRMC HV} \qquad (4.11)$$

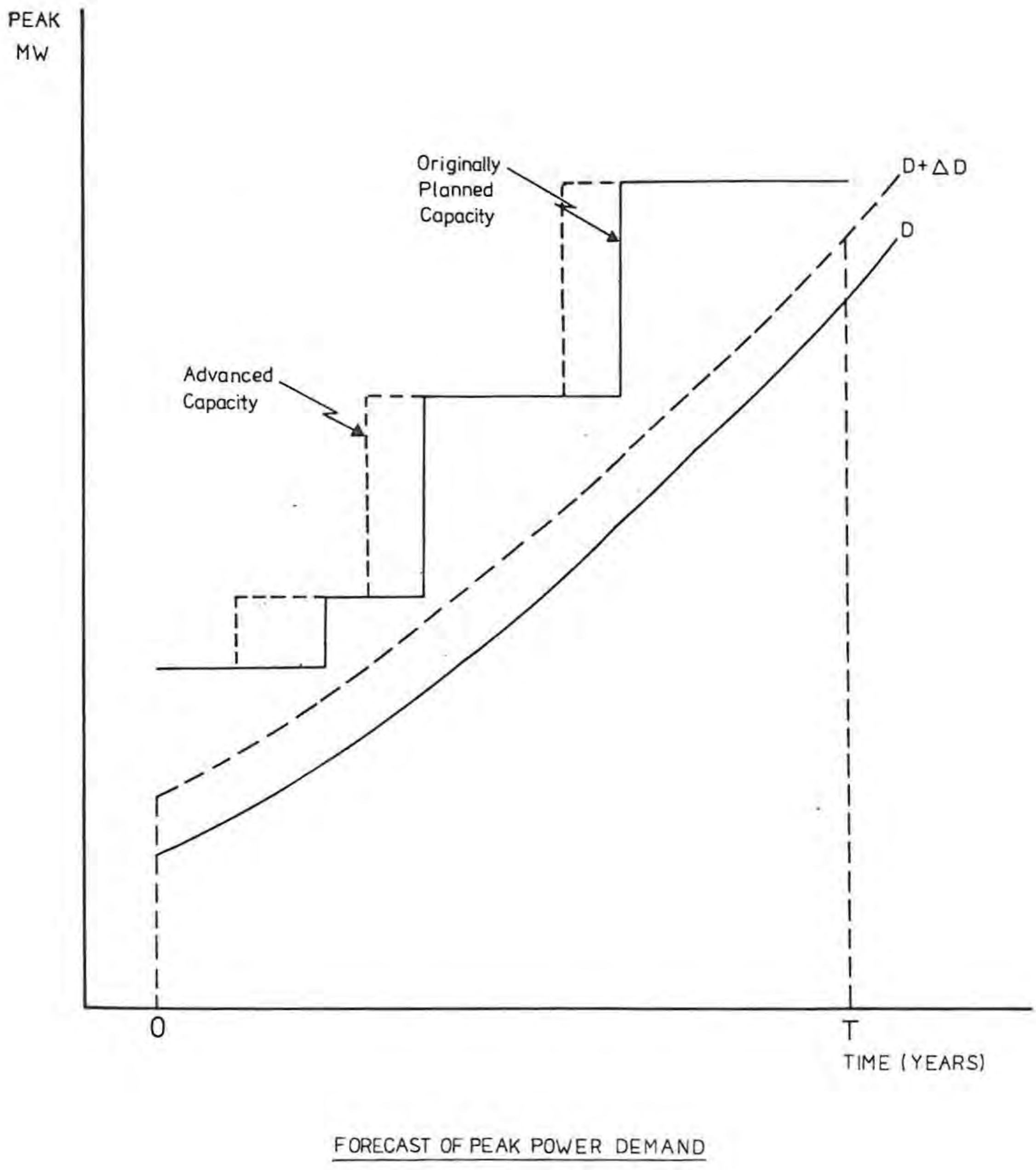


FIGURE 4:4 The Effect of a Sustained Increment in Long Run Peak Demand on System Capacity Costs

where Δ LRMC HV is the average incremental costs of the extra high voltage (EHV) and HV transmission facilities annuitized over the lifetime of the plant and LHV% is the percentage of incoming peak power that is lost in the EHV and HV network. Similarly the LRMC of capacity to medium voltage consumers could be calculated as follows:

$$\text{LRMC MV Cap} = \text{LRMC HV Cap.} / 1 - \text{LMV}/100 + \Delta \text{ LRMC MV} \quad (4.12)$$

where ΔLRMC MV is the annuitized average incremental cost of incremental medium voltage capacity such as distribution substations and primary feeders while LMV% is the percentage of incoming peak power that is lost at the MV level.

In attributing these capacity costs to consumers it may be useful to distinguish, as Boiteux and Stasi have done, between the costs of the "collective", "semi individual" and "individual" network. The collective network includes the facilities which are located so far upstream from an individual consumer's connection that the random characteristics of the consumer's behaviour play an insignificant role in determining their capacity since these facilities will be supplying a large number of consumers. The capacity of this section of the network will therefore depend primarily on the average consumption of consumers at the time of the collective peak so that the costs of this capacity should be allocated according to the average kWh consumed during the peak period. The semi-individual network usually serves a small number of consumers of similar size so that the uncertainty of consumption of an individual consumer may be a significant factor in determining the capacity of these facilities. Furthermore the load peaks of those local distribution networks may differ from the system peak. Boiteux and Stasi suggested that the effect of local peaks and irregular individual demand on the capacity of the semi-individual network can be best allowed for when the cost of this capacity is allocated according to the contracted maximum power demand of the customers. The individual network includes those facilities which are specifically provided for individual consumers. The costs of these facilities will form part of the consumer-related costs to be discussed in the next sub-section.

C. CONSUMER COSTS

Consumer costs are those costs which are incurred to supply individual consumers. They are therefore allocated directly to the individual consumer for whom they are incurred. These costs consist of non-recurrent expenses such as service drop lines, meters and the labour used for installation which may be charged to the customer as a lump sum or distributed payments over several years and recurrent expenses such as those incurred for meter reading, billing and administration. These expenses are in the nature of a fixed cost which does not vary significantly with either the kWh consumption or contracted kW demand of consumers, so that they form a relatively small proportion of the total costs incurred in supplying large consumers.

D. SHADOW PRICING

The market prices of the inputs used by an electricity utility may diverge from their marginal social costs due to monopoly practices, external economies and diseconomies and interventions in the market process through taxes, import duties and subsidies. It may thus be necessary to use the "shadow prices" of these inputs which measure their economic resource cost, both to calculate the LRMC of electricity and to determine an optimal investment programme. It is usually necessary to adopt a common yardstick or numéraire which provides a measure of economic value in order to derive a consistent set of shadow prices for the inputs used by the utility. Little and Mirrlees[16] have proposed that an appropriate numéraire should be a unit of uncommitted public income at border prices, expressed in

---

16. Little, I.M.D. and Mirrlees, J.A., Project Appraisal and Planning for Developing Countries, Basic Books, New York, 1974, Chapter 9.

real terms, which is basically equal to a unit of freely disposable foreign exchange available to the government but expressed in terms of units of local currency converted at the official exchange rate and deflated by an appropriate price index. By using such a numéraire the distortions caused by inflation, taxes, duties and subsidies are removed.

To facilitate the estimation and use of shadow prices, the inputs used by the utility should be categorized into tradable and non-tradable items. The most important tradable inputs used by an electricity utility are usually capital goods and petroleum based fuels. The shadow pricing of such inputs presents little problem, since either the world market or border prices may be used where appropriate. In the case of non-tradable items which are locally purchased inputs whose values are known only in terms of domestic market prices, it is necessary to convert these market prices into border shadow prices by multiplying them by appropriate conversion factors. Before the conversion factor can be applied it may first be necessary to adjust the domestic market prices of these inputs, so that they reflect their marginal social costs of production. For example, the resource costs of extraction should be used for coal which has no significant export market while the market price of urban land may be a good indication of its economic value unless it is used for agricultural or recreational purposes. It should be noted too that the rate of interest used by the utility should be adjusted in the manner indicated by (2.20) which takes into account the fact that the tax exemption of public enterprise income results in investment by these enterprises displacing private consumption and private enterprise investment at different rates. In other words, the interest rate used by the utility should reflect the opportunity cost of capital.

In the case of non-tradables which are not important enough to merit individual attention or for which there is insufficient data, a standard conversion factor (SCF) may be used. The SCF is equal to the official exchange rate divided by a shadow exchange rate appropriately defined. A useful approximation of this shadow exchange rate may be the free trade exchange rate when the country is moving towards a free trade regime, since this will reflect the average level of import duties and export subsidies. The SCF calculated on the basis of this exchange rate will usually be less than unity. It follows that the division of LRMC calculated according to the border prices of inputs by this standard conversion factor will result in an optimal market price for electricity which is greater then this border priced LRMC.

Once the LRMC of electricity production has been calculated according to the principles described in this section, it is necessary to devise a tariff structure which will result in the recovery from each consumer or consumer group of the LRMC incurred in supplying their electricity tariffs. The principles underlying the formulation of a tariff schedule according to the criterion of economic efficiency will be discussed in the next section.

4:3:2 THE FORMULATION OF AN OPTIMAL TARIFF STRUCTURE

It will be recalled from the specification of the peak load pricing model in Chapter Two, that the optimal utilization of the capacity of an electrical power system will be attained when the recovery of capacity costs through tariffs is limited to those periods when the charging of a price equal to marginal running costs would result in the demand for electricity being greater then the capacity of the system. This is illustrated in Figure 4.5 which is drawn on the assumption that demand fluctuates on a regular daily basis between peak and off-peak periods. It can be seen from Figure 4.5 that if a marginal cost pricing policy is adopted by the electricity utility then the price of electricity will be set in the off-peak period at p2 which is equal to unit running cost r, and allowed to rise to p1 to restrict demand to fixed capacity $\bar{q}o$, in the peak period. The recovery of capital costs, k, will therefore only be recovered in peak periods and total capital costs will only be fully recovered if an investment policy is followed whereby capacity is adjusted to the optimal level indicated at $\bar{q}o$ where the peak price, p1, is equal to LRMC. The basic implication of this peak load pricing model for the tariff policy of an electricity utility is therefore that "the off peak consumer should not make any contribution to capacity

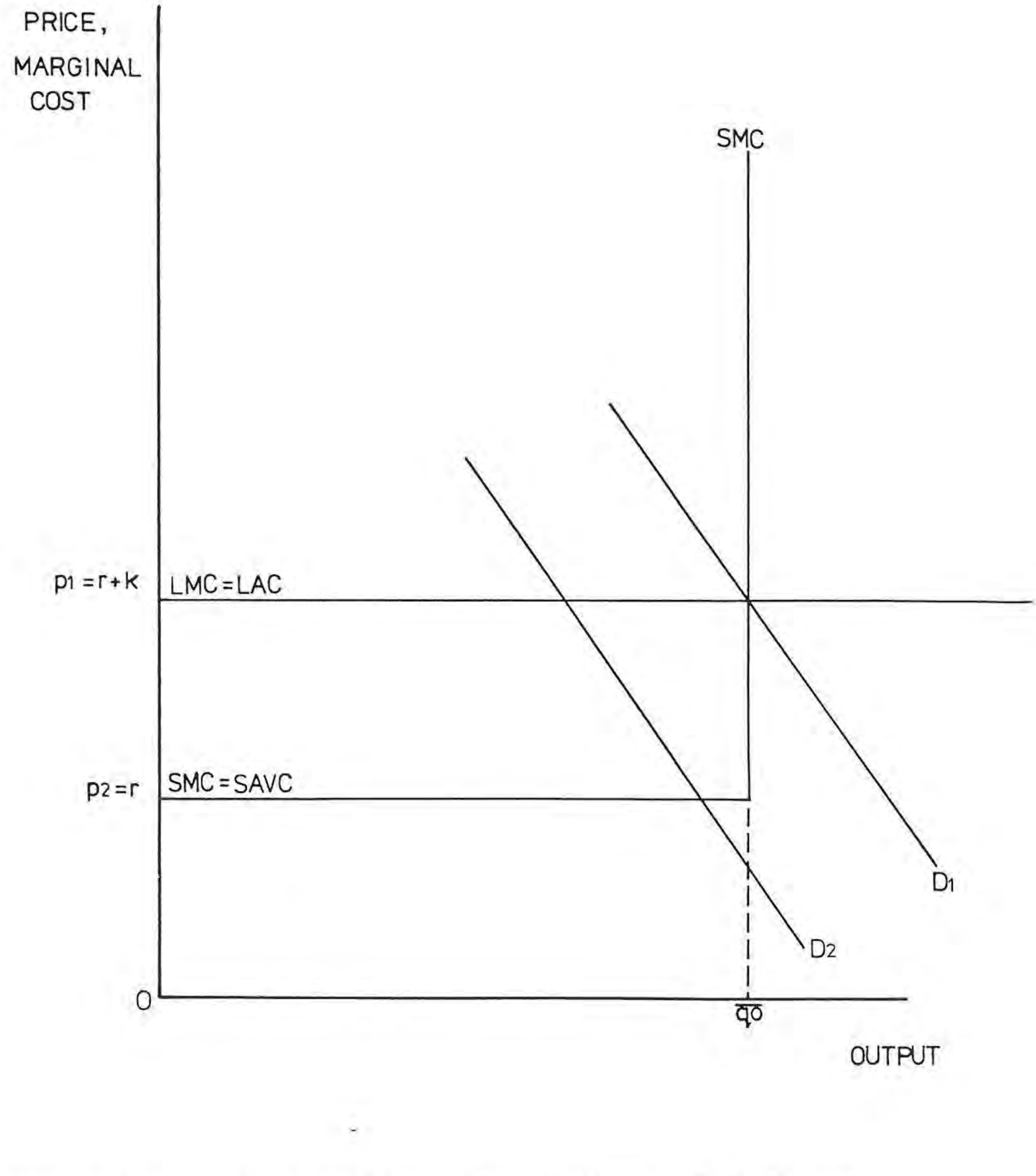


FIGURE 4:5 Recovery of Capacity Costs Through Peak Period Tariff Differentiation

cost, which should be apportioned entirely to the peak consumer; and if more than one consumer uses the total capacity available, the several consumers should contribute to the cost of capacity in proportion of the intensities of their demands."[17] This conclusion is, however, crucially dependent on the following assumptions which underly the simple model illustrated in Figure 4.5:

(i) running costs are constant over the finite output of the system;

(ii) capacity costs are constant per unit of capacity;

(iii) finite increments in capacity are perfectly divisible;

(iv) demand is independent as between different consumers and time periods; and

(v) LRMC is constant.

It will be shown that the relaxation of these assumptions makes it necessary to reassess the implications of the peak load pricing model for electricity tariff policy.

The assumption that running costs are constant over the finite output of the system appears to be unrealistic, particular in view of the discussion of the generating system in section 4.1, where it was suggested that an optimal mix of generating plant would contain a number of units with differing operating costs and that such a system should be operated in merit order with the plant with the lowest unit running costs being used first to meet the demand for power. The short run marginal cost of generation (SRMC gen) curve can therefore be more realistically depicted as a rising stepped function as shown in Figure 4.6. It is assumed that demand fluctuates on a regular daily basis between peak, full-use and slack periods represented by the demand curves D1, D2, and D3 respectively. It can be seen from Figure 4.6 that D1 intersects the SRMC gen curve in its vertical portion which indicates that if a marginal cost pricing policy is adopted the price of electricity should be set at p1 to limit the load

---

17. Hermann, G.F.K., op.cit., p. 7-2.

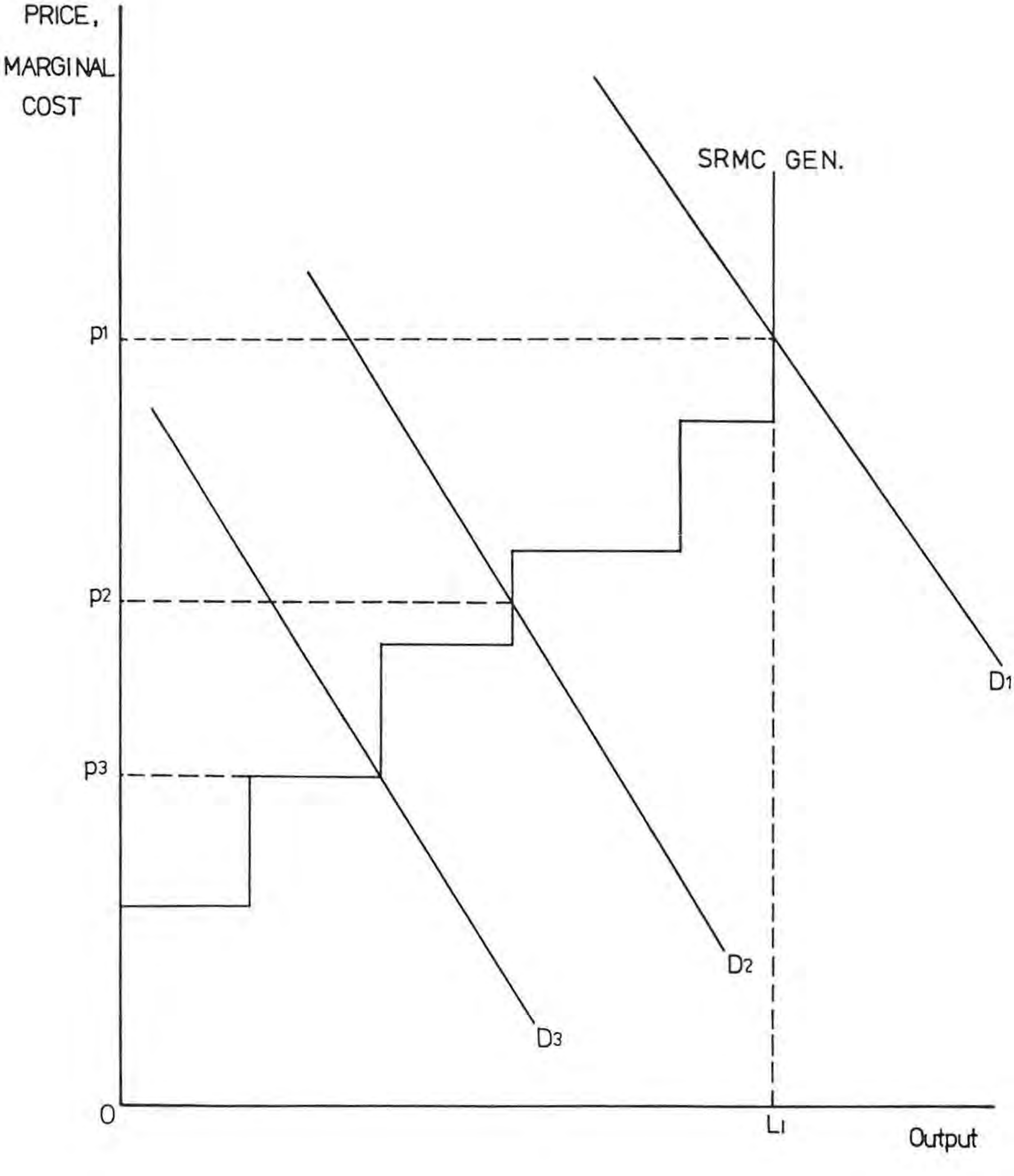


FIGURE 4:6 The SRMC gen Curve as a Stepped Function

to its limit L1 which is defined by the capacity of the system. It is clear that the greatest contribution to generating capacity costs will be made in the peak period but it no longer follows that there will be no recovery of these costs in the other periods if the marginal cost pricing rules are applied. It can be seen from Figure 4.6 that, in both the full-use and slack periods, the setting of price equal to SRMC gen results in the generation of revenue which is greater that the operating costs incurred so that a contribution to capacity costs is made in both periods. The relaxation of the assumption that there are constant unit running costs may therefore strengthen the case for time-of-day differentiation of tariffs, since it means that capacity cost recovery is no longer restricted to one period but may, in fact, be spread out over other periods if marginal cost principles are applied. Furthermore if the other assumptions of the peak load pricing model hold, it still follows that an optimal investment policy will result in capacity being adjusted to the level where total revenue covers total costs. It will now be shown, though, that the relaxation of some of these assumptions may prevent an electricity utility from preserving a balance between total revenue and total cost if it pursues a marginal cost pricing and investment policy.

If the capacity of the system can only be expanded in indivisible units then it is clear that there may be times when total revenue will not cover total cost if the marginal cost pricing rules are applied. However, as Hermann has noted:

> "In the case of generating plant the assumption of perfect divisibility is not important. In a reasonably large utility capacity is added by constructing additional power stations with several generating units, each of which, if only from stand by considerations will represent a demand increment of perhaps three or four percent. With a growth rate of eight to ten per-cent per annum, any surplus capacity due to indivisibility will therefore be of very short duration in relation to the life of the plant."[18]

18. <u>Ibid</u>., p. 7-4.

It follows that a marginal cost pricing policy will result in the recovery of most of the capacity costs associated with generating plant. This, however, is unlikely to be the case in the case of the facilities used in the transmission and distribution network. This equipment is much more likely to be characterized by indivisibility since it is usually designed to allow for several years' load growth and may therefore have a substantial amount of excess capacity for much of its life. As was explained in the previous section, the cost of carrying an incremental load along an underutilized transmission and distribution network is usually limited to the cost of energy losses along the power lines. If only this cost is charged to the consumer it is clear that a substantial portion of the cost of transmission and distribution capacity will not be recovered from tariff revenue. It can thus be seen that where indivisibility exists, a small proportion of generating and a relatively larger portion of transmission and distribution costs may not be recovered.

The application of marginal cost pricing rules is also likely to lead to a residual of unrecovered capacity costs where the utility allows for uncertainty by providing for more capacity than is necessary to meet future forecasted load requirements and where LRMC is a declining rather then a constant function due to economies of scale. The question which now arises is whether these residual capacity costs which result from the presence of indivisibility, uncertainty or economies of scale should be apportioned to peak consumers as an extension of the basic principle which underlies the peak load pricing model. The main difficulty associated with allocating these residual costs entirely to peak period consumers is that the level of price in one period may influence the demand in other periods so that this may create a "shifting peak" problem. In other words if, say, residual capacity costs are charged to peak consumers whose demand function is represented by D1 in Figure 4.6, then this may result in an inward shift of the curve D1 and an outward shift of D2 with the result that period 2 may become the new peak period. To obviate this shifting peak problem it has been suggested in a report by the British Electrical Research Association that the optimal method of allocating residual capacity costs between time periods may be as follows:

> "There will be certain times when, no matter how small a charge is made (even zero) in respect of demand related cost or whatever likely changes occur due to other causes, some plant will still be lying idle. This is the off-potential peak period, and there is then no liability for plant expenses, and no demand related cost need be allocated. During all other times there is some potential of an absolute peak arising, and cost must be allocated according to this liability. Throughout this potential-peak period there is a graduated peak responsibility or potentiality, measured by the price differentiation which would tend to produce a constant height load curve."[19]

In other words, the objective of the tariff maker in allocating residual capacity costs to different time periods should be to smooth out demand fluctuations over time. The desired effect of this method of cost apportionment is shown in Figure 4.7 where the consumer's load curve attains a plateau over the main hours of daily use instead of the typical peaking of electricity requirements for a short period around 8.00 a.m. and 5.00 p.m.

These basic principles of marginal cost pricing provide a useful indication of the main factors to be taken into account by an electricity utility in structuring its tariffs, It is now necessary to examine how a practicable tariff structure can be devised which is based on these principles. The following discussion will frequently refer to the tariff-making practices of Electricitè de France which, perhaps more than any other electricity utility in the world, has attempted to base the structure of its tariffs on marginal cost pricing principles. Meek has pointed out with respect to this tariff policy that "the traditional concept of marginal cost pricing, already somewhat

---

19. British Electrical and Allied Industries Research Association Report No. K/T 109, 1945: An improved method of allocating to classes of consumers the demand related portion of the standing cost.

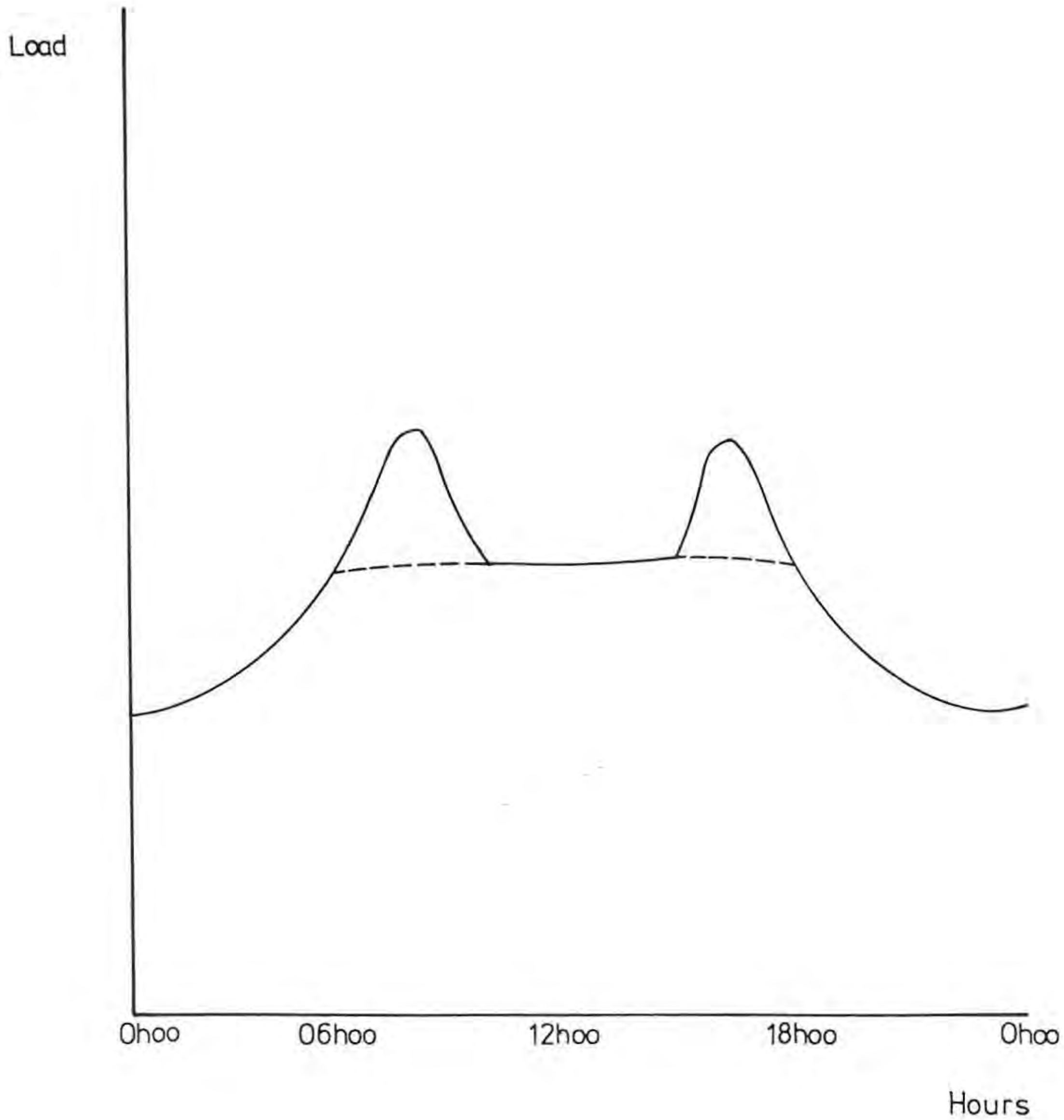


FIGURE 4:7 Typical Daily Load Curve of an Electricity Undertaking

attenuated in certain respects in Boiteux's theory, becomes much more attenuated in practice – so much so, indeed, that one wonders whether it is proper to call the final result marginal cost pricing at all".[20] In spite of this, the view has been taken that the purpose of a marginal cost approach to pricing policy is not to attain strict conformity with the conditions of a theoretical optimum since this is usually not completely practicable but rather to indicate the direction in which the pattern and structure of rates should be rearranged to achieve greater allocative efficiency. As Vickrey has pointed out:

> "The dominant issue is one of whether the pattern of rates should be based on tradition, inertia and happenstance, or whether it is to be developed by a careful weighing of the relevant factors with a view to guiding consumers to make as efficient use of the facilities which are available".[21]

The structuring of tariffs by Electricitè de France is largely acceptable in terms of this approach.

In many electricity utilities consumers are charged according to a three-part tariff with a fixed service charge per consumer covering consumer related costs, a charge which is fixed per kWh consumption based on the maximum demand in kW of the consumer which is designed to cover at least a portion of the demand related capacity costs, and a charge per kWh which covers energy costs and a portion of demand related capacity costs. A marginal cost based tariff schedule is also likely to take the form of a three-part tariff although there are likely to be additional parameters taken into account in structuring the demand and, in particular, the energy component of the tariff so that they vary according to, say, the marginal costs of serving demands:

---

20. Meek, R.L., "The New Bulk Supply Tariff for Electricity", Economic Journal, Vol. - 78, 1968, p. 55.
21. Vickrey, W., "Some objections to marginal cost pricing", Journal of Political Economy, June 1948.

(i) by different consumer categories;
(ii) in different geographical areas;
(iii) at different hours of the day;
(iv) in different seasons;
(v) by different voltage levels; and
(vi) with different load factors.

It is clear that the marginal cost of supplying electricity will vary between consumers on an individual basis. However, since an electricity utility serves a large number of consumers it is usually necessary to classify consumers into different categories with basically similar consumption patterns so as to minimize the deviation from the norm in the averaging processes that are required. An initial basis for such a classification would probably be a regional one, since the marginal cost of electricity distribution will often be higher in geographical areas where a relatively smaller number of customers are served per square kilometre of distribution network than in other areas. Regional differences in load density are likely to affect all the components of the electricity tariff with the result that a different tariff schedule will have to be set for each area served by a transmission network and the average selling prices of electricity will vary between the different areas. For example, for the purposes of applying its Tarif Vert on a national scale, the Electricitè de France has divided France into 23 areas in which different rates apply.

The marginal energy costs of generation are jointly incurred for all distribution undertakings served by an integrated national grid so that a load increment will cause the same increment in the energy costs of generation regardless of the area in which it arises. It follows that in all geographical areas, the energy charge should be differentiated on the same basis to reflect time of day and also, possibly, seasonal fluctuations in the system's load requirements. This time-of-day and seasonal differentiation is necessary if power stations are operated in merit order to meet the load requirements of the system, since the unit energy cost of the station which is operated last to meet a peak period demand for power will be higher than that of the station which is marginal in the off-peak period. The main problem involved in introducing time-of-day differentiation of the energy charge is that the metering costs involved in providing for even two daily time periods are so much higher than those involved in simple kWh metering that there are a very large number of consumers whose consumption is so small that this additional metering cost is not economically justified. Furthermore, even if time-of-day differentiation is introduced for large power users, it is often the case that where only two time periods are used, then during the peak period of the day there is a very wide variation in marginal energy cost, so that the pricing of all consumption during this large proportion of the day at peak marginal energy cost will result in consumers in near-peak periods having to pay an energy charge that may be greater than the marginal cost incurred in generating electricity to meet their demand. For this reason Hermann has suggested that "if there are only two time periods .... it is probable that the loss of efficiency resulting from pricing energy in the 'peak period' of average variable cost is less than that resulting from pricing at short run marginal cost." He notes however, that "the addition of a third time period involves relatively little more expenditure, in fact, no more than an additional meter register, since the same time switch can usually be adopted for more than two time periods."[22] It therefore appears that once the decision to provide time-of-day metering has been made it is likely to be economically justifiable to provide for an additional time period. The Tarif Vert actually makes provision for 5 time periods with the energy charge in each distribution undertaking being differentiated according to "peak", "normal", and "valley" hours in winter, and "normal" and "valley" hours in summer.

---

22. Hermann, G.F.K., op.cit., p. 6-8

Although the marginal energy costs incurred in power generation are likely to be the same for a given load increment in different distribution undertakings, it is still necessary to have some regional differentiation of the energy charge to reflect the cost of the higher energy losses incurred in transmitting power to distribution undertakings located at a greater distance than others from generating stations. Once a tariff maker has ensured that the structure of the energy component of the tariff set for each undertaking broadly reflects the factors affecting the marginal energy cost incurred by each undertaking, he must then establish the main parameters which should determine the allocation of the residual capacity costs to the different consumers.

It is clear from the discussion of the LRMC of capacity in the previous section that the allocation of these residual capacity costs should vary with the voltage level of the consumers supplied, since the lower the voltage level the greater the incremental capacity costs of transmission and distribution that have to be incurred to supply power to the group of consumers involved. It follows that the tariff schedule for each distribution undertaking should be differentiated according to voltage, with the allocation of residual capacity cost varying inversely with voltage. This principle has been applied by the Electricitè de France with the structuring of tariffs according to voltage being determined by the intervals at which high voltage current has to be transformed into current of a lower voltage.

Once residual capacity costs have been allocated to the distribution undertakings according to voltage, it is necessary to determine for the consumer groups at each voltage level within the undertaking, the basis on which the allocated capacity costs are to be apportioned between the fixed demand charge and the energy charge.

As has already been mentioned, a portion of residual capacity costs should be covered by the energy charge in all periods except the off-peak period in relation to the peak responsibility of the consumers concerned. The aim of this transfer from demand-related to energy-related costs should be to smooth out fluctuations in the load curve. Consequently the main parameter determining this transfer should be the degree of utilization by consumers of their contracted maximum demands for power. This can be illustrated by reference to the the structive of tariffs applicable to the 220 kilovolt (KV) group under the Tarif Vert in France in 1976 as shown in Table 4:1. It can be seen that in the case of consumers with very high utilization factors there is a relatively small transfer from demand related to energy related costs. This is because such consumers already have fairly flat daily load curves so that it would be difficult to induce them to shift their demand from peak to normal or valley periods by differentiating the energy change according to time of day by a greater extent than that which is necessary to reflect periodic fluctuations in the marginal cost of energy. It follows that the residual capacity costs allocated to such consumers are mainly recovered from the demand charge which will burden consumers at an equal average rate whatever the time period.

TABLE 4:1

	Demand charge (a) F/kW/an	Energy charge (centimes/kWh)				
		Winter			Summer	
		Peak	Normal	Valley	Normal	Valley
Very High Utilization	276,76	7,18	7,13	5,79	6,40	5,68
High Utilization	131,15	12,92	10,02	6,38	8,83	6,25
General Tariff	100,19	19,67	11,10	6,45	9,80	6,30
Low Utilization	40,08	31,41	17,13	6,45	11,75	6,30

a) French franc per kW per annum

Table 4:1 indicates, however, that the magnitude of the difference between the energy charge in valley and peak periods increases as the utilization factor of consumers decreases. A high proportion of demand-related residual capacity costs can be transferred to energy-related costs, recoverable during peak and normal periods in the case of consumers with low utilization factors since this provides these consumers to improve their utilization factors by, where possible, shifting their demand to valley periods. The fixed portion of the demand charge levied on these consumers is therefore correspondingly lower than for consumers with higher utilization factors. This categorization of consumers within each voltage class according to their utilization factors is therefore a means of increasing the efficiency of the utilization of the national transmission network as a whole. The utilization of each consumer would be determined by his subscription at the beginning of the year in order to encourage consumers to plan in advance their consumption pattern.

In summary, therefore, a practicable marginal cost pricing policy would provide for three-part tariffs systematically structured according to geographical area, voltage level and utilization factor with a time-of-day and seasonal differential in the energy charge. There are, however, a number of factors which would have to be allowed for in a special manner, since they cannot be taken into account in the standard tariff schedule structured in the above manner.

Firstly, there will be a large number of consumers whose power requirements are too small to justify the sophisticated metering of their demand required for marginal cost pricing. Since the energy charge can no longer be used as an instrument to reduce daily and seasonal fluctuations in the demand of such consumers, it is the usual practice to limit their demand in peak periods in a quantitative manner through load-limiting devices. The residual capacity costs allocated to such consumers are recovered by means of block tariffs which impose on consumers with very low electricity requirements a lower unit cost of power than if a simple two-part tariff had been applied.

Another case where it may be necessary to provide for the special treatment of certain consumers would be where consumers are located at substantially varying distance from the transformation node of a distribution line of given voltage. Here the fact that all such consumers would be charged the same tariff would mean that consumers located near the transformation node might be subsidizing consumers located at a greater distance from it. A subsidized station-terminal tariff should therefore be charged to customers located near the point where energy enters the system while a special contribution should be levied on consumers located far away from this point. There is likely to be a number of these instances where it is necessary to charge supplementary tariffs to reflect the special characteristics of the cost of supplying consumers who would be inefficiently charged according to the standard tariff sechedule.

4:3:3 ADJUSTMENTS TO A MARGINAL COST BASED TARIFF STRUCTURE

Once a tariff structure based on marginal cost principles has been formulated it may be necessary to modify it to take into account economic second best, financial, political and other constraints. If an electricity utility is not required to meet any socio-political objectives then the main adjustments which should be made to the tariff structure are (i) those which compensate for the misallocation of resources resulting from deviations from marginal cost in the pricing of substitutes and complements for electricity and (ii) those necessary to ensure that electricity tariffs generate sufficient revenue to meet the financial targets of the utility.

A. SECOND-BEST CONSIDERATIONS

It will be recalled from Chapter One that it has been suggested by Turvey that a public utility should only depart from marginal cost pricing to take account of non-optimalities which are known, significant and which can be corrected by adjusting its pricing policy. For example, Munasinghe noted that "the subsidies for imported generators and/or diesel fuel, which exist in many developing countries, may make it advantageous for users to set up their own captive plant, even though to the economy as a whole this is not the least cost way of meeting the demand."[23] He suggested that where it

---

23. Munasinghe, M., op.cit., pp. 35-36.

is not politically feasible to remove these subsidies "the low cost of (subsidized) private generation may require the setting of an optimal grid supplies electricity price which is below strict LRMC." It is difficult to precisely determine the electricity price which would satisfy second best optimality considerations, but these conditions do indicate the direction of the adjustment and the factors which should determine the magnitude of the adjustment. Generally, the magnitude of the adjustment would be determined by the size of the subsidy and the degree of substitutability of alternative energy sources. Another example, of a known and significant deviation from marginal cost in the price of an electricity substitute is the paraffin subsidy which in many developing countries is aimed at providing basic energy requirements for low income consumers at prices they can afford. Here, too, electricity tariffs should be reduced below marginal cost in accordance with the main factors determining second best optimality.

Although many electrical appliances can be regarded as complements of electricity, the effect of departures from marginal cost in the prices of these goods on the demand and cost structure of the electricity utility is usually not significant enough to warrant adjusting its pricing policy to reflect second best considerations.

B. FINANCIAL TARGETS

A financial target is usually set for the management of a public utility by the ministry to whom it is responsible. The minimum financial target for a public utility usually requires it to generate sufficient revenue to just cover total costs so that it does not have to be subsidized by the State. However, it is often the case with electricity utilities that they need to earn a surplus in order to generate the internal funds necessary to make a contribution to the financing of future capital development. This occurs when they cannot rely entirely on borrowed funds to finance their expansion programme. It is important for the purpose of regulating the monopolistic powers of an electricity utility that this financial target be determined by a body independent of its management. A marginal cost based tariff will usually generate revenue which is either above or below this financial target with the result that it needs to be adjusted to satisfy this revenue constraint. This involves the allocation of the surplus or deficit between consumer categories in a manner consistent with the objectives of the utility. There are three basic criteria according to which the excess or shortfall may be allocated; equity, efficiency and equality. In the discussion which follows it will be assumed that the revenue generated by a marginal cost pricing policy falls short of the financial target although the same principles could be applied if tariffs had to be reduced to remove a surplus:

(i) THE EQUITY CRITERION

It is usually considered equitable to apportion a deficit to consumer groups according to their responsibility for the expanded financial requirements of the utility. Thus where the main determinant of the internal financing requirement is the expected growth of capital investment, then it would be equitable for the responsibility of the different consumer categories for meeting the financial target to vary according to their expected growth in maximum power demanded.

(ii) THE EFFICIENCY CRITERION

According to the efficiency criterion the purpose of allocating a deficit between different consumer categories should be to minimize the misallocation of resources resulting from the failure to set electricity tariffs equal to marginal cost. This usually involves discriminating between the various consumer categories, so that the greatest divergence from the marginal cost price occurs for the consumer group with the lowest price elasticity of demand, and *vice versa*, so that the deviations from the "optimal" levels of consumption consistent with the strict marginal cost pricing rule are minimized.

(iii) THE EQUALITY CRITERION

To satisfy the equality criterion the profit target should be allocated according to the relative merit or marginal social utility of income of the consumer categories. The pricing policy of the electricity utility could thus to some extent take account of the income redistribution objectives of public policy.

It will be recalled that the condition for optimal pricing under a revenue constraint may be expressed as:

$$pj\left(1 - \frac{kj}{ej}\right) = MC \qquad (2.12)$$

where $kj = 1 + \frac{Dj}{Bj}$ (2.13)

and where ej denotes the price elasticity of demand Dj the distributional characteristic and B, the Lagrangean multiplier associated with the revenue constraint. This condition clearly takes account of both the efficiency and equality criteria, since consumer categories with both a high price elasticity of demand and marginal social utilities of income should have a lower price/marginal cost ratio than those with opposite characteristics. If it is assumed that public policy is mainly concerned with the regional distribution of income then the allocation of the deficit among the consumer categories provided for by a marginal cost-based tariff structure may take place as follows:

(i) the total amount by which the revenue earned by the utility is expected to fall short of the financial target should be allocated to the different distribution undertakings in accordance with their "area need indicators" which can be determined in the manner described in Chapter Three. This basis of allocation should satisfy the equality criteria;

(ii) the allocation by each undertaking to consumer groups within each voltage and utilization level should be determined by an estimate of the average price elasticity of demand within these categories. It is interesting to note here that the Electricitè de France allocates its shortfall in inverse proportion to the length of utilization time. The greatest contribution per kWh towards the recovery of the deficit will therefore have to be made by the low utilization category in each undertaking, since this group of consumers is likely to have the lowest price elasticity of demand.

4:3:4 EVALUATION OF MARGINAL COST BASED TARIFF STRUCTURE

The main purpose of a marginal cost based tariff structure is to improve the allocation of resources to the electricity sector. It is clear that time-of-day and seasonal differentiation of the energy charge may achieve this purpose if they induce consumers to shift their demand to off-peak periods, since this will reduce the future level of peak system demand and thereby result in savings in expenditure on the capital expansion planned to meet this forecasted demand. Furthermore this peak/off-peak price differentiation may also encourage greater utilization of idle plant in off-peak periods. The differentiation of tariffs according to voltage may also lead to savings in capital expenditure since consumers will be encouraged to receive their electricity at a higher voltage level, thereby reducing the need for the electricity utility to invest in the transformer and distribution equipment necessary to supply low voltage consumers. Finally, the differentiation of tariffs by distribution undertaking in relation to area load density may be a factor encouraging greater centralization of economic activity and thereby saving investment in areas with high transmission and distribution costs per square kilometre. This effect of a marginal cost tariff structure will clearly have to be weighed against the welfare losses attributed to increasing the long term cost disadvantage of locating in less developed areas. In this case there is a trade-off between equality and efficiency, although it should be noted that it is generally inadvisable for public utilities to pursue public policy objectives other than economic efficiency. The marginal cost based tariff has the additional advantage that it is flexible enough to provide for the special treatment of consumers where factors other than demand fluctuations, load density, voltage and utilization affect the marginal cost of supplying them and to take out of second-best considerations and financial objectives.

The significance of the benefits obtainable from structuring electricity tariffs according to marginal cost principles depends to some extent on the responsiveness of consumers to the signals conveyed by the tariff structure of an electricity utility. If tariffs are to be adequately structured according to marginalist principles, they may become too complicated to induce the desired behavioural changes in the majority of consumers. Furthermore, as was explained in section 4:2, it is often the case that consumers have fairly fixed electricity requirements, so that they may be unable to take advantage of the opportunity to reduce their electricity costs by altering their consumption pattern in the manner signalled by the electricity tariff structure. It has been pointed out, for example that initially the introduction of the Tarif Vert had little significant impact on the load factors of electricity consumers in France. This is indicated in Table 4:2 below, which sets out the trend in the percentage load factor (% LF) in France over the period 1949-1972.

TABLE 4:2

Year	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
% L.F.	62,8	57,8	60,7	61,1	61,6	62,3	63,6	64,1	62,7	65,5	64,0	63,7
Year	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
% L.F.	63,4	64,4	64,2	66,6	66,6	68,0	66,7	64,8	67,2	68,5	68,2	66,2

Between 1956, when the new tariff was in the process of being introduced, and 1965 "the variation in the load factor is no more than one would expect in consequence of climatic and business cycle variations, and there is, surprisingly no evidence of the reduction in contracted demand."[24] There appears to be a definite change after 1965 and this may be an indication that there has been a delayed adjustment to the

24. Hermann, G.F.K., op.cit., p. 6-11.

charge in the tariff structure. Consumers may, therefore, require time to respond correctly to the pricing signals of a marginal cost tariff policy.

Nevertheless regardless of whether or not the introduction of a marginal cost pricing policy produces a significant reduction in the load factor of an electricity network, it can still be argued that the structuring of tariffs according to the cost of supply in different areas, at different voltages and at various times of the day and year, provides a substantially improved basis of economically efficient allocation of resources than a tariff policy based on more arbitrary methods of cost allocation.

4:4 DESCRIPTION AND EVALUATION OF THE TARIFF POLICY OF THE ELECTRICITY SUPPLY COMMISSION IN SOUTH AFRICA

4:4:1 INTRODUCTION

A PROFILE OF ESCOM

The bulk of electricity in South Africa is supplied by the Electricity Supply Commission (Escom). This is a public corporation established as an electricity utility by the Electricity Act No. 42, of 1922. This Act was repealed and replaced by in 1958 by the Electricity Act No. 40. The functions, powers and duties of the Commission are laid down in section four of the Act. These basically require Escom to "stimulate the provision, wherever required, of a cheap and abundant supply of electricity."[25] In its 1982 Annual Report the basic objective of Escom was stated to be "to provide an adequate supply of

---

25. Electricity Supply Commission, Annual Report 1982, Megawatt Park, Sandton, p. 10.

electricity at cost price so that it can be used for the economic advancement of South Africa."[26] Escom's performance is therefore, measured, not in terms of profitability, but rather "by criteria such as the amount and price of electricity provided for consumption."

The organization is directed by a chairman and six other commissioners, all appointed by the State President. Management vests in the Chairman who is assisted by a senior general manager and corporate and regional management teams. For administrative purposes, Escom is divided into six regions each headed by a regional manager who is responsible for the operation of power stations and the distribution system in his region. For accounting purposes Escom is divided into seven distribution undertakings and one generation undertaking.

By 1982 Escom's power stations had a total sent-out capacity of 20523 MW and its electricity sales represented more than 93 per cent of all electricity sold in South Africa and about 60 per cent of that sold on the African continent. At the end of 1982, plant with a total sentout capacity of 21035 MW was under construction or had been ordered including South Africa's first muclear power station at Koeberg near Cape Town. It is estimated that electricity provides 23 per cent of the total net energy usage in South Africa and it is expected that by the end of the century, Escom will supply nearly 40 per cent of the country's energy requirements.

In 1982 South Africa had 20 coal-fired stations of which 3 were in the Cape, 4 in Natal and 13 in the Transvaal and Orange Free State. In addition there were two hydro-electric stations, Hendrik Verwoerd and Van der Kloof, on the Orange River and two gas-turbine stations in the Cape, while Escom also has a contract for the provision of power by the Cabora Bassa scheme in Mozambique. Coal-fired

---

26. Ibid., p. 10.

stations have always generated the vast bulk of electricity sent out by Escom with 96 per cent of the electricity sent out by Escom coming from coal-fired stations in 1982. The fact that coal deposits are the main source of power for Escom generating stations has been an important influence on the development of the national grid, since most of the power stations are located near the Transvaal coalfields. The electricity which is generated at these power stations is transmitted along 400 kilovolt (KV) cables to regional undertakings which are responsible for the regional distribution of power. Electricity generation for the national grid is centrally controlled from Simmer pan near Germiston in the Transvaal. The system of using stations in their merit order to meet hourly load requirements is effected through the control system at Simmer pan. It can thus be seen that Escom operates a fully developed national transmission network and is therefore able to obtain all the benefits, described in the previous section, which are associated with an interconnected system of electricity generation and transmission.

Escom follows the general practice of most electricity undertakings in the world of providing reserve plant capacity to meet unpredictable increases in demand. It appears to be Escom's policy to maintain a reserve plant margin of at least 17 per cent.

Escom has no share capital. Capital expenditure and loan repayments are financed from internal and external sources in the manner prescribed by the Electricity Act. External finance is obtained by raising loans on local and overseas capital markets and through trade finance arranged with suppliers of capital equipment. Most of the external finance is used to fund capital expenditure but a portion is also used to refinance loans which are of too short a duration to be amortized over their lives without undue strain being placed on electricity tariffs. Internal finance, which is obtained by the retention of tariff income, is the other source of funds available to Escom and is controlled by the provisions of the Electricity Act. This legal framework for the determination of Escom's tariffs will now be examined in greater detail.

4:4:2 ESCOM'S TARIFF POLICY

4:4:2:1 THE CONSTRAINT ON ESCOM'S PRICING POLICY

The legal framework for Escom's pricing policy is laid down in the Electricity Act. Section fourteen provides that, except under the special conditions stipulated in section fifteen, the prices charged by Escom should cover the costs of production, distribution, maintenance, administration, contributions to the Capital Development Fund, The Reserve Fund and the amount required for the redemption and repayment of interest on loans raised to finance capital development. Section Fifteen provides that, in certain cases, Escom may supply electricity at a loss subject to ministerial approval after an investigation has been conducted in terms of section eight.Escom must be fully compensated by the Treasury in these circumstances. The general principle that Escom must set its prices to equate total revenue with total costs is made explicit in section 14(4) where it is stated that Escom ".... shall, as far as practicable, be carried on neither at a profit nor at a loss and that its charges shall be adjusted accordingly from time to time." This principle is also applied to each regional undertaking since it is stipulated in section 16(1) that "each undertaking carried on by the commission shall be separately taken into account when the prices to be charged ... are being assessed or adjusted. To comply with these statutory requirements in formulating its tariffs, Escom must each year forecast its total costs and total sales and set its prices to generate sufficient revenue to cover total costs. Escom's pricing policy is thus subject to a revenue constraint. Escom's record in meeting its annual revenue target during the period 1967 to 1982 is illustrated in Table 4:3 which compares the total revenue earned with "total cost" and measures the surplus or deficit (in parantheses) incurred during these years.

TABLE 4:3

YEAR	TOTAL REVENUE	TOTAL COST	SURPLUS /(DEFICIT)
	R,000	R,0000	R,000
1967	146783	146928	(145)
1968	161475	161993	(518)
1969	176106	175374	732
1970	193475	195866	(2391)
1971	219584	219639	55
1972	254394	258021	(3627)
1973	302034	306162	(4128)
1974	358768	364055	(5287)
1975	460073	487149	(27076)
1976	656381	656322	59
1977	1030552	997097	33455
1978	1301829	1234468	67361
1979	1529474	1511686	17788
1980	1772000	1869967	(97967)
1981	2140689	2219063	(77374)
1982	1695422	2753342	(57920)

One should be careful to interpret the surpluses or deficits recorded in Table 4:3 as a measure of Escom's profitability since "total cost" is made up of those items which must be covered by revenue in terms of section 14 of the Electricity Act and includes a number of items which would not generally be taken into account in determining the net profit before tax of a company incorporated under the Comapnies Act. It can be seen from Table 4:3 that at no time did Escom generate sufficient revenue to cover its total costs. This target was nearly achieved in 1971 when it made a deficit of R55 000 and in 1976 when a surplus of R59 000 was earned. Since 1977, however, the size of the annual surplus or deficit has increased considerably with the record deficit of R97 967 000 being incurred in 1980. This is indicative of the planning problems generally faced in the South African economy which has experienced rapid inflation, and flexible exchange and interest rates during recent years. During periods of rapid inflation

Escom comes under increasing pressure to moderate its tariff increases and thus makes it difficult for the Commission to achieve its goal of equating total revenue with total cost.

Although the size of the deficits or surpluses incurred by Escom may appear large in money terms, it forms a fairly insignificant proportion of the total revenue earned by Escom, so that the average selling price of electricity has tended to bear a fairly close relation to its average cost per kilowatt-hour as is shown in Table 4:4. It follows that much of the trend in the average selling price of electricity during the period 1960 to 1980, which is illustrated in Figure 4:8, can be explained by changes in the size and composition of the items making up the unit cost of electricity. It can be seen from Figure 4.8 that the average selling price of electricity rose very gradually up to about 1974. From 1975 onwards there has been a sharp rise in the rate of increase of the average selling price of electricity, with the compunded annual percentage rate of increase during the period 1975 to 1982 being approximately 17 per cent. Figure 4.8 indicates, though, that in real terms the rate of increase in the average selling price of electricity has been much less dramatic. In real terms (at 1960 prices) the price actually declined from 0,508 cents in 1960 to 0,446 cents in 1970. The 1980 real price includes contributions to the Capital Development Fund which it did not in 1960 and 1970. If these contributions were excluded, the 1980 price in real terms would have been 0,365 per kWh. It is apparent from Figure 4.8, then, that the substantial price increases during the latter half of the 1970s could be regarded as an adjustment of the electricity price from an artificially low level. The average selling price of electricity in South Africa is still fairly low when one compares it with the price charged in other countries. A report by the Energy Research Institute of the University of Cape Town, based on 1979 figures, found that only Norway, Iceland, Canada, New Zealand and Turkey have cheaper power than South Africa. These countries all have a high proportion of hydro-electric power so that South Africa compares particularly favourably with those countries which rely mainly on the thermal generation of electricity. A sample of the results of this study is given in Table 4:5 on the following page.

TABLE 4:4

Year	(1) Average Selling Price Cents/kW/hr	(2) Average Cost Cents/kW/hr	Difference (1) - (2)
1967	0,5506	0,5512	(0,0006)
1968	0,5590	0,5608	(0,0018)
1969	0,5590	0,5566	0,0024
1970	0,5545	0,5614	(0,0069)
1971	0,5772	0,5774	(0,0002)
1972	0,6108	0,6195	(0,0087)
1973	0,6484	0,6573	(0,0089)
1974	0,6822	0,6923	(0,0101)
1975	0,7950	0,8418	(0,0468)
1976	1,0360	1,0360	0,000
1977	1,5353	1,4854	0,0499
1978	1,7887	1,6961	0,0926
1979	1,8980	1,8759	0,0221
1980	2,0242	2,1361	(0,1119)
1981	2,2811	2,3636	(0,0825
1982	2,2811	2,3636	(0,0825)

TABLE 4:5

INTERNATIONAL COMPARISON OF AVERAGE SELLING PRICE OF ELECTRICITY IN 1974

Country	Average Selling Price Cents/kWh	% Hydro-electric
Norway	0,77	100
Iceland	0,88	100
Canada	1,42	100
Turkey	1,86	42
South Africa	1,90	8
Spain	2,58	10
Australia	2,60	48
United Kingdom	4,95	1

Source - Financial Mail, March 5 1982.

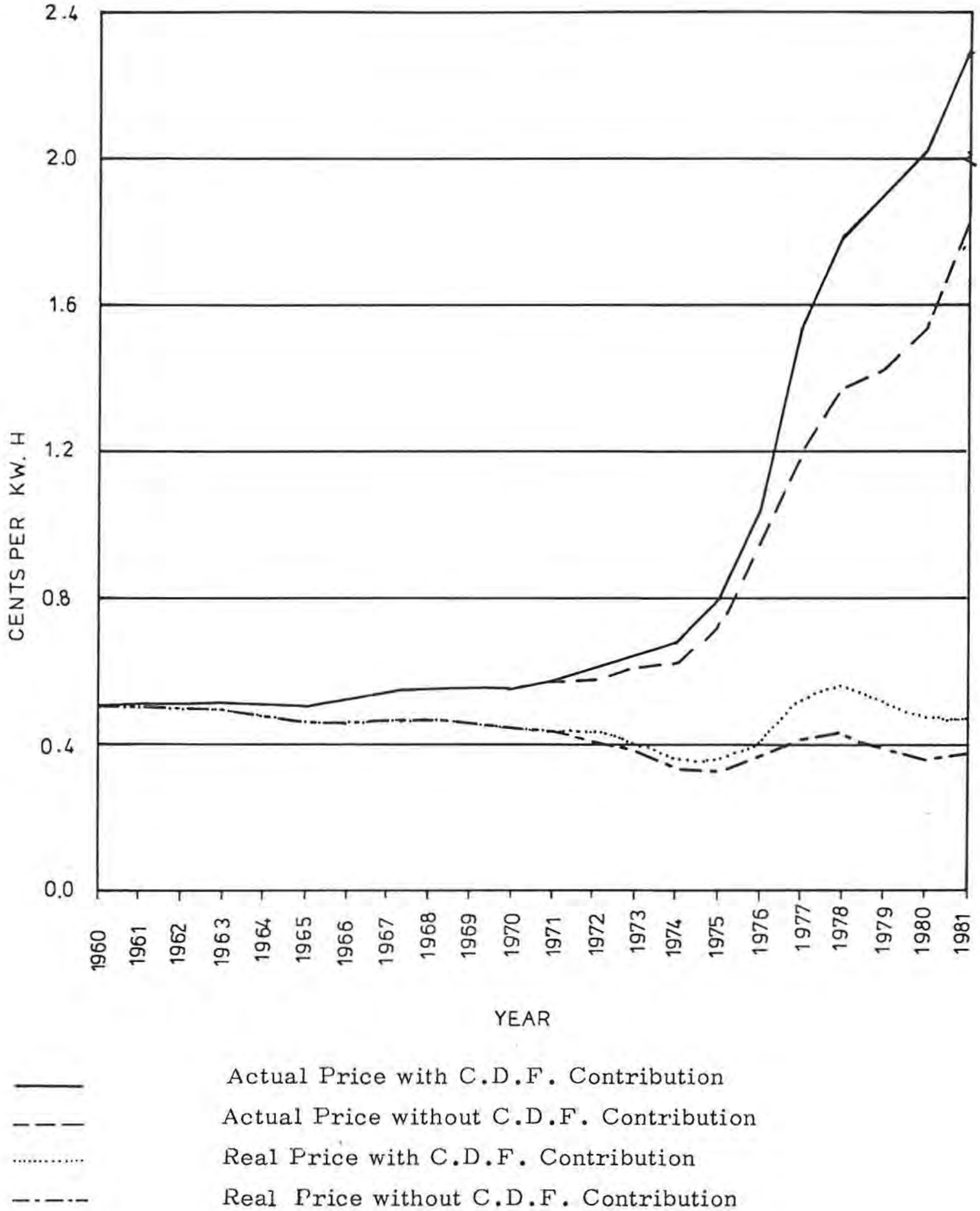


FIGURE 4:8 Escom Average Selling Price of Electricity Deflated by Production Price Index, 1960-1981

To determine the reasons underlying the trend in the average selling price of electricity, it is necessary to conduct a times series analysis of the components of "total cost" which must be covered by tariff revenue. Thus Table 4:6, which has been directly reproduced from Escom Annual Reports, sets out the breakdown of total Escom costs during the period 1967 to 1982. It is now necessary to examine the trend over this period in each item in greater detail.

A. CONTRIBUTIONS TO INTEREST AND REDEMPTION FUNDS

An Interest Fund and Redemption Fund are established in terms of paragraph 11 of the First Schedule of the Electricity Act "for the payment of interest and the redemption and extinction of securities." The annual contribution to the interest fund is made out revenue in terms of paragraph 12(1) of the first schedule:
"there shall be paid and transferred to the interest fund in each year for the payment of interest in respect of the securities a sum equal to the aggregate amount of all the interest payable in that year on the securities outstanding."

It can be seen from Table 4:6 that contributions to the interest fund have formed a substantial proportion of total cost during the period 1967 to 1982: in 1967 the contributions to the interest fund formed 25,39 per cent of total cost. This proportion rose to 33,58 per cent in 1972 although it fell steadily to 22,51 per cent in 1977 as the contributions to the Capital Development Fund constituted an increasing share of total cost. In recent years, however, the share of the interest contribution has again risen with interest forming 26,22 per cent of total cost in 1982. This can be attributed to the risng interest cost of new loan which is indicated in Table 4:7 which sets out the annual percentage rate of increase in interest payments during the period 1968 to 1982.

			Total Escom costs			
Year	Total Escom mill. kW.h sold		Interest	Redemption and other provision for loan repayment	Reserve Fund	Capital Development Fund
1967	26 657,1	R(000)	37 312	24 536	9 912	—
		c/kW.h sold	0,140 0	0,092 0	0,037 2	—
		% of total cost	25,39	16,70	6,75	—
1968	28 885,0	R(000)	43 282	23 884	12 300	—
		c/kW.h sold	0,149 8	0,082 7	0,042 6	—
		% of total cost	26,72	14,74	7,59	—
1969	31 505,6	R(000)	50 943	20 809	13 605	—
		c/kW.h sold	0,161 7	0,066 0	0,043 2	—
		% of total cost	29,05	11,87	7,76	—
1970	34 890,6	R(000)	59 484	23 654	15 202	—
		c/kW.h sold	0,170 5	0,067 8	0,043 6	—
		% of total cost	30,37	12,08	7,76	—
1971	38 040,0	R(000)	70 266	30 928	8 568	—
		c/kW.h sold	0,184 7	0,081 3	0,022 5	—
		% of total cost	31,99	14,08	3,90	—
1972	41 648,9	R(000)	86 631	30 575	3 056	13 596
		c/kW.h sold	0,208 0	0,073 4	0,007 3	0,032 6
		% of total cost	33,58	11,85	1,18	5,27
1973	46 578,4	R(000)	101 858	34 200	3 760	15 366
		c/kW.h sold	0,218 7	0,073 4	0,008 1	0,033 0
		% of total cost	33,27	11,17	1,23	5,02
1974	52 585,1	R(000)	114 308	27 151	66	28 114
		c/kW.h sold	0,217 4	0,051 6	0,000 1	0,053 5
		% of total cost	31,40	7,46	0,02	7,72
1975	57 869,2	R(000)	136 963	30 814	1 400	40 730
		c/kW.h sold	0,236 7	0,053 2	0,002 4	0,070 4
		% of total cost	28,12	6,33	0,29	8,36
1976	63 355,7	R(000)	173 829	41 470	1 700	53 584
		c/kW.h sold	0,274 4	0,065 5	0,002 7	0,084 6
		% of total cost	26,49	6,32	0,26	8,16
1977	67 125,4	R(000)	224 418	63 403	900	224 000
		c/kW.h sold	0,334 3	0,094 5	0,001 3	0,333 7
		% of total cost	22,51	6,36	0,09	22,47
1978	72 780,4	R(000)	308 970	76 036	900	300 000
		c/kW.h sold	0,424 5	0,104 4	0,001 2	0,412 1
		% of total cost	25,03	6,16	0,07	24,30
1979	80 582,8	R(000)	373 718	88 800	900	380 000
		c/kW.h sold	0,463 7	0,110 1	0,001 1	0,471 5
		% of total cost	24,72	5,87	0,06	25,14
1980	87 539,3	R(000)	504 732	101 629	900	426 400
		c/kW.h sold	0,576 6	0,116 1	0,001 0	0,487 1
		% of total cost	26,99	5,44	0,05	22,80
1981	93 844,0	R(000)	603 546	117 088	900	435 478
		c/kW.h sold	0,643 1	0,124 8	0,001 0	0,464 0
		% of total cost	27,21	5,28	0,04	19,63
1982	96 135,9	R(000)	721 948	154 758	26 000	450 000
		c/kW.h sold	0,751 0	0,161 0	0,027 0	0,468 1
		% of total cost	26,22	5,62	0,95	16,34

Table 4:6 Escom Costs and Revenue, 1967 to 1982

Total Escom costs

Sub-total capital related costs	Purchase of electricity	Fuel	Other power station operating and mainte-nance costs	Distribution operation and maintenance costs	General expenses	Total costs	Total revenue
71 760	313	42 488	14 618	7 146	10 603	146 928	146 783
0,269 2	0,001 2	0,159 4	0,054 8	0,026 8	0,039 8	0,551 2	0,550 6
48,84	0,21	28,92	9,95	4,86	7,22	100,00	99,90
79 466	121	45 117	17 016	8 097	12 176	161 993	161 475
0,275 1	0,000 4	0,156 2	0,058 9	0,028 0	0,042 2	0,560 8	0,559 0
49,06	0,07	27,85	10,50	5,00	7,52	100,00	99,68
85 357	102	48 035	19 038	9 264	13 578	175 374	176 106
0,270 9	0,000 3	0,152 5	0,060 4	0,029 4	0,043 1	0,556 6	0,559 0
48,67	0,06	27,39	10,86	5,28	7,74	100,00	100,42
98 340	89	49 440	21 955	10 594	15 448	195 866	193 475
0,281 9	0,000 3	0,141 7	0,062 9	0,030 4	0,044 3	0,561 4	0,554 5
50,21	0,05	25,24	11,21	5,41	7,89	100,00	98,78
109 762	82	53 587	26 276	11 492	18 440	219 639	219 584
0,288 5	0,000 2	0,140 9	0,069 1	0,030 2	0,048 5	0,577 4	0,577 2
49,97	0,04	24,40	11,96	5,23	8,40	100,00	99,97
133 858	95	57 259	31 586	13 486	21 737	258 021	254 394
0,321 4	0,000 2	0,137 5	0,075 8	0,032 4	0,052 2	0,619 5	0,610 8
51,88	0,04	22,19	12,24	5,23	8,42	100,00	98,59
155 184	117	68 634	38 685	17 082	26 460	306 162	302 034
0,333 2	0,000 3	0,147 4	0,083 1	0,036 7	0,056 8	0,657 3	0,648 4
50,69	0,04	22,42	12,64	5,58	8,64	100,00	98,65
169 639	86	92 530	48 572	20 617	32 611	364 055	358 768
0,322 6	0,000 2	0,176 0	0,092 4	0,039 2	0,062 0	0,692 3	0,682 2
46,60	0,02	25,42	13,34	5,66	8,96	100,00	98,55
209 907	114	141 913	*44 980	*18 477	*71 758	487 149	460 073
0,362 7	0,000 2	0,245 2	0,077 7	0,031 9	0,124 0	0,841 8	0,795 0
43,09	0,02	29,13	9,23	3,79	14,73	100,00	94,44
270 583	2 399	208 316	62 477	19 712	92 835	656 322	656 381
0,427 1	0,003 8	0,328 8	0,098 6	0,031 1	0,146 5	1,036 0	1,036 0
41,23	0,37	31,74	9,52	3,00	14,14	100,00	100,01
512 721	15 501	239 228	76 294	19 859	133 494	997 097	1 030 552
0,763 8	0,023 1	0,356 4	0,113 7	0,029 6	0,198 9	1,485 4	1,535 3
51,42	1,55	23,99	7,65	1,99	13,39	100,00	103,36
685 906	26 364	271 222	89 193	23 677	138 106	1 234 468	1 301 829
0,942 4	0,036 2	0,372 6	0,122 5	0,032 5	0,189 7	1,696 1	1,788 7
55,56	2,14	21,97	7,22	1,92	11,19	100,00	105,46
843 418	36 061	319 428	95 887	28 689	188 203	1 511 686	1 529 474
1,046 6	0,044 7	0,396 3	0,118 9	0,035 6	0,233 5	1,875 9	1,898 0
55,79	2,39	21,13	6,34	1,90	12,45	100,00	101,18
1 033 661	35 806	405 630	117 968	36 824	240 078	1 869 967	1 772 000
1,180 8	0,040 9	0,463 3	0,134 8	0,042 1	0,274 2	2,136 1	2,024 2
55,28	1,91	21,69	6,31	1,97	12,84	100,00	94,76
1 157 012	4 106	569 949	170 206	43 034	273 756	2 218 063	2 140 689
1,232 9	0,004 4	0,607 3	0,181 4	0,045 9	0,291 7	2,363 6	2,281 1
52,16	0,19	25,70	7,67	1,94	12,34	100,00	96,51
1 352 706	3 615	693 979	261 842	59 852	381 348	2 753 342	2 695 422
1,407 1	0,003 7	0,721 9	0,272 3	0,062 3	0,396 7	2 864,0	2,803 8
49,13	0,13	25,21	9,51	2,17	13,85	100,00	97,90

Table 4:6 (Continued)

TABLE 4:7

ANNUAL GROWTH INTEREST COSTS: 1968 - 1982

Year	Annual % in increase in interest cost
1968	16,0
1969	17,7
1970	16,8
1971	18,1
1972	23,3
1973	17,5
1974	12,2
1975	19,8
1976	26,9
1977	29,1
1978	37,6
1979	20,9
1980	35,1
1981	19,6
1982	19,6

The rate of growth in interest payments remained fairly steady during the period 1968 to 1975, when the average annual rate of increase was 17,6 per cent. After 1976 the annual rate of growth of the interest burden showed a sharp increase with the average rate of increase during this period being about 27,0 per cent.

The current contributions from revenue to the interest fund are significantly inflated by the fund accounting techniques adopted by Escom. The investments of the redemption, capital development and reserve funds largely consist of Escom stock. At 31st December 1982, these funds held 40,3 per cent of the book value of Escom's internally registered stock. They thus form an important market for Escom stock over and above that which can be sold to local buyers or raised in foreign countries. The interest earned on this stock is reinvested in these funds. It follows that a substantial proportion of the total interest payable which is charged to revenue, is retained internally by the Commission and thus provides a source of funds for the internal financing of capital development. This differs from the generally accepted accounting practice of companies incorporated under the Companies Act in terms of which transfers to internal funds are not treated as a charge against revenue but as an appropriation from net profit. In 1982 the interest credited to the Capital Development Fund which holds all its investments in Escom stock, amounted to R350 163 000. This constituted 48,5 per cent of the total interest charge of R721 948 000. When one considers that the reserve and redemption funds hold 97,7 per cent and 99,95 per cent, respectively, of their investments in Escom, stock it follows the major proportion of the amounts of R15 216 000 and R76 315 000 of interest credited to these funds, respectively, was charged to revenue. Table 4:8 sets out the ratios of interest credited to internal funds to the total interest charge for the period 1977 to 1982. Table 4:8 provides a fair indication of the proportion of the total interest charge which is providing a source of internal finance to Escom. It can be seen from this table that this proportion has grown substantially during the last five years. This may be attributed to the rapid growth in the interest earned by the capital development fund. Table 4:9 indicates the extent to which the amount of interest credited to this fund has grown in recent years.

TABLE 4:8

Year	% Interest Credited / Total interest charge
1977	35,6
1978	40,5
1979	46,2
1980	44,4
1981	51,3
1982	61,2

TABLE 4:9

INTEREST CREDITED TO CAPITAL DEVELOPMENT FUND

Year	R'000	Annual Percentage Increase	% Credit Total Interest Charge
1977	33169		14,8
1978	70091	111,3	22,7
1979	111264	58,7	29,8
1980	159171	43,1	31,5
1981	233720	46,8	38,7
1982	350163	49,8	48,5

It is clear that as the investment base of the Capital Development Fund grows, the interest payable out of revenue into this fund will also grow substantially with the result that the revenue constraint on Escom's pricing policy will become increasingly steeper. It can be concluded that the fund accounting techniques employed by Escom result in an overcharging of interest against revenue and an overstatement of the total costs which must be covered by revenue if Escom is to "break even".

B. CONTRIBUTIONS TO THE REDEMPTION FUND

Escom has adopted the following accounting policy in respect of its fixed assets in commission:

"Fixed assets in commission are not depreciated but are reflected at historical cost. Long term loans are raised to finance these assets. Because of the correlation between the loans so raised and fixed assets, the charge to revenue for loan amortisation takes the place of depreciation."[27]

---

27. Electricity Supply Commission, Annual Report 1982, Megawatt Park, Sandton, 1982, p. 35.

Provision for depreciation of these assets thus takes place by means of the Redemption Fund for local loans and the "Provision for Repayment of Foreign Loans". The Redemption Fund is established in terms of the Electricity Act and is maintained on a sinking fund basis to amortize the loan debt of long and medium term loans. When fixed assets are placed in commercial operation a loan is allocated as being the financing source of these assets. An annual contribution is determined on a sinking fund basis for the lifetime of the loan, which at present is about 23 years. This contribution is based on a projected interest earning rate to equate the redemption value of the loan. To appropriate the income necessary to make these annual contribution, in terms of paragraph 12(2), the following amounts are paid and transferred to the redemption fund:

"(a) the net proceeds of any sales of fixed property belonging to the Commission;

(b) all other incomings of the Commission in respect of any sales of rights or interests in the nature of or analogous to fixed property or servitudes;

(c) the income of the investments of the redemption fund .....".

In terms of paragraph 16(3) the investments of the Redemption Fund are revalued on an annual basis to determine whether they are sufficient to meet the requirements of the sinking fund. If there is a shortfall, then a contribution to the sinking fund is made out of revenue. The amounts shown in Table 4:6 as charges from revenue to the redemption fund are thus residual items, so that it is difficult to discern any significant trend in these amounts. Certainly during the period 1967 to 1975, the contribution out of revenue to this fund fluctuated in their amount. For example in 1968, 1969, 1972 and 1974 the contributions fell from the previous annual amount, while in 1970, 1971, 1973 and 1975 there was an annual increase in this contribution. During the period 1976 to 1982 the contribution to the Redemption Fund showed a steady annual growth indicating that this item of "total cost" has placed an increasing burden on Escom's tariff revenue. Although the share of "total costs" constituted by this item has decreased steadily from 16,70 per cent in 1967 to 5,62 per cent in 1982 it appears to have levelled out at around 5,55 per cent during the last four years.

The contribution to the Redemption Fund together with the internal interest on the investment of the fund and the annual amounts set aside on a straight line basis for the repayment of foreign loans are charged to revenue instead of depreciation on fixed assets in commission. Depreciation is, however, charged on movable plant and equipment at rates considered appropriate to reduce cost to estimated residual value over the useful lives of the assets. The question thus arises as to what extent Escom's system gives results comparable to a system of depreciation based on historical cost of assets taking into account that the present average life of Escom's assets is approximately 35 years. The Board of Trade and Industries addressed this problem and found "... both in respect of local and overseas loans the redemption period of the loans is shorter than the average lifetime of the assets resulting in accelerated depreciation."[28] To illustrate this, the Board compared the straight line historical cost depreciation of an asset costing R1 000 with a life of 35 years with the contributions to the Redemption Fund to redeem a local loan raised to finance the purchase of the asset at 10 per cent interest redemable after 23 years Figure 4:9 depicts the results of this comparison. It could be argued that if loans were repaid in full after 23 years, Escom's position would be the same as as in the case of historical value depreciation. A portion of the investments of the Redemption fund is, however, always held in internal Escom stock so that only a certain portion of its loans have to be repaid at their redemption date while the remaining redeemable portion is reinvested in Escom stock. The terminal value of this internal portion not repaid, after 12 years for different proportions of internal financing is shown in Table 4:10.

---

28. Republic of South Africa, Board of Trade and Industries Report No. 1889, Investigation into the tariff policy and tariff structure of the supply of electricity in South Africa, Pretoria, 1977, p. 41.

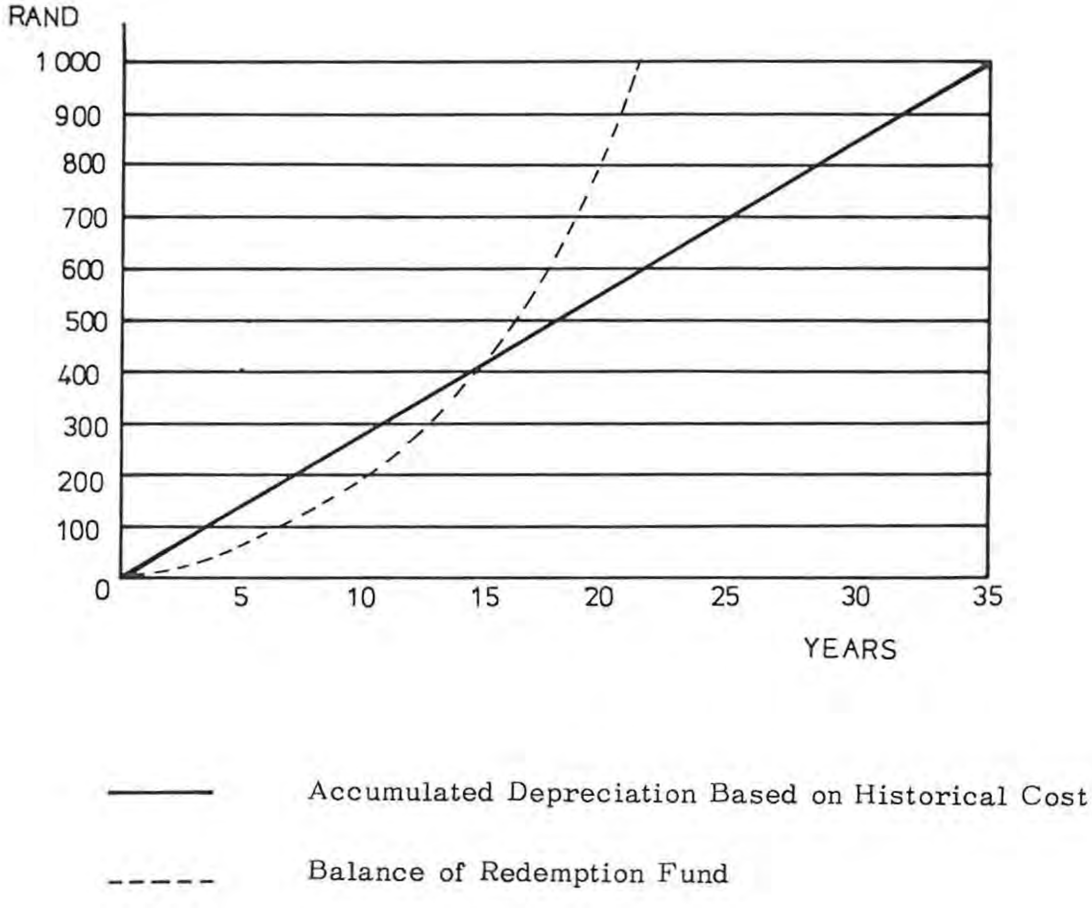


FIGURE 4:9 Accumulated Amounts Charged to the Electricity Supply Account Under a System of Straightline Historical Depreciation and Under Escom's Present System (Assuming an Initial Investment of R1 000

TABLE 4:10

	50% Internal	60% Internal
Terminal amount (of R500 and R600 respectively) assuming interest rate of 10%	R1 569	R1 883
Amount to be borrowed	R1 569	R1 255
Cost of asset to be financed	R3 138	R3 138

Escom's policy can be justified to some extent in an inflationary period, since the cost of replacing an asset will rise substantially over the life of the asset. In the example above, Escom's policy provides for higher replacement cost resulting from an average inflation rate over 35 years amounting to 3,32 percent per annum.

To support the contention by the Board of Trade and Industries that Escom's policy results in a greater charge against revenue than if its assets were depreciated on a historical cost basis, the charge against revenue which would have resulted during the period 1977 to 1982 if fixed assets in commission had been depreciated over 35 years on a straight line basis has been calculated. This charge has also been compared with that which resulted under the present Escom policy of charging the contribution and interest on the Redemption Fund and the provision for the redemption of foreign loans against revenue.

It can be seen from Table 4:11 that during the period 1977 to 1979 and in 1982 the charges under the current Escom system were greater than those which would have resulted if depreciation had been charged on a historical basis. This, however, was not the case in 1980 and 1981. Over the entire period, however, the Escom charges were R31,5 million greater than historical cost depreciation so that there appears to be some validity in the conclusion of the Board of Trade and Industries Report.

TABLE 4:11

COMPARISON BETWEEN CHARGES AGAINST REVENUE UNDER PRESENT ESCOM POLICY AND WITH HISTORICAL COST DEPRECIATION

Years	(1) Charge on Current Basis	(2) Charge on Historical Cost Basis	Difference (1)-(2)
	R'000	R'000	R'000
1977	94153	81460	12693
1978	110774	101846	8928
1979	132489	120729	11760
1980	150022	160115	(10093)
1981	177503	180659	(3156)
1982	231073	219697	11376

C. THE CAPITAL REDEMPTION AND RESERVE FUNDS

Since 1971 Escom has been empowered to generate funds for future capital expenditure through the Capital Development Fund. In terms of the 1971 amendment to the Act, the annual contributions to the Capital Development and Reserve Funds, excluding interest earned by these funds, was limited to 3 per cent of the value of unredeemed loans, provided that the accumulated amount excluding interest earned by these funds did not exceed 15 per cent of unredeemed loans in the case of the Capital Development and 7,5 per cent in the case of the Reserve Fund. In 1977 the Act was again amended to raise the upper limit on annual contributions to 6 per cent and the ceiling on the accumulated amount of the capital development fund to 30 per cent.

For the first five years after introduction of the Capital Development Fund during the period 1972 to 1976, the annual contributions to this fund were well below their legal limits. For example, it was estimated that in 1976 total Capital Development Fund and Reserve Fund contributions amounted to only 1,8 per cent of unredeemed loans.[29] Table 5:6 shows that, compared to later years, the contributions to the Capital Development Fund made up a relatively small proportion of total cost. In 1972 contributions made up only 5,27 per cent of total costs, in 1973 5,02 per cent, in 1974 7,72 per cent, in 1975 8,36 per cent and in 1976 8,16 per cent. In 1977, however, there was a change in Escom policy regarding the internal financing of capital development and the contributions to this fund increased dramatically from R53 584 000 to R224 000 000. The background to this development was the dramatic decline in foreign sources of finance in 1977 which virtually forced Escom to rely on internal funds if it was to continue its capital development programme. The importance of the Capital Development Fund as a stable and reliable source of finance has been recognized by Escom in recent years, particularly in the light of the highly volatile nature of the foreign capital market. This volatility of foreign borrowing is indicated by the changing ratios of foreign to local loans as shown in Table 4:12.

TABLE 4:12

RATIO LOCAL TO FOREIGN LOANS, 1976 TO 1981

Year	1976	1977	1978	1979	1980	1981
Local %	40	70	50	35	40	60
Foreign %	60	30	50	65	60	40

The progress which Escom has made during the period 1977 to 1982 in increasing the amount of the capital development up to its legal maximum is reflected in Table 4:13.

29. *Ibid.*, p. 44.

It can be seen from column (4) that the difference between the accumulated amount of the Capital Development Fund, excluding interest, and its legal maximum steadily declined during the period 1977 to 1981. This trend is not only accounted for by the annual increases in the contribution to the Capital Development Fund but also by the reinvestment of interest income into the Escom stock held by the fund. In 1982, however, the accumulated amount of the Capital Development Fund fell further below the legal limit than in 1981, largely due to the relatively small increase in annual contributions in that year. Columns (5) and (6) compare the actual annual contributions to the Capital Development and Reserve Funds with the legal maximum which is set at 6 per cent of unredeemed loans. It can be seen that during the period 1977 to 1980 the annual contributions actually exceeded the legal limit. This indicates that it was a deliberate policy of Escom to build up the Capital Development Fund as quickly as possible (the annual contributions to the Reserve Fund remained steadily at R0,9 million during this period). In 1981 and 1982, however, the annual contributions fell significantly short of the legal limit. In particular, it can be seen that in 1982 it would have been possible to increase the annual contribution by R195,5 million. This would have had the effect of increasing total costs by 7,1 per cent to R2 948,8 million and the average cost per kWh of electricity by 0,2 cents/kWh to 3,07 cents/kWh. It is possible that Escom refrained from burdening its consumers with this additional cost due to the recessionary conditions in the economy and because it may have been considered that the amount of internal financing had reached a satisfactory level.

The Reserve Fund was established by the Electricity Act for "the replacement of obsolete machinery or plant and generally for the betterment of plant owned by the Commission or in lieu of insurance, or for exceptional repairs or emergencies, but not for ordinary maintenance." Contributions to this fund have never been as important a component of total costs as contributions to the Capital Development Fund. Table 4:6 shows that the contributions to this fund rose from R9,9 million in 1967 to R15,2 million in 1970 and then fell to R1,7 million in 1976. From 1977 to 1981 the contributions were held

TABLE 4:13

TREND IN THE CAPITAL DEVELOPMENT FUND (C.D.F.) 1977 TO 1982

Year	(1) Unredeemed Loans	(2) 30% x Unredeemed	(3) Accumulated CDF Lees Interest	(4) Shortfall from legal limit (2)-(3)	(5) 6% x Unredeemed	(6) Actual Contributions to CDF and RF	(7) Potential Increase In contributions to CDF and RF
	R'000	R'000	R'000	R'000	R'000	R'000	R'000
1977	3 717 595	1 115 279	405 661	709 618	223 056	224 900	-
1978	4 710 009	1 413 003	738 907	674 906	282 601	300 900	-
1979	5 750 696	1 755 209	1 189 142	536 067	345 042	380 900	-
1980	6 785 436	2 035 631	1 727 162	308 469	407 126	427 300	-
1981	8 500 843	2 550 253	2 323 140	227 113	510 051	436 378	73 673
1982	11 191 648	3 357 494	3 006 860	350 634	671 499	476 000	195 499

steadily at R0,9 million, although in 1982 they were increased again to R2,6 million. Since 1970 the proportion of total costs constituted by these contributions fell from 7,76 per cent in 1970 to 0,04 per cent in 1981.

Although the contributions to the Capital Development and the Reserve Fund are both charged to revenue as part of "total cost", in terms of the generally accepted accounting practice of companies incorporated under the Companies Act these items would both be shown as an appropriation of profits.

D. CAPITAL RELATED COSTS

To summarize the analysis of the capital-related component of "total costs" so far, it would seem that a fairer measure of Escom's profitability could be obtained by adopting the following procedure:

Add: (a) Interest credited to funds in respect of Escom Stock.
(b) Contributions to Redemption Fund.
(c) Contributions to Capital Development Fund.
(d) Contributions to Reserve Fund.

Deduct: Historical Cost Depreciation.

Table 4:14 sets out the results of using this procedure to calculate the adjusted surpluses/earned by Escom during the period 1977 to 1982.

It can be seen that in each of the above years a substantial portion of the capital related costs charged to the Electricity Supply Account consists of internally retained profits and that Escom earns a substantial surplus in each year. It is apparent therefore that Escom does not formulate its pricing policy subject to a strict break even constraint.

In Table 4:14 the return on sales and return on total assets was calculated as the percentage of the adjusted surplus to total revenue and total assets respectively. It can be seen that during the period 1977 to 1979 when tariffs were raised to provide the revenue for the

TABLE 4:14

ADJUSTED ESCOM SURPLUS 1977 TO 1982

	1977	1978	1979	1980	1981	1982
Surplus/Deficit (R'000)	33 455	67 361	17 788	(97 967)	(77 374	(57 920)
Interest credited to funds	79 763	125 060	172 348	224 024	309 559	441 694
Contributions to funds	288 303	376 936	469 700	528 929	553 466	630 758
	401 521	569 357	659 836	654 986	785 651	101 532
Less Depreciation	(81 460)	(101 846)	(120 729)	(160 115)	(180 659)	219 627
Adjusted Surplus (R'000)	320 061	467 511	539 107	494 871	604 992	794 835
Returns on Sales (%)	31,1	35,9	42,4%	27,9%	28,3%	29,5%
Returns on Total Assets (%)	7,0	8,0	7,1	5,5	5,4	5,4

expansion of the Capital Development Fund the return on sales reached the fairly high figures of 31,1 per cent in 1977, 35,9 per cent in 1978 and 41,4 per cent in 1979. The return on sales and total assets showed a marked decline during the period 1980 to 1982. If Escom had, however, raised its tariffs to generate the revenue required to cover the maximum possible contributions to the Capital Development and Reserve Funds the adjusted surplus would have been R678 665 000 in 1981 and R990 334 000 in 1982, with the return on sales being 31,7 per cent in 1981 and 36,7 per cent in 1982. The return on total assets in these years would have been 6,0 per cent and 6,7 per cent, respectively. It appears, then, that Escom has considerable amount of discretion within the legal constraints imposed by the Electricity Act to use its monopoly power to earn a substantial return on its investment. The question of whether it is justifiable for Escom to have this discretionary power, will be examined in the next section.

Although Table 4:6 indicates that capital related charges form a major portion of total costs, the above analysis shows that the capital related charges should be divided between those which relate to current consumption of capital and those which provide for future capital investment. Table 4:15 shows the effect of deducting those charges which are in effect retained profits from both the capital related charges and the "total costs" on the cost structure of Escom. The annual current capital consumption charge after the adjustment is made up of the net interest paid to outside stockholders and historical cost depreciation. The effect of the adjustment is to reduce the average share of total costs contributed by capital costs for the period 1977 to 1982 from 53,2 per cent to 31,7 per cent. Escom, therefore, is clearly required to generate revenue from its tariffs more for future capital expansion than to cover the current capital cost of production.

E. CURRENT OPERATING COSTS

Tables 4:6 and 4:15 indicate that each year a substantial portion of Escom's total costs is made up of current operating costs. The average percentage of total costs contributed by these running costs

TABLE 4:15

EFFECT OF ADJUSTMENT ON ESCOM COST STRUCTURE

Year		Before Adjustment			After Adjustment			
		Capital related charges	Running Cost	Total Cost	Current Capital Consumption	Running Cost	Total Costs	Retained Profits
1977	R (000)	51 272	484 376	997 097	226 115	484 376	710 491	286 606
	% of total cost	51,42	48,58	100	31,8	68,2	100	40,3
1978	R (000)	685 906	548 562	1 234 468	285 756	548 762	834 518	400 150
	% of total cost	55,56	44,44	100	34,2	65,8	100	47,9
1979	R (000)	843 418	668 268	1 511 686	322 099	668 268	990 367	521 319
	% of total cost	55,79	44,21	100	32,5	67,5	100	52,6
1980	R (000)	1 033 661	836 306	1 869 967	440 823	836 306	1 277 129	592 838
	% of total cost	55,28	44,72	100	34,5	65,5	100	46,4
1981	R (000)	1 157 012	1 061 051	2 218 063	474 646	1 061 051	1 535 697	682 366
	% of total cost	52,16	47,84	100	30,9	69,1	100	44,4
1982	R (000)	1 352 706	1 400 636	2 753 342	499 951	1 400 636	1 900 587	852 755
	% of total cost	49,13	50,87	100	26,3	73,7	100	44,9

during the period 1977 to 1982 was 46,8 per cent before and 68,3 per cent after the abovementioned adjustment. There can be no objection to charging these costs to revenue, since they are directly incurred in the production of electricity. It follows that every year Escom should aim to generate sufficient revenue from its tariffs to at least cover these costs.

By far the most important component of operating costs is the cost of fuel which is largely made up of the cost of coal, since the bulk of electricity geneneration in South Africa is performed by coal-fired power stations. This is illustrated in Table 4:16 which is directly reproduced from Escom's Annual Report for 1982, where it can be seen that the total output of coal-fired stations in 1982 of 100 217 million kWh made up 97,5 per cent of the total output of 102 769 million kWh. Furthermore, the total coal cost of R648,5 million made up 93,5 per cent of the total fuel costs of R694 millions and 23,4 per cent of the total costs indicated in Table 4:6.

A major benefit of the national grid network operated by Escom is that most of its coal-fired power stations are situated near a colliery specifically established to meet their coal requirements. Such collieries are established as a consequence of the acceptance by Escom, of a tender for the supply of coal for a new power station. The coal supplied normally has a lower calorific value than the standard value of that for manufacturing industry, so that it generally cannot be used for any purpose other than electricity generation. For example, in 1982 the average calorific value of coal consumed by Escom was 21,39 milojoules/kg (MJ/kg) while the most recent contract was for the supply of coal at an average heat content of 17 MJ/kg whereas the standard calorific value of coal consumed by manufacturing industry is 24 MJ/kg. Escom therefore enjoys a degree of monopsonistic market power with respect to the purchase of coal of low calorific content and is able to tender for coal at prices which are generally lower than its controlled price.

Table 4:16 indicates that during the period 1960 to 1972 the unit cost of coal actually fell from 0,1466 cents/kWh sent out in 1960 to 0,1285

	Coal-fired power stations						
	Coal used					Coal cost	
Calendar year	Thousands of tons	Average heat content (as received) MJ/kg	kg/kW.h sent out	Average heat rate MJ/kW.h sent out	Overall thermal efficiency sent-out basis %	Total R000	Average rand/ton
1950	6 323,4	22,72	0,869	19,74	18,2	5 302	0,84
1951	6 662,9	22,72	0,855	19,43	18,5	6 553	0,98
1952	7 113,4	22,75	0,865	19,68	18,3	8 520	1,20
1953	7 393,9	23,08	0,837	19,32	18,6	9 862	1,33
1954	8 024,9	23,06	0,805	18,56	19,4	11 329	1,41
1955	8 999,7	22,89	0,788	18,04	20,0	13 709	1,52
1956	9 688,5	22,96	0,765	17,56	20,5	13 653	1,62
1957	10 220,6	22,79	0,750	17,09	21,1	17 256	1,69
1958	10 784,1	22,73	0,743	16,89	21,3	19 039	1,77
1959	11 548,7	22,44	0,732	16,43	21,9	20 970	1,82
1960	12 512,6	22,52	0,723	16,28	22,1	25 373	2,03
1961	13 194,9	22,39	0,722	16,17	22,3	27 713	2,10
1962	13 955,5	22,22	0,719	15,98	22,5	29 230	2,09
1963	14 721,1	22,15	0,708	15,68	23,0	31 009	2,11
1964	15 654,7	22,15	0,692	15,33	23,5	32 367	2,07
1965	16 726,7	22,39	0,680	15,23	23,6	34 986	2,09
1966	16 982,3	22,20	0,666	14,79	24,4	37 901	2,23
1967	18 307,7	22,44	0,645	14,47	24,9	42 053	2,30
1968	19 133,9	22,63	0,620	14,03	25,6	44 604	2,33
1969	19 982,9	22,73	0,595	13,52	26,6	47 453	2,37
1970	21 630,6	22,97	0,580	13,32	27,0	48 807	2,26
1971	23 416,2	23,30	0,576	13,42	26,8	52 705	2,25
1972	24 952,8	22,89	0,571	13,07	27,5	56 113	2,25
1973	27 907,9	22,47	0,563	12,65	28,5	66 837	2,39
1974	30 891,4	22,42	0,560	12,56	28,7	90 269	2,92
1975	34 231,7	22,21	0,567	12,59	28,6	138 592	4,05
1976	37 257,4	21,87	0,579	12,66	28,4	200 781	5,39
1977	37 505,6	21,78	0,576	12,55	28,7	233 229	6,22
1978	39 589,5	21,61	0,574	12,44	28,9	263 880	6,67
1979	43 264,9	21,22	0,580	12,33	29,2	301 273	6,96
1980	46 755,0	21,34	0,568	12,16	29,6	379 942	8,12
1981	53 903,7	21,25	0,563	12,01	30,0	523 663	9,71
1982	55 198,4	21,39	0,551	11,82	30,5	648 550	11,75

TABLE : 4:16 Operating Costs for Escom, 1982

Cents/ kW.h sent out	Total power stations output million kW.h sent out: Coal-fired stations	Hydro-electric stations	Pump-storage stations	Diesel and gas turbine stations	Total power station output	Total power station capacity assigned sent-out rating MW as at 31 December	Average power station plant load factor sent-out basis %
0,072 9	7 276	7	—	4	7 287	1 290	64,7
0,084 0	7 797	6	—	3	7 806	1 361	66,1
0,103 7	8 220	6	—	1	8 227	1 454	66,9
0,111 6	8 838	7	—	—	8 845	1 635	65,5
0,113 6	9 971	6	—	—	9 977	1 846	66,4
0,120 1	11 419	6	—	—	11 425	2 145	65,9
0,123 6	12 663	7	—	—	12 670	2 498	61,2
0,126 6	13 634	6	—	—	13 640	2 555	61,1
0,131 2	14 511	5	—	—	14 516	2 748	62,0
0,132 9	15 774	3	—	—	15 777	2 983	62,6
0,146 6	17 306	2	—	—	17 308	3 091	65,2
0,151 6	18 282	2	—	—	18 284	3 226	66,2
0,150 7	19 401	3	—	—	19 404	3 406	65,8
0,149 2	20 789	4	—	—	20 793	3 788	65,7
0,143 0	22 634	5	—	—	22 639	4 077	65,2
0,142 3	24 583	—	—	—	24 583	4 181	67,4
0,148 6	25 504	—	—	—	25 504	4 377	67,1
0,148 2	28 371	—	—	—	28 371	5 328	66,8
0,144 6	30 843	—	—	—	30 843	5 800	62,9
0,141 2	33 598	—	—	—	33 598	6 441	62,1
0,130 8	37 321	—	—	—	37 321	7 060	62,9
0,129 7	40 645	94	—	—	40 739	8 373	61,3
0,128 5	43 662	813	—	—	44 475	8 849	59,6
0,134 8	49 570	189	—	—	49 759	9 482	62,5
0,163 7	55 141	1 110	—	—	56 251	10 002	66,3
0,229 5	60 400	1 098	—	—	61 498	10 522	68,6
0,312 2	64 309	1 853	—	26	66 188	11 688	66,8
0,358 2	65 114	1 924	—	12	67 050	12 756	61,9
0,382 4	69 004	1 887	—	11	70 902	13 595	60,7
0,404 5	74 485	1 144	—	14	75 643	15 056	60,9
0,461 4	82 342	992	—	28	83 362	17 339	57,8
0,547 3	95 675	1 653	415	81	97 824	18 989	62,2
0,647 1	100 217	1 016	1 519	17	102 769	20 523	59,3

TABLE 4:16 (Continued)

cents/kWh sent out in 1972. This trend can be attributed to the rather slow rate of increase in the average price of coal from R2,03 per ton in 1960 to R2,25 per ton in 1972, and in the substantial improvement in the efficiency of coal use in electricity generation which is indicated by the increase in weight of coal per kWh sent out from 0,723 kg/kWh in 1960 to 0,571 kg/kWh in 1972. The improved efficiency in coal consumption by Escom power stations resulted from the shift from small scattered power stations in the various undertakings to the supply of electricity from the large stations located on the Transvaal coalfields. This had two major effects on the cost of coal consumption:

(i) the amount of coal consumed per kWh sent out of the larger stations is substantially lower than that for smaller stations due to economies of scale and improved technique. For example, in 1982, the giant Arnot, Duvha, Kriel and Matla power stations consumed 0,498 kg/kWh, 0,459 kg/kWh, 0,509 kg/kWh and 0,569 kg/kWh, respectively, whereas the small Klip and Vaal power stations consumed 1,171 kg/kWh and 1,071 kg/kWh respectively. It is only on an inter-connected power grid that the economies of scale of large power stations can be used for the benefit of the whole system; and

(ii) with the increased use of a national network stations located far from the coalfields can be used to a lesser extent, with the result that savings in the costs of coal railage to these stations can be achieved.

After 1972, however, the cost of coal consumed per unit of electricity sent out rose from 0,1348 cents per kWh in 1973 to 0,6471 cents per kWh in 1982. This can be partly attributed to the dramatic increase in the average price of coal from R2,39 per ton in 1973 to R11,75 per ton in 1982. It is also significant that during the period 1973 to 1979 the efficiency of coal consumption by Escom actually decreased, which is indicated by the increase in the weight of coal used per kWh from 0,563 kg/kWh in 1973 to 0,58 kg/kWh in 1979, although coal consumption appears to have become more efficient during the last three years with the weight of coal used per kWh sent out being 0,551 kg in 1982. The slight decrease in the efficiency of coal use during the mid-1970s can probably be attributed to the following factors:

(i) the scale benefits associated with an integrated power grid may have largely been exhausted;

(ii) there has been a slight decrease in the quality of coal used with its calorific value falling from 23,47 MJ/kg in 1973 to 21,22 MJ/kg in 1979; and

(iii) the shortage of external finance during the mid-1970s appears to have caused Escom to fall behind in its capital development programme, with the result that there has been a greater use of the capacity of national power grid. This is indicated by the fall on the reserve plant margins during this period as shown in Table 4:17 below.

TABLE 4:17

RESERVE PLANT MARGIN PERCENTAGES 1970 TO 1982

Year	%
1970	25,6
1971	36,9
1972	33,4
1973	29,0
1974	17,0
1975	14,6
1976	15,9
1977	18,8
1978	18,3
1979	17,1
1980	26,8
1981	29,4
1982	32,1

According to Escom policy the reserve plant margin should not be allowed to fall below 17 per cent. It can be seen that from 1974 to 1979 the reserve plant margin either fell below or remained fairly close to this minimum, indicating that the utilization of capacity must have fairly high during this period. The resultant greater use made of smaller power stations must have contributed to the decline in the efficiency of coal consumption during this period. It is significant that the easing of the capacity bottleneck during the last three years has been accompanied by an improvement in coal efficiency.

There appears to be a fairly close correlation between the average selling price of electricity and the average coal cost during the period 1960 to 1982. From 1960 to 1972, Escom was able to cause the average selling price of electricity to actually decrease in real terms through economies in coal consumption. Since then, however, the average price of electricity has risen since, on the one hand, the requirement to raise revenue to finance future capital expansion has substantially risen while, on the other hand, it has become increasingly difficult to improve the efficiency of coal consumption.

Having looked at the nature of the revenue constraint facing Escom and how it has changed over time, it is now necessary to examine the actual structure of the tariffs which Escom sets to raise the revenue to meet this constraint.

4:4:2:2 ESCOM'S TARIFF STRUCTURE

Escom's tariff policy is based on the following principles:

(i) the tariffs set for each regional undertaking should be based on the costs incurred in supplying electricity by the particular undertaking. No cross-subsidization between the different undertakings is therefore permitted.

(ii) the tariffs charged to the different groups of consumers supplied by each undertaking should be based on the different costs involved in supplying each group. One consumer group should therefore not subsidize another.

(iii) the structure of tariffs should reflect the different functional relationships affecting the various categories of cost incurred by the undertaking;

(iv) no undertaking should earn a surplus above or a deficit below its total allocated cost. Each undertaking is therefore permitted to adjust its tariffs to eliminate any such surplus or deficit;

(v) the tariffs charged may depart from the standard schedule of tariffs to take account of special cost considerations in particular cases.

Table 4:18 reflects how these principles are actually applied by Escom. The application of principle (i) is indicated by the fact that a separate schedule of tariffs is set for each regional undertaking. A schedule of standard tariffs is specified in the licence of each undertaking. An amendment of the licence is required to change the standard tariffs and the procedure for such amendment is prescribed in section 33 of the Act. As will be explained, below, it is not necessary to undertake this procedure in order to effect a periodic revision of tariffs in accordance with changes in cost levels. The effective schedule of tariffs prevailing at any point in time may thus differ from the schedule of standard tariffs. This is clearly shown in the schedule in Table 4:18 which applies to the first quarter of 1983.

Table 4:18 also shows how in each undertaking tariffs are differentiated between different consumer groups in accordance with principle (ii). Escom has four classes of consumer tariffs;

Tariff A - large user tariffs, generally applicable to loads in excess of 100 kW or kVA;

Tariff B - small user tariffs, generally applicable to consumers other than farmers with loads not exceeding 100 kW or kVA;

Tariff C - domestic tariffs, applicable to consumers residing within a proclaimed township or within an area considered by Escom to be similar to such a township; and

Tariff D - small user tariffs, generally applicable to farmers with loads not exceeding 100 kW or kVA.

TABLE 4:18 SCHEDULE OF ESCOM'S STANDARD AND EFFECTIVE TARIFFS (First Quarter 1983)

	Border		Orange River		Cape Northern		Cape Western	
Surcharge Discount	3% discount		4% surcharge		1% discount		NIL	
	Standard	Effective	Standard	Effective	Standard	Effective	Standard	Effective
TARIFF A								
Basic Charge	25,00	24,25	25,00	26,00	25,00	24,75	25,00	25,00
Demand Charge								
Above 380 V	12,35	11,98	12,00	12,48	9,75	9,65	9,90	9,90
380/220 V	12,60	12,22	12,25	12,74	10,00	9,90	10,15	10,15
132 K.V.								
Energy Charge (C/kWh)	1,72	1,752	1,67	1,820	1,40	1,469	1,44	1,524
TARIFF B								
Basic Charge	12,00	11,64	12,00	12,48	12,00	11,88	12,00	12,00
First 500 kWh C/kWh	8,75	8,56	8,96	7,10	7,10	7,70	7,70	7,78
Balance of kWh C/kWh	5,45	5,36	5,25	5,53	4,40	4,43	4,40	4,48
TARIFF C								
Basic Charge	6,00	5,82	6,00	6,24	6,00	5,94	4,00	4,00
First 300 kWh	8,75	8,56	8,55	8,96	7,10	7,10	7,20	7,28
Balance kWh	5,45	5,36	5,25	5,53	4,40	4,43	4,40	4,48
TARIFF D								
Basic Charge	12,00	11,64	12,00	12,48	12,00	11,88	12,00	12,00
First 800 kWh	8,75	8,56	8,55	8,96	7,10	7,10	7,70	7,78
Balance kWh	5,45	5,36	5,25	5,53	4,40	4,43	4,40	4,48

TABLE 5:18 (Continued)

	Eastern Transvaal		Natal		Rand and O.F.S. (estbl)		Rand and O.F.S. (Extn)	
Surcharge	36% surcharge		13,5% discount		33,5% surcharge		47,5% surcharge	
	Standard	Effective	Standard	Effective	Standard	Effective		
TARIFF A								
Basic Charge	25,00	34,00	35,00	30,27	25,00	33,37	25,00	33,31
Demand Charge								
Above 380 V	5,75	7,82	10,10	8,99	7,00	9,34/Kw	7,00/Kw	9,34/Kw
380/220 V	5,90	8,02	10,95	9,47	7,00/KW	9,34/kW	7,00/Kw	9,34/Kw
132 K.V.	-	-	10,40	9,25	-	-	-	-
Energy Charge (C/kWh)	1,03	1,484	1,62	1,259	1,114	1,614	1,114	1,614
TARIFF B								
Basic Charge	10,00	13,60	12,00	10,38	12,00	16,02	12,00	16,02
First 500 kWh C/kWh	4,95	6,80	8,00	6,77	5,35	7,21	5,35	7,21
Balance of kWh C/kWh	2,95	4,08	4,70	3,92	2,45	4,00	2,95	4,00
TARIFF C								
Basic Charge	5,00	6,80	6,00	5,19	6,00	8,01	6,00	8,01
First 300 kWh	4,95	6,80	-8,00	6;77	5,35	7;21	5,35	2,21
Balance kWh	2,95	4,08	4,70	3,92	2,95	4,00	2,95	4,00
TARIFF D								
Basic Charge	10,00	13,60	12,00	10;38	12,00	16,02	15,00	16,02
First 800 kWh	4,95	6,80	8,00	6,77	5,35	7,21	5,35	7,21
Balance kWh	2,95	4,08	4,70	3,92	2,95	4,00	2,95	4,00

In practical terms the application of principle (iii), that tariffs should be structured according to the variables functionally related to cost, means that the total cost allocated to each undertaking is divided into consumer, demand and energy-related costs.

A. <u>CONSUMER-RELATED COSTS</u> constitute the portion of total costs which vary with the number of consumers, for example, the cost of meters, meter reading, billing and connection. Since these costs are incurred for each consumer supplied by the undertaking, they are recovered by means of a basic or service charge which is levied on each consumer.

B. <u>DEMAND-RELATED COSTS</u> are the fixed costs related to the capacity constructed to provide for the demand, for example, interest and depreciation. Demand related costs also include those operating and maintenance expenses which do not vary with the quantity of service supplied in kWh or with the number of consumers, as well as part of the administrative and general expenses. The demand charge is that component of the tariff in which these costs are recovered. It can be seen from Table 4:18 that for consumers charged according to Tariff A, the fixed demand charge is based on the maximum demand in a particular month to cover the main portion of the demand-related costs. Escom's undertakings have, however, adopted a policy of including part of the demand related costs in the energy charge per kWh, referred to below, to "obtain a more equitable application of the tariff to consumers with either abnormally high or abnormally low load factors." The Board of Trade and Industries also found that "there appears to be a considerable measure of subjectivity in calculating the transfer" and that "the magnitude of the transfer ... differs between undertakings."[30]

---

30. Republic of South Africa, Board of Trade and Industries, <u>op.cit.</u>, p. 21.

Consumers who fall within tariff classes B, C and D are not levied a specific demand charge in their standard tariffs, since where the expenditure on distribution and transmission in supplying a group of such consumers exceeds the amount covered in the standard tariffs, they are required to pay extension charges. In the case of a farming scheme, common extension charges are applied based on consideration of the group as a whole. These charges are related to the demand which the consumers require and are subject to increase if the limit is exceeded, In calculating the costs of supplying these users, Escom distinguishes between the cost of "personal" and "common" equipment. "Personal" equipment is the terminal equipment provided at each consumer's point of supply. The cost of this equipment is charged directly to each consumer and is therefore deducted from the total cost of extension to determine the cost of "common" equipment, which is allocated to consumers according to the maximum demand to be supplied. The excess of estimated monthly expenditure over revenue has to be recovered through extension charges which are quoted to potential users. If a reasonable number, say 75 to 80 per cent, are prepared to sign agreements, the extension charges are fixed and the scheme proceeds. New consumers are required to pay either the standard extension charge for their level of use or, if the cost of connecting them is above the standard level, a higher charge based on actual cost.

Each undertaking differentiates its demand and extension charges according to the voltage at which electricity is supplied. This is because an additional cost is incurred to provide the equipment required to transform electricity from one voltage to another. It can be seen from Table 4:18 that in each undertaking, other than the Rand and Orange Free State (O.F.S.) undertaking, a lower demand charge is levied when the supply is furnished at a nominal voltage above 380/220 volts than when it is furnished at the nominal voltage of 380 volts "between phases and 220 volts between phase and neutral."[31]

---

31. Electricity Supply Commission, Statement of Changes for 1983, Megawatt Park, Sandton, p. 2.

In the Natal undertaking a further reduction in the demand charge is made when the supply is furnished "at a nominal voltage of 132 000 volts between phases."[32] The setting of the demand charge by the Rand and O.F.S. undertaking, however, differs from that applied by the other undertakings in that there is no differentiation of the demand charge according to voltage, and a demand charge of R7,00 per month per kilowatt of the maximum simultaneous demand supplied in the month at all the points of supply is levied. The reason for this difference is historical and this undertaking has not changed to the system of demand charges applied by the other undertakings, since it is believed that the administrative costs of effecting the change will outweigh any benefits that may be derived.

C. <u>ENERGY-RELATED COSTS</u> constitute that part of cost which varies proportionally with energy supplies, for example, costs of coal, fuel, fuel handling and water. These costs are recovered by means of an energy related charge per kWh of consumption. It can be seen from Table 4:18 that the energy charge is expressed as a constant rate per kWh consumed for Tariff A users, whereas for Tariff B, C and D users a high rate is charged for the first block of 500, 300 and 800 kWh, respectively, and the balance of consumption is charged at a constant rate. The aim of charging a decreasing block rate to this latter group of users seems to be to encourage users to increase their loads and thereby obtain a greater utilization of the capacity individually connected to them.

In summary, then, it can be seen that the application of principle (iii) results in Escom charging a three-part tariff which generally encourages users to increase their load factors, since with this tariff structure the ratio of fixed to total charges decreases if kWh consumption increases and discourages consumers with small loads from applying for an Escom connection at all.

---

32. *Ibid.*, 00305, p. 2.

Principle (iv), that each undertaking should adjust its tariffs to eliminate surpluses or deficits, is applied according to section 14(2) of the Electricity Act which states that "... the commission shall decrease or increase its prices for all classes of consumers in equal proportions when making any adjustment of prices in accordance with the provisions of this section." Escom interprets this section as indicating that each undertaking can apply surcharges or discounts in equal proportions for all classes of consumers, in order to adjust revenue to cover the total cost incurred by the undertaking without having to amend the schedule of standard tariffs contained in its licence in accordance with section 33 of the Act. Table 5:18 indicates how the surcharges or discounts levied by each undertaking cause the effective tariffs to differ from the standard tariffs. In addition to these surcharges and discounts, a coal price adjustment is applied exclusively to the energy charges of the tariffs. The cost of coal is calculated quarterly and compared with a standard price stated in the licence of the undertakings. Any difference results in a change in the kWh rates in the following quarter. This adjustment forms part of the standard tariff and is therefore subject to any surcharge or discount applied to the standard prices. As explained in the last section, the average price of coal has risen sharply in recent years, so that the effect of this adjustment can be significant.

According to principle (v) the tariffs set by Escom's undertakings are modified in certain cases to take account of special cost considerations:

(a) Although the standard schedule of tariffs does make provision for the differentiation of demand charges according to voltage, Escom also has special arrangements to charge additional voltage transformation costs to certain large municipalities such as Johannesburg, Pretoria, Cape Town, Port Elizabeth, East London and Durban.

(b) A minimum annual payment is calculated for the annual demand charge to municipalities which generate a portion of their consumption themselves according to the formula:

$$\frac{\text{Purchases from Escom in overall annual maximum demand month}}{\text{Overall monthly maximum demand (purchased and generated)}}$$

$$\times \; \Sigma \text{ Monthly maximum demand (purchased and generated)}$$

If a municipality purchases, say, two-thirds of its maximum monthly consumption from Escom, then, according to this formula, it should buy in total two-thirds of its annual consumption from Escom. The minimum annual payment is thus aimed at a sharing of the generating municipality's summer load between the municipality and Escom in the same proportion as that in which the winter load is shared.

(c) To obtain optimum utilization of its standby plant, Escom has entered into standby agreements with generating municipalities in terms of which, at all times except during the peak months of June, July and August, "... these municipalities are allowed to increase their maximum demand without paying additional demand charges, to the extent of largest set on their system .... provided Escom has adequate reserve plant."[33] If Escom does not have sufficient reserve plant, this concession is withdrawn and the municipalities can also be requested to reduce their load by reducing their standby plant. The municipalities benefit from these agreements in that they are induced to use power at a charge below the coal cost, while Escom benefits by using its spare capacity to sell power at a rate above marginal cost and by being able to call for assistance if it is short of generating capacity.

(d) In addition, the tariff policy of the Rand and O.F.S. undertaking differs from that adopted by the other undertakings in the following respects:

---

33. Board of Trade and Industries, op.cit., p. 20.

(i) the demand charge is set per kilowatt of electricity supplied, rather than per kilovolt as in the other undertakings;

(ii) the maximum demand is 60 minutes and not 30 minutes;

(iii) rather than differentiating the demand charge according to voltage, this undertaking offers a discount of 10 per cent of the amount by which the total monthly electricity charge exceeds R1 500;

(iv) a higher surcharge is levied on consumers located in the extension area of this undertaking;

(v) where electricity is furnished at more than one point of supply to consumers in this undertaking, an additional charge calculated on the difference between the sum of the non-simultaneous maximum demands taken in the month at each point of supply, and the maximum simultaneous demand is levied at the rate of R1,10 per month per kilowatt; and

(vi) only in this undertaking is S.A.T.S. charged at a straight rate per kWh without any service and demand charges.

4:4:2:3 ADMINISTRATIVE ASPECTS OF ELECTRICITY TARIFFS

The responsibility for the formulation of tariffs is vested in the Commercial Department at Escom Head Office. This department is concerned with the calculation of standard tariffs, discounts or surcharges and with research into tariffs. It bases its tariff calculation on estimates of cost and revenue for the distribution undertakings.

Changes in tariffs may be effected in the following ways:

(i) by changes in the Electricity Act;

(ii) by changes in standard tariffs; or

(iii) by changes in surcharges or discounts.

It is thus possible for Escom's tariffs to be changed via an amendment to the Electricity Act. To apply for such a change Escom would need to channel its proposal through the Department of Mineral and Energy Affairs. An example of this occurred in 1977 when the maximum contribution to the Capital Development Fund was raised from 3 to 6 per cent of unredeemed loans, on the basis of a claim by Escom that it needed to increase its internal funds to finance its expansion plans, since it was having problems in obtaining overseas loans.

Escom can also change its standard tariffs with the approval of the Electricity Control Board. To obtain this approval the proposed changes must be published in a number of prescribed newspapers, a prescribed time period must elapse during which the public is given the opportunity to lodge objections to the tariff increase with the Board, a public hearing of objections received must be arranged by the Control Board and the Board must determine whether or not to grant its approval.

Escom can, however, change its tariffs without obtaining the approval of the Control Board by adjusting the surcharges and discounts for the different undertakings. Although it has been Escom practice to change the surcharges as its own discretion, the Board of Trade and Industries reported that:
"The Electricity Control Board, however, feels that it should also approve changes in surcharges. One of the reasons for this attitude is that the Electricity Control Board should approve of new capital investments of the CGU (Central Generating Unit) ... and such investments directly affect tariffs. It came to the attention of the Board that in the case of the Koeberg nuclear power station application for approval was made after Escom had been fully committed on the main plant contract ... and also the associated fuel fabrication contract."[34]

---

34. Ibid., p. 33.

4:4:2:4 COST ALLOCATION

The process of cost allocation involves two steps:

A. the allocation of costs to the different regional undertakings; and

B. the allocation of costs within the undertaking to consumer groups.

A. ALLOCATION OF COSTS TO UNDERTAKINGS

The Electricity Supply Account for 1982, shown in Table 4:19, which is directly reproduced from the Escom Annual Report for that year, indicates that the total costs of any undertaking consist of the following items:

(i) the operating expenditure actually incurred by the undertaking;

(ii) the loan charges of the undertaking which are linked to its assets in commercial operation;

(iii) contributions to the Capital Development Fund which are allocated in proportion to the kWh sales of the undertaking;

(iv) a share of the corporate burden or head office expense which is determined in proportion to the capital in commercial operation in the undertaking at the end of the year;

(v) credits to the undertaking in respect of transmission circuits used by the Central Generating Unit (C.G.U.) for interconnecting purposes;

(vi) credits and debits in respect of the use of transmission circuits for reciprocal assistance between undertakings;

(vii) credits and debits in respect of electricity supplied by an undertaking to a consumer in another undertaking; and

(viii) allocated C.G.U. costs.

It can be seen from Table 4:19 that for each undertaking, by far the largest item of cost is the allocated C.G.U. cost. In an interconnected

Electricity supply account

for the year ended 31 December 1982

1981		R000 1982								
				Central	Distribution					
Total		Total	Corporate Services	Generating	Total	Cape Western	Cape Northern	Border	Orange River
2 140 689	**Electricity sold**	2 695 422	—	—	2 695 422	253 513	102 970	44 728	48 038
710 025	Industrial	874 582	—	—	874 582	83 648	12 484	6 090	3 942
680 031	Bulk	908 289	—	—	908 289	109 341	18 315	36 287	43 875
563 983	Mining	695 392	—	—	695 392	—	56 089	—	—
139 261	Traction	161 893	—	—	161 893	27 208	15 680	—	—
47 389	Domestic and lighting	55 266	—	—	55 266	33 316	402	2 351	221
1 061 051	**Operating expenditure**	1 400 636	34 789	1 145 972	219 875	39 516	13 016	8 425	7 686
645 270	Operations	793 387	1 084	789 466	2 837	346	284	88	88
137 919	Maintenance	222 286	810	164 461	57 015	8 393	2 895	2 221	2 092
4 106	Electricity purchased	3 615	—	3 615	—	—	—	—	—
273 756	Administration and general expenses	381 348	32 895	188 430	160 023	30 777	9 837	6 116	5 506
720 634	**Loan charges**	876 706	(3 269)	636 888	243 087	31 379	19 154	5 040	9 223
603 546	Interest and finance charges	721 948	(3 948)	506 456	219 440	28 936	17 644	4 656	7 805
70 493	Redemption of local loans	106 714	679	82 538	23 497	2 443	1 510	384	1 418
46 595	Repayment of foreign loans	48 044	—	47 894	150	—	—	—	—
900	**Contribution to Reserve Fund**	26 000	—	26 000	—	—	—	—	—
—	**Distribution of costs**	—	(31 520)	(1 808 860)	1 840 380	156 167	63 506	23 693	25 525
—	Corporate burden	—	(31 520)	23 753	7 767	1 072	684	189	205
—	Interconnectors	—	—	2 821	(2 821)	—	—	—	(1 178)
—	Use of circuits	—	—	—	—	—	178	91	(91)
—	Transmission costs	—	—	(38 292)	38 292	17 324	7 508	1 180	1 824
—	Pooled generation	—	—	(1 797 142)	1 797 142	137 771	55 136	22 233	24 765
1 782 585	**Total charges against revenue**	2 303 342	—	—	2 303 342	227 062	95 676	37 158	42 434
358 104	**Operating surplus for the year**	392 080	—	—	392 080	26 451	7 294	7 570	5 604
435 478	**Amount set aside to Capital Development Fund**	450 000	—	—	450 000	32 836	12 486	4 843	6 467
(77 374)	**Surplus/(Deficit) for the year**	(57 920)	—	—	(57 920)	(6 385)	(5 192)	2 727	(863)
(18 365)	**Accumulated surplus/(deficit) at beginning of year**	(95 739)	—	—	(95 739)	(14 072)	764	2 126	(7 919)
(95 739)	**Accumulated surplus/(deficit) at end of year**	(153 659)	—	—	(153 659)	(20 457)	(4 428)	4 853	(8 782)

TABLE 4:19 Escom Electricity Supply Account

R000

1981

Undertakings: Natal	Undertakings: Eastern Transvaal	Undertakings: Rand and O.F.S.	Corporate Services	Central Generating	Distribution Undertakings: Total	Cape Western	Cape Northern	Border	Orange River	Natal	Eastern Transvaal	Rand and O.F.S.
435 154	388 255	1 422 764	—	—	2 140 689	197 005	92 637	35 277	38 740	325 615	316 915	1 134 500
149 980	238 540	379 898	—	—	710 025	66 708	9 370	4 754	3 478	111 002	205 032	309 681
219 437	47 241	433 793	—	—	680 031	79 211	14 149	28 729	35 045	161 621	32 314	328 962
10 938	80 332	548 033	—	—	563 983	—	51 189	—	—	9 003	62 020	441 771
47 750	20 649	50 606	—	—	139 261	22 913	16 717	—	—	37 921	16 369	45 341
7 049	1 493	10 434	—	—	47 389	28 173	1 212	1 794	217	6 068	1 180	8 745
37 120	31 436	82 676	14 419	879 750	166 882	30 727	10 991	6 191	5 484	29 950	22 699	60 840
283	486	1 262	782	642 225	2 263	233	199	74	74	416	355	912
9 310	11 720	20 384	712	96 436	40 771	6 637	2 378	1 606	1 363	6 261	7 973	14 553
—	—	—	—	4 106	—	—	—	—	—	—	—	—
27 527	19 230	61 030	12 925	136 983	123 848	23 857	8 414	4 511	4 047	23 273	14 371	45 375
27 101	38 318	112 872	8 738	494 800	217 096	28 475	19 938	5 744	5 714	27 727	35 277	94 221
24 715	35 421	100 263	7 869	399 834	195 843	25 558	18 014	5 261	5 168	24 938	31 700	85 204
2 236	2 897	12 609	869	48 522	21 102	2 917	1 924	483	546	2 638	3 577	9 017
150	—	—	—	46 444	151	—	—	—	—	151	—	—
—	—	—	—	900	—	—	—	—	—	—	—	—
308 385	252 543	1 010 561	(23 157)	(1 375 450)	1 398 607	115 933	51 807	17 739	19 715	229 079	199 078	765 256
949	1 223	3 445	(23 157)	17 051	6 106	854	565	158	169	757	933	2 670
—	(171)	(1 472)	—	2 970	(2 970)	—	—	—	(1 202)	—	(182)	(1 586)
—	(223)	45	—	—	—	—	296	91	(91)	—	—	(296)
9 826	84	546	—	(35 254)	35 254	15 408	7 451	1 011	1 813	8 951	33	587
297 610	251 630	1 007 997	—	(1 360 217)	1 360 217	99 671	43 495	16 479	19 026	219 371	198 294	763 881
372 606	322 297	1 206 109	—	—	1 782 585	175 135	82 736	29 674	30 913	286 756	257 054	920 317
62 548	65 958	216 655	—	—	358 104	21 870	9 901	5 603	7 827	38 859	59 861	214 183
74 768	65 781	252 819	—	—	435 478	30 986	12 667	4 507	6 346	69 073	67 678	244 221
(12 220)	177	(36 164)	—	—	(77 374)	(9 116)	(2 766)	1 096	1 481	(30 214)	(7 817)	(30 038)
(2 393)	(15 348)	(58 897)	—	—	(18 365)	(4 956)	3 530	1 030	(9 400)	27 821	(7 531)	(28 859)
(14 613)	(15 171)	(95 061)	—	—	(95 739)	(14 072)	764	2 126	(7 919)	(2 393)	(15 348)	(58 897)

TABLE 4:19 (Continued)

system of electricity generation and transmission such as Escom, the total cost incurred by the central generating undertaking is a common cost which is jointly incurred by all the distribution undertakings. Since it is not possible to determine what proportion of this common cost is directly attributable to each individual undertaking, Escom has accepted that the fairest method of allocation of C.G.U. costs is to pool them between the different undertakings in proportion to their electricity requirements. It so happened, however, that when the different undertakings were connected to the national transmission system. Escom had to take over the Salt River and Hex River stations operated by the Cape Western Undertaking; the Colenso, Congella, Ingagene and Umgeni stations operated by the Natal Undertaking; and the West Bank Station operated by the Border Undertaking. The generating costs per kW capability and per kWh of all these power stations was higher than that of the C.G.U., and the basic principle was then established that the excess generating cost of these power stations must continue to be carried by the undertakings concerned.

Although the original permit, which applied from 26 January 1972 to 4 August 1976, stated that the total excess generating cost should be allocated to the above coastal undertakings, this was amended firstly, on 4 August 1976, to exclude the excess generation costs of coastal power stations with annual load factors of more than 20 per cent and again, on 19 January 1977, to limit the excess generation costs to the interest and redemption charges of these power stations. The burden of excess generation costs borne by the coastal undertakings has thus declined considerably since 1972.

The basis for allocating contributions to the Capital Development Fund to the undertaking was changed in 1977 from the average monthly maximum demand of the undertakings to their kWh sales. This had the effect of reducing the burden on newer undertakings where the historical cost of assets in service is higher in relation to sales than in the older undertakings. The Board of Trade and Industries reported that "the change in the allocation methods was to the benefit of all undertakings except the Rand and O.F.S. undertaking."[35]

The loan charges for each undertaking are based on the capital in commercial operation in the undertaking. In undertakings where the loans used to finance assets still in service have been redeemed, the historical cost of assets in service will exceed the unredeemed loans in respect of these assets. The loan costs of undertakings will therefore be determined by the costs of servicing these outstanding loans. Since there has been an upward trend in interest rates, it follows that the overall average interest rates applicable to the unredeemed loans of older undertakings will be lower than those of the younger undertakings. There will be no such regional difference in the allocated cost of overseas loans, however, since these loans have been mainly used to finance C.G.U. assets and are therefore allocated in the same "equitable" proportion as other C.G.U. costs.

In allocating transmission costs to the undertakings, it has been necessary to distinguish between the cost of interconnectors which are required by the C.G.U. for the optimum utilization of the power resources available to it, and the cost of transmission lines which are required mainly for the transmission of electricity to specific undertakings. A C.G.U. interconnector circuit contributes to the reliability and economy of supply of all undertakings and its cost is therefore included in the total C.G.U. costs, which are pooled in an equal proportional manner between all undertakings. The cost of a transmission circuit, however, can be attributed to a particular undertaking since the outage of such a circuit would only affect the transmission of power to a particular distribution undertaking. It follows that the undertakings are required to bear the transmission costs specifically attributable to them.

---

35. R.S.A., Board of Trade and Industries, op.cit., p. 50.

B. COST ALLOCATION TO CONSUMERS

Once costs have been allocated to the distribution undertakings, the next step in the tariff formulation process is for each undertaking to allocate costs to its consumers according to their tariff class. Escom tends to have a comparative advantage in the bulk supply of electricity, so that the main part of Escom's sales is to Tariff A consumers. The following procedure has been adopted to allocate costs to Tariff A consumers:

(i) establish the total costs of the undertaking;

(ii) deduct income receivable from Tariff B, C and D consumers, as well as extension charges payable by Tariff A consumers;

(iii) deduct service charge revenue receivable from Tariff A consumers; and

(iv) divide the balance of total cost remaining between demand related cost (including consumer related cost) and energy related cost.

As was mentioned in the previous section, the demand-related cost is not completely recovered by means of demand charges, since a portion of it is transferred to the energy-related cost to be recovered through the energy charge. The object of the transfer is to prevent the overcharging of low load factor consumers who make a larger contribution to diversity. The Board of Trade and Industries Report found that this transfer "... is determined subjectively for each undertaking to result in an acceptable relationship between undertakings."[36] The following factors seem to affect the magnitude of the transfer:

(i) the load factor of the undertaking since the higher the load factor, the lower is the diversity, and, therefore, the lower should be the transfer;

---

36. *Ibid.*, p. 61.

(ii) the similarity of the cost structure and the regional proximity of undertakings. It is, for example, considered desirable by Escom "to set the tariffs of two adjoining undertakings with similar cost structures but with different load and diversity factors ... on such a level that average costs per kWh are the same;"[37]

(iii) the generating cost of municipalities which are generating electricity themselves seems to set a ceiling on the extent to which the energy charge can be raised by transferred demand cost;

(iv) the extent of the incentive which Escom wants to give its consumers to improve load factors, since this must be traded off against the benefits of the transfer; and

(v) the combined effect of the coal prices adjustment and the periodic adjustment of the overall surcharge.

A similar method is used to allocate costs to consumers in tariff classes B, D and D. This can be illustrated with reference to the domestic consumers falling within Tariff class C. Since domestic consumers are generally supplied by municipalities, Escom determines the tariffs of the domestic consumers it supplies in lieu of certain municipalities, by considering the costs of distribution and reticulation which would have been borne by a typical municipal area if it had undertaken the responsibility to supply these consumers. Thus the rate for the second block of Tariff C is "... designed to cover the average cost per kWh sold of electricity purchased in bulk at Tariff A by the hypothetical municipality, i.e. the maximum demand charge expressed per kWh at the after diversity load factor plus the energy charge." No exact calculation is performed in respect of the first block and the general practice adopted by Escom's undertakings has been to set this rate at double the rate of the second block. The estimated shortfall of the revenue expected from the first and second blocks from the

37. *Ibid.*, p. 63.

estimated cost of domestic supply, is shared on a pro rata basis between consumers as a service charge. The difference between the first and second blocks also contributes to the recovery of costs incurred in the distribution and reticulation of electricity from bulk input points onwards.

4:4:3 EVALUATION OF ESCOM'S TARIFF POLICY

4:4:3:1 THE REVENUE CONSTRAINT

In section 4.2.2, Escom's financial record and the composition and method of formulation of some of the important charges against revenue were examined. It was found that although Escom purports to set its tariffs to earn sufficient revenue to just cover its costs, its tariff policy has in fact resulted in substantial profits being earned in recent years. This has occurred despite the fact that public enterprises are not usually permitted to use their monopoly powers to earn profits, but are required to pursue pricing policies which maximize net social benefits. However, as was explained in Chapter Three, the pricing policies adopted by public enterprises are often constrained by a requirement to generate sufficient revenue to cause the enterprise to remain financially viable and independent of Treasury support through subsidization of losses. In evaluating the effect that Escom's pricing has had on its profitability, it is necessary to consider the following questions:

A. Is it possible to justify the actual level of profitability which has resulted from Escom's pricing policy? and

B. Are there adequate controls to prevent Escom from using its monopoly powers in a socially suboptimal manner?

A. THE JUSTIFIABILITY OF THE LEVEL OF INTERNAL FINANCING PURSUED BY ESCOM

The profits earned by Escom from its tariff revenue cannot be distributed but must be ploughed back into the funding of the Commission's operations. In considering whether the level of profits recently earned by Escom can be justified, it is therefore necessary to evaluate Escom's internal financing requirements.

It is important to recognize that Escom does not view its financial viability in the sense of just breaking even in the short term, but rather assesses its viability according to its ability to meet the long term financing requirements of its capital development programme. Escom's forms of financing are external borrowing and statutory funds, reserves and provisions. External borrowings consist of four sources:

(i) Escom stock issued less Escom stock held internally;
(ii) import financing facilities;
(iii) short term loans; and
(iv) net current liabilities.

Statutory funds, reserves and provisions consist of:

(a) funding comparable to accumulated depreciation in a manufacturing concern, namely the following items which appear in Escom's financial statements:
  (i) Redemption Fund;
  (ii) Provision for Repayment of Foreign Loans; and
  (iii) Capital Reserve; and
(b) financing comparable to equity in a manufacturing concern, namely:
  (i) Capital Development Fund;
  (ii) Reserve Fund;
  (iii) unrealized surplus on Escom stock held internally; and
  (iv) Surplus/Deficit.

Table 4:20 shows the equity/total asset ratio for Escom during the period 1967 to 1982. To calculate this ratio, the funds comparable to accumulated depreciation have been deducted from total assets.

TABLE 4:20

ESCOM : EQUITY/TOTAL ASSET RATIOS, 1967 TO 1982

Year	Per cent
1967	11,8
1968	10,9
1969	10,9
1970	11,2
1971	11,0
1972	11,6
1973	12,3
1974	12,8
1975	13,1
1976	13,1
1977	17,5
1978	22,2
1979	24,4
1980	26,8
1981	28,1
1982	27,4

Now, the optimal capital structure in any undertaking will be determined by the business and financial risks associated with its operations. Business risk will be affected by the type of assets employed, the type of industry and the demand for the product (s). Natural monopolies such as Escom are generally characterized by low business risk, due to the stable demand for their services and their stable earnings pattern. In the long term, Escom is characterized by a particularly low business risk due to the following factors:

(i) electricity has a particularly stable demand since it is a small input into a large number of business undertakings;

(ii) the price elasticity of demand for electricity is low;

(iii) the Electricity Act permits Escom to set its tariffs in order to cover the items included in "total cost"; and

(iv) as a public enterprise, Escom is effectively underwritten by the State.

Financial risk is inversely related to the equity/total asset ratio since the smaller the proportion of total assets financed by equity capital, the larger the financial risk of the undertaking defaulting on its interest and redemption payments. An enterprise with a low business risk such as Escom will, in the long run, be able to bear a high financial risk by having a capital structure characterized by a low equity/total asset ratio. In the short run, however, the ratio between equity and borrowed capital will depend on the availability of loan capital, interest rates, the stage of the business cycle and the monetary growth rate of fixed assets. It follows that this ratio may fluctuate from year to year. For example during an economic upswing, the bulk of funds required might be generated internally by raising tariffs, while during a recession prices may not be increased and a greater use be made of loan capital which, in any case, should be readily available during this phase. Furthermore, if there is a sharp increase in the growth rate of fixed assets, a large proportion of the financing requirements will have to be met by raising loans, since it would be undesirable to implement a massive increase in tariffs.

Table 4:20 indicates that during the period 1967 to 1976, Escom had an average equity/total asset ratio of 11,9 per cent which is low in comparison to the widely accepted "normal" ratio for a manufacturing concern of 50 per cent. As mentioned above, this capital structure results from the low business risk characterizing Escom's operations. In 1970 the Franzsen Commission recommended that public enterprises in South Africa should make greater use of internal financing for the following reasons:

"(i) to reduce the burden of interest payments;

(ii) to reduce the claims made by public enterprises on the capital market;

(iii) to provide the additional funds necessary to finance the massive capital development programmes required by developing countries; and

(iv) to provide for the rising cost of replacing fixed assets in inflationary times."[37]

To this end, the Franzsen Commission recommended that the tariffs of public enterprises be "... adjusted so that (1) provision for depreciation can be made on the basis of replacement instead of original costs of assets and (2) 40 per cent, or such a percentage as may be determined by the Cabinet Committee for Finance, of the net or new investment, i.e. the difference between the gross capital expenditure and the provision for depreciation, can be provided from current revenue."[38] The Franzsen Commission thus recommended that the ratio of equity to loans be determined on a marginal basis. The problem with this marginal approach is that the debt/equity ratio will be generally affected by short term economic fluctuations. This conflicts with the principle that this ratio should be determined on the basis of a long term view of the relationship between business and financial risk in the enterprise.

Escom introduced the Capital Development Fund in 1972 as a result of the recommendations of the Franzsen Commission. However, during the period 1972-1976, the contributions to this fund out of revenue were below the legal maximim and there was thus no significant increase in the equity/total asset ratio. From 1977 to 1980, however, these contributions were stepped up to their legal limit, the effect of which is indicated by the increase in equity/total asset ratio from 13,1

---

37. It should be noted that an increase in internal financing by Escom may not significantly reduce the total interest charge, since the investments associated with Escom's internal funds are held in Escom stock, the interest of which is still chargeable to revenue.

38. R.S.A., Third Report of the Commission of Enquiry into Fiscal and Monetary Policy in South Africa, R.P. 87/1970, p. 50.

per cent in 1976 to 26,8 per cent in 1980. It was on the basis of this apparent change in policy with respect to internal financing that the Board of Trade and Industries predicted that the equity/total asset ratio would increase in the manner shown in Table 4:21 below.

TABLE 4:21

Year	Projected %	Actual %
1978	20,2	22,1
1979	23,2	24,4
1980	27,2	26,8
1981	31,9	28,1
1982	37,4	27,4
1983	43,9	-
1984	51,3	-
1985	57,2	-

It can be seen from Table 4:21 that from 1978 to 1980, the equity/total asset ratio increased at approximately the rate predicted in the Board Report, but that in 1981 and 1982 the actual ratio was significantly below the projected ratio. This may indicate that the substantial increase in financing from tariff revenue during the period 1977 to 1980 was the result of short term factors, such as the need to rapidly expand the investment base of the Capital Development Fund, and not indicative of a deliberate policy to increase Escom's equity/total asset ratio to a level similar to that in manufacturing. The Board of Trade and Industries' conclusion that the degree of internal financing planned by Escom is excessive may not be completely justified, since it appears that Escom may allow its equity/total asset ratio to stabilize at approximately 30 per cent. According to the Board Report, this level broadly reflects the minimum degree of financial risk which can be borne by Escom. In 1981 and 1982, the contributions to the Capital Development Fund fell below the legal limits. This indicates that a certain amount of "slack" may have arisen in Escom's internal financing, which can be used when it is necessary for Escom to once again increase its reliance on funds generated from revenue sources. This raises the question of whether the current legal framework provides adequate control over Escom, which will be examined in part (b) of this section.

It can be concluded that the level of profit retention by Escom in recent years may have been justified, if it is accepted that the equity/ total asset ratio needed to be raised to reduce the financial risk associated with Escom's overdependence on borrowed funds. The proportion of debt to equity has, however, now reached an apparently acceptable level, and further tariff adjustments to increase the share of internal financing may therefore be unnecessary.

B. THE ADEQUACY OF THE SYSTEM OF CONTROL OVER ESCOM

Escom is, in effect, a monopoly supplying a service with a low price elasticity of demand. It is not difficult, therefore, for Escom to increase its revenue by raising its tariffs. To prevent Escom from setting prices excessively higher than social opportunity cost and from tolerating "X-inefficiency"[39] in its own organization, it is necessary that there be some control over its tariff policy. As was explained in Chapter Two, it is preferable that control be exercised in a decentralized rather than a centralized manner, to avoid excessive administrative costs and the duplication of decision making. This decentralized type of control is normally exercised through the basic constitution of the enterprise which delineates its duties, responsibilities and privileges, and through the pricing and investment rules which the enterprise is required to observe in its decision making.

---

39. The concept of X-inefficiency is discussed in Chapter One, section 1:3:1.

In the case of Escom, the Electricity Act clearly defines the costs which must be covered by its tariff revenue and stipulates that Escom should neither earn a surplus above, nor incur a deficit below, these costs. The question still remains as to whether it is possible for Escom to still exercise a certain amount of discretion, within these legal constraints, as to

(i) the level of internal financing it is going to generate through its tariffs, and

(ii) the amount of X-inefficiency it is going to tolerate within the organization.

In section 4:4:2:3 it was explained how Escom could raise its tariffs by adjusting the surcharges or discounts applied with respect to the distribution undertakings without the approval of some outside body. The purpose of these adjustments in the surcharges and discounts is to cause each distribution undertaking to cover its costs. There is thus clearly a limit to the extent to which tariffs can be raised by an adjustment in the surcharges, since the undertakings are required to make neither a surplus nor a deficit. The real issue, then, in evaluating the degree of control over Escom's monopolistic power to overcharge its users in terms of marginal cost pricing principles is whether there is any constraint on Escom to contain increases in some of the items included in "total cost", since if costs rise, Escom is obliged to raise its tariffs to generate the additional revenue needed to cover these costs. In this respect the method of fund accounting employed by Escom can result in a persistent upward drift in costs and prices. This arises from the fact that the income earned on the investments held by Escom's internal funds is generally reinvested in the internal stock of Escom. Thus, although the contributions to the Capital Development Fund are limited to 6 per cent of the value of unredeemed loans, these unredeemed loans include the internal stock held by Escom, so that each annual contribution not only adds to the accumulated amount of the fund but is invested in Escom stock and therefore also adds to the amount of unredeemed loans, and thereby pushes up the ceiling on further annual contributions. Furthermore, as the accumulated amount of this fund rises, the interest earned on the Escom stock held by the fund places an increasing burden on the users of electricity, since they must pay higher tariffs to cover these rising interest costs. Tables 4:8 and 4:9 indicate the escalating nature of the interest charge on the Escom stock held by the Capital Development Fund, which rose from R33,1 million and 14,8 per cent of Escom's total interest charge in 1977 to R350,2 million and 48,5 per cent of the total interest charge in 1982. In section 4:4:2:1 it was suggested that both contributions to funds and the interest earned on internal stock were retained profits rather than costs. It follows that since the method of fund accounting used by Escom allows this undertaking to retain an increasing amount of its profits, there is scope for Escom to exercise its monopoly powers to both raise the price of its services above their opportunity costs and to shift the burden of its X-inefficiency on to its consumers.

Another important area of decision making on the part of Escom management, not subject to external approval or control, is investment. Escom's management appears to adhere very strongly to the view that its function is to ensure that the supply of electricity is adequate to meet the economy's electricity requirements, and that management should be responsible for determining the nature and volume of the capital expenditure required to fulfil this function. The following comment by Dr. Straszacker, a past chairman of Escom, clearly expresses this attitude:

> "It is important for Escom to be independent from controls that have nothing to do with power demand ... Escom is so much a pillar of economic development that it must not be diverted from its main purpose ... If the demand for electricity grows well, the economy creates what is necessary to produce that growth ... There is a circular effect. If you start curbing the supply of electricity, not only do you create unemployment but you make it difficult for the economy to get going again for lack of power."[40]

---

40. Financial Mail, April 6, 1979, p. 24.

He comments further:

> "The worst thing that could happen to South Africa now, I believe, would come about if Escom was prevented from safeguarding electricity supply for the future."[41]

Since the capital related charges to revenue, which in most years constitute more than half of total costs, are directly related to the volume of capital expenditure required for Escom's capital development programme, it follows that the discretion given to Escom's management in making investment decisions may affect the price charged for electricity, if there are factors which motivate the management to incur capital expenditure in excess of the minimum amount necessary to meet the forecasted demand for electricity. In other words, is it in the management's interest to pursue an investment policy which minimizes cost? To answer this question, it may be helpful to considerthe "managerial" theories of Williamson[42] and Marris[43], which seek to explain decisionmaking in enterprises where there is a separation between ownership and management. According to these theories, the main factor which constrains the management to take account of the interests of the owners of the enterprises is the fact that they are accountable to these owners and may be dismissed by them. In the case of Escom, its management is accountable to a Minister who evaluates their overall performance in terms of their ability to ensure that Escom fulfils its basic function to "provide a cheap and abundant supply of electricity wherever required."[44] Escom's management is therefore constrained

---

41. Financial Mail, July 27, 1979, p. 322.
42. Marris, R., "A Model of the Managerial Enterprise", Quarterly Journal of Economics, 1963.
43. Williamson, O., "Managerial Discretion and Business Behaviour", American Economic Review, 1963.
44. Electricity Act., 54(b).

both to contain the cost and to safeguard the supply of electricity. Often, however, there is a trade-off between these two functions. An example of this occurs in the reserve plant margin decision. As has been explained, it is the policy of most electricity undertakings to provide capacity in excess of the amount required to meet forecasted peak demand, to avoid possible power cuts as a result of unplanned increases in demand or breakdown in generating capacity. In deciding on the amount to invest in reserve capacity, the undertaking has to balance the cost of the capital investment in reserve plant against the benefits of reducing the risk of power failure. It is possible that Escom's management might attach a greater weight to the benefits obtainable from reducing the risk of power failure than would be the case in an independent cost-benefit analysis of the reserve plant investment decision. The reason for this is that the provision of an adequate supply of electricity is a direct responsibility of Escom's management, so that any power failures would adversely reflect on their planning ability whereas it is much more difficult to hold the management directly responsible for tariff increases caused by excessive capital expenditure. Consumer resistance to such a tariff increase is likely to be relatively weak at present due to the following factors:

(i) the price of electricity in South Africa is still relatively low by international standards;

(ii) in an inflationary period consumers become conditioned to expect regular increases in the prices of most goods and services;

(iii) the cost of electricity normally forms a small part of the total expenditure of most consumers; and

(iv) the establishment of a national transmission network by Escom has meant that there is a substantial gap between the cost of power supplied by Escom and that which could be supplied by municipal generating stations.

An indication of this relative weakness of consumer resistance to tariff increase is provided by the fact that in 1977 Escom was able to "get away with" an increase in the average price of electricity of 48 per cent. Finally, it can be suggested that the high preference which the Escom management may well have for minimizing the risk of power failure is likely to be reflected in a bias toward overinvestment in capacity, which is in turn likely to place an upward pressure on costs and tariffs.

This view was expressed by the Board of Trade and Industries Report, which pointed out that the investment by Escom in reserve plant may have been excessive. Table 4:22 sets out the reserve plant margins attained by Escom during the period 1970 to 1982. When one considers that it is Escom policy to aim at a reserve plant margin of 17 per cent, then it can be seen that, during the periods 1970 to 1973 and 1980 to 1982, the actual margin was significantly greater than this target.

TABLE 4:22

ESCOM : RESERVE PLANT MARGIN, 1970 TO 1982

Year	Percent
1970	25,6
1971	36,9
1972	33,4
1973	29,0
1974	17,0
1975	14,6
1976	15,9
1977	18,8
1978	18,3
1979	17,1
1980	26,8
1981	29,4
1982	32,1

The Board's Report stated that "a projected reserve of thirty per cent appears relatively high in the light of the burden imposed upon the country's capital resources as well as the impact on the economy by way of higher tariffs"[45] and suggested that "careful cost benefit analysis should be undertaken by Escom to determine whether enlargement of the maintenance staff, the introduction of maintenance shifts and/or carrying on of maintenance work in both summer and winter would not be less costly and more productive than substantial enlargement of the reserve margin as a means of overtaking maintenance backlogs."[46]

However, if it is accepted that Escom management has too high a risk aversion to investigate such cost-reducing possibilities, a greater control over capital expenditure may be effected if the authority to approve investment decisions is vested in a body "possessing both the overall insight into the present and probable future internal and external economic situation of the Republic and also the analytical capability necessary to make sound, considered judgements upon the merits of such capital expenditure from the national point of view."[47] Acceptance of the Board of Trade and Industries' recommendation that a "Capital Projects Evaluation Group" should be appointed in the Department of Finance to evaluate the investment merits and monitor the financing of the capital expenditure plans of Escom and other state corporations, might therefore play an important role in containing the Escom's capital expenditure and the escalating cost of electricity.

4:4:3:2 EVALUATION OF ESCOM'S TARIFF STRUCTURE AND RATE MAKING PRINCIPLES

In this section the tariff structure and rate making principles of Escom will be evaluated according to three basic criteria:

---

45. R.S.A., Board of Trade and Industries, Report No. 1889, op.cit., p. 153.
46. Ibid., p. 153.
47. Ibid., p. 153.

A. The "equity" criterion according to which it is fair or equitable for electricity prices to be set equal to the accounting costs attributable to particular groups of consumers, so that no group of consumers should subsidize any other group;

B. the "efficiency" criterion according to which electricity tariffs should reflect the opportunity costs of supplying electricity, so that they act as signals to induce the most efficient allocation of resources to the electricity sector; and

C. the "equality" criterion according to which it is considered proper for all consumers to pay the same rate per kWh for electricity, regardless of the costs incurred in supplying them. In particular the effects of removing regional differences in the average selling price of electricity will be examined.

It is clear that conflicts will often arise between these criteria, so that the rate maker will have to decide which criterion should be preferred in the formulation of electricity tariffs. In all cases, however, his choice will be constrained by the costs of implementing a tariff policy on the basis of the principles implied by any particular criterion.

A. THE EQUITY CRITERION

It is clear from section 4:4:2:2, that Escom's tariff policy is largely based on the equity criterion. Before the effect of Escom's tariff policy on efficiency and regional inequality can be evaluated, it is first necessary to evaluate this policy on its own terms according to the equity criterion. In other words, it is necessary to consider whether historical accounting costs have been "fairly" allocated to the different distribution undertakings and the consumer groups within each undertaking. In general, Escom's practice of identifying the costs attributable to each distribution undertaking and then allocating these costs to consumers according to whether the costs are related to the number of consumers, their maximum demand for capacity or their energy consumption appears to be fair according to this criterion. It is necessary, however, to examine in more detail the equity of the allocation of individual cost items.

The costs incurred by the Central Generating Undertaking are in the nature of joint costs and cannot be easily attributed to the individual distribution undertakings. It is therefore equitable that these costs are pooled between the different undertakings according to their electricity requirements. It is also acceptable according to the equity criterion, that the coastal undertakings are allocated the excess capital costs of their generating stations which were taken over by Escom, since such costs can definitely be attributed to these particular undertakings.

Escom has also adhered to the equity principle by charging the costs of transmission to the receiving undertakings, since the outage of a transmission circuit would only affect the power input of a specific distribution undertaking and the cost of the circuit is thus specifically attributable to it.

Escom's policy of linking loans to the commercial assets in operation at an undertaking, appears to be an effective means of attributing the finance charges on these loans to the undertaking concerned.

With respect to the fairness of the method of allocation of the Capital Development Fund (C.D.F.) contribution, the Board of Trade and Industries made the following criticism:

"The initial objective of the C.D.F. contributions was to provide for expansion (growth). One would have expected that the calculation of the contributions would be linked to a measure of growth. If not, the high growth consumer could make too small a contribution to the expansion needs of Escom compared to the low growth consumer."[48]

---

48. Ibid., p. 146.

A further criticism of the method of allocation of C.D.F. contributions is that, although they are essentially demand-related costs, they are treated in the same manner as energy related costs since they are allocated on the basis of kWh sales. This method of allocation is to the disadvantage of high load factor undertakings.

Another problem with Escom's method of cost allocation arises from the fact that the boundaries of the undertakings have developed over the years with the development of Escom's system, so that in some cases the difference between the costs allocated to adjacent undertakings may be the result of historical rather than economic factors. This may give rise to certain inequities in cost allocation. For example, Bloemfontein, which is not very much more favourably in terms of distance from the C.G.U. situated than Durban, has the benefit of the Rand and O.F.S. undertakings tariff. Furthermore, since Port Elizabeth municipality is served by the Orange River undertaking it pays a lower demand charge than customers served by the adjacent Border undertaking. It has become particularly important for Escom to reassess the boundaries of its undertakings now that regional development policy in South Africa is being pursued within the framework of the eight regions set out by the "Good Hope" plan.[49]

With respect to Escom's method of allocating costs to consumers, it may be inferred that it is acceptable that each undertaking apportions its total costs to consumer classes rather than individual consumers due to the very large number of consumers who are supplied. If the allocation of costs to consumer classes is to satisfy the equity criterion, then these classes must consist of consumers with basically similar consumption patterns, so that the deviation from the norm in the averaging process is minimized. It is however, submitted that Escom's Tariff A class is too large since it accounts for approximately 97 per cent

---

49. Department of Foreign Affairs and Information, <u>op.cit</u>.

of total sales, and that Escom's tariff structure should be revised to make provision for more consumer groups in respect of Tariff A. It can be argued, for instance, that a consumer who has a maximum demand of 800 Megawatt (MW) obtains it at the same tariff as the consumer with one of 0,1 MW, and that the saving due to economies of scale is therefore ignored, especially in respect of lower voltage distribution cost.

With respect to the demand charges levied on Tariff A consumers, it is further submitted that the existing differentiation of these charges according to voltage is equitable, since the additional cost of voltage transformation should be borne by the low voltage consumers on whose behalf this cost is incurred.

It is also equitable for Escom to transfer a portion of the demand-related costs to the costs recoverable by energy charges, in order to compensate small consumers for their contribution to diversity in the maximum demand for system power. The method of determination of this transfer is largely arbitrary and it is submitted that a more accurate recovery of these demand related costs by energy charges is recovered under marginal cost pricing. This issue will be examined in more detail in the next section.

Tariff classes B, C and D are not levied a specific demand charge, but demand-related costs are recovered from these consumers by means of block tariffs. For example, it can be seen from Table 4:18 that in the Border undertaking, Tariff B consumers are charged a basic levy of R12,00 per month to recover consumer-related costs and that the first 500 kWh of consumption is charged at 8,75 cents per kWh while the second block of consumption is charged at 5,45 cents per kWh. For any consumer using more than 500 kWh the effect of this tariff is the same as demand charge of R16,50 per month plus all consumption charged at 5,45 cents per kWh. It may be argued that this type of tariff is discriminatory in its effect on any consumer using less than 500 kWh. In other words, it will cause larger consumers to be subsidized by smaller ones who pay higher marginal and average prices. This argument, however, overlooks the fact that such a tariff is similar in effect to a fixed demand charge and accomplishes the similar purpose of inducing consumers to improve their load factors. Furthermore if, for example, a consumer charged on the basis of this tariff uses only 200 kWh he would pay a marginal price of 8,75 cents per kWh and would have a total monthly bill of R29,50 out of which only R6,60 of the allocated fixed demand costs of R16,50 would be recovered. There is consequently an under-recovery of allocated average fixed demand costs which may necessitate the recovery of disproportionately larger amounts from the larger consumer, to the extent that the decreased consumption of the smaller consumer does not also result in a corresponding decrease in capacity cost responsibility. It is therefore possible that under the block tariff system the larger consumer may be subsidizing the smaller, and not the reverse.

Moreover, if the relationship between demand and energy related charges established by the standard tariff is considered fair, then the effect of Escom's practice of adjusting the overall surcharge or discount applied by each distribution undertaking to cover the total costs actually incurred by the undertaking and to adjust the energy charge for increase in the coal price, would be to distort the equitable pattern of relationships embodied in the standard tariff structure. As the Board of Trade and Industries noted:

"The magnitudes of the transfer from demand-related to energy related charges vary considerably from what Escom evaluated as being fair during the last restructuring of tariffs."[50] There is also a distortion of the regional differences in the energy charge, when one compares the standard and the effective electricity tariffs. This is illustrated in Table 4:23 which is derived from Table 4:18.

The Board of Trade and Industries submitted that the fairness of the relationship between the different components of the electricity tariff

---

50. *Ibid.*, p. 27.

could be better maintained, if Escom followed the practice adopted in countries such as England, France and Germany of requiring a complete annual revision of the tariff structure to effect any necessary adjustment of tariff revenue. The unwillingness by Escom to pursue such a policy, is perhaps an indication that it does not consider there to be significant benefits to be obtained from the extension of the equity principle. This view is reflected in the following arguments which were raised by Escom against a complete annual revision of the tariff structure:

"(i) The imbalance between the components of a tariff does not affect the overall total revenue as defined in section 14 of the Electricity Act.

(ii) The subdivision of total costs into demand-related, energy-related and consumer-related categories involves many uncertain areas and cannot be in principle totally 'correct'.

(iii) A large majority of consumers having load factors near the average of all Tariff A consumers in an undertaking are not significantly affected by the ratio between energy related and demand related charges..."[51]

TABLE 4:23

ENERGY CHARGES EXPRESSED AS PERCENTAGE OF RAND AND O.F.S. CHARGE (FIRST QUARTER 1983)

Undertaking	Standard	Effective
Border	151	109
Orange River	146	113
Cape Northern	123	91
Cape Western	126	94
Eastern Transvaal	90	92
Natal	142	78
Rand and O.F.S.	100	100

51. Ibid., p. 27.

Despite its various drawbacks, it can be concluded that if Escom's tariff policy is evaluated on its own terms according to the equity criterion, then one must concur with the conclusion reached by the Board of Trade and Industries that "... generally, Escom's methods of allocating costs ... are acceptable." However, it is clear from the discussion of Escom's tariff policy that it departs to some extent from marginal cost pricing principles. It is therefore also necessary to consider the effects on economic efficiency of the failure by Escom to explicitly incorporate aspects of marginal cost pricing in its tariff policy.

B. THE EFFICIENCY CRITERION

Although Escom's policy of structuring its tariffs according to voltage level and region appears to be consistent with the principle that it is economically efficient for consumers to be charged the costs of supplying their particular region at their particular voltage, there are a number of respects in which Escom fails to apply the efficiency criterion in structuring its tariffs in the manner indicated in the discussion of a marginal cost based tariff structure in section 4:3. This matter is now discussed under the three headings that follow:

(i) THE USE OF ACCOUNTING RATHER THAN ECONOMIC COSTS

In setting its tariffs, Escom follows the traditional accounting approach which is concerned with the recovery of sunk costs rather than the resources used or saved by consumer decisions. Since prices are the amounts paid for unit increments of consumption, they generally reflect the extra cost incurred as a result of these increments of consumption. Such increments to cost will occur if either existing consumers increase their demand or if new consumers are connected to the system. It follows that if prices are to act as signals inducing consumers to make decisions which ultimately lead to an efficient allocation of resources, then present prices should also be related to the economic value of the resources to be used in the future to meet such consumption changes. The accounting approach may therefore be inefficient since it measures costs in terms of the historical value of assets and thus implies that future economic resources will be as cheap or as expensive as in the past. Furthermore, the deviation of accounting costs from economic resource costs is likely to be exacerbated by the fact that the accounting approach will make no adjustment for distortions in input prices caused by taxes, subsidies, import duties or monopolistic practices, in the manner suggested in the discussion of shadow pricing in section 4:3:1

(ii) THE FAILURE TO APPLY A PEAK/OFF-PEAK DIFFERENTIATION OF THE ENERGY CHARGE

In the discussion of a marginal cost based tariff structure it was established that the energy charge by a distribution undertaking should be based on the running cost of the generating plant to be used last in merit order in a given period, adjusted for the marginal cost of transmission to the undertaking. The application of this principle necessarily leads to some peak/off-peak differentiation of the energy charges since the generating units used last in a peak period will clearly have a higher operating cost then those used last in an off-peak period. By basing its energy charge on the average energy costs, Escom is probably setting this energy component of its tariff below the opportunity cost of meeting an increment in consumption in the peak period and above that for the off-peak period. The effect of this is illustrated in Figure 4:10. The SRMC of energy used in generation is depicted as a stepped function to indicate that five generating units are used in merit order. If the energy component of the tariff is set at PA which is equal to the average energy cost in both peak and off-peak periods, it can be seen that in the off-peak period qAo will be consumed and that the value of the last unit consumed will be greater than its opportunity cost. There is therefore a net economic loss equivalent to the area of the triangle abc which results from the failure to set the off-peak energy charge equal to the short run marginal cost of energy. The diagram also indicates that

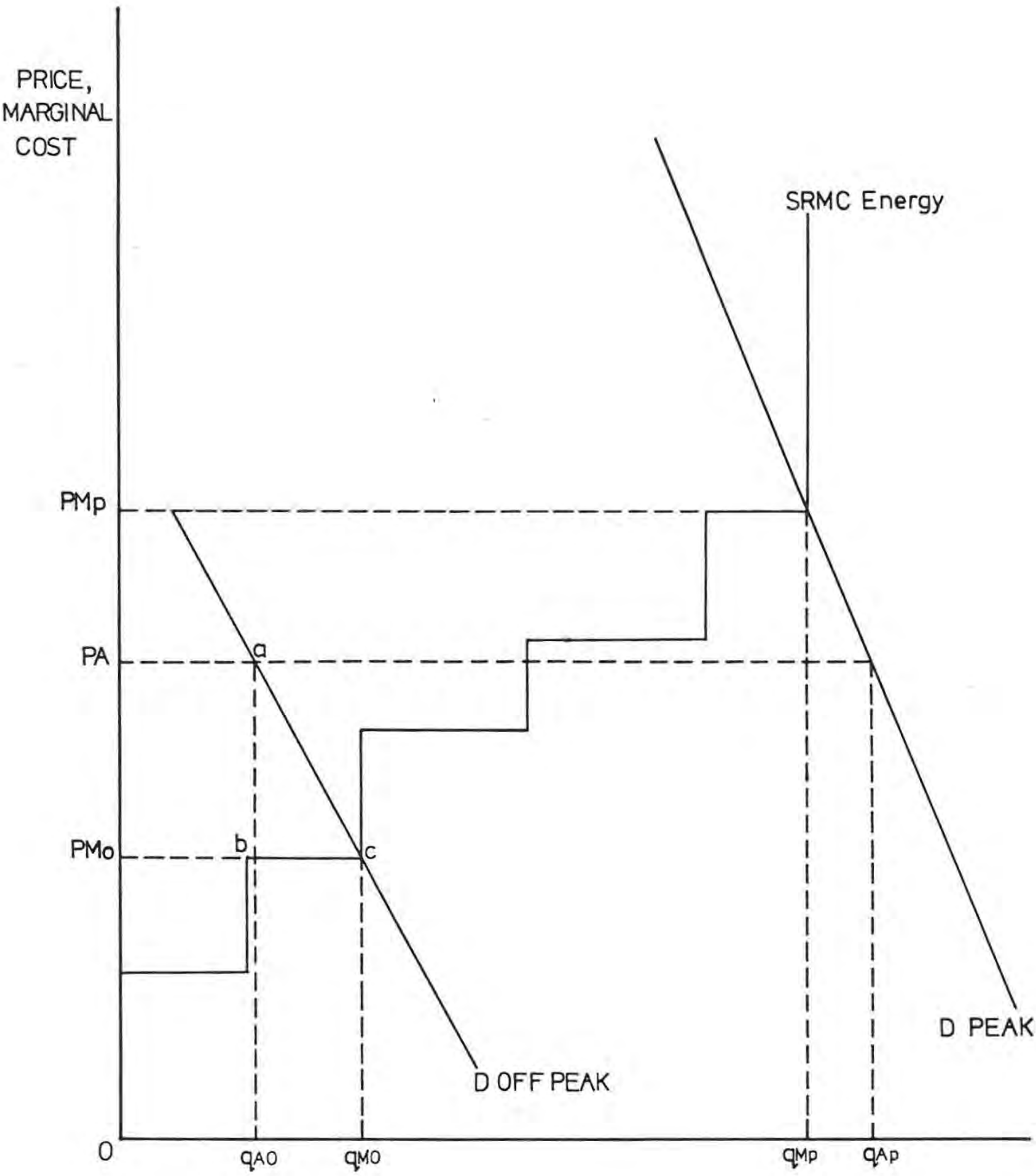


FIGURE 4:10 The Effect of Escom Failing to Apply to Peak Period Price Differentiation

the failure to differentiate charges between periods leads to underutilization of base load capacity. Figure 4.10 shows that in the peak period, if the energy charge was set at Pmp, the generating plant would be fully utilized. However a policy of setting the energy charge equal to unit energy costs, PA, would clearly result in an excess demand and place pressure on the electricity utility to expand its capacity by advancing the installation of new plant in terms of its investment programme. The economic loss associated with the failure to set a peak period energy charge can therefore be established by calculating the difference between the present value of the investment programme under an average and a peak-load pricing policy, and setting off against this amount the reduction in operating costs resulting from bringing forward the installation of more efficient generating units and the gain in consumers' surplus resulting from charging a lower price in peak periods under the average cost pricing policy. It is clear, therefore, that Escom's failure to differentiate its energy charge results in losses in economic efficiency in both peak and off-peak periods.

A further loss in economic efficiency may be caused by the fact that Escom treats transmission costs allocated to receiving undertakings as a demand-related cost and does not adjust the energy charge for the power losses which occur along the transmission lines. The energy charges of undertakings located further from the central generating undertaking may therefore be understated since they are not adjusted for these transmission losses.

(iii) THE EFFECT OF THE TRANSFER FROM DEMAND TO ENERGY-RELATED COSTS ON THE INCENTIVE FOR CONSUMERS TO IMPROVE THEIR LOAD FACTORS

Both Escom and the Electricitè de France provide for a transfer from demand to energy related costs. The transfer provided for by Escom, however, is likely to result in a lesser inducement for consumers to improve their load factors, since it is not limited to peak and full-use periods and therefore does not give consumers the opportunity to shift their consumption to off-peak periods. Furthermore there is no differentiation in the proportion of the transfer between consumers with high utilization factors and those with lower utilization factors. As it was explained in section 4:3, this differentiation is necessary since high-utilization consumers have little opportunity to alter the daily distribution of their load as a result of having to pay higher energy charges in peak periods, whereas low-utilization consumers often have much greater scope for shifting their loads to off-peak periods. It appears, therefore, that Escom could improve its overall system load factor by adopting the French practice of transferring demand-related costs to the energy-related portion of the tariff in peak periods according to the utilization factors of the consumer groups served by the distribution undertakings. The determination of the amount of the transfer by Escom is currently rather arbitrary in nature and it would seem that a more efficient utilization of its facilities could be obtained if this transfer was used as an instrument to encourage consumers to improve their load factors.

In summary therefore, it seems that Escom could promote greater economic efficiency by changing its tariff structure along the lines followed by the Electricité de France. It would be necessary, however, to compare this gain in economic efficiency with the higher consumer-related costs which would be associated with introducing this more complicated tariff structure.

C. THE EQUALITY CRITERION

The application by Escom of the equity criterion according to which each distribution undertaking sets its tariffs to cover the costs specifically attributable to it, results in significant regional differences in the average selling price of electricity, as shown in Table 4:24. It can be seen from this table that the average selling

TABLE 4:24

REGIONAL DIFFERENCES IN AVERAGE SELLING PRICE OF ELECTRICITY BY CONSUMER CATEGORY, 1982

Undertaking	Bulk	Domestic street lighting	Industrial	Mining	Traction	Total
		Cents Per kWh				
Border	4,013	7,151	6,256	-	-	4,323
Cape Northern	3,446	5,887	5,059	3,645	4,562	3,860
Cape Western	2,804	5,846	4,308	-	4,505	3,614
Eastern Transvaal	2,754	4,840	2,709	2,731	3,683	2,763
Natal	2,591	4,743	2,647	3,274	3,524	2,724
Orange River	3,310	9,248	7,356	-	-	3,477
Rand and O.F.S.	2,798	4,564	2,685	2,430	3,650	2,804
Total	2,808	5,418	2,825	2,541	3,650	2,804

price of electricity in the Border undertaking is, for example, 1,6 times as much as the price charged in the Rand and O.F.S. undertaking. These regional differences are more accentuated when they are examined on an intersectoral basis. Thus, the average selling price of electricity used for industrial pruposes in the Border undertaking is 2,3 times as great as for that used in the Rand and O.F.S. undertaking.

There exist various reasons for these regional differences in electricity prices, among which are the following:

(i) EXCESS CAPACITY COSTS

As was explained in the discussion of Escom's tariff policy, the finance charges in respect of the excess generating capacity of the Natal, Cape Western and Border undertakings are added to the costs which must be recovered by means of tariffs calculated for these undertakings, and this clearly inflates the average selling price of these coastal undertakings in relation to the other undertakings which are only supplied by the Central Generating Unit (C.G.U.).

(ii) TRANSMISSION COSTS

Escom's practice of allocating the expenditure incurred in the transmission of electricity to the undertakings concerned will obviously place a relatively greater cost burden on undertakings located a greater distance from the C.G.U. Table 4:25 compares the regional differences in transmission cost per kWh.

According to Table 4:25 the undertakings with the highest transmission cost per kWh in 1982 were Border, Cape Western, Cape Northern and Orange River. It should be noted that these figures are influenced by the load factors of the under takings.

TABLE 4:25

TRANSMISSION COST PER KWH, 1982

Undertaking	Cents per kWh
Border	0,11
Cape Western	0,25
Cape Northern	0,28
Orange River	0,13
Natal	0,06
Eastern Transvaal	0,001
Rand and O.F.S.	0,001

(iii) AGE OF UNDERTAKINGS

Escom's practice of allocating loan charges in relation to the assets in commercial operation, burdens the newer undertakings where the historical cost of assets in service is higher in relation to sales than in the older undertakings.

(iv) OPERATING EXPENSES

There are significant differences in the nature and characteristics of the demand for electricity imposed by consumers in the different distribution undertakings. Distribution, reticulation, and administration costs are understandably higher in sparsely populated regions and when a large part of the total supply is to a large number of small consumers. This is illustrated in Table 4:26 where it can be seen that in 1982 the combined operations, maintenance and administration expenditure per kWh was significantly higher in the Cape Western, Border, Cape Northern and Orange River undertakings than in the Natal, Eastern Transvaal and Rand and O.F.S. undertakings.

TABLE 4:26

OPERATING EXPENDITURE IN CENTS PER KWH IN 1982

Border	0,81
Cape Northern	0,49
Cape Western	0,56
Orange River	0,56
Eastern Transvaal	0,23
Natal	0,23
Rand and O.F.S.	0,15

(v) MIX OF CONSUMERS

The costs of an undertaking are affected by whether sales are predominantly to a few large consumers or to a large number of smaller consumers. The costs per unit consumption in the Border and Cape Western undertakings are detrimentally affected by the fact that these undertakings have a far lower consumption per consumer than in the Orange River, Rand and O.F.S. and Eastern Transvaal undertakings. The figures for the Cape Northern and Natal undertakings are near the average for Escom as a whole.

(vi) LOAD FACTOR

If the largest part of the total sales of a regional undertaking is to large consumers operating at a high load factor, the average selling price will necessarily be lower than in an undertaking supplying a large number of small consumers with low load factors. The average consumer's load factor has an important bearing on the undertaking system load factor. The 1982 system load factors of the different undertakings are shown in Table 4:27.

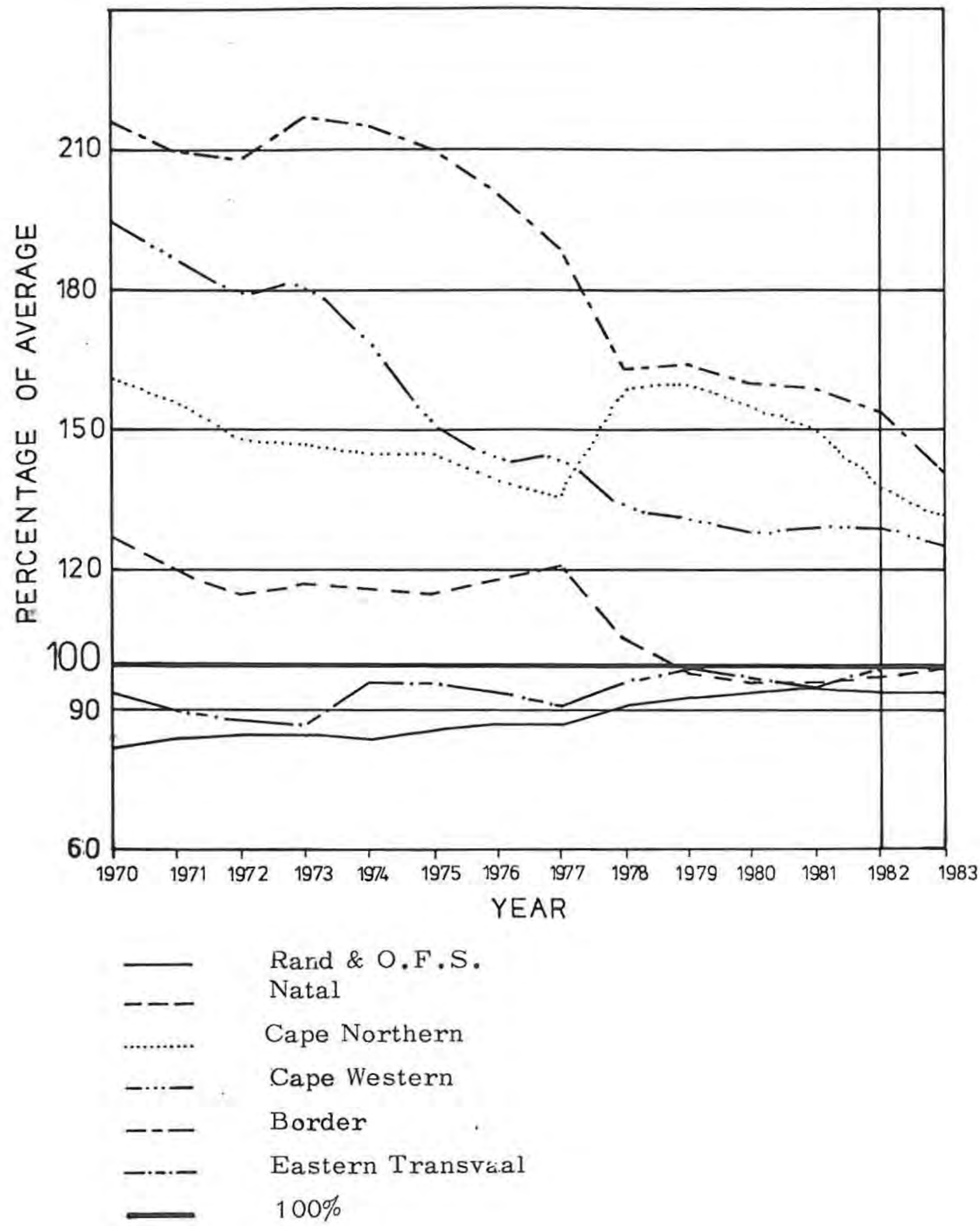


FIGURE 4:11 Escom Regional Undertaking Price Ratios Average Undertaking Price in Cents per kWh as a Percentage of the Average Escom Price in Cents per kWh

TABLE 4:27

UNDERTAKING'S SYSTEM LOAD FACTORS IN 1982

Undertaking	Annual System Load Factor
Border	58,8
Cape Northern	69,4
Cape Western	71,9
Orange River	71,9
Natal	73,1
Eastern Transvaal	71,8
Rand and O.F.S.	71,2

The relatively high system load factors in Rand and O.F.S., Eastern Transvaal and Natal undertakings can be ascribed to the large, 24 hour, steady loads taken by large mining consumers and industrial consumers. The system load factors of Cape Western and Orange River undertakings are favourably influenced by the relatively large and sustained partial bulk supplies taken by the municipalities of Cape Town and Port Elizabeth, to augment the outputs from the municipal power stations in these cities. In contrast, there are no mining loads and few large, 24 hour industrial loads in the Border undertaking and the system load factor of this undertaking is significantly lower.

Despite the fact that there are currently significant regional differences in the average selling price of electricity, there has been a sustained trend over the past decade in narrowing the difference between the price levels of the separate undertakings. This is illustrated in Figure 4.11. A major factor underlying this narrowing of regional price differentials has been the development of the national transmission system which now interconnects all major power resources in South Africa. With the interconnection of all Escom's power stations, the placing of new power stations is no longer primarily determined by regional power requirements but rather by the availability and cost of the required natural resources. With the pooling of generation costs between the distribution undertakings in an equal manner, proportional to their respective electricity requirements, regional differences in tariffs are no longer caused by regional differences in electricity cost. Furthermore the amount of excess capacity costs allocated to the coastal undertakings has shown a declining trend, not only due to the 1977 amendment to the licences of these undertakings which limited the apportionment of these costs to interest and redemption charges, but also because the loans tied to the generating plant of these undertakings are not being renewed so that the amount of these excess capital costs is continually diminishing over time.

The Board of Trade and Industries investigated the effects of a further equalization of cost allocation between the undertakings. It was found that if the interest charges to the undertakings were based on an average interest, there would be cost per unit decreases in the Border (- 1,3 per cent), Cape Western (- 0,1 per cent) and Orange River (- 2,8 per cent) undertakings, and increases in the Cape Eastern (+ 1,1 per cent), Cape Northern (+ 0,2 per cent), Eastern Transvaal (+ 0,4 per cent) and Natal undertakings, while there would be no change in the Rand and O.F.S. undertaking. Furthermore, it was reported that the full pooling of C.G.U. costs would result in an increase in the cost per unit in the Cape Eastern (+ 3,2 per cent), Eastern Transvaal (+ 1,3 per cent), Orange River (+ 1,5 per cent) and Rand and O.F.S. undertakings while the cost per unit in the Border (- 7,5 per cent), Natal (- 3,6 per cent) and Cape Western (- 3,1 per cent) undertaking would decrease. The marginal effect of pooling transmission costs were, however, found to be much smaller with the Eastern Transvaal and Rand and O.F.S. undertakings experiencing increases in the cost of electricity per unit, while all the other undertakings would be liable to lower unit costs. It was thus found that there were considerable opportunities to achieve greater equalization of tariffs by altering the method of allocation for certain tariffs.

It is now necessary to consider the economic effects of such measures to promote the regional equalization of tariffs. It is clear that in those understanding where there is a relative decrease in tariffs there will be an upward shift in the forecast for future electricity demand and an increase in the present value of the investment programme planned to meet this demand, while in those undertakings where there is a relative increase in tariffs there will be a downward shift in the forecast for future electricity demand and a decrease in the present value of the investment programme planned to meet this demand. The revised investment programme is likely to depart further from optimality than the existing programme, since there is likely to be excessive investment to meet the shift in demand towards those regions which are characterized by relatively higher costs of transmission, distribution and reticulation. If the costs and benefits of such a revision in the investment programme are to be correctly evaluated, then one must offset the present value of the future gain in consumer's surplus against the increased investment cost in those undertakings where tariffs are reduced, and the loss in consumers' surplus against the reduction in investment cost in those undertakings where tariffs are raised. Furthermore, these costs and benefits should be weighted by an area need indicator, in the manner suggested by MCGuire and Garn.[52]

It was not possible to quantify the costs and benefits mentioned above, since Escom has no information on the sensitivity of regional load forecasts and the capital development programme to the equalization of tariffs. In July 1983 I interviewed officials in the Escom Statistics Department. They expressed the view that the equalization of tariffs would not have much effect on Escom's load forecasts for the different undertakings, due to the low elasticity of electricity demand. They also indicated that the price of electricity was not explicitly taken into account in the load forecasts which are used by Escom in planning its future capital expenditure. They did not perceive there to be a conflict between efficiency and equality, but saw the main trade-off as being between equity and equality. The equity criterion has been embodied in the provisions of the Electricity Act and Escom has been required to accept it as the basis for formulating its tariffs, It can be argued that the adoption of the equality criterion in its place would create a conflict between the objectives of Escom management similar to that which has been

---

52. This was discussed in Chapter Three, Section 3:5.

suggested to affect S.A.T.S., since on the one hand Escom would be required to operate according to commercial principles, while on the other hand it would be required to sacrifice these commerical principles to the interests of regional policy. If regional differences in the price of electricity are considered to be a significant factor discouraging industrial decentralization, then it would seem preferable for the State to directly compensate the electricity users affected, rather than to require Escom to ignore regional differences in the average cost of electricity per kWh and equalize its tariffs.

In fact, this approach appears to have been accepted by the South African Government in its latest industrial decentralization policy to assist the Board for the Decentralization of Industries in its determination of subsidies. Escom calculates the electricity tariffs of its consumers at industrial decentralization points according to the tariff schedule applied by the Eastern Transvaal undertaking, and then subtracts this from its current electricity charge to determine the amount of the subsidy which should be paid.

4.4 CONCLUSION AND RECOMMENDATIONS

In conclusion, the following recommendations may be made on the basis of the preceding analysis with a view to promoting the objective of economic efficiency:

(a) Provision should be made in the Electricity Act for the earning of profits by Escom. This would obviate the need to treat transfers to reserves as costs chargeable against revenue, when they actually constitute an internal retention of profits. The recommendations by the Board of Trade and Industries that "... the Redemption Fund, the Provision for Redemption of Foreign Loans and the Capital Reserve should be abolished and the amounts concerned transferred to Accumulated Depreciation" and that "the system of Fund Accounting should be discontinued and outstanding Escom loans should be reduced by the amount of the transfers referred to"[53], should therefore be implemented.

---

53. Board of Trade and Industries, op.cit., p. 157.

(b) Escom's profitability, and therefore the general level of tariffs, should be subject to control by an independent body, such as the Department of Industries, which would set profit targets on the basis of its evaluation of the internal financing requirements of Escom.

(c) The establishment of a Capital Project Evaluation Group along the lines proposed by the Board of Trade and Industries would provide a measure of control over the tendency towards excessive investment and growth of capital-related costs, which can arise where Escom management has full responsibility for investment decisions.

(d) If Escom decides to adjust its tariffs, then it should only be able to do this through a complete revision of its tariff structure, approved by the Electricity Control Board, and should no longer have the option of being to change its tariffs without Board approval by merely adjusting its surcharges and discounts.

(e) The basic principle that tariffs should reflect costs so that there is no cross-subsidization between any undertaking or consumer-group should be retained, but Escom should investigate methods of relating costs more closely to the economic value of resources used, such as "shadow pricing" and replacement cost accounting.

(f) Escom should investigate the possibility of structuring the tariffs charged to large power users according to marginal cost principles, in the way which has been implemented by Electricite´de France.

(g) To the extent that the regional differential in the price of electricity conflicts with industrial decentralization policy, this problem should continue to be dealt with according to the current policy of State compensation to consumers at industrial development points, rather than by requiring Escom to adjust its methods of cost allocation in a manner which leads to a greater regional equalization of tariffs. Certain inequities in the regional price differential could, however, be removed if a more functional definition of the boundaries of the electicity undertakings was applied.

CHAPTER FIVE

WATER

5:1 ECONOMIC FEATURES OF WATER SUPPLY

5:1:1 WATER IN MODERN SOCIETY

An adequate supply of water is an essential requirement for most of the activities carried on in a modern economy. Practically every modern household has a basic requirement of running water although this is often exceeded as water is consumed for less essential uses such as lawnsprinkling. In the industrial sector water is an important input in many production processes such as electricity generation while in the agricultural sector water is needed for stockwatering and crop irrigation. There has been a rapid growth in water consumption in the agricultural sector, in particular, as a result of the introduction of cultivation practices that are aimed at achieving better crops by retaining as much as possible of the rain that falls on the ground and the sharp increase in the use of supplementary irrigation to raise crop yields even in areas where normally a reasonable harvest can be expected without irrigation.

If a modern economy is to sustain its development and economic growth, water has to be provided in rising quantities for all these purposes. Furthermore water has a profound effect on the overall quality of life in a society. It has provided the setting, in many cases for gratification of aesthetic and recreational needs and in a modern society this must be recognized as a legitimate water use.

5:1:2 WATER RESOURCES

Many countries frequently face the problem that the pattern of demand for water does not coincide with the pattern of availability of water resources. For example, the seasonal distribution of precipitation may not coincide with seasonal water requirements. While little precipitation may occur in the summer months when there is the heaviest demand for water because crops are growing, a heavy runoff in winter and early spring may produce floods. Furthermore, irregularity in precipitation may not only be seasonal but may also be cyclical with a country or region experiencing periodic droughts. Riverflow may thus be highly erratic in contrast to the steadily increasing demand for water. Finally, the geographic distribution of water resources may not match the spatial distribution of activities. In many countries there are some regions with a more than adequate water supply whereas in other regions there is a prospect that economic advancement will be affected by water shortage.

When a country is plagued with this problem of a maldistribution of water resources - seasonally, annually and geographically - it is necessary to engage in water development activities which are directed at adapting and utilizing water resources to meet the water requirements of the economy.

5:1:3 WATER RESOURCE DEVELOPMENT

From a historical point of view the process of water resource development usually goes through three distinct phases in a region or country.

(a) PRIVATE, SMALL-SCALE DEVELOPMENT

In the early period of settlement, water resource development usually takes the form of stream diversions to irrigate fertile land which is within reach of the main rivers. These diversion structures were usually constructed by the settlers individually, with an expenditure of time and effort, but without the need to raise funds from financial markets.

(b) LOCAL DEVELOPMENT PROGRAMMES

As the process of regional settlement continues, communities start to grow up with diverse water needs. It now becomes necessary to supplant the crude diversion structures with central water supplies, developed by impounding surplus flows in retention basins for release during the dry season. Furthermore the distribution of water may be extended to areas having no access to surface sources of supply, by building a system of large reservoirs and canals to convey increased supplies of water to these areas. The most suitable institution for promoting local water resource development usually turns to out to be the public district authority, endowed with quasi-governmental powers to levy assessments on residents for recovery of financial expenditure.

As water requirements expand, local water resource development programmes may start to have substantial external effects on water supplies in other districts. For example, the growing usage of water in one district may reduce downstream flows and thereby increase the concentration of pollution and salinity in these water supplies. Also pumping from subsurface aquifers may reduce water tables in some areas at an alarming rate. As the process of water resource development reaches this stage, it becomes necessary to adopt an integrated approach transcending local levels of jurisdiction.

(c) INTEGRATED WATER RESOURCE DEVELOPMENT BY A PUBLIC AGENCY

An integrated approach to water resource development would have to be planned at either the national or regional as opposed to the district level. An integrated development programme might involve the following features:

(i) long distance transfers of water from surplus to deficit area;

(ii) the capture of flood flows during wet seasons to replenish subsurface storage depleted during dry seasons;

(iii) co-ordinated releases from surface storage to maintain flows in established stream beds for abating pollution and repelling the intrusion of salinity into river mouths; and

(iv) the multiple purpose exploitation of storage reservoirs. For example, a large dam may not only be a storage reservoir but can also be used for hydro-electric power, and to provide recreational facilities;

An integrated plan such as this would differ in a number of significant respects from locally sponsored development programmes:

(i) from a technical point of view, it represents a higher order of interdependent system. Whereas local supply and distribution systems are limited to explciting water resources in a particular district, an integrated system which pools a network of streams and combines ground and surface water management, would add a great deal to the technical potential of water resource development and utilization; and

(ii) large scale economies may be realized by the integration of water resource development into a single system since storage facilities required to accomplish any one of a number of separate purposes can be used to achieve equally well other common purposes.

A public agency endowed with powers of planning, co-ordinating and controlling water projects at the national level and with access to funds from general tax revenue or with the power to levy assessments upon its beneficiaries would seem to be the most suitable institutional instrument for promoting this integrated programme of water resource development. Such an institution would be able to "internalize" the externalities which would otherwise plague local water resource development.

A highly successful example of such an integrated system is the national water grid which was developed in Israel and which interconnects the country's major water resources and regulative storage facilities and caters to practically all the major water needs of the country. This system is characterized by both spatial and intertemporal interdependence of water projects. There is thus a hierarchical layout pattern of projects where each project generation integrated the projects of the preceding generation and was, in turn, recombined into larger project units by the subsequent project generations. Furthermore, improved utilization of Israel's water resources has not only been achieved by connecting local resources to the national grid but also by a nationally conceived pattern of water management. In humid years, surface water is used to the maximum possible extent while the exploitation of groundwater resources is low. Reduction of groundwater use and concurrently conducted recharge operations will, in such years, increase the underground storage and cause a rise in groundwater levels. In dry periods on the other hand, when surface flows are scarce, the groundwater accumulated in previous years is drawn upon.

Wiener has made the following comment on Israel's progress in the field of water resource development:

> "Between 1948 and 1966, the utilization of the country's proven water resources rose from 17 per cent (representing mainly the use of shallow local groundwater formations) to almost 90 per cent (including the utilization of Israel's only two rivers, the Yarquan and the Jordan) ...... The high rate of agricultural productivity and the extremely high overall plant factor of water project capacity that were achieved in spite of the lack of experience of the new farmers were due mainly to the close co-ordination between water resource development and the provision of complementary inputs - material human and organizational"[1]

---

1. A. Wiener; The Role of Water in Development, McGraw-Hill, New York, 1972, p. 405.

In subsequent years, however Israel has increasingly found itself in the position of having developed and used most of its natural water resources so that it has had to depend on new methods of water-resources management for future economic growth. It thus shares the problem, with many arid countries that although water supply limitations are not absolute, they have become economic. Wiener describes this problem of aridity as follows:

> ".... limitations in national water resources will .... no longer constitute an absolute (Malthusian) barrier to growth; the development of man-made fresh water will have become technologically feasible, albeit at still relatively high costs. Supplementing natural water by man-made supplies will, however, steeply increase the cost of marginal quantities of water and thereby, to a marked extent, the average water costs; the limitations imposed by the lack of furthermore accessible natural resources will thus be of an economic (or Ricardian) nature"[2].

As will be discussed, in the next section, the economic growth trends predicted for South Africa, indicates that it may have to face a growing problem of aridity, in the economic sense defined above, towards the close of the twentieth century.

5:1:4 THE EFFECT OF ARIDITY ON PUBLIC WATER SUPPLY POLICY

Conventional problem solving in the field of water supply policy usually assumes the demand for water to be an exogenous variable. According to this approach, the aim of water resource development and management is to harness water resources with a view to complying with the requirements of such a demand. In situations of aridity, this problem solving path will be blocked since available

---

2. Ibid., p. 408.

resources will not be sufficient to comply with the requirements of an exogenous demand. Conditions of aridity therefore result in greater emphasis being placed on the following elements of a water resource development policy:

(a) the available natural water resources are no longer treated as a constraint upon meeting the growing water requirements of the economy, since it is technologically possible to incorporate man-made water into a water supply system through desalination of sea-water, although this process is at present costly. Furthermore, the natural hydrological cycle may be modified either by increasing the conservation of runoff once precipitation has occurred, or by attempting to increase precipitation by means of cloud-seeding;

(b) an emphasis may be placed on maximizing water conservation subject to the constraint that marginal costs of the water conserved does not exceed pre-determined cost ceilings Industrial and agricultural processes may thus have to be modified to economize on the use of scarce water through, for example, recirculation and reclamation techniques. Pollution control also becomes important; and

(c) in most countries water, and particularly water supplied for agricultural purposes, is seriously underpriced. Since the marginal cost of water to consumers is generally below its marginal opportunity cost to society, it tends to be wasted by being employed in a number of uses where it has a low marginal value relative tc its marginal social cost. It is clear, therefore, that greater conservation of water resources can be attained by a realistic pricing policy which equates the price paid by consumers to the marginal social cost of supplying water. Before a marginal cost based pricing policy for water can be examined however, it is necessary to consider a number of the special characteristics of water supply which would complicate the application of such a rule.

5:1:5 SPECIAL CHARACTERISTICS OF WATER SUPPLY

In a number of ways the production and cost conditions facing a water utility differ from those assumed in the perfectly competitive model. Figure 5:1 illustrates the following characteristics of the cost conditions which typically face a water utility:

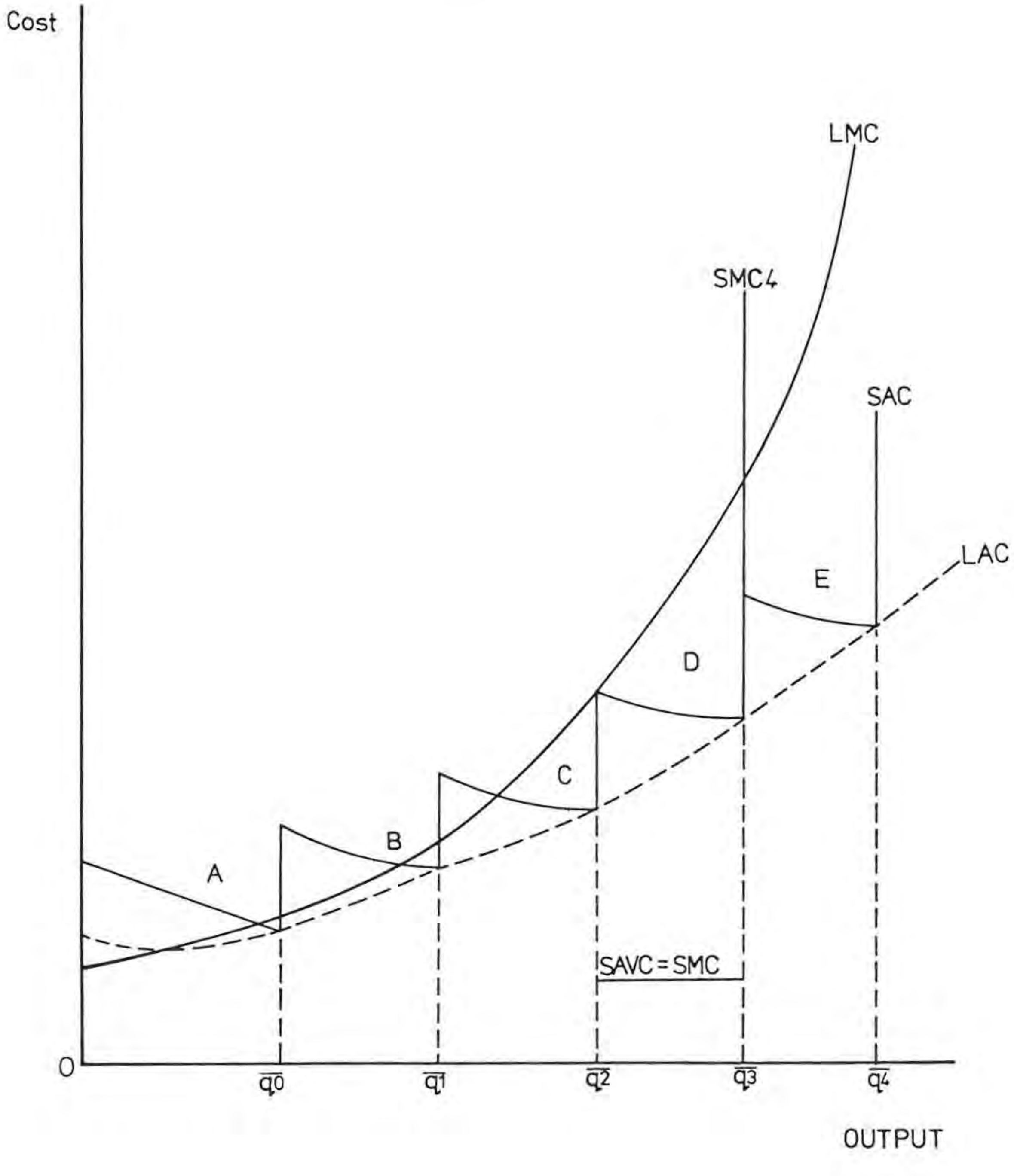
(i) due to plant indivisibilities the water utility can only choose between a finite number of plant sizes. In the figure five scales of plant, A, B, C, D and E are represented;

(ii) there is likely to be a long run rising trend in unit costs due to the fact that, in meeting local water requirements, a water utility will first make use of the most readily available water sources and then go on to develop the more distant and expensive supplementary supplies;

(iii) once plant of a particular scale has been installed, the short run marginal cost (SMC) of an output increment is likely to be substantially below unit cost if plant capacity is not fully utilized; as a result the short run average cost (SAC) curve associated with each plant size declines until the capacity output level of the plant is reached; and the SMC curve will be horizontal at those levels of output over which capacity is underutilized and vertical at full capacity if it is assumed that unit variable costs are constant and that there are technical limitations to plant capacity. It is often the case, though that unit variable costs tend to increase as designed "capacity" is approached and that it may be possible to operate beyond this "capacity" at high but not infinite cost.

These special features of the cost functions confronting water utilities have a significant effect on the application by water utilities of the marginal cost based pricing and investment rules discussed in Chapter Two as will be shown in a later section.

It is also often the case that there are a number of externalities which



5:1 Cost curves typically facing a water utility are associated with the supply of water. These may be of the nature of technological external economies resulting from physical interdependence in resource use such as would occur when irrigation leads to the recharging of subsurface aquifers and thereby reduces the costs of pumping groundwater. They may also arise because benefits such as flood prevention, salinity control, pollution abatement, and recreation arise from water schemes which are not directly paid for by water consumers. Krutilla and Eckstein have described the economic effects of these externalities as follows:

> "At any stage in a river's seaward movement, its conversion into water derivatives may entail gains (external economies) for complementary uses, or inflict losses (external diseconomies) for competitive uses, outside the intermediary of the market. Moreover, the gains or losses for complementary or competitive uses downstream are not independent of the product mix at upstream plants during low-flow seasons or adverse hydrologic periods, when common storage must be used outside the range of complementary production. Accordingly, rather extreme, direct interdependence prevails among the complementary facilities representing a multiple purpose river basin development."[3]

The effect of this interdependence is to result in a divergence between the private and social valuation of the factors employed in water resource development. It would therefore be suboptimal to strictly apply a marginal cost pricing rule which failed to take account of externalities. Such externalities would have to be incorporated into a pricing rule which leads to the closest possible approximation between marginal <u>social</u> benefit and marginal <u>social</u> cost.

---

3. Krutilla, J.V. and Eckstein, O., <u>Multiple purpose river basin development</u>, John Hopkins Press, 1958, p. 70.

5:1:6 MARKET IMPERFECTIONS ACTIVITIES RELATED TO IRRIGATION

An irrigation project will obviously result in a substantial increase in the value of the irrigated land. The enterprise, whether public or private, which is responsible for the employment of water on reclamation farms may fail to appropriate the returns associated with its investment, for the following reasons:

(a) the agricultural processing industries who purchase the output of the irrigation farms are often in a monopsonistic market position because of:

(i) the interrelation of resource uses combined with indivisibility in production which contribute to internal economies of large scale so that the supply area is often not big enough to accommodate a large number of these enterprises;

(ii) the conversion of raw materials into processed output is often accompanied by substantial weight losses so that the supply area is limited to a short radius around the processing plant and can therefore only accomodate one processor.

Consequently, the increase in physical yields per acre with irrigation, may become a processor's surplus rather than a higher financial return to the water enterprise providing the advantage.

(b) When an area is opened up to irrigation, public utilities such as the railways which serve the area will benefit from a higher volume of freight. Freight rates, however, are unlikely to be reduced since they are generally set for areas wider than that under irrigation, so that these "pecuniary external economies" will accrue to the railways and not the water enterprise.

(c) Irrigation farmers are often in a superior bargaining position <u>vis-à-vis</u> the irrigation enterprise due to their political power. During agricultural depression periods, they may therefore rely on the political process to scale down water charges while during prosperous periods their improved incomes are often capitalized in the value of the land and associated investments when ownerships are transfered. This accounts for the fact that irrigation water is underpriced in most countries of the world.

5:2 WATER PROVISION IN SOUTH AFRICA

5:2:1 HISTORICAL BACKGROUND

The historical evolution of water resource development in South Africa tended to follow the typical path described in the previous section with the scale and regional impact of water projects growing over time. There has also been a trend towards the increasing involvement of the State in water resource development which has been accommodated by a number of changes in the underlying legal framework:

(i) before Union, the State played a minor role in water resource development and the riparian principle of Roman Dutch Law was applied in terms of which the rights to the use of water flowing in a public stream are vested primarily in the owners of the land adjoining the banks of the river. This system usually operates satisfactorily in an agrarian economy but becomes increasingly untenable once secondary and tertiary uses of water become significant;

(ii) at Union, the machinery was created which enabled the State to extend its involvement in water resource development by the merging of the irrigation activities of the different provinces in the Department of Irrigation in terms of the Irrigation and Conservation of Waters Act No 8 of 1912. Although this legislation was largely concerned with the development of water resources for agricultural purposes, it did take account of the external effects of water use in a particular area by making it an offence to render water unfit for irrigation or industrial purposes but the relevant penalty clauses were so vague that prosecution was virtually impossible;

(iii) with the South African economy growing rapidly after the Second World War, the competition for water supplies among industrial municipal and agricultural interests developed to such an extent that periodic ammendments to the Irrigation Act proved to be quite inadequate to cope with the changing circumstances while the application of the riparian principle began to exercise increasingly detrimental effects. Accordingly the Water Act No 54 of 1956 was introduced to modify the riparian principle by granting the State power to intervene at any time to exercise control over public water although it still stipulated that, as a general principle, the rights to use water are vested in the riparian owners. This meant that the State could assume responsibility for planning, developing and apportioning water in the integrated manner required during the advanced stage of water resource development.

During the 1960's there were fears that with the rapid rate of industrial development and economic growth in South Africa, the demand for water would eventually outstrip available water resources so that South Africa would be faced with an increasing problem of aridity. Accordingly the Viljoen Commission was appointed to investigate ways of preserving a balance between water supply and water demand in South Africa.

5:2:2 THE FINDINGS OF THE VILJOEN COMMISSION

An important finding of the Commission was that:
"The indications are that about 40 per cent of the Republic's potentially usable water is already in use and, if the rate of increase of consumption over the 35 years persists, all will have been committed by the end of the century."[4]

---

4. Republic of South Africa, Report of the Commission of Enquiry into Water Matters, R.P. 34/1970, Pretoria, p.26.

In examining the distribution of water resources in South Africa, the Commission noted that this was highly irregular in a number of different respects:

(a) regional irregularity : water resources are clearly "not uniformly distributed geographically over the Republic. About 85% of the surface runoff originates from 36% of the area, located mainly within a 250 mile wide strip along the east coast and a narrow strip along the south coast of the subcontinent";[5]

(b) annual irregularity : droughts occur repeatedly in South Africa and, as a result, riverflow is highly erratic compared to other countries. The Viljoen Report notes that:

> ".... variability of streamflow, and the fact that prolonged droughts can occur in many regions, means that water must often be held over for a number of years in storage reservoirs. Evaporation loss can then become so high that an increase in storage capacity is no longer accompanied by an improvement in the assured yield. To secure maximum water yield during drought the ratio of magnitude of storage provision to mean runoff at a dam site depends on climatic conditions and shape of storage basin: for South African rivers this ratio usually lies between $1\frac{1}{2}$ and 3"[6]; and

(c) seasonal irregularity : although there is a variety of climatic conditions in South Africa, nearly all areas experience wet and dry seasons, with the result that riverflow varies seasonally and that the construction of storage reservoirs is therefore essential.

---

5. Ibid., p. 29.
6. Ibid., pp. 27-29.

Unlike other arid countries South Africa does not benefit from appreciable sources of groundwater since the rock formations in South Africa are generally highly impermeable. Furthermore the outflow of groundwater would not be available to sustain riverflow if it was extracted in large quantities from boreholes.

With respect to the growth of water consumption, the Viljoen Commission found that although the annual rate of increase in the use of water for irrigation had remained practically constant since 1910, "the quantity of water used for industrial and municipal purposes rose gradually till about 1933 but since then growth rate has been very rapid."[7] If this rapid rate of growth continues in the future then this would "mean that even before the end of the century all available water resources will already have been committed to meet our total needs."[8] Furthermore although on an aggregate basis it appears that there is a real danger of water shortage in South Africa, the problem is largely regional since there are parts of the country where future development is never likely to be hindered by water shortage, while in other areas economic growth is already being adversely affected on account of limited water.

To obviate this problem of future aridity the Viljoen Commission emphasizes the necessity of improving the efficiency of water management and utilization in the country. The Report reflects an awareness that the problem of aridity is an economic one, and that its solution requires an approach based on sound economic principles. For example, while the Commission stresses "the need for realistic action by the authorities to provide adequate funds for the development of our water resources"[9] it also recognizes

---

7. Ibid., p. 29.
8. Ibid., p. 29.
9. Ibid., p. 27.

that "..... due care must nevertheless be exercised to ensure that maximum possible benefits accrue from water schemes, that expenditure on water development can be justified in the face of demands on funds for other purposes and that the necessary balance among the industrial domestic and agricultural sectors is achieved and maintained."[10]

In examining the various measures to avert the possibility of a water crisis in South Africa which were recommended by the Viljoen Commission, it is interesting to note that they reflect the various characteristics of a water supply policy in an arid country described in section 5:1:4 above. For example, although the Commission recommends conventional measures to balance water supply and demand such as "raising the potentially useable component of surface runoff",[11] developing groundwater sources, improving the efficiency of water application in irrigation and eliminating "the numerous small shallow storage works in which practically all the water evaporates and almost no beneficial use is made of the stored water"[12], they also consider further measures which are particularly applicable to arid countries, such as:

(a) boosting available natural water resources through desalination processes and artificial stimulation of rain;

(b) maximizing water conservation through variable draft operation techniques, reduction of evaporation by providing a protective chemical layer on water surfaces and construction of larger, deeper storage dams; and recirculation and reuse of water in production processes;

---

10. Ibid., p. 27.
11. Ibid., p. 32.
12. Ibid., p. 33.

(c) implementing "a realistic water price policy" based on marginal cost pricing principles. This last recommendation is particularly interesting since it reflects a recognition that water scarcity is simply an instance of the general economic problem of scarcity, and that the price system provides an efficient means of balancing supply and demand. It is encouraging to note that the Commission reflects a trend to move away from a technocratic or "engineering" approach to the problem of "meeting water requirements" towards a more market orientated approach.

5:2:3 THE PLANNING OF WATER RESOURCE DEVELOPMENT IN SOUTH AFRICA

The planning of water resource development is of particular importance to the State since, not only may water shortages limit overall economic growth, but the sectoral or spatial misallocation of water resources may prevent economic growth from proceeding in a balanced manner. Furthermore as the guardian of the people's health and recreational facilities, it is interested in the intangible benefits that accrue from the planned promotion of water resources while as the custodian of the country's natural resources, it is interested in the conservation and unimpaired transmission to future generations of the country's water resources.

The Viljoen Report noted a number of deficiencies in the planning of water resource development in South Africa:

(a) The planning of new projects tended to be initiated in an ad hoc manner: "..... the comprehensive planning of the country's water resources development to meet growing needs, was neglected. Instead of ascertaining for itself which were the most urgent needs and planning objectively to meet them, the Department tended to tackle schemes under pressure from outside and frequently only when these had actually become urgent."[13]

(b) There appears to have been insufficient coordination of the functions of the water authority with those of the other Government authorities concerned with comprehensive economic and physical planning and with water conservation and related use: "As the Department concentrated mainly on the provision of water for irrigation, the only departments that were regularly consulted .... were the Departments of Agriculture and the then Department of Land."[14]

(c) New projects were generally not evaluated on a sound economic basis: "As agriculture was regarded as providing social services that to some extent transcended economic values, prior cost - benefit studies were seldom undertaken to justify new irrigation schemes, although as a matter of course sophisticated unit cost analyses were carried out for the fixing of tariffs for water from regional water supply schemes."[15]

The Viljoen Report therefore made a number of recommendations to correct these deficiencies:

(a) Water development proposals should be integrated with the National Economic Development Programme since "these schemes can be effectively planned only after the national framework into which they are to fit has taken shape."[16] In this way they "would find their logical place in a programmed sequence of priorities."[17]

---

13. Ibid., p. 156.
14. Ibid., p. 156.
15. Ibid., p. 156.
16. Ibid., p. 157.
17. Ibid., p. 157.

(b) Since the comprehensive planning of the country's physical resources is the task of the Prime Minister's Planning Advisory Council, the Viljoen Commission recommended that "the Department of Water Affairs should in future liaise closely with the Council in regard to all new individual projects, on plans for the development of river basins and in the regional planning of the country's water resources, and that there should be regular consultations between the two bodies with a view to ensuring co-ordination of water resource planning with overall planning."[18] The Viljoen Commission also recommended that to achieve a closer co-ordination of water development proposals with national planning objectives a Standing Committee for Water Affairs should be set up from among its own members, with co-opted members representing other official organisations interested in water development. The proposed junctions of this Committee were:

"(i) to co-ordinate the plans of the Department of Water Affairs with the overall economic and physical planning of the country and to fix general priorities;

(ii) to promote close liasion and co-operation among the departments and other organisations concerned with the conservation and utilisation of the country's water and affiliated resources and with research in these fields;

(iii) to give advice on the implementation of a suitable price policy for water."[19] This committee would thus functions as an effective instrument for the control of water resource development and planning.

---

18. Ibid., p. 158.
19. Ibid., p. 158.

(c) The Viljoen Commission also came out strongly in favour of cost-benefit analysis of proposed water schemes: "..... economic analyses for the justification of water projects should not be based solely on the monetary income that could be expected from a scheme but should take into account the total socio-economic value of the benefits flowing from the scheme, irrespective of whether such benefits are paid for directly or not. ..... no scheme for which the benefit - cost ratio is less than unity should be undertaken. In general, a scheme with a high-ratio should enjoy priority over one with a low ratio. Such a policy would ensure that only economically sound schemes that will be to the ultimate advantage of the country and yield a positive contribution to national development would be constructed."[20]

The machinery for detailed cost-benefit analyses of proposed water schemes already existed, in that the 1956 Water Act required that a White Paper should be tabled before Parliament to obtain its approval for any proposal in excess of R500 000.

5:2:4 ADMINISTRATION OF WATER RESOURCE DEVELOPMENT IN SOUTH AFRICA

The central role of the State in water resource development in South Africa arises from the following factors:

(a) the peculiar characteristics of water which lend an overriding national interest to its development and utilization:

(i) it is a resource that is scarce in relation to national needs;

(ii) it is very unevenly distributed between different areas and in relation to different potential users, which often requires transmission over long distances;

---

20. Ibid., pp. 139-140.

(iii) it can normally be effectively stored and distributed only on a very large scale;

(iv) its storage and distribution necessitates the purchase and expropriation of considerable areas of land; and

(v) its exploitation by one user often has considerable external effects on other users; and

(b) the stage of water resource development which has been reached in South Africa where individual projects often have considerable external effects so that centralization of authority is required to plan, operate and utilize water resources in an integrated manner.

The private sector therefore tends to play a subsidiary role in the supply of water services and is mainly involved at the technical level of design and construction. The role of local and regional authorities is largely restricted to meeting the water needs of urban areas and in water distribution.

The Water Act No. 54 of 1956 defined the functions of the Department of Water Affairs "to provide, build, extend or control water works; to obtain water from underground' sources; to assemble information about occurrences of water in the Republic; to make investigations into irrigation possibilities and the extent of irrigable land; to generate hydroelectric power; to give advice to farmers, irrigation boards, local authorities and others on water works and to assist with the establishment of works; to take such other steps as may be deemed necessary for the development, control and use of water; and to give effect to the terms of the Water Act, No 54 of 1956 as detailed in section 2 and other sections of the Water Act."[21]

---

21. Ibid., p. 158.

cf

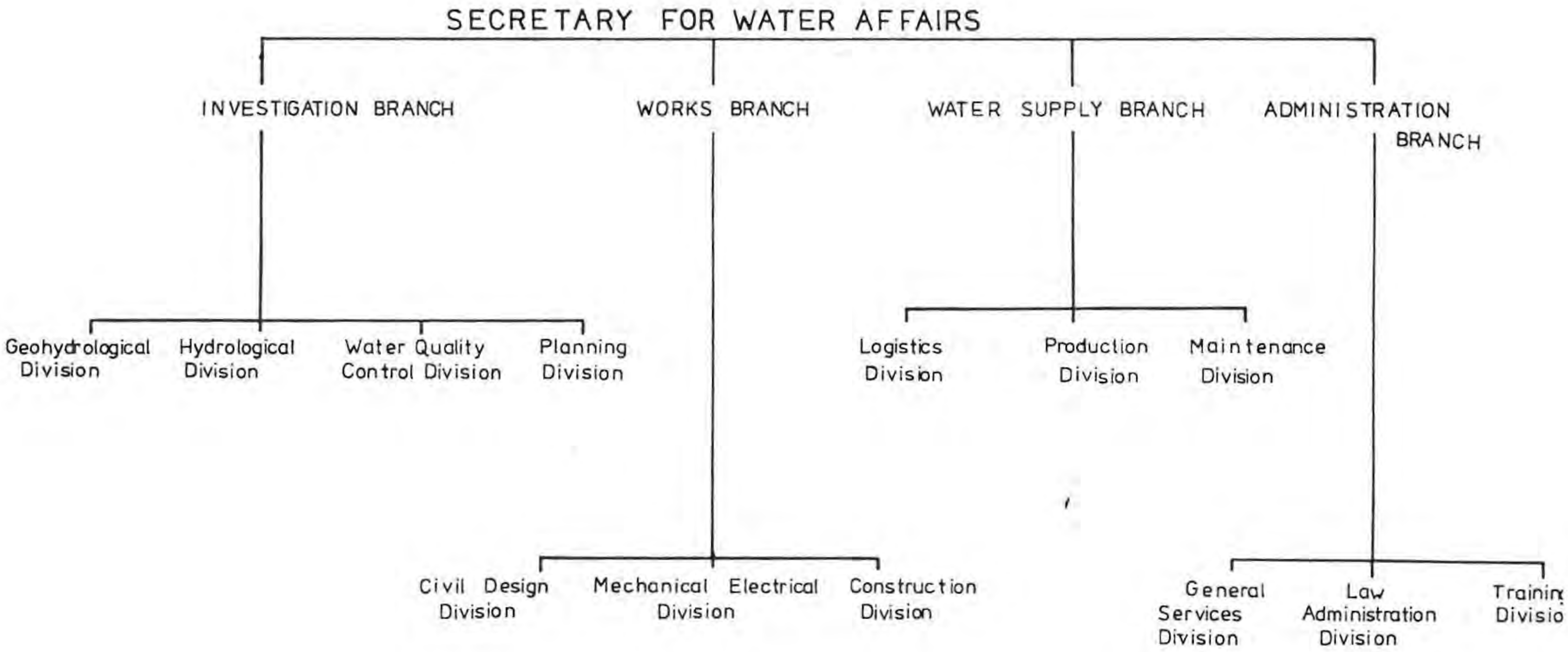


FIGURE 5:2 Organizational Structure of the Department of Water Affairs

It is clear, therefore, that the Department is concerned with both professional, scientific and technical work and with administrative legal and financial duties. The organizational structure of the Department is reflected in Figure 5:2. As can be seen, the Secretary of Water Affairs directs the affairs of the Department which is divided into four branches: the Investigation Branch; the Works Branch; the Water Supply Branch and the Administration Branch; with each branch being required to fulfil the following duties and responsibilities:

(i) the Investigation Branch is required to carry out investigations into surface water sources through its Hydrological Division and into underground water sources through its Geohydrology Division, plan water schemes on a national basis through its Planning Division and test and control the quality of water through its Water Quality Control Division;

(ii) the Works Branch has three division dealing with the civil design, the mechanical and electrical design and the construction of water works;

(iii) the Water Supply Branch operates and controls water scheme with its Logistics Division compiling data in respect of operating factors with a view to the optimization of the operation of water schemes, its Production Division operating water schemes and controlling consumption and its Maintenance Division maintaining operating equipment; and

(iv) the Administration Branch provides administrative auxiliary and training services through its General Services, Law Administration, Personnel and Functional Training Divisions.

It is interesting to contrast the organizational structure of the Department of Water Affairs with the public transport and electricity undertakings described in Chapters 3 and 4, since unlike these undertakings, the water authority does not provide the division of authority on a regional basis. This is because local water needs are usually met by municipal and regional authorities, so that the Department of Water Affairs is responsible for the co-ordination of these bodies at a national level as well as the supplementary provision of water supplies where they are not provided by these bodies.

5:3 WATER PRICE POLICY IN SOUTH AFRICA

5:3:1 SYSTEM OF SUBSIDIZATION

Although water prices are generally kept artificially low in South Africa there is a considerable difference in the degree of subsidization of water rates between the agricultural and non-agricultural sector and also within the agricultural sector itself.

The Viljoen Report found that for government irrigation schemes "the revenue from water rates and sales of water covered less than one-third of the annual operating expenses."[22] In effect this amounted to a subsidy of 100 per cent on capital costs while operating and maintenance costs were subsidized to an average extent of about 69 per cent.

The Viljoen Report also found that "about 29% of the irrigated area in the country was controlled by irrigation boards and taking into account loans written-off, these boards in effect enjoyed a 60.2 per cent subsidy on capital costs from the State but paid in full their own running costs."[23]

Although the maximum subsidy payable by the Government to private irrigators is 331/3% of the capital cost, many private schemes have been constructed without subsidy aid so that the subsidization of approximately 40% of the irrigable land which is under private irrigation, in all probability amounted to less than 30 per cent of the capital cost. Furthermore there were no subsidies on operating cost.

---

22. Ibid., p. 144.
23. Ibid., p. 10.

It follows that although the policy for irrigation water supply means that the irrigation sector of agriculture enjoys a hidden subsidy in the form of exceptionally low prices for water, "there is also anomalously extreme discrimination against private irrigators and those controlled by irrigation boards in favour of irrigators under Government water schemes."[24] Although the irrigation board schemes and private irrigation works in general embraced the more favourable schemes this did not appear to justify the extent of the preferential treatment accorded to irrigators on Government water schemes in comparison with those who have, on their own or through irrigation boards, brought their land under irrigation.

In the case of regional water schemes supplying municipal and industrial areas the full operating costs were generally covered as well as a substantial proportion of interest on capital. The Viljoen Commission thus found that revenue from sales of water on 9 schemes amounted on average to about 379% of the running expenses incurred on these schemes. Furthermore after deduction of operating expenses, the revenue available for meeting interest on capital and redemption expressed as a percentage of total capital expenditure was 3,88%. It was noted, however, that "in view of the fact that not all these schemes were yet in full operation and as revenue tends to rise with increasing consumption"[25], the actual recovery of capital and interest was in all probability even more favourable than this figure of 3,8 per cent.

The Viljoen Report accepted this water pricing policy and local authorities have continued to set their tariffs on the basis of the full cost recovery principle.

---

24. *Ibid.*, p. 145.

25. *Ibid.*, p. 144.

The Government accepted and the Department of Water Affairs has implemented as far as practicable the following recommendation by the Viljoen Report that, "..... in principle, water from Government water schemes should be supplied to all consumers, whether for consumption or processing, at a tariff such that after taking account of all direct, indirect and intangible benefits, as well as assurance of supply, is sufficient to cover capital and operating costs. This presupposes that unit prices will be scheme-related as is currently the practice, i.e. that a unit price will continue to be recommended for each scheme."[26]

In calculating the unit price for each scheme the Department of Water Affairs takes into consideration the following factors:

(i) capital cost must be redeemed over the useful life of the works;

(ii) interest on outstanding capital should be charged at the rate that it costs the State to provide the capital, and

(iii) the amount of water used does not remain constant from year to year, but normally increases as the demand rises.

It is therefore considered "unreasonable that, while the demand for water is low, the initial users should have to pay very high tariffs for water to cover the full annual interest on and redemption of capital for facilities that are really intended for enjoyment at a later stage when the full yield of the scheme is required."[27] The Department of Water Affairs has therefore worked out a method of calculation of the unit cost of water, according to which a uniform cost per unit of water is calculated for the full useful life of the

---

26. Ibid., p. 23.

27. Department of Water Affairs, Annual Report, 1977-1978, Pretoria, 1978, p. 165.

scheme by allowing initial annual deficits, when consumption is low, to accumulate for redemption when revenue increases due to increased use. The unit cost is then so determined that all expenditure will be redeemed by the estimated end of the useful life of the scheme. With rising operation costs and extensions to the scheme the unit cost figure is revised from time to time to conform to the actual demand for water and the actual expenditure. The tariffs that must be paid for water are based on this calculated unit cost.

It should be borne in mind when evaluating the profitability of any scheme that, according to this system, a scheme may during its initial years accumulate large deficits and yet still be in a fundamentally sound position. On the other hand a scheme which has been operating for many years may appear to be highly profitable on the basis of only the annual expenditure on operation and maintenance, whereas it is in fact falling into arrear with the payment of interest and redemption.

As far as the practical implementation of a realistic water price policy for Government water schemes was concerned, the Department of Water Affairs accepted the following recommendations by the Viljoen Commission:

"(i) that, for each specific scheme of the Department of Water Affairs, water intended for domestic or industrial use be supplied at a price that covers the full cost and that the artificial lowering of prices to industries and other urban consumers be not permitted by the Department, local authorities or other distributors;

(ii) that water rates on new irrigation schemes cover the full running cost as well as a percentage of the interest and redemption costs .....

(iii) that in so far as is consistent with socioeconomic conditions, the water rates on existing Government irrigation schemes be gradually raised to cover at least the operating costs.

(iv) that water rates for each irrigation scheme be determined by the Department of Water Affairs after investigation by and consultation with the Department of Agricultural Economics and Marketing;

(v) that an annual financial reviews for each Government water scheme be published."[28]

In calculating the cost of Government water schemes there are, however, a number of factors which make it impracticable to redeem in full all past expenditure. Furthermore in the case of Government irrigation schemes it is also considered to be unfeasible to redeem current capital and operating costs. Firstly, the Department of Water Affairs considers that the subsidy granted to these schemes should not be less than that granted to irrigation boards and private schemes in terms of the Water Act. This means that there should be a write-off of at least one-third of capital cost. It is also considered reasonable that the subsidy on capital cost should be in broad agreement with that applicable to local authorities.

Apart from these amounts, the Department also considers it unfair to recover in full the following costs:

(i) deficits which accumulated before the adoption of the recommendations of the Viljoen Commission on 1 April 1970 since during, this period the State had fixed tariffs that were too low to recover the full cost. Consequently only the actual unredeemed capital expenditure of older schemes is carried forward as a liability after 1 April 1970 and all accumulated deficits up to and including 31 March 1970 are regarded as irrecoverable;

---

28. R.P. 34/1970, <u>op.cit</u>., p. 23.

(ii) deficits incurred after 1 April 1970 which arose in circumstances where irrigation tariffs could not be raised sufficiently to cover operating costs and annual interest; and

(iii) interest accrued on the capital outstanding on the completion of each major unit of a large scale project such as the Orange River Project, where water is not yet being fully used consumptively, is regarded by the Department as a contribution to the general development of the country and is accepted as irrecoverable.

The recent performance of Government water schemes is reflected in Table 5:1 and Figure 5:3 which set out the operating results for the period 1976 to 1981. It should be pointed out that any trend analysis based on these results will not be wholly correct since revenue is not always received in the same year that financial disbursements are made.

With respect to irrigation schemes it can be seen that some progress has been made towards meeting the recommendation of the Viljoen Commission that "water rates on existing Government irrigation schemes should be gradually raised to cover at least the operating costs."[29] The percentage of operating costs covered by revenue earned on water used for irrigation purposes rose from 20 per cent for the year ended 31st March 1977 to 40 per cent for 1980-1. The main improvement in the operating performance of irrigation schemes occurred during the years 1977-8 and 1978-9, particularly during the latter year when revenue increased by 83,4 per cent and the operating deficit fell by 29,7 per cent compared to the previous year. The Annual Report for 1978-9 of the Department of Water Affairs attributed this

---

29. Department of Water Affairs, <u>op.cit.</u>, p. 169.

TABLE 5:1

SUMMARY OF OPERATING COSTS AND REVENUE FOR GOVERNMENT WATER SCHEMES

	OPERATING COSTS			REVENUE			OPERATING SURPLUS (DEFICIT		
	Industrial & Domestic	Irrigation	Total	Industrial & Domestic	Irrigation	Total	Industrial & Domestic	Irrigation	Total
	R1000	R1000	R1000	R1000	R1000	R1000	R1000	R1000	R1000
1976-77	11043,3	8929,0	19972,3	18481,1	1800,9	20282	8877,3	(7127,7)	1749,6
% Increase									
1977-78	15228,3	9610,6	24838,9	24765,6	2429,5	27195,1	11399,0	(7099,1)	4299,9
% Increase	37,9	7,6	24,4	34,0	34,9	34,1	28,4	0,4	145,7
1978-79	18609,8	9723,5	28333,3	33834,3	4456,2	38290,5	17428,4	(4992,5)	12435,9
% Increase	22,2	1,2	14,1	36,6	83,4	40,8	52,9	29,7	189,2
1979-80	17275,8	10891,8	28167,6	37266,8	4933,8	42200,6	19991	(5958)	14033
% Increase	7,2	12,0	0,5	10,1	10,7	10,2	14,7	19,3	12,8
1980-81	19333,3	12922,8	32256,6	53896,6	5486,3	59382,9	34563,3	(7436,5)	27126,8
% Increase	11,9	18,6	14,5	44,6	11,2	40,7	72,9	24,8	93,3

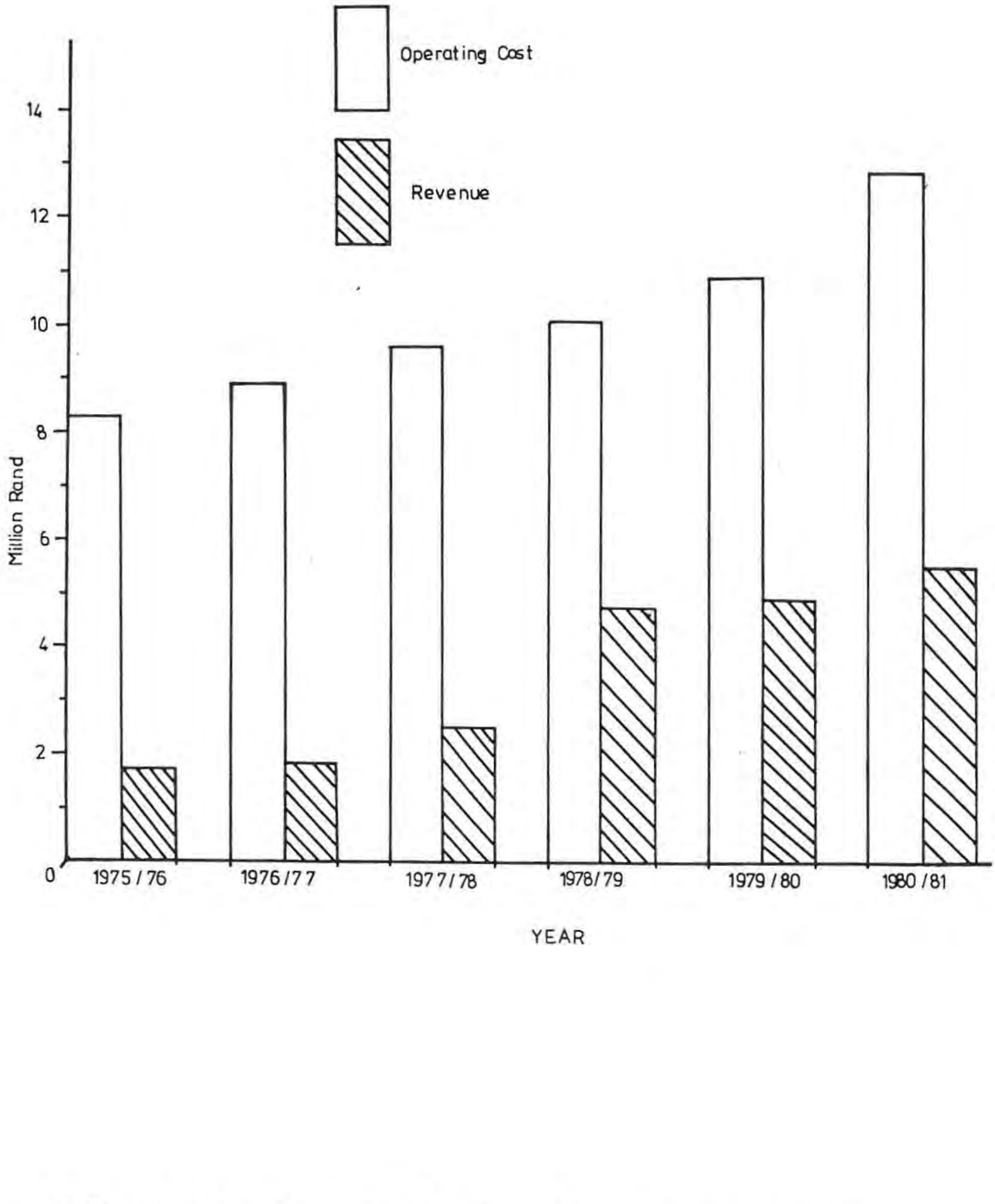


FIGURE 5:3 Operating Results on Government Irrigation Schemes 1975-76 to 1980-81

this improvement to the fact that scheduling was made more complete, that collection of various amounts was more strictly controlled and that water rates and tariffs were raised. However, during the years 1979-80 and 1980-1 revenue increased at the much slower rate of about 11 per cent per annum, while the operating deficit actually increased during both these years. This may be an indication of a return by the Department to the view that irrigation farmers merit special treatment in the form of being charged artificially low prices for water. In a recent investigation of water tariff policy, the Claassen Committee[30] recommended that the Department should discontinue its practice of gradually reducing the gap between revenue and running costs on irrigation schemes since it was considered that this would place an undesirable burden on irrigation farmers. It can be seen that this represents something of a reaction against the trend towards a more cost-orientated tariff policy which was supported by the Viljoen Commission.

Table 5:2 indicates the revenue/operating cost ratio for each individual irrigation scheme which generated revenue during the year 1980-1. It thus excludes all schemes which did not generate any revenue during this year. It can clearly be seen from this table that there are considerable variations around the average revenue/operating cost ratio of 42 per cent. For example the van der Kloof Canals have a ratio of 206,5 per cent whereas it is only 4,7 per cent for Doorn River. This is also demonstrated by the fact that even if one excludes the particularly high value for Krokodilpoort, the standard deviation still obtains the very high value of 41,9. The fact that the application of the Viljoen Commission's recommendations to schemes built after 1970 had led to an improvement in

---

30. Departement van Omgewing, Verslag van die Komitee van Ondersoek na die Prysbeleid met betrekking tot die Bepaling van Watertariewe, Pretoria, Maart 1982.

their operating performance is reflected by the fact that the average revenue operating cost ratio for schemes built after 1970 is 68,8 per cent while that for schemes built before 1970 is 36,1 per cent. Furthermore 9,5 per cent of the schemes built before 1970 earn operating surpluses as compared to 23,5 per cent of schemes built after 1970. The improved results of the newer schemes have occurred despite the fact that a number of these schemes are not yet being used to full capacity and can therefore be expected to incur large deficits according to the method of unit cost calculation employed by the Department. There should thus be a greater variation in the operating results of the newer schemes and this is borne out by the fact that the standard deviation of the ratios for schemes built after 1970 is 56,1 whereas it is 28,3 for schemes built before 1970 (excluding Krokodilpoort).

With respect to water used for industrial and domestic purposes it can be seen from Table 5:1 and Figure 5:4 that the revenue earned on Government water schemes has been making an increasing contribution towards the recovery of capital costs. This is in line with the Department's policy that the rate set for each scheme should cover the unit cost of the scheme. There was a particularly impressive improvement in the performance of industrial and domestic water schemes during the year 1980-1 when revenue increased by 44,6 per cent compared to the previous year while the operating surplus rose by 72,9 per cent from R2 million to R34,6 million.

TABLE 5:2

REVENUE/OPERATING COST PERCENTILES ON IRRIGATION SCHEMES 1980-81

A. IRRIGATION SCHEMES BUILT PRIOR TO 1970 WHICH SHOULD RECOVER AT LEAST MAINTENANCE AND OPERATING COSTS

		%
1.	Albasini Dam	26,4
2.	Blood River	21,5
3.	Bushmans River	8,6
4.	Buffeljachts River	65,8
5.	Buffalo River	13,5
6.	Doorn River	4,7
7.	Duivenhoks River Dam	40,3
8.	Elands River (Lindleyspoort Dam)	12,2
9.	Elands River (Rust de Winter Dam)	24,0
10.	Gamka River	49,4
11.	Hluhluwe River	155,1
12.	Kafferkuils River	9,7
13.	Kat River	65,0
14.	Kingna River	68,3
15.	Klein Marico River	11,3
16.	Klippaat River	58,6
17.	Konings River	56,25
18.	Korente River	35,0
19.	Koste River	19,6
20.	Krokodilpoort	1330,5
21.	Crocodile River (Hartebeespoort Dam)	37,5
22.	Leeu River	27,5
23.	Leeu Gamka Dam	113,7
24.	Magalakwin River	47,0
25.	Mapochsgronde	15,7
26.	Mnyamvubu River	5,1

TABLE 5:2 (Continued)

27.	Mooi River	46,3
28.	Njelele River	143,2
29.	Ohrigstad River	26,8
30.	Olifants River (Loskop Dam)	52,6
31.	Orange River (Buchuberg Dam)	21,5
32.	Orange River (Vioolsdrift)	54,9
33.	Pongola River	95,5
34.	Rhenoster River (Koppies Dam)	13,9
35.	Riet River	19,8
36.	Sand River	12,9
37.	Sand-Vet River	38,2
38.	Schoonspruit	32,2
39.	Sterk River	19,6
40.	Tarka River	8,6
41.	Vaal River	45,3
42.	Vaalharts Irrigation Scheme	58,9
B.	IRRIGATION SCHEMES BUILT AFTER 1970	
43.	Brand River	9,9
44.	Gamtoos River (Paul Sauer Dam)	45,0
45.	Great Letaba (Fanie Botha Dam)	52,5
46.	Crocodile River (Bushveld)	27,4
47.	Modder and Lower Riet River	8,0
48.	Nwanedzi River	68,7
49.	Olifants River (Stompdrift Dam)	48,6
50.	Orange River (H.Verwoerd & P.K. Le Roux Dam	130,4
51.	Orange River (Kakamas)	17,9
52.	Orange River (Upington Islands)	10,4
53.	Orange River (van der Kloof Canals)	206,5
54.	Politsi River	60,0
55.	Sanddrift River	177,5
56.	Fish-Sundays River	50,8
57.	Watervals River	63,7
58.	White River	117,6
59.	Witwaters River	74,3

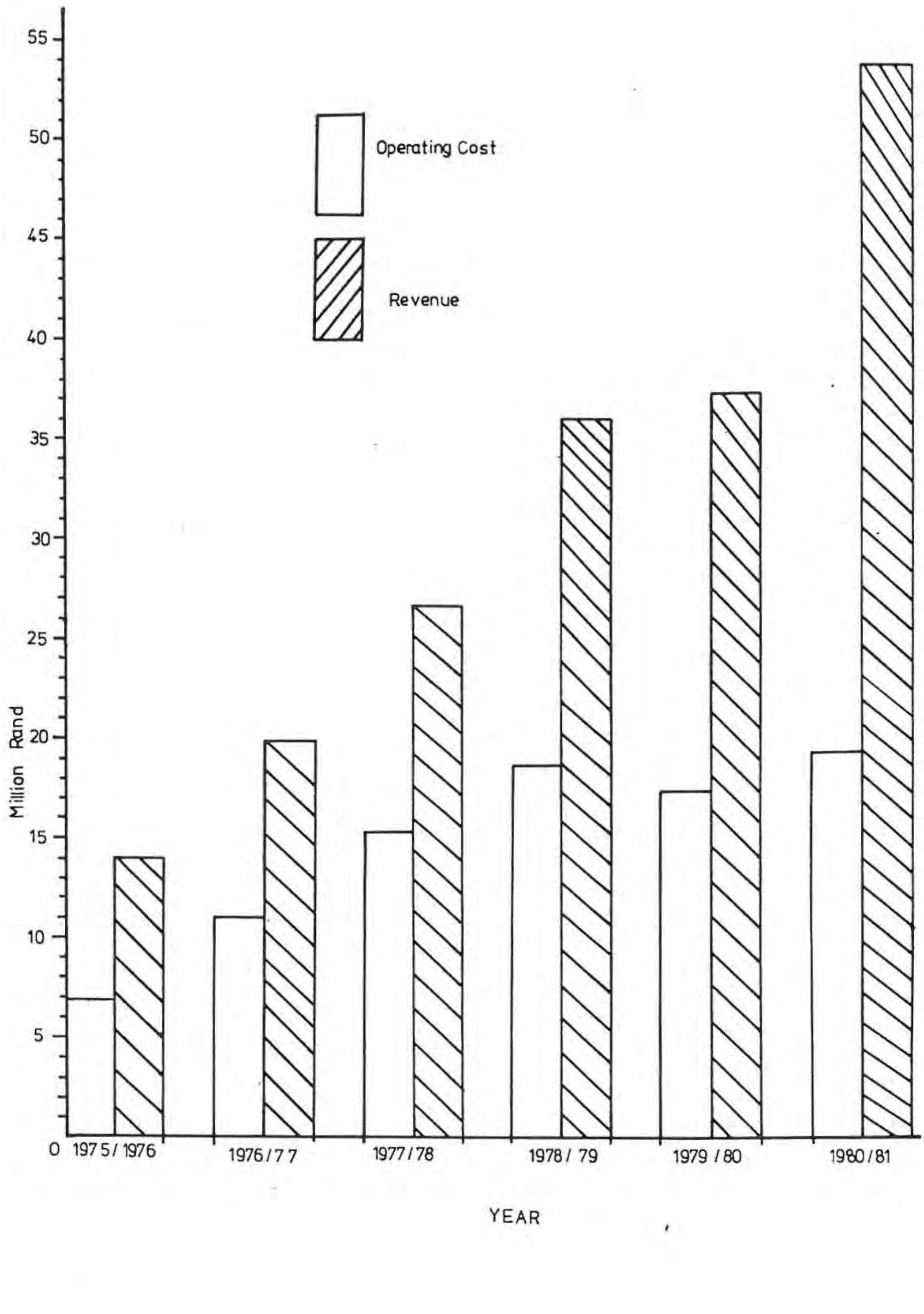


FIGURE 5:4 Operating Results on Industrial and Domestic Schemes, 1975-76 to 1980-81

5:3:2 WATER RATES ON GOVERNMENT IRRIGATION SCHEMES

The prices of water for agricultural purposes on Government irrigation schemes are fixed in various ways for different agricultural purposes and may fluctuate not only from scheme to scheme, but also from year to year at the same scheme. An indicative water price in the form of a basic water levy is normally stated in the relevant White Paper. These prices normally vary considerably from scheme to scheme according to the potential yield of the irrigated land. The highest water levies are normally indicated for those schemes which are expected to have the greatest potential yield.

Section 66 of the Water Act endows the Minister with the power to levy water rates and to increase or decrease them from time to time. These levies are announced, as necessary, by proclamation in the Government Gazette. The water rates are usually adjusted downwards during difficult drought periods as a measure of relief for irrigation farmers.

Table 5:3 below is reproduced from the Viljoen Report.[31] It presents data on water rates on three irrigation schemes which typify the structure of water prices prevailing on government schemes in South Africa.

---

31. R.P. 34/1970, op.cit., p. 149.

TABLE 5:3 DATA ON IRRIGATION SCHEMES, 1966-7

	SCHEME		
	Vaalharts	Hartbeespoort	Pongola
Average water rate paid, per morgen, 1966-7	R 1,91	R 1,83	R 4,70
Operating and maintenance costs per morgen,[32] 1966-7	R 1,30	R 6,73	R 4,85
Interest on capital per morgen	R 29,0	R 28,40	R 21,80
Average water rate as percentage of annual operating and maintenance costs	26,2%	27,1%	
Gross return per morgen	R296	R 395	R 366
Cost of production (excluding interest)	R162	R 237	R 218
Net return	R134	R 158	R 148
Average water rates as percentage of gross return	0,65%	0,46%	1,18%
Operating and maintenance costs as percentage of gross return	2,46%	1,70%	1,33%
Net income per predominent size plot.	R402	R1896	R0460

32. 1 morgen is equivalent to 0,857 hectares.

Table 5:3 indicates that the structure of rates between the different schemes failed to reflect their relative costs. For example, although the Pongola scheme had the lowest operating and maintenance costs, the farmers involved in this scheme are, at the same time, charged the highest water rates.

In a survey of the tariff schedules proposed in the White Papers which were published in respect of government irrigation schemes proposed during the period 1979-83, the following was found:

(i) In most schemes the farmer is limited to a fixed annual quota of water consumption;

(ii) In about half the schemes the tariff is charged as a flat rate per hectare;

(iii) In the other schemes the tariff is set according to a sliding scale with a minimum quota being charged at a flat rate per hectare and additional units of consumption being charged at a rate per unit of water supplied. Three examples of this sliding scale, viz., Upington Islands, Elands River and Vaalharts, are given in Table 5:4 below.

TABLE 5:4

UPINGTON ISLANDS

Total annual quota	1500 m^3/ha
Fixed rate up to 500 m^3	R20/ha
Variable rate for 501 m^3 to 1000 m^3	2c/m^3
Variable rate for 1001 m^3 to 1500 m^3	4c/m^3

ELANDS RIVER

Total annual quota	6200 m^3/ha
Fixed rate up to 2500 m^3	R6,50/ha
Variable rate 2501 - 5000 m^3	0,35c/m^3
Variable rate 5001 - 6200 m^3	0,45c/m^3

TABLE 5:4 (continued)

VAALHARTS

Total annual quota	7 700 m^3/ha
Fixed rate up to 3800 m^3	R17,00/ha
Variable rate 3801-5800 m^3	0,49c/m^3
Variable rate 5801 m^3 - 7700 m^3	0,65c/m^3

Although the Viljoen Commission indicated its approval of this type of sliding scale and recommended its extension to other schemes on the grounds that it encouraged more efficient use of water by acting as a disincentive against overirrigation, the Claassen Committee[33] criticized this incremental block rate pricing policy on the grounds that it discriminated against those users who consume their full quota of water. It considered that if users do not consume their full quotas, the surplus water has no alternative uses and is therefore simply wasted.

5:3:3 WATER RATES ON MUNICIPAL AND INDUSTRIAL WATER SCHEMES

An average cost pricing policy appears to have been adopted for municipal and industrial water schemes. The following factors are usually taken into account in calculating unit costs:

(a) interest on and redemption of capital over the useful life of the scheme;

(b) annual maintenance costs;

33. Departement van Omgewing, op.cit., p. 6.

(c) supervision and labour expenses;

(d) costs of chemicals for water treatment and power supplies for pumping water;

(e) contributions to contingency and replacement funds; and

(f) anticipated demand for water; this is an important aspect of unit cost computation since during the first few years of a scheme the proportion of delivery capacity used is relatively small since it rises gradually over a number of years to full demand.

If the full annual costs of a water supply scheme are required to be paid off over a number of years this means that during the initial years, the tariff should be relatively high when water demand is comparatively low, but that it diminishes to a much lower figure as water consumption rises. To obviate such undesirable fluctuations in water rates, the Department of Water Affairs normally computes a more or less constant tariff over the life span of the scheme. Effectively this means that losses accumulate during the initial years when water consumption is low but that they are liquidated in later years when water consumption is at a higher level. Consequently, during the early years it is necessary to have additional working capital over and above that required for the construction and operation of the scheme, in order to finance the accumulated losses. In view of this, the Viljoen Commission recommended that instead of subsidizing water rates charged on regional water schemes the government should grant the local authorities loans to provide working capital during the early years of their water schemes.

The basic water rate is calculated <u>for each scheme separately</u> taking account of anticipated consumption, capital costs, extent of the water distribution system, degree of purification, pumping costs, and estimated costs of administration and maintenance according to the type of scheme. Where only relatively small quantities of water are drawn from an irrigation scheme for industrial or other purposes, a uniform basic rate based on costs averaged over a large number of schemes is universally applied.

The Viljoen Commission[34] compared the water rates charged in the different cities and towns in South Africa and found that while there were not substantial differences in the water rates charged in the main urban centres of Johannesburg, Cape Town, Pretoria, Port Elizabeth and Bloemfontein, there are also a few smaller towns such as Calvinia, Ugie, Louis Trichardt, Carolina, Brand vlei and Springbok where tariffs are much higher.

The Viljoen Commission also predicted that:

> ".... water tariffs in most urban areas will have to rise with increasing costs of water supply as water has to be brought in over greater and greater distances .... Rising prices must however, be recognized as a normal development and at the present stage are by no means so high that general subsidisation can be justified as a means of maintaining reasonable health and living standards."[35]

The regulations of the Department do, however, make provision for a subsidy on capital cost according to a sliding scale if the unit cost of water exceeds $22c/m^3$ for towns with a population of more than 5000 white inhabitants.

It is important to note that the price charged for water supplied to industrial and domestic consumers is based on the accounting costs of the individual scheme. No provision is made for inflation although the unit cost of water is revised from time to time to take account of cost increases above the levels estimated. The interest charge for each scheme is, however, calculated by applying the Treasury interest rate to the historical capital cost of the scheme and will only rise during the life of the scheme if there are increases in the Treasury rate.

---

34. <u>Ibid</u>., p. 152.
35. <u>Ibid</u>., p. 152.

Since the capital costs of water schemes tend to rise over time the unit capital cost of older schemes will tend to be lower than that for newer schemes and this is likely to cause the price of water to vary from scheme to scheme. In cases where one consumer such as the Rand Water Board is supplied by a number of schemes of varying ages, any increment in consumption will not be charged with the unit cost of the newest scheme but the price of water will be set to cover a weighted average of the unit costs of all the schemes. The Claassen Committee, however, has recommended that users should be charged the incremental average cost of the most recently constructed scheme. The implementation of this recommendation would imply that the water authorities devise some method of attributing the costs of new schemes to the users "responsible" for their construction .

5:4 EVALUATION OF EXISTING WATER PRICE POLICY

5:4:1 GOVERNMENT IRRIGATION SCHEMES

A. DEGREE OF SUBSIDIZATION

According to the basic principles of economic efficiency discussed in the first two chapters of this thesis, an optimal pricing policy will cause the cost to a consumer of an additional unit of consumption to reflect the cost to society of supplying that extra unit. Now, if this marginalist principle is applied to water pricing it means that, in the absence of externalities, the price should not be set below the level of running costs, since even when capacity is underutilized, these expenses must still be incurred to provide an additional unit of consumption.

In South Africa there is clear evidence that the water supplied from goverment irrigation schemes is substantially underpriced as, on average, water rates do not cover running costs. This artificially low water price tends to encourage excessive water consumption in the sense that water is put to uses with a value to society below the opportunity cost of the resources committed to these uses.

Not only is the artificial underpricing of water not conducive to the efficient utilization of this scarce commodity, but it leads also to the overpricing of the complementary production factor, namely the soil. This is a clear hindrance to efficiency in irrigation farming since it hampers the optimum combination of these two production factors. The Viljoen Commission was aware of this problem as can be seen from its statement that:

> ".... in the opinion of the commission, the position can be rectified only by establishing realistic prices that recognize the inherent scarcity of both factors in the national economy."[36]

If water is underpriced, it follows that water requirements will grow at a faster rate than if water prices were set according to marginal cost pricing principles, with the result that the need to invest in new capacity will arise sooner. The cost of bringing forward investment projects in this sector represents another loss to society of a sub-economic pricing policy.

Where external economies are associated with the supply of water for irrigation purposes, it may be optimal to set price below unit running costs when capacity is under-utilized, since price may not reflect all the marginal net social benefit obtained from an additional unit of

---

36. Ibid., p. 11.

consumption. This may often be the case where an irrigation scheme serves a number of other uses which are not explicitly charged for, such as flood control, pollution abatement or recreation. There is little to suggest, however, that the subsidization of water rates in South Africa is in any way directed at causing prices to reflect marginal social benefits. The pattern of subsidization varies from scheme to scheme and is characterized by extreme discrimination against irrigation boards and private irrigators in favour of farmers on government irrigation schemes,which is unlikely to be related in any way to the external benefits associated with water use on these different schemes. In fact the degree of subsidization appears to be related to an evaluation of the merit of the irrigation farmers on a scheme in terms of their need for state assistance. Subsidization therefore serves the redistributive rather than efficiency objective of economic policy.

The Viljoen Commission was highly critical of the fact that the subsidization of the irrigation sector by charging subeconomic water rates is largely hidden from the public "under the current budgeting system, whereby the revenue from sales of water is credited to general revenue and no separate record is kept of interest and capital redemption for individual schemes"[37] and that no attempt was being made to conduct intersectoral investigations of the effects of this subsidization on the national economy. There is therefore nothing to suggest that the policy of water price discrimination in the irrigation sector was being harmonized with regional policy objectives. The Commission therefore

---

37. Ibid., p. 146.

recommended that in motivating the schemes, there should be a shift in emphasis from "a political-agricultural-economic orientation to a national-economic orientation ... (since) ... application of a policy of price discrimination or of subsidization of certain sectors of the national economy has, apart from socio-political implications, an economic function requiring consultation with the relevant departments such as the three Departments of Agriculture and the Departments of Industries and of Finance."[38]

Although the Department has accepted the recommendations of the Viljoen Commission that for irrigation schemes constructed before 1970, water tariffs should be raised until revenue covers operating costs, and that the revenue generated from water supplied by schemes constructed after 1970 should also make some contribution to capital costs, Table 5:2 indicates that operating deficits are still incurred on the majority of water schemes. It is clear, then, that there still exists a substantial degree of over-subsidization and that the policy of raising tariffs to reduce operating deficits should be continued in the interests of economic efficiency. The Claassen Committee, however, has recommended that the objective of economic efficiency should be subordinated to the socio-political objective of encouraging the development of the agricultural sector by granting special assistance to irrigation farmers in the form of low water prices. If this recommendation is is accepted by the Department it should be borne in mind that any special treatment of irrigation farmers in this manner does have a cost to the economy in the form of a loss of economic efficiency.

---

38. Ibid., p. 146.

It is now necessary to evaluate the manner in which rates are actually charged to individual farmers on government irrigation schemes.

B. STRUCTURE OF WATER RATES

In section 5:3 it was shown that there are basically two types of tariff schedule used for pricing irrigation water in South Africa:

(i) a flat rate charged for a fixed quota of water consumption; and

(ii) a flat rate charged for a minimum quantity of water demanded followed by unit price increases as consumption increases over discrete ranges.

On water schemes where only a flat rate is charged, the marginal price of water is effectively zero. The price of water cannot therefore be used as an instrument to induce irrigators to economize on water consumption. The charging of a flat rate may thus be viewed as a means of generating revenue rather than as a tool for regulating demand.

On water schemes where a flat rate policy (FRP) is pursued, water consumption is usually limited to the capacity of the schemes by the specification of a maxima quota for each individual irrigator. The effect of a FRP and the imposition of quotas on the efficiency of production can be illustrated by means of insoquant analysis as in Figure 5:5. It is assumed that there are only two irrigators A and B. The quantities required by A and B of all factor inputs other than water are denoted by ya and yb respectively. The quantities of water required by A and B are denoted by xa and xb respectively. If a FRP is pursued then the isocost lines for consumers A and B will be the lines Ca and Cb

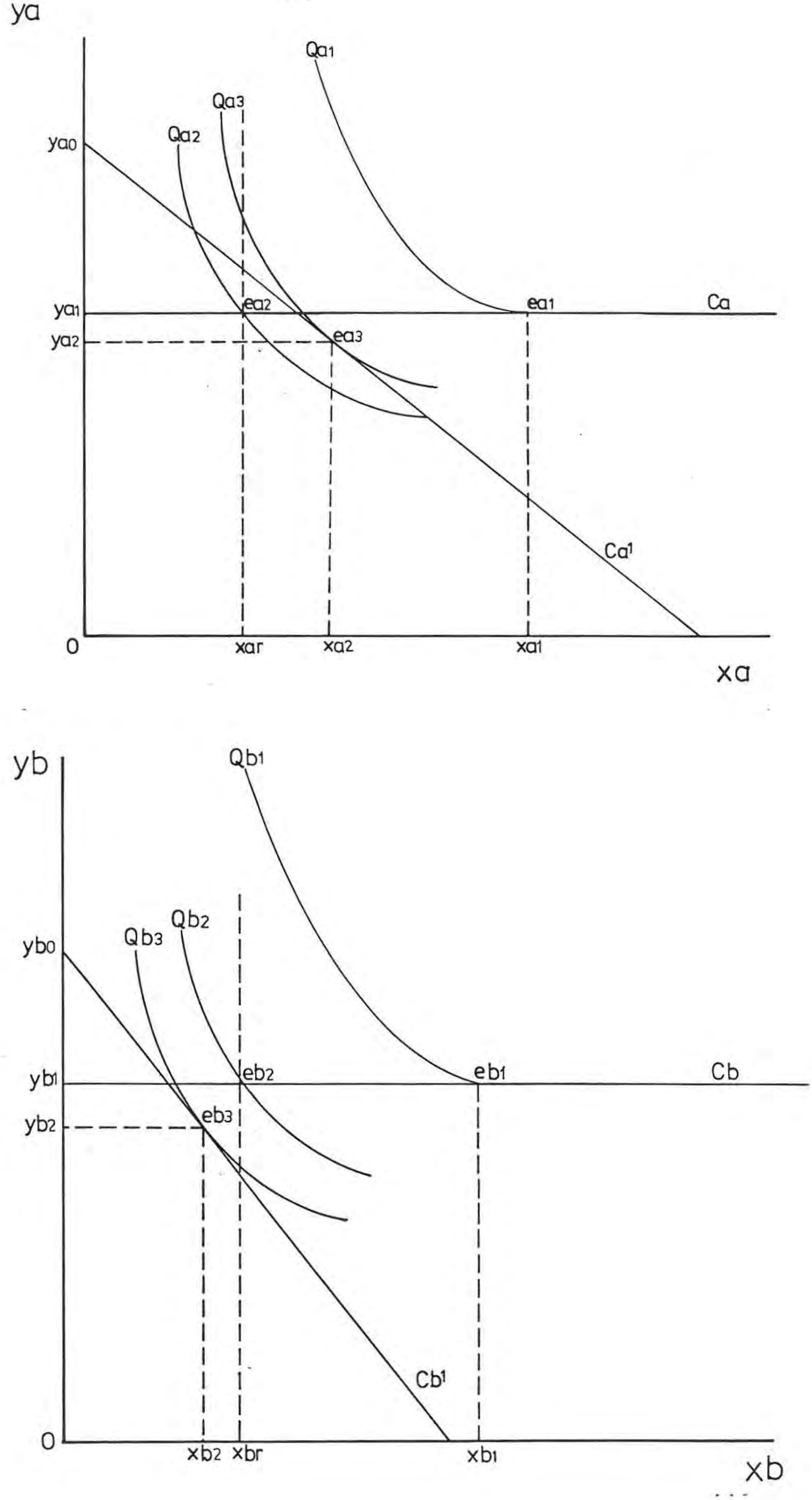

ya
Qa1
Qa3
Qa2
ya0
ya1
ea2
ea1
Ca
ya2
ea3
Ca1
0
xar
xa2
xa1
xa
yb
Qb1
Qb2
Qb3
yb0
yb1
eb2
eb1
Cb
yb2
eb3
Cb1
0
xb2
xbr
xb1
xb

FIGURE
5:5

which are horizontal since the price of x is zero. After paying the water levy, it is assumed that the farmers will spend the rest of their incomes on the other factor inputs, in this case buying ya1 and yb1 respectively. If no quotas are imposed the irrigators will demand the quantities xa1 and xb1 which are related to the equilibrium points ea1 and eb1 at which the isoquants Qa1 and Qb1 are tangential to the respective isocost lines. The imposition of quotas, however, results in both farmers reducing their water requirements to the specified maximum levels xar and xbr respectively. It is assumed that the sum of these two quotas is equal to the maximum capacity of the water scheme. .

Now, instead of rationing demand to capacity by means of quotas the water authority could set its water rates such that demand could rationed to the level of maximum capacity. In Figure 5:5 the isocost line change to Ca' and Cb' as a result of the abandonment of the flat rate and the charging of a positive unit price for water, with the result that new equilibrium positions are attained at ea3 and eb3 with the quantities of water demanded by A and B, xa_2 and xb_2, being such that their sum equals the maximum capacity of the water scheme, in other words,

$$xa2 + xb2 = xar + xbr$$

It can be seen that if water is rationed by price rather than by quotas, the quantities required by individual farmers will vary according to their willingness and ability to pay for water. In this case farmer A will use an amount greater than the fixed quota while consumer B will use an amount less than the fixed quota. According to the condition for Paretian efficiency in production discussed in Chapter 1, the quota system is likely to be inefficient since it is unlikely to cause water to be allocated between irrigators such that the marginal rates of technical

> substitution between water and other commodities will be equalized. If the rates at which the irrigators are prepared to substitute water for other commodities differ, then the reallocation of water between them could result in a net gain for one of them and no loss to the other provided the gainer fully compensated the loser for any loss he might suffer. Under a quota system, however, this mutually beneficial exchange could not take place. It therefore appears that the non-price method of rationing used by some water schemes may be inefficient.

The validity of this conclusion can be shown to depend on the assumption that water is continuously substitutable for other factor inputs. It may, however, be the case that irrigators can only use a single production process which combines water with other inputs in fixed proportions for each level of output. The implications of fixed factor proportions for water pricing policy can be analysed in terms of the simple model illustrated in Figure 5:6. The slope of the process ray OR indicates the fixed proportion in which water can be combined with other inputs to produce any given output level. If the quota is set equal to the fixed water requirement, oxr, for the output level indicated by the isoquant Qr, then it can be seen that even if water is rationed by price rather than by quota, the same quantity of water will still be required. It follows that in this fixed proportions case a marginal-cost-based water pricing policy will provide no greater incentive for irrigators to increase the efficiency of their water utilization than a system of flat water rates and quotas.

The problem of whether irrigation farmers can choose between irrigation techniques involving different intensities of water use was discussed with Department officials who conceded that there may be a limited number of irrigation processes available although they were sceptical that changes in the price of water relative to the prices of other inputs could induce irrigators to adopt less water-intensive techniques. Figure 5:7 which represents an intermediate case between the extremes of perfect factor substitutability illustrated in Figure 5:5 and fixed factor proportions shown in Figure 5:6, may thus be the closest

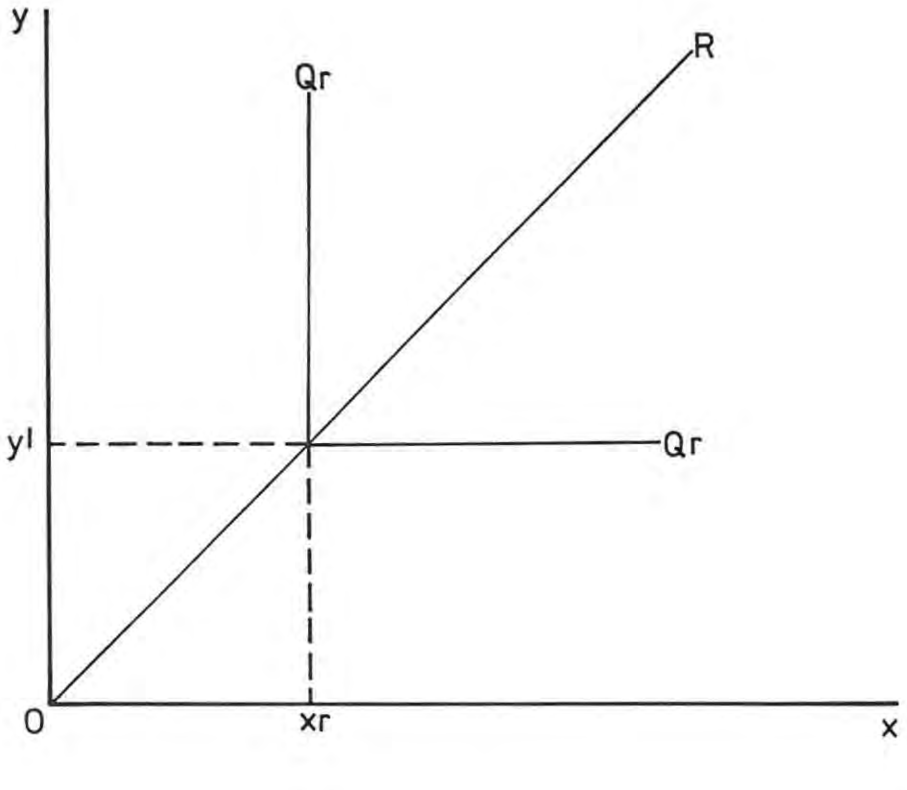


FIGURE 5:6 The Effect of Fixed Factor Proportions on Water Pricing Policy

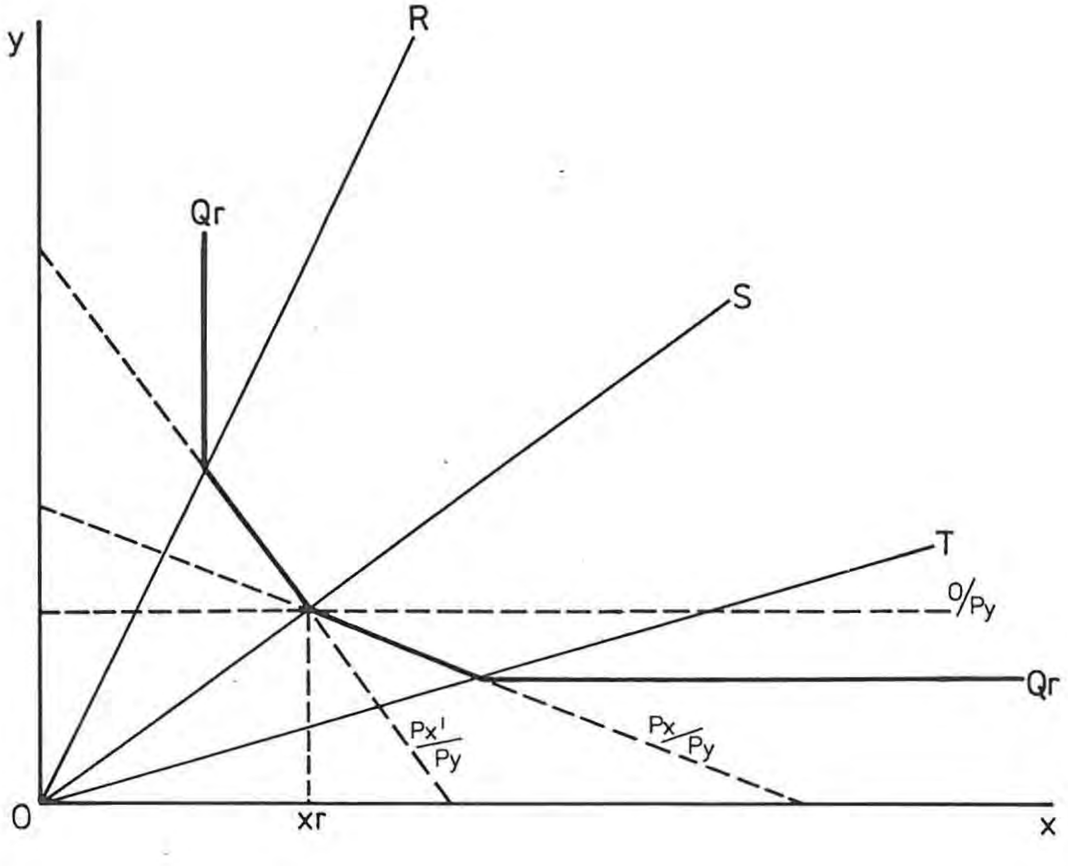


FIGURE 5:7 The Effect of Water Pricing Policy on Process Substitution where there a limited number of Irrigation Techniques

approximation to the actual set of technical opportunities facing a typical irrigation farmer. It can be seen in this figure that there are three processes available to the irrigator which combine water with the other inputs in the proportions indicated by the slopes of the process rays OR, OS and OT. If a flat rate is charged for the fixed quota of water Oxr, the process OS will be used. However, even if the price of water is allowed to rise to Px to ration demand to the capacity of the scheme this will not be sufficient to change relative factor prices to the ratio $\frac{Px1}{Py}$ at which the irrigator is induced to switch to the less water-intensive irrigation process indicated by the ray OR. Thus although it is possible for process substitution to take place it takes significant shifts in relative input prices to bring this about.

A greater degree of flexibility has been introduced into the quota system of allocating water on those schemes where the supply of water has been charged according to a sliding scale as described in the previous section. Table 5:4 indicates how this sliding scale has been applied by the Department for the Upington Islands, Elands River and Vaalharts Government irrigation schemes. It can be seen that a total annual quota has been fixed for each scheme. Thus the maximum level of water consumption is limited to 1500 m^3/Ha, 6200 m^3/Ha and 7700 m^3/Ha for the Upington Islands, Elands River and Vaalharts schemes respectively. It is therefore evident that the Department uses these maximum quotas, rather than pricing policies, as the means to ration water consumption to the level of existing capacity. It follows that there is still likely to be a certain measure of inefficiency with respect to the allocation of water between different consumers.

The sliding scale of water tariffs does, however, act to encourage water conservation by irrigation farmers in a manner which varies between each individual irrigating farmer according to the quantity of water he demands. As shown in Table 5:4, for each water scheme, the total annual quota is divided into three blocks which are each charged for at a different rate:

(i) The first block is charged out at a flat rate per hectare.

(ii) The second block is charged at a rate per cubic metre of water consumption.

(iii) The third block is charged at a rate per cubic metre of water consumption which is higher than that charged in (ii).

If the quantity demanded by an irrigation farmer falls within the first block, then he will pay for his water at a flat rate so that the marginal price of water will be zero and there will be no incentive for him to take measures which conserve water. If, however, the level of consumption by an irrigation farmer falls within either the second or third block, then he will be charged for each unit of water consumed so that there will be an incentive for him to introduce water saving measures. This incentive will clearly be greater for irrigators charged at the third block rate rather than the second block rate.

This can be illustrated by comparing the percentage savings in the cost of water per hectare for each scheme which results when water consumption is reduced by ten per cent from the upper limits of the first, second and third blocks of the total annual quota for the following types of water pricing policy:

(i) a flat rate (FR) policy under which a flat rate is charged per hectare;

(ii) a constant rate (CR) policy according to which a constant unit price is charged per cubic metre of water consumption; and

(iii) an increasing block rate (IBR) policy according to which water is currently being charged on these schemes.

To make this comparison it is assumed that on each of the schemes the total cost of purchasing the total annual quota will be the same regardless of which pricing policy is adopted. The alternative tariff schedules for the same three schemes as selected before are set out in Table 5:5 below.

TABLE 5:5

	UPINGTON ISLANDS	ELANDS RIVER	VAALHARTS
Total Annual Quota	1500 m^3/Ha	6200 m^3/Ha	7700 m^3/Ha
Flat Rate	R50,00/Ha	R20,65/Ha	R39,80/Ha
Constant Rate/m^3	3,33c/m^3	0,33c/m^3	0,52c/m^3
Increasing block rate			
First block	20	6,50	17,00
Second block	10	8,75	9,80
Third block	20 R50,00/Ha	5,40 R20,65/Ha	13,00 R39,80/Ha

The results of the comparison are set out in Table 5:6

TABLE 5:6

COMPARISON BETWEEN THE CONSERVATION INCENTIVES OF DIFFERENT PRICING POLICIES

PERCENTAGE DECREASE IN COST/HECTARE

	FIRST BLOCK			SECOND BLOCK			THIRD BLOCK		
	CR	FR	IBR	CR	FR	IBR	CR	FR	IBR
Upington Islands	10	0	0	10	0	6,7	10	0	12,0
Elands River	10	0	0	10	0	11,5	10	0	13,5
Vaalharts	10	0	0	10	0	10,6	10	0	14,2

The following observations can be made about the results of the above the table:

(i) an irrigation farmer will experience no savings in the cost of water per hectare if he introduces conservation measures when water is charged at a flat rate per hectare. For example, if a flat rate of R20,65/Ha is charged on the Elands River scheme, farmers will continue to pay this amount to the water authority regardless of how for they are able to reduce their consumption below the total annual quota of 6200 m^3/Ha;

(ii) if an irrigation farmer introduces a measure which results in the reduction of his water consumption when water is charged out at a constant unit price per cubic metre of consumption then he will be able to reduce the cost per hectare of water consumption in equal proportion to the quantitative water saving. This can be seen in Table 5:6 where in all cases a ten per cent reduction in water consumption results in a ten per cent saving in water costs; and

(iii) the conservation incentives under an IBR policy clearly vary according to the block into which the quantity demanded by the individual irrigator falls. For example if an irrigation farmer on the Elands River scheme reduces his water consumption by ten per cent from 2500 m^3/Ha to 2250 m^3/Ha he will continue to pay the flat rate of R6,50/Ha and thus will enjoy no saving in water costs. If he reduces his consumption level from 5000 m^3/Ha to 4500 m^3/Ha the cost of water per hectare will fall by 11,5 per cent from R15,25 to R13,50/Ha although the incentive to conserve water will be greater if consumption is reduced from 6200 m^3/Ha to 5580 m^3/Ha since in this case the cost of water per Hectare will fall by 13,5 per cent from R20,65 to R17,86 per hectare.

It may thus be concluded that the incremental block rate policy currently being adopted on some water schemes may be an effective means of curbing excessive consumption and encouraging water conservation. Whether this leads to an improvement in economic efficiency, however, depends very much on whether the water saved as a result of this pricing policy has alternative uses. It may actually be economically inefficient to encourage farmers to reduce their consumption below their maximum quotas if this water saved simply flows off to the sea. If this is the case then, as the Claasen Committee has pointed out, an IBR policy would discriminate against large users who consume their full quotas. If such discrimination is to be avoided it must be possible for the surplus water to be either resold to irrigation farmers or reallocated to industrial or domestic uses. Where the surplus water has alternative uses, it follows that since water is becoming an increasingly scarce resource in South Africa it may be desirable to extend the application of the IBR policy to government water schemes where water is currently being charged out at a flat rate. In conclusion though it should be noted that rationing water through price is a pricri more efficient than the quota system currently being used although there are a number of practical problems which must be considered before an optimal pricing policy is introduced. These will be discussed in section 5:5.

C. DISTRIBUTIONAL EFFECTS

As was explained in Chapter Two no evaluation of a pricing policy according to the principles of welfare economics is complete if it is only evaluated from the viewpoint of economic efficiency. It is also necessary to consider the effect of the policy on the distribution of income in the country. It is obvious that the construction of an irrigation scheme will substantially raise the value of the land settled under irrigation. Table 5:7 is reproduced from the Viljoen Report [39] and shows how the average cost per morgen of irrigated land can be deduced.

---

39. Ibid., p. 150.

TABLE 5:7

	Schemes comprising diversion weirs and canals without storage	Schemes with water supply secured by storage
A. Cost of unirrigated land, say	R 50	R 50
B. Cost of irrigation scheme per morgen	R750	R1200
C. Preparation, debushing and levelling of land	R100	R 100
D. Interest on investment for period until income is earned	R 50	R 75
	R950	R1425

At the time this report was published, the prices paid for developed irrigated land varied from R800 to R1500 per morgen depending on the locality and crops that were grown. This implies that the higher price of the irrigation land is mainly attributable to the cost of the irrigation scheme that supplies the water. If society has no preference between the users and suppliers of irrigation water, then it would clearly be equitable if the increase in land value were to accrue to the water supplier. The water rates charged on these schemes are too low, however, to appropriate the increase in land value to the water supplier and this means that this appreciation in value accrues to the farmers who owned the land at the time the scheme was introduced. They realize this value as a capital gain when they sell the land. Unless the original landowners are regarded as a particularly deserving section of the community, it would seem that the current pricing policy results in an arbitrary distribution of income which cannot be conclusively justified on the grounds of equity.

It may be concluded that the pricing policy for irrigation water which existed at the time of Viljoen Report suffered from some serious deficiencies according to both efficiency and equity criteria. The Viljoen Commission's recommendation that this pricing policy be changed to take account of the relative scarcity of water and soil resources in South Africa can thus be considered to be justified.

5:4:2 MUNICIPAL AND INDUSTRIAL WATER SCHEMES

As described in the previous section an average cost pricing policy has generally been adopted by municipal and industrial water schemes. Table 5:8 is reproduced from the Viljoen Report[40] except for the last two columns which have been calculated from the other figures. It contains figures calculated for a scheme comprising treatment works, a pumping installation, a pipeline delivering an average of 1 million gallons per day (m.g.d.) to a reservoir of one million gallons capacity. The capital cost was R600 000 and the useful life was estimated at 30 years for the pipeline and reservoir and at 15 years for the pumping unit which cost R25 000. Demand was forecast to be initially 0,4 m.g.d. rising at $10\frac{1}{2}\%$ to 1 m.g.d. over the first ten years and remaining constant thereafter. Any subsequent increase in demand will have to be met by an extension of the scheme. No provision was made for contingency or replacement funds.

As discussed in the previous section, a tariff calculated according to strict average cost pricing principles will be high during the initial years of the project where annual consumption has not risen to full capacity but will fall during later years as consumption rises. This pattern of price fluctuation is in complete contrast

40. Ibid., p. 133.

TABLE 5:8

ANNUAL COSTS

Year	Water Consumption m.g.d.	Interest at $6\frac{3}{4}\%$ and redemption on capital	Supervision and maintenance	Costs of chemicals and power	Total cost	Average cost-cents per thousand gallons	Adjusted marginal running cost
		R	R	R	R	cents	cents
1	0.40	47880	11400	7300	66580	45.7	
2	0.44	47880	11400	8050	67330	42.0	6,8
3	0.49	47880	11400	8940	68220	38.2	7,0
4	0.54	47880	14400	9850	72130	36.7	7,1
5	0.60	47880	14400	10950	73230	33.5	6,2
6	0.67	47880	14400	12200	74280	30.4	6,2
7	0.74	47880	14400	13500	75780	28.1	6,4
8	0.82	47880	19900	14970	82750	27.6	6,6
9	0.90	47880	19900	16400	84180	25.7	4,9
10-30	1.00	47880	19900	18250	86030	23.6	5,1

to that which would arise if marginal cost pricing principles are applied. During the early years when water consumption is below full capacity, the most efficient utilization of the capacity can be attained, in the absence of externalities, by charging a price equal to marginal running cost. In Table 5:8 marginal running cost has been calculated as the sum of the increase in the unit cost of chemicals and power, which are variable expenses and of the average cost of supervision and maintenance, which is a semi-variable expense. It can be seen that the difference between the price calculated to cover marginal running costs is substantial during the first few years when there is excess capacity. This implies that if the marginal cost pricing principle is adopted, water consumption should rise much more rapidly to full capacity than the ten years it takes to do so when average cost pricing is applied. It can thus be seen that average cost pricing will prolong the period of underutilization of capacity since water will be overpriced when there is excess capacity and the opportunity cost of water is consequently low.

Once water consumption has risen to the level where it equals the capacity of the water scheme, then, according to marginalist principles,given lumpy inputs,price should rise to ration demand to existing capacity and make a contribution to unit capital costs until it becomes economically desirable to extend the scheme. Thus, under marginal cost pricing, water prices will be relatively higher in later years while under the average cost method of pricing they are relatively lower. It can happen then that the price required to limit demand to capacity is higher than the price determined according to average cost pricing principles so that the former price will make a contribution not only to the capital costs allocated to the current year but also to those capital costs which were not recovered in previous years. If this is the case then average cost pricing may not be consistent with the fullest possible recovery of capital costs in each years when capacity is fully utilized and may therefore prolong the period of utilization of existing capacity.

In section 5:3 it was explained how most Provincial Administrations obviate the problem of price fluctuations which arise under an average cost pricing policy by charging a more or less constant tariff over the life span of the scheme. The problem of price fluctuations is likely to be particularly acute under a marginalist policy, as explained in the next section, although these fluctuations will occur in an opposite direction to those under a strict average cost pricing policy, with low prices during early years and high prices during the later years of the scheme. Consequently if marginal cost pricing is regarded as the standard of efficiency, the policy of setting a constant tariff will result in water being overpriced and the water scheme underutilized during the early years of the scheme, and in water being underpriced and capital costs underrecovered during the later years of the scheme.

In setting the prices to be charged for water supplied by industrial and domestic schemes the Department is clearly more concerned with the recovery of accounting costs than with signalling to consumers the economic value of the resources used or saved as a result of their decisions. The present departmental practice of calculating the interest charge by applying the Treasury rate to the historical capital cost of each particular scheme means that the difference between the revenue derived from water tariffs and the variable costs of the scheme will not cover the current replacement cost of the capital works of the scheme. This also means that the consumers of water supplied from more recently constructed schemes are likely to be discriminated against in favour of consumers supplied from older schemes. Furthermore the practice of setting the price for consumers supplied by a number of schemes of different ages on the basis of a weighted average of their unit costs implies that consumers are not being charged the incremental costs incurred as a result of their increments in consumption with the result that water is likely to be over-consumed in relation to its relative scarcity. In this respect the recommendation by the Claassen Committee that users should be charged the incremental average cost of the newest scheme is likely to improve allocative efficiency by leading to savings in the resources required for the provision of water.

The implications of the pricing policy for industrial and domestic water schemes on the Government's policy of industrial decentralization will be considered in the next section.

5:4:3 THE IMPACT ON INDUSTRIAL DECENTRALIZATION POLICY OF THE WATER PRICING POLICY FOR INDUSTRIAL AND DOMESTIC CONSUMERS

The departmental practice of setting the price of water supplied to industrial and domestic consumers to recover the full costs of each scheme over its useful life can lead to significant differences in the price of water from scheme to scheme as shown in Table 5:9

TABLE 5:9

AVERAGE PRICE OF WATER ON GOVERNMENT WATER SCHEMES IN REGIONS A TO H IN 1980-81

REGION A	cents/m^3
Berg River - Saldanha Water Project	43,5
Berg River (Swartland Region)	48,3
Berg River (Voelvlei Dam)	2,0
Gamka River (Beaufort West Dam)	6,1
Hex River (Bospoort Dam)	3,3
REGION B	
Upper Molopo	10,4
REGION C	
Caledon - Blóemfontein Pipeline	16,2
Lesaka Regional Water Supply	15,5
REGION D	
Buffalo River (Rooikrantz Dam)	4,3
Kubusi River (Gubu Dam)	11,6

TABLE 5:9 (continued)

REGION D	
Middle Buffalo River (Laing Dam)	19,2
Nahoon River	14,2
REGION E	
Ngagane River	3,8
Sterkspruit (Hammarsdale)	27,6
Umgeni River (Midmar and Albert Falls Dam)	2,2
REGION F	
Blyde River Poort	1,8
Duiwelskloof Regional Water Supply	29,0
Komati River (Nooitgedacht and Vygeboom Dams)	1,8
Usutu River	10,8
REGION G	
Mogol River (Hans Strijdom Dam)	7,9
Pietersburg Regional Water Supply	19,9
Potgietersrus Regional Water Supply	24,6
Springbok Regional Water Supply	44,3
REGION H	
Pienaars River - Hammanskraal (Temba)	13,2
Pienaars River (Roodeplaat Dam)	19,5
Usutu - Vaal (Grootdraai Dam)	8,1
Vaal River and Tugela - Vaal Conduicit	1,8

REGIONAL CLASSIFICATIONS

A: Western Cape
B: Western Transvaal/Northen Cape
C: OFS /Que Que / Bophuthatswana
D: Eastern Cape / Ciskei / Transkei (South)
E: Natal / Kwazulu / Transkei (North)
F: Eastern Transvaal / KaNgwane / Parts of Lebowa / Gazankulu
G: Northern Transvaal / Venda / Parts of Lebowa and Gazankulu
H: PWV area

The following factors appear to cause the price of water to differ from scheme to scheme:

(i) AGE OF THE SCHEME

Since the Second World War there has been a consistent upward pressure on the costs of dam construction with the result that, other things being equal, the price charged for water supplied from more recently completed water schemes tends to be higher than that charged in respect of older schemes. For example the unit cost of water supplied from the original Vaal Dam was estimated at 0,04 cents / m^3 whereas the average unit cost of the water supplied from the more recent extensions to the project was calculated as 3,0 cents / m^3 in 1982. It is apparent, therefore, that if a policy of industrial decentralization results in the water requirements at industrial development points growing beyond the capacity of existing schemes so that new schemes have to be constructed, then there is likely to be a substantial increase in the price of water in these areas.

(ii) EXPECTED LEVEL OF CONSUMPTION

The unit cost of water will clearly decline as the fixed capital costs of a scheme are spread over a greater level of consumption. It follows, therefore, that the price of water will be lower for those schemes supplying a greater volume of water. An inspection of the Department's annual report for 1980-1 reveals that those schemes which are indicated as having a low water charge in Table 5:9 such as the Voëlvlei Dam in Region A, the Midmar and Albert Falls Dams in Region E, the Nooitgedacht and Vygeboom Dams in Region F and the Vaal River and Tugela-Vaal Conduit in Region H have all had to supply a relatively higher annual volume of water. The importance of this factor is particularly evident when one compares the price of 2,2 cents / m^3 for the 64,5 million m^3 supplied by the Midmar Dam with the 27,6 cents / m^3 charged for the 4,6 million m^3 supplied by the Sterkspruit scheme which is also in Region E.

(iii) THE LOCAL BALANCE BETWEEN THE SUPPLY AND DEMAND OF WATER

Even if differences between the age and consumption levels of schemes are taken into account, the price charged to industrial and domestic consumers may still vary from scheme to scheme due to differences in the input requirements for dam construction, water distribution, pumping or water purification. Generally higher unit capital costs are incurred when there are insufficient readily available water resources to meet local water requirements. For example, in Region A (Western Cape) the unit capital cost of water supplied by the Berg River - Saldanha Water Project was R18,28 / m^3 in 1980-81 which was significantly higher than the R0,28 / m^3 incurred for the Voëlvlei Dam. This is reflected by the fact that the price of water supplied by the Saldanha Water Project was 43,5 cents / m^3 whereas that supplied by the Voëlvlei Dam only cost 2 cents / m^3.

It is usually due to a combination of these factors that the price of water at industrial development points is often higher than that charged to industrial and domestic consumers in the main metropolitan areas. For example, Table 5:9 indicates that the average selling price of water supplied by the Laing Dam which serves consumers at the industrial development points of East London, Berlin and King WilliamsTown in Region D is 19,2 cents / m^3 whereas the charge for water supplied by the Vaal - Tugela River Project which serves the Witwatersrand area is 1,8 cents / m^3. Although the Department does provide for a subsidy on a sliding scale if the unit cost of water exceeds 22c / m^3 for towns with a population of more than 5000 White inhabitants, it appears to be the case that many industrial development points do not qualify for this subsidy. In the last two chapters it was shown that the Decentralization Board compensates firms located at industrial development points for the long term locational disadvantage they suffer as a result of having to pay higher costs for electricity and railage than would be incurred if they were located in the main urban areas. No such subsidy is, however, available to such firms in respect of higher water costs. This is probably due to the fact that the cost of water usually forms a small percentage of the sale value of products even for industries that use relatively large quantities of water. This is indicated in Table 5:10 which shows the cost of water as a percentage of the market value of product for a number of water intensive industries.

TABLE 5:10

COST OF WATER AS A PERCENTAGE OF MARKET VALUE OF PRODUCT

PRODUCT	PERCENTAGE
Steel	0,6
Paper Pulp	3,0
Cotton Textiles	1,16
Electricity	3,0
Oil from Coal	1,61

It is apparent, therefore, that a relatively higher charge for water is not likely to result in a serious locational disadvantage for firms located at industrial development points and there is probably not a strong case for an industrial decentralization policy package to make use of a water subsidy.

The availability of adequate supplies of water at industrial development points is, however, likely to be an important consideration in the formulation of an industrial decentralization policy since inadequate supplies may be a limiting factor to industrial development. It is for this reason that the areas chosen as decentralized growth points usually have an established water supply system. If industrial growth takes off at these points then it is likely that water requirements will grow beyond the capacity of existing schemes. It is for this reason that the water authorities must ensure that their programme of water resource development takes account of the expected effects of industrial decentralization policy.

In the next section some of the practical difficulties associated with marginal cost pricing will be discussed as well as the methods by which tariff schedules can be structured to approximate optimal pricing principles.

5:5 THE IMPLEMENTATION OF OPTIMAL WATER PRICING POLICIES

5:5:1 METERING WATER CONSUMPTION

If a water authority decides to use the price of water as an instrument to adjust consumption to the level necessary to preserve a balance between water supply and demand this means that it must be able to charge consumers for each unit of water consumed. To do this, a system for metering water consumption must be installed. The main benefit to be derived by the water authority from metering arises from the fact that when there is excess demand under a flat rate pricing policy, the introduction of metering and the raising of unit prices to ration demand to the level of existing capacity may be viewed as an alternative measure to implementing an investment programme to expand capacity to meet the requirements of water users. Once it is accepted that water is an economic good which may be sold at a price per unit in the same way as other commodities, there are usually two main objections to the installation of metering:

(a) the quantity of water demanded is unlikely to be markedly reduced by the introduction of a commodity charge since the demand for water is highly price-inelastic;

(b) the installation of metering results in a temporary fall in consumption, followed by a fairly rapid return to the original level.

The argument that metering will have a negligible impact on water consumption is essentially an empirical matter which appears to be contradicted by the results of a number of studies conducted overseas, particularly in the United States.[41,42,43]

It appears, then, that metering is likely to result in a substantial reduction in the level of water consumption although the extent to which this occurs will depend on the elasticity of demand for the different uses of water. Certainly there is nothing to suggest that water consumption is perfectly price inelastic and will not be affected at all by metering.

Empirical evidence also appears to contradict the "generally accepted hypothesis ... that water meters will initially reduce the quantity of water consumed, but that their original impact will gradually wear off and consumers will begin consuming more water until a new equilibrium is reached at a level that approaches the original flat rate use."[44] Hanke has surmised that the misconception that metering has no long-term effect on water consumption might be based on a confusion between the demand for water as a functional relation at a point in time and as a trend over time. This is illustrated in Figure 5:8. The demand for water for residential uses in 1980 is represented by D 1980. If flat rates exist the quantity of water demanded will be Qfr. The installation of meters in 1980 will reduce the quantity demanded from Qfr to Qm when the metered rate is Pm. By 1983, the demand function has shifted to the right as

---

41. Hanke, S.H., "Demand for Water Under Dynamic Conditions", Water Resources Research, 1970, pp. 1255-1261.
42. Cloonan, E.T., "Meters Save Water" in Modern Water Rates, New York, Battenheim Publishing Company, 1965.
43. Howe, C.W., and Linaweaver, F.F., "The Impact of Price on Residential Water Demand and Its Relation to System Design", Water Resources Research, No. 1, 1965.
44. Hanke, S.H., op.cit., p. 1254.

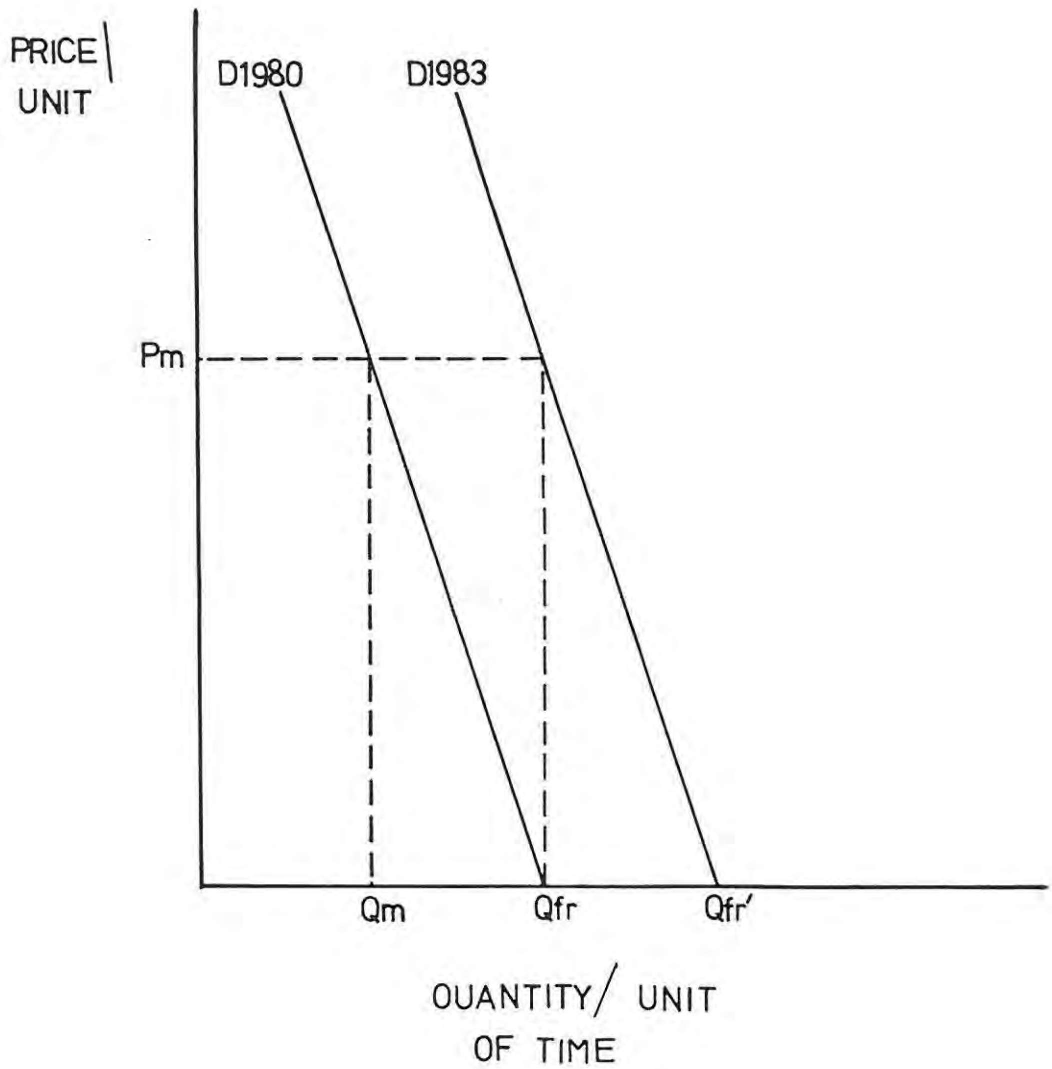


FIGURE 5:8 The Change in Water Demand Over Time

represented by D 1983. This shift could have been caused by changing tastes, increased incomes, population increases, alterations in habits, or changes in other parameters of the demand function. If metered rates are maintained at Pm the quantity of water demanded will increase from Qm to Qfr.

The increase in water used from 1980 to 1983 should not be regarded as evidence that metering is ineffective after three years. If flat rates were again imposed in 1983, the quantity of water demanded would equal Qfr which is considerably greater than the 1983 use under metered rates. Hanke thus attributes the argument that metering has only a temporary effect on water consumption to a failure to recognize the elementary economic principle that price changes induce movements along a demand function, whereas other factors cause the locus of the demand function to change.

If the water authorities accept that metering will be an effective tool for restraining water consumption, they should then carry out a cost-benefit analysis to determine whether there will be a net economic benefit to be obtained from the installation of meters. The main benefit to be obtained from metering is that by encouraging economy in water use it delays the implementation of investment programmes designed to expand the capacity of the water supply system. This is shown in Figure 5:9 in which the growth of consumption in the absence of metering is represented by curve C1, while C2 shows the consumption growth path if metering already exists and an optimal pricing and investment policy is being pursued in which capacity is adjusted to its optimal level in each period so that price is always equal to long run marginal cost. Now if metering is introduced at time tc, at which time it is assumed that existing capacity can supply no more than OA units of consumption, price will be raised from zero to a level equal to short run marginal cost and consumption will fall from OA to OB. Consumption will not fall to OC since if metering is installed and consumption falls, there will temporarily be excess capacity. The only opportunity costs now involved are running cost so that according to optimal pricing principles price will be set equal to short run marginal cost which is less than long run marginal cost. Price will thus be lower, and consumption higher, than if metering had always been in operation. The dotted line in Figure 5:9 represents the new consumption path. As demand grows over time, consumption is shown to increase from OB, until at t5 consumption returns to the level OA and capacity is again being fully utilized. Investment in new capacity will not take place until the price consumers are willing to pay covers both short run and long run marginal cost. This will occur at t 15 , since curve C2 shows the level of consumption if an optimal pricing and investment policy is being followed. Consumption will then continue to follow C2 until t20 which denotes the end of the meters life. If the meters are not replaced at t20 consumption will return to the original unmetered growth path.

At time to, existing capacity could supply no more than OA units of consumption, so the effect of metering in this example is to defer the necessity for additional investment to t10. What actually happens is that a series of investment projects which would have been undertaken in the absence of metering are now delayed for a number of years. The economic benefit to be obtained from this deferment can be measured as the difference between the present worth of the series in the absence of metering and the present worth of the series if it is delayed by the introduction of metering. This benefit should be compared with the costs of metering which can be measured as the sum of:

(i) the costs of installing, maintaining and operating the meters themselves; and

(ii) the loss of consumer's surplus resulting from the reduction of consumption below its flat-rate level.

The recommendation by the Viljoen Commission that "steps be taken to have all domestic and industrial consumers individually metered

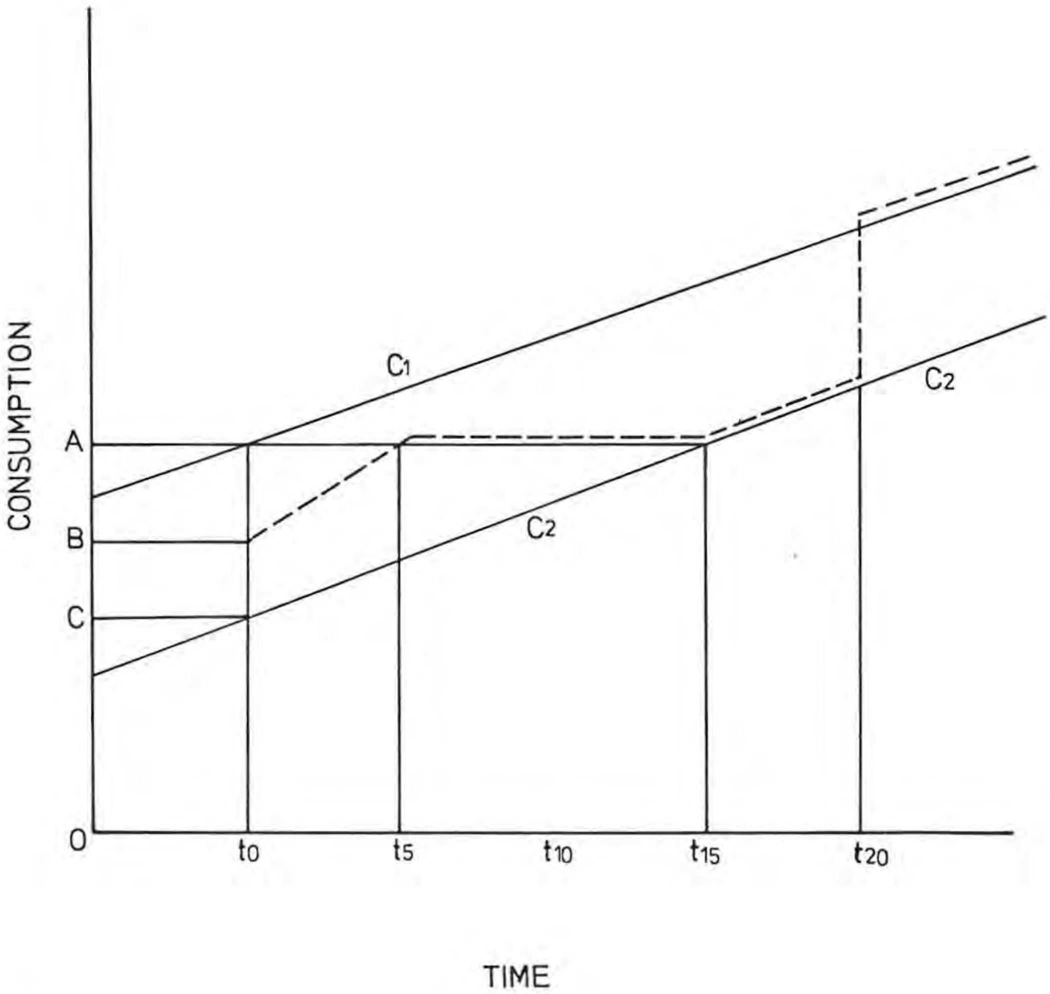


FIGURE 5:9 The Effect of Metering on the Time Path of Consumption

and charges levied according to the quantity used"[45] is not strictly economically valid since the metering decision should according to widely accepted economic principles, be based on a cost-benefit analysis and there may clearly be situations where it is uneconomic to introduce metering. However, as Warford[46] points out:

> "A decision not to install meters at time to does not preclude doing so at a later date; thus it may be better to defer metering a number of years, while building up extra capacity for the intervening period."

At present industrial and domestic water consumption is metered throughout South Africa. However, in the case of flatdwellers, the usual practice is for the apartment house owner to pay for water so that there is no incentive for flatdwellers to conserve water. It is therefore suggested that a cost-benefit analysis should be undertaken to determine whether it is economically desirable for meters to be fitted to individual apartments so that each flatdweller is responsible for his own water raters. Water supplied for agricultural uses is not metered since as has been explained, it is largely rationed by means of the quota system.

5:5:2 THE PROBLEMS ASSOCIATED WITH MARGINAL COST PRICING

Once a system of metering water consumption has been installed, the water authority must decide on the pricing policy on which it will base its charges per unit of water used. The main problems associated

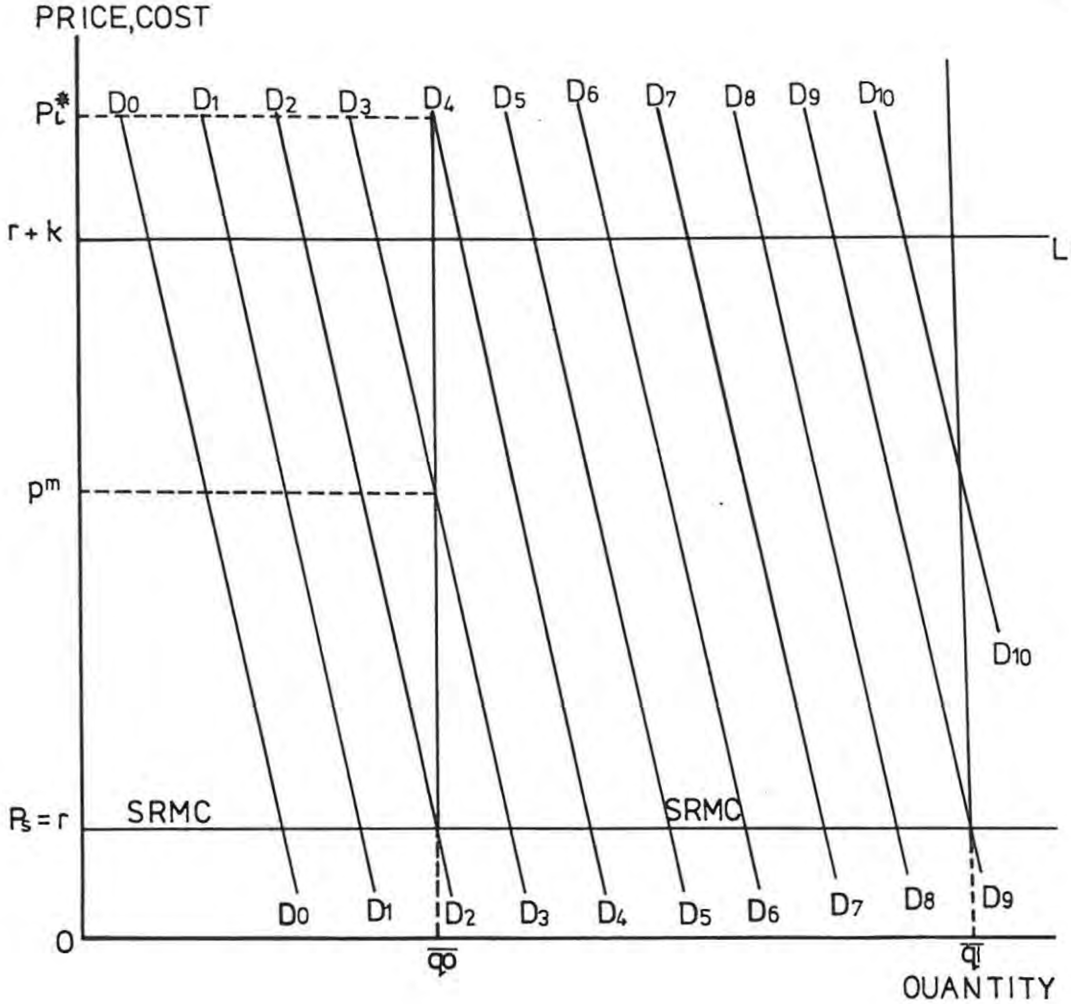
---

45. R.S.A., op.cit., p. 22.
46. Warford, J.J., op.cit., p. 105. In the appendix to his paper Warford conducts a formal analysis of the timing of the introduction of meters.

with the application of marginal cost pricing and investment rules arise because investment in water supply projects can usually only be undertaken in large indivisible lumps and that running costs are often a very small proportion of fixed capital costs for most water schemes. If water consumption is expected to grow at a steady rate then the application of marginal cost pricing in the presence of indivisibilities and a low running to capital cost ratio is likely to result in dramatic price fluctuations as shown in Figure 5:10 It is assumed in this figure that there are constant returns to scale and fixed capacity which can only be increased in an indivisible lump from $\overline{qo}$ to $\overline{q1}$.

The demand curves for the period from to to t10 are represented by Do to D10. It can be seen that during the period to to t2, there is excess capacity and opportunity cost is equal to running costs, r. According to optimal pricing principles price is set equal to short run marginal cost which in this case is equivalent to unit running costs. From period t2 to t4 price is allowed to rise from ps to pl to limit demand to the fixed capacity output $\overline{qo}$. It can be seen that where there are indivisibilities price is allowed to rise higher than unit running and capital costs, since where there are indivisibilities the undertaking will not always expand capacity once it has started to earn a surplus since there is likely to be excess capacity after the new investment is completed. It is assumed that after period t4 there is a net economic benefit to be obtained from the expansion of capacity from $\overline{qo}$ to $\overline{q1}$. During the period t4 to t9 there is thus once again excess capacity and it can be seen that if the short-run marginal cost pricing rule is applied, price should be dropped from pl back to ps where it is equal to running cost. It can thus be seen that where there are indivisibilities and running costs are a small proportion of total costs, a marginal cost pricing policy is likely to result in dramatic cyclical fluctuations in the price of water according to the degree of utilization of the water scheme.

Fluctuations in the price of a public utility such as a water authority which supplies a broad range of users are usually considered undesirable for the following reasons:


PRICE,COST
P_l^*
$r+k$
p^m
$P_s = r$
0
D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10
LRMC
SRMC
SRMC
D10
D0 D1 D2 D3 D4 D5 D6 D7 D8 D9
$\overline{q_0}$
$\overline{q_1}$
QUANTITY

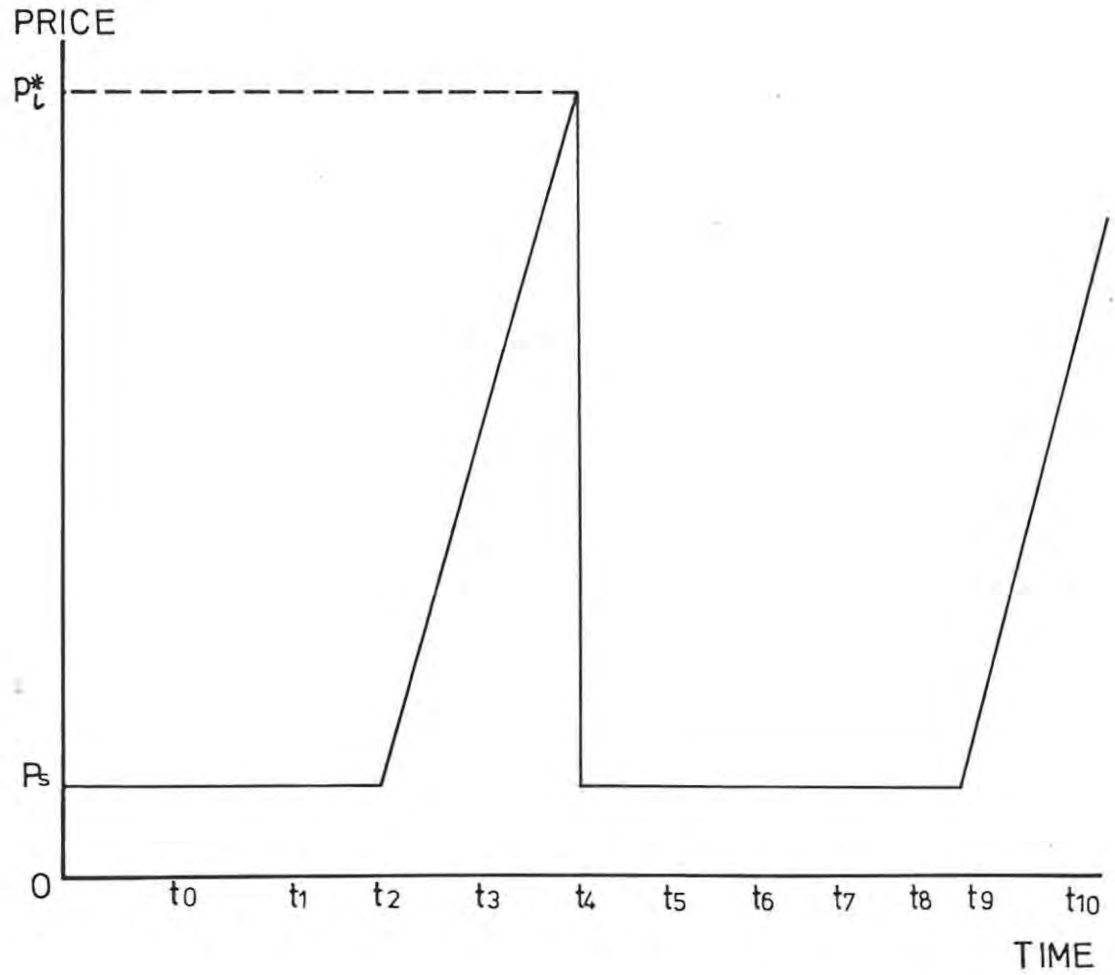

PRICE
P_l^*
P_s
0
t0 t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
TIME

FIGURE 5:10

(i) an additional element of uncertainty is introduced into the plans of water users;

(ii) water users may have to consider making periodic adjustments to changing water prices which, in certain cases, may be costly. For example, Triebel has pointed out that "higher irrigation water tariffs would either lead to higher food prices or force irrigation farmers out of the market. Neither is acceptable as the former will hit the poor, whereas the latter would remove that sector of agriculture which stabilizes food production during periods of drought"[47];

(iii) large deficits will be incurred during periods when there is excess capacity and the water authority sets price equal to running costs. These deficits will need to be financed by short term loans which will be repaid once capacity is fully utilized and price rises to ration demand to capacity. Given the low proportion of the total unit costs of a water scheme which are usually made up of running costs, it follows that the short term financing requirements will be substantial.

A further problem in implementing marginal cost pricing involves the accurate determination of the marginal cost and of the cost increment to be used. There are likely to be differences in the marginal cost of supplying the different groups of users served by a single water scheme and these differences should be reflected in the tariff structure of the scheme. If, for example, a uniform rate is charged to all the users of a scheme then as Baumann and Dworkin point out "customers who reside in high-density areas closest to the water scheme subsidize those who reside in low-density areas farthest

---

47. Triebel, C., "Water resources", Financial Mail, April 1 1983, p. 19.

from load centers. This is the result of the costs of extra pumping and piping required to serve remote areas."[48]

Although it may be impractical to implement a water pricing policy whereby each user is charged exactly the marginal cost of supplying him, it is nonetheless possible for marginal cost pricing principles to be reflected in the tariff policy of water authorities. If it is assumed that there is a steady growth in water consumption over time as reflected in Figure 5:10 then the price charged for water should be allowed to change as the degree of capacity utilization changes over the life of a water scheme:

(i) during the early stages of a scheme when demand has not yet grown to the level where capacity is fully utilized price should be set at a relatively lower level to promote water consumption since the cost of supplying additional consumption is likely to be relatively low; while

(ii) once capacity is fully utilized price should be allowed to rise to induce greater water conservation and delay the necessity for further investment.

Generally, then, price should be allowed to follow an upward trend over the life of a water scheme, in order to encourage the most efficient use of the given water resources.

In setting the charge per unit to be levied on metered users, a water authority also has to consider how the charge per unit should be varied as the user increases his consumption. There are basically four types of price schedules which can be used when water consumption is metered.

---

48. Baumann, D., and Dworkin, D., "Water Resources for our Cities", Resource Papers For College Geography, Illinois, 1978, p.

(a) Declining block rate or promotional pricing

This schedule uses a minimum charge for a small initial quantity followed by lower unit prices as the volume of water purchased increases in discrete ranges. This is the most common type of pricing policy applied in the United States although it is not used in South Africa. Declining block rates may be introduced to increase the revenue a water supplier derives from a given quantity of water sold. The increase in producer revenue is extracted from consumer surplus with this pricing policy since the price a consumer pays for each block of consumption will bear a closer relation to his willingness to pay than would be the case if a constant unit price was charged.

The main problem with declining block rate pricing is that it may encourage wasteful patterns of water use since the average cost of water to the consumer will decrease with increasing consumption. In the long term the growth of water consumption may be greater under a declining block rate schedule then if a constant unit price is charged. This type of price schedule should only be introduced in areas where there are potential abundant water resources and it is expected that there will be a low probability of water shortages occuring in the future.

(b) Constant unit price schedule

Under this method of pricing, a constant rate is charged for each unit of consumption. With this price schedule, then, no attempt is made to influence the consumption behaviour of individual users although, quite clearly, the constant price can be varied upwards or downwards to accomplish the pricing objectives of the water authority.

(c) Incremental block pricing

This is the opposite of declining block rate pricing since it is characterized by succeedingly higher unit prices in discrete ranges of quantities consumed as the amount of water purchased increases. Incremental block pricing is principally intended to encourage water conservation as was shown in section 5:4:1 where the application of this policy by certain irrigation schemes in South Africa was evaluated. The fact that the average cost of water to the individual user rises with increasing consumption may induce users to adopt economical patterns of water use which, in the long term, lead to a slower growth of water consumption than would be the case with the other pricing policies. Incremental block pricing may thus be a particularly useful pricing policy for a water authority to adopt areas suffering from the problem of aridity in the sense described in section 5:1:4.

(d) Peak period differential rates

Peak period differentiation may be introduced into the water pricing schedule where there is a seasonal irregularity in water. For example, if the demand for water for lawn sprinkling is higher in summer then winter, a higher rate may be charged for water in summer in order to lessen peak demands. Summer differential rates may be used in conjunction with incremental block pricing to encourage water conservation.

In choosing between the different types of water price schedules, a water authority should also consider their distributional effects. The constant price policy may seem to meet the test of complete fairness since under this policy all users pay the same price at all times. Incremental block pricing is more progressive, however, since with this policy low use (and probably low income) consumers will be charged a lower average unit price than high use consumers. This seems to reflect the payment by ability-to-pay criterion of distributional policy and may thus be considered fairer than either a constant price policy or the regressive declining block rate pricing policy under which low use consumers are charged a higher average rate than high use consumers.

Gysi and Loucks[49] have compared the long run effects of these pricing policies in terms of consumer benefits and system costs. They have assumed that both the declining block rate and increasing block rate pricing policies have three block rates and that the final block rate in both cases is the same as that charged under a constant price policy. Furthermore they assumed that prices were so low that all consumers bought their water in the third price range of the increasing and decreasing block rate policies which equalled the constant rate charged under the constant price policy. As a result the optimal expansion paths and rationing figures remained identical for corresponding runs of all three policies, so that Gysi and Loucks could base their comparison on the average price paid by consumers under the three different policies. It can be seen that the increasing block rate policy resulted in higher average and minimum percentage of summer demand received and lower long run average unit cost and present value of capital consturnction cost for the same long run average price then the constant price policy. Conversely the declining block rate policy has the opposite effects. Gysi and Loucks thus conclude:

49. Gysi, M., and Loucks, D.P., "Some Long Run Effects of Water Pricing Policies", Water Resources Research, Vol. 7, No 6, Dec. 1971.

> "For comparable long run average prices the increasing block rate policy appeared superior to the constant price policy which in turn appeared superior to the decreasing block rate policy." [50]

It can thus be concluded that there are both distributional and conservational advantages to be obtained from a tariff structure characterized by increasing block rates and a seasonal rate differential.

5:6 CONCLUSION AND RECOMMENDATIONS

The recommendations for the pricing of water in South Africa which emerge from the analysis in this chapter may now be summarized:

(i) since there is evidence that water resources are becoming increasingly scarce in South Africa, there is a need to use pricing policy not only as an instrument to raise revenue but also as a tool for restraining water demand. This however, makes it necessary for the metering of water consumption to be used as widely as is feasible in this country;

(ii) the recommendation by the Viljoen Commission that water tariffs should be raised to at least cover the operating costs of irrigation schemes should continue to be implemented since although there is evidence of improvement, the water supplied by government irrigation is still substantially underpriced in South Africa;

(iii) a more efficient utilization of water capacity may be achieved if the unit price of water is allowed to increase over the life of a water scheme than under the current policy of charging a constant tariff;

(iv) the practice of charging individual users on an increasing block rate basis which is applied by some government

---

50 *Ibid*., p. 1381.

irrigation schemes should be extended not only to other irrigation schemes but also to the water supplied for industrial and domestic uses, since this type of pricing policy tends to encourage water conservation by consumers. The possibility of introducing seasonal price differentiation should also be investigated since this may be an effective way of curbing demand in peak periods;

(v) the possibility of extending the departmental practice of providing a subsidy on capital cost according to a sliding scale if the unit cost of water supplied by a scheme exceeds a certain specified limit should be investigated, since this may be a way in which water pricing policy can be made consistent with industrial decentralization policy; and

(vi) although the price of water at industrial development points is often higher than that charged to industrial and domestic consumers in the main metropolitan areas, this is not likely to result in a serious locational disadvantage for firms located at these points so that there is not a strong case for an industrial decentralization policy package to make use of a water subsidy. The availability of adequate supplies of water may, however, be a limiting factor to industrial development so that it is important for the water authorities to ensure that their programme of water resources development takes account of the expected effects of industrial decentralization policy.

CONCLUSION

In Chapters One and Two a theoretical framework for public utility pricing was formulated in fairly general terms, with the object of applying a common set of principles to the determination of efficient pricing policies for rail transport, electricity and water services. Having examined both present and proposed certain alternative pricing policies for these services, the thesis can now be concluded by pointing to some of the differences and similarities which emerged from the preceding analysis.

Unlike electricity and rail transport, water services cannot be provided by public enterprises operating on a national scale. The pricing policies of the various local water suppliers are, however, controlled by the Directorate of Water Affairs. Despite certain institutional similarity between S.A.T.S. and Escom, the pricing policies followed by these two public enterprises differ substantially because they are required to promote different objectives. Whereas Escom sets its tariffs to generate sufficient revenue to cover the items of total "cost" specified in the Electricity Act, S.A.T.S. is not only required to attain financial equilibrium as an organization but also to provide various uneconomic social services. S.A.T.S. is therefore obliged to depart from the principle of full cost recovery in charging for some of its remunerative services in order to cross-subsidize its unremunerative social services. Although the users of electricity are liable to such cross-subsidization, Escom does set its tariffs at a level sufficient to ensure that its users make a contribution to future capital development since the total "costs" which must be covered by tariff revenue include items which would be regarded as retained profits is an ordinary commercial undertaking. The users of water supplied by individual schemes also do not have to cross-subsidize other schemes since the price charged for water supplied to industrial and domestic consumers is set to recover the full accounting costs of a scheme over its useful life, while the deficits incurred by irrigation schemes are directly subsidized by the State.

Escom also differs from S.A.T.S. in that, in formulating its tariffs, it follows a procedure of allocating its costs first to different regions, and then to different classes of consumers within each region. Electricity tariffs are thus structured primarily according to differences in the cost of supplying different classes of consumers. In the same way the price of water varies from area to area according to the differences between the unit costs incurred by the local water schemes. S.A.T.S., on the other hand, applies the "collective principle" according to which the same rate per tonne kilometer is charged for a commodity regardless of the route on which it is carried. This effectively implies that remunerative routes cross-subsidize unremunerative routes.

Furthermore, the structure of tariffs for different commodities is determined not only according to differences between their costs of conveyance but also takes into account differences in their value per unit of weight. This implies that the revenue earned on commodities with a high value is used to cross-subsidize commodities with a low value. As was explained in Chapter Three this policy of cross-subsidization between different services, routes and commodities largely arose because S.A.T.S. is required to reconcile different policy objectives.

It might, at first, seem that if the pricing policies of the different public utilities were required to promote the objective of economic efficiency by following the optimal rules discussed in Chapter Two, then the differences between them would largely disappear. However, as was explained in Chapter Two, these pricing rules substitute for the functions normally performed by a competitive market process and the way in which they should be applied will depend upon the competitive pressures facing the different utilities. Although S.A.T.S. can be regarded as a monopolistic supplier of rail services, it does face competition from other modes of transport. It currently enjoys a measure of protection from such intermodal competition through the permit system. As was explained in Chapter Three, the permit system interferes with users' freedom of modal choice and may be causing a misallocation of resources within the transport sector. The objective of economic efficiency could therefore by promoted if the transport sector were deregulated. This would, however, probably necessitate some relaxation of the controls over the pricing and investment policy of S.A.T.S. to enable it to compete on equal terms with other modes of transport by, for example, being able to negotiate rates with users below the published scales and to discontinue unremunerative services. The Railways could apply marginal cost principles by negotiating lower rates to improve the utilization of capacity which is underutilized in off-peak periods and on low density routes, and higher rates to ration demand to capacity in peak periods and on high density routes. The combined effects of deregulation, rate negotiation and a flexible investment policy by S.A.T.S. should then improve the allocation of resources within the transport sector.

The suppliers of electricity and water are not subject to the same degree of competitive pressure as that which faces the Railways. The optimal pricing rules can therefore be used as instruments of decentralized control to ensure that resources are efficiently allocated to these sectors. The adoption of these rules would actually required a tightening of control over Escom's pricing and investment policies. An independent external body would have to control the rates of return and the level of investment undertaken by Escom since there might be a tendency for Escom's management to be excessively risk averse and to over-invest in reserve plant capacity. Furthermore the ambt of external control over Escom's tariff policy would have to be extended so that Escom could no longer impose surcharges and discounts at its own discretion. Escom's tariffs could also be structured according to marginal cost principles in the same manner as has been applied in France. In the case of water services, there already exists an adequate system of external control over pricing and investment decisions. There does, however, exist scope for structuring water tariffs according to marginalist principles by extending the application of incremental block rates and generally raising the price of water on irrigation schemes to more economic levels.

Finally, it appears that industrial decentralization would be more affected by regional differences in the availability than in the cost of public utility services. This is particularly the case with electricity and water since inadequate supplies of these servies is likely to be a factor limiting industrial development at any point. Although an adequate transport system is also necessary for industrial development to take place, regional transport needs can be met by road as well as rail transport. It is for this reason that it has been proposed that Regional Transport Authorities (RTA) should be established to monitor regional transport needs where the pursuit of an efficient investment policy by S.A.T.S. would result in the closure of unremunerative lines. The R.T.A. could determine, in such cases, whether there might be a net economic benefit to be obtained from subsidizing the continued operation of rail services.

It is likely that the pricing policies recommended for electricity, water and rail services would lead to regional differences in the cost of these services. In cases where this would cause a long term cost disadvantage for firms locating at industrial development points they could be compensated in the manner currently followed by the Board for the Decentralization of Industries. This approach reduces the conflict between industrial decentralization policy and the promotion of efficient pricing policies by public utilities in South Africa, since it recognizes the general principle that instruments and objectives should balance in the application of economic policy.

BIBLIOGRAPHY

1. Abouchar, A., Transportation Economics and Public Policy with Urban Extensions, J.Wiley and Son, Toronto, 1977.

2. Aharoni, Y., "Managerial Discretion" in Aharoni, Y., and Vernon, R., eds., State Owned Enterprise in the Western Economics, Croom Helm, London, 1981, pp. 184-193.

3. Alonso, W., "Location Theory" in Alonso, W., and Friedmann, J., Regional Development and Planning: A Reader, Massachussets Institute of Technology Press, 1964, pp. 78-106.

4. Archibald, G.C., and Allingham, M.G., "Second Best and Decentralization", Journal of Economic Theory, 1975, pp. 157-173.

5. Arrow, K.J., "On Finance and Decision Making", in Aharoni, Y., and Vernon, R., eds., op.cit., pp. 63-69.

6. Balassa, B., Policy Reform in Developing Countries, Pergamnon Press, Oxford, 1977.

7. Bator, F.M., "The Anatomy of Market Failure", Quarterly Journal of Economics, August 1958, pp. 351-79.

8. Baumann, D. and Dworkin, D., "Water Resources in Our Cities", Resource Papers for College Geography, Southern Illinois University, Washington, 1978.

9. Baumol, W.J. and Bradford, D.F., "Optimal Departures from Marginal Cost Pricing", American Economic Review, Vol. 60, No. 3, June 1970, pp. 265-283.

10. ________ , et.al., "The Role of Cost in the Minimum Pricing of Railroad Services", Journal of Business, vol. 35, pp. 357-66.

11. ________ , Welfare Economics and the Theory of the State, G.Bell and Sons Ltd., London, 2nd. ed., 1965.

12. Berrie,T.W., "The Economics of System Planning in Bulk Electricity Supply", in Turvey,R., ed., Public Enterprise, Penguin Harmondsworth, 1968.

13. Black,P.A., Theories of Economic Development: A General Equilibrium Analysis, Unpublished Phd. thesis, Rhodes University, December 1981.

14. Blauwens,G.J., "The Spatial Theory of the Demand for Transport: A Survey", in Polak, J.B., and van der Kamp,J., (ed.s), Changes in the Field of Transport Studies: Essays in Honour of Prof. Dr J.P.B. Tissot van Patot, Marthinus Nijhoff Publishers, The Hague, 1980, pp. 14-25.

15. Blonk,W.A.G., ed., Transport and Regional Development: Introduction, Saxon House, Farnborough, 1979.

16. Boiteux,M., "The 'Tarif Vert' of Electricite` de France" in Nelson, J.R., ed., Marginal Cost Pricing in Practice, Prentice-Hall, New York, 1964, pp. 124-149.

17. Boiteux,M., and Stasi,P., P., "The Determination of the Costs of Expansion of an Interconnected System of Production and Distribution of Electricity", in Nelson,J.R., ed., op.cit., pp. 95-123.

18. Bonem, G.W., "On the Marginal Cost Pricing of Municipal Water", Water Resources Research, Feb., 1968, vol. 14 No.1, pp. 191-3.

19. Bonnafous, A., "Underdeveloped Regions and Structural Aspects of Transport Infrastructure", in Blonk, W., ed., op.cit., pp. 45-62.

20. Borts, G.H., "Production Relations in the Railway Industry", Econometrica, January 1952.

21. ________, "The Estimation of Rail Cost Functions", Econometrica, vol. 28, January 1960.

22. Boulding, K.E., "Welfare Economics", in Haley, B.F., ed., A Survey of Contemporary Economics, R.D.Irwin, Illinois, 1952.

23. British Electrical and Allied Industries Research Association, An Improved Method of Allocating to Classes of Consumers the Demand Related Portion of the Standing Cost, Report No. K/T 109, 1945.

24. Buchanan, J.M. and Stubblebine, W., "Externality", Econometrica, vol. 29, 1962, pp. 371-84.

25. Cassese, S., "Public Control and Corporate Efficiency", in Aharoni, Y., and Vernon, R., eds., op.cit., pp. 145-156.

26. Collard, D., Prices, Markets and Welfare, Faber and Faber Ltd., London, 1972.

27. Comanor, W.S., and Leibenstein, H., "Allocative Efficiency and X-Inefficiency and the Measurement of Welfare Losses", Economica, August 1969, pp. 304-9.

28. Crew,M.A., "Electricity Tariffs", in Turvey,R., op.cit., pp. 258-283.

29. Crew,M.A., and Kleindorfer, P.R., Public Utility Economico, Macmillan, 1979.

30. Crew,M.A. and Rowley,C.K., "On Allocative Efficiency, X-Efficiency and the Measurement of Welfare Losses", Economica, May 1971, pp. 199-203.

31. Cyert,R.M., and March,J.G., A Behavioural Theory of the Firm, Prentice-Hall, 1963.

32. Davis,O., and Whinston,A., "Externalities, Welfare, and the Theory of Games", Journal of Political Economy, vol. 70, June 1962.

33. ________, "Piecemeal Policy in the Theory of Second Best", Review of Economic Studies, July 1967.

34. Davis,S., Mullins,D., Louw,T.J., and Botha,J.L., The Financing of Transport Infrastructure in South Africa - History and Outlook for the Future, Paper delivered at the Annual Transport Convention, Pretoria, August 1982.

35. Demsetz,H., "Why Regulate Utilities?", Journal of Law and Economics, 1964.

36. Departement van Omgewingsake, Verslag van die Komitee van Ondersoek na die Prysbeleid met betrekking tot die Bepaling van Watertariewe, Pretoria, March 1982.

37. Department of Foreign Affairs and Information, The Promotion of Industrial Development: An Element of a Co-ordinated Regional Development Strategy for South Africa, S.A. Digest, April 2 1982.

38. Department of Water Affairs, Annual Reports , April 1977 to March 1981.

39. Du Toit, P.J.D., Economic Principles Involved in Transport Subsidies , Paper delivered at the Annual Transport Convention, Johannesburg, July 1983.

40. ______ , Framework for the Formulation of National Transport Policy , Unpublished paper, National Institute of Transport and Road Research, March 1982.

41. Electricity Supply Commission, Annual Reports , 1977-1982, Megawatt Park, Sandton.

42. Farrell,M.J., "In Defence of Public Utility Price Theory", Oxford Economic Papers, vol. 10, 1958, pp. 109-123.

43. Floor,B.C., Considerations Underlying the Formulation of a National Transport Policy for South Africa, Unpublished paper, Transport Research Centre, University of Stellenbosch, April 1982.

44. Freeman,P.N.W., The Recovery of Costs from Road Users in South Africa, unpublished D. Comm. Thesis, Unisa, December 1981.

45. Graaff,J. de V., Theoretical Welfare Economics, Cambridge University Press, 2nd ed., 1963.

46. Gysi,M., and Loucks,D.F., "Some Long Run Effects of Water Pricing Policies", Water Resources Research, vol. 7, No. 6, December 1971, pp. 1371-1382.

47. Hamerslag, R., "Spatial Development: Developments in Traffic and Transportation and Changes in the Transport System", in Polak,J.B., and van der Kamp,J., eds., op.cit., pp. 60-80.

48. Hanke,S.H., "Demand for Water Under Dynamic Conditions", Water Resources Research, October 1970.

49. Harberger,A.C., "Three Basic Postulates for Applied Economics", Journal of Economic Literature, 1971.

50. Hay,D.A. and Morris,D.J., Industrial Economics: Theory and Evidence, Oxford University Press, 1st ed., 1979.

51. Henderson,A.M., "The Pricing of Public Utility Undertakings" in American Economic Association, Readings in Welfare Economics, Allen and Unwin, London, pp. 541-560.

52. Hermann,G.F.K., Electricity Tariffs, unpublished D. Comm. Thesis, University of Stellenbosch, 1974.

53. Heydenrych, J.A., and Kennedy, T.L., "Economic Efficiency of Rail Transport", paper delivered at the Annual Transport Convention, July 1983, Johannesburg.

54. Hirschleiffer,J., Price Theory and its Applications, Prentice-Hall, New Jersey, 2nd ed., 1980.

55. Holtgrefe,A.A.I., "Towards a Rational Model for Passenger Transport Policy", in Polak,J.B. and van der Kamp,J., ed.s, op.cit., pp. 89-97.

56. Hotelling,H., "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates", in American Economic Association, op.cit., pp. 280-324.

57. Howe,K.M., and Rasmussen,E.F., Public Utility Economics and Finance, 1st ed., Prentice-Hall Inc., Eaglewood Cliffs, New Jersey, 1982.

58. International Bank for Reconstruction and Development, World Bank Report 1983, Oxford University Press, July 1983.

59. Joy, S., "British Railways Track Costs", Journal of Industrial Economics, vol. 13, 1964.

60. ________, "Pricing and Investment in Railway Freight Services", Journal of Transport Economics and Policy, September 1971, pp. 231-246.

61. Kennedy, T.L., Costs of Road and Rail Transport, Technical Note TT 23/82, National Institute of Transport and Road Research, Pretoria, 1982.

62. ________, Economic Allocation of Transport Resources, paper delivered at the Annual Transport Convention, Johannesburg, July 1983.

63. ________, Economic Analysis of Goods Transport by Road and Rail Study Proposal, Technical Note TT/7/82, National Institute of Transport and Road Research, February 1982.

64. Klaassen, L.H., "Optimum Use of Transport Networks and its Possible Contribution to Welfare Maximization", in Polak, J.B. and van der Kamp, J., op.cit., pp. 75-88.

65. ________, and Bourdrez, J.A., "Integrated Transport Planning", in Blonk, W.A.G., op.cit., p. 63-77.

66. Koutsoyiannis, A., Modern Microeconomics, 2nd ed., Macmillan, 1979.

67. Krutilla, J.V., and Eckstein, O., Multiple Purpose River Development, John Hopkins Press, 1958.

68. Leftwich, R.H., The Price System and Resource Allocation, Holt, Rhinehart and Whinston, New York, 1960.

69. Leibenstein, H., "Allocative Efficiency vs. X-Efficiency", American Economic Review, vol. 56, 1966, pp. 392-415.

70. ________, The General Theory of X - Efficiency and Economic Underdevelopment, Oxford University Press, 1978.

71. Lerner, A.P., "On Optimal Taxes with an Untaxable Sector", American Economic Review, 1970, pp. 284-294.

72. Lewis, W.A., Overhead Costs, Allen and Unwin, 1949.

73. L'Huillier, D. and Domonach, O., "Transport and Regional Policy in France", in Blonk, W.A.G., ed., op.cit., pp. 121-140.

74. Lintner, J., "Economic Theory and Financial Management", in Aharoni, Y., and Vernon, R., eds., op.cit., pp. 8-23.

75. Little, I.M.D., A Critique of Welfare Economics, 2nd ed., Oxford University Press, 1957.

76. Littlechild, S.J., "The Fallacy of the Mixed Economy", Hobart Paper No. 40, Institute of Economic Affairs, London, June 1978.

77. Loubser, J.G.H., "Transport and Regional Policy in South Africa", in Blonk, W.A.G., ed., op.cit., pp. 241-255.

78. McCrone, G., Regional Policy in Britain, Allen and Unwin Ltd., London, 1971.

79. McGuire, M.C. and Garn, H.A., "The Integration of Efficiency and Equity Criteria in Public Project Selection", Economic Journal, December 1969, pp. 882-93.

80. Meade, J.E. and Fleming, M.J., "Price and Output Policy of State Enterprise", Economic Journal, vol. 54, 1944, pp. 321-39.

81. Meek, R.L., "The New Bulk Supply Tariff for Electricity", Economic Journal, vol. 78, 1968.

82. Mera, K., "The Trade-off Between Aggregate Efficiency and Inter-regional Equity: A Static Analysis", Quarterly Journal of Economics, 1967, pp. 658-74.

83. Mishan, E.J., "A Reappraisal of the Principles of Resource Allocation", Economica, November 1957, pp. 324-42.

84. ________, "Economic Evaluation of Allocative Changes", Journal of Transport Economics and Policy, September 1970, pp. 221-34.

85. ________, "Second Thoughts on Second Best", Oxford Economic Papers, vol. 14, October 1962, pp. 205-217.

86. ________, Welfare Economics: An Assessment, de Vriess Lectures, North Holland Publishing Company, Amsterdam, 1969.

87. Morgan, E.V. and Morgan, D.M., The Economics of Public Policy, Edinburgh University Press, 1972, pp. 169-185.

88. Morlok, E.K. and Viton, P.A., "Self-sustaining Public Transportation Services", in Transport Policy and Decision Making, vol. 1, No. 2/3, 1980, pp. 169-194.

89. Munasinghe, M., Electric Power Pricing, Staff Working Paper, International Bank for Reconstruction and Development, Washington, July 1979.

90. Nash, C.A., Public vs. Private Transport, Macmillan, 1976.

91. Nath, S.K., A Reappraisal of Welfare Economics, Routledge and Kegan Paul, London, 1969.

92. Nove, A., Efficiency Criteria for Nationalised Industries, Allen and Unwin Ltd., London, 1973.

93. Oettle, K., "Obligations of Transport for Underdeveloped Regions", in Blonk, W.A.G., ed., op.cit., pp. 17-34.

94. Okun, A.M., Equality and Efficiency: The Big Tradeoff, The Brookings Institution, Washington, 1975.

95. Oliver, J.M. and Webster, G.H., Public Policy and Economic Theory, Hutchinson, 1970.

96. Peacock, A.T. and Rowley, C.K., Welfare Economics: A Liberal Restatement, 1st ed., Martin Robertson, London, 1975.

97. ________, "Welfare Economics and the Public Regulation of Natural Monopoly", Journal of Public Economics, 1972, pp. 227-244.

98. Pegrum, D.F., Transportation Economics and Public Policy, Homewood, Illinois, R.D. Irwin Inc., 3rd ed., 1973.

99. Phelps Brown, E.H., A Course in Applied Economics, Pitman and Sons Ltd., London, 2nd ed., 1964.

100. Ponsonby, G.J., "Transport Policy: Co-ordination through Competition", Hobart Paper No. 49., Institute of Economic Affairs, 1969.

101. Raiffa, H., "Decision Making in State-owned Enterprise", in Aharoni, Y., and Vernon, R., eds., op.cit., pp. 54-62.

102. Redwood, J. and Hatch, J., Controlling Public Industries, Blackwell, Oxford, 1982.

103. Rees, R., Public Enterprise Economics, Weidenfeld and Nicolson, London, 1978.

104. ________, "Second Best Rules for Public Enterprise Pricing", Economica, August 1968, pp. 260-274.

105. Republic of South Africa, Board of Trade and Industries, Investigation into the Tariff Policy and Tariff Structure of the Supply of Electricity in South Africa, Report No. 1889, Pretoria, 1977.

106. ________, Report of the Commission of Enquiry into the Co-ordination of Transport in South Africa, R.P.32/1969, Pretoria.

107. ________, Report of the Committee of Enquiry into Socio-economic Rail Passenger Services, Pretoria, May 1979.

108. ________, Report of the Commission of Enquiry into Water Matters, R.P. 34/1970, Pretoria.

109. ________, Report of the Committee on Railway Rating Policy and Industrial Location in South Africa, Johannesburg, April 1984.

109. Rimmer, P.J. and Dick, H.W., "Improving Public Transport in South East Asian Cities: Some Reflections on the Conventional and Unconventional Wisdom", in Transport Policy and Decision Making, vol. 1, No. 2/3, 1980, pp. 97-120.

111. Ruggles, N., "Recent Developments in the Theory of Marginal Cost Pricing", in Turvey, R., ed., op.cit., pp. 11-43.

112. Russell, C.S., "Restraining Demand" in Furrell, B.M. and Hay, R.D., eds., The Management of Water Resources, Saxon House, Lexington, 1974.

113. Salter, W.E.G., Productivity and Technical Change, 2nd ed., Cambridge 1966.

114. Samuelson, P.A., "Social Indifference Curves", Quarterly Journal of Economics, February 1956.

115. Santoro, F., "Planning of Railway Transport and of Land Use: Experiences and Perspectives in Italy", in Polak, J.B., and van der Kamp, J., op.cit., pp. 175-189.

116. Scherer, F.M., Industrial Market Structure and Economic Performance, Rand McNally College Publishing Company, 2nd ed., 1980.

117. Scitovsky, T., "A Reconsideration of the Theory of Tariffs", Review of Economic Studies, vol. IX, 1942, pp. 89-110.

118. Sharp, C.H., "Transport and Regional Development - with Special Reference to Britain", in Transport and Decision Making, vol. 1, No. 1. 1980, pp. 1-11.

119. Shepherd, W.G., Public Enterprise, Economic Analysis of Theory and Practice, Heath-Lexington, Massachussetts, 1976.

120. South African Railways and Harbours, "Annual Reports", 1977-1980.

121. _______ , Board Report, R.P. 79/80, Pretoria.

122. South African Transport Services, Annual Reports, 1981-1982.

123. _______ , Memorandum by the Minister of Transport Affairs Setting Out the Estimated Results of Working for the Financial Year 1982/83 and Anticipated Revenue and Expenditure for Year 1983/84 Together with Latest Traffic and Other Statistics, W.P.A-'83, Pretoria.

124. Stillwell, F.J.B., Regional Economic Policy, Macmillan, London, 1972.

125. Talley, W.K., "Determining Fully Allocated Cost Prices For Regulated Transportation Industries", in International Journal of Transport Economics, vol. IX, No. 1, April 1982, pp. 25-42.

126. Thomson, J.M., Modern Transport Economics, Penguin Modern Economics Texts, Penguin, Harmondsworth, 1974.

127. Turvey, R., Economic Analysis and Public Enterprises, London, Allen and Unwin Ltd., 1st ed., 1971.

128. ________, "Marginal Cost", Economic Journal, June 1969, pp. 282-297.

129. ________, Optimal Pricing and Investment in Electricity - An Essay in Applied Welfare Economics, Allen and Unwin Ltd., London, 1968.

130. Ullman,E.L., "The Role of Transportation and the Bases for Interaction", in Blunden,J., ed., Regional Analysis and Development, Open University Press, Harper and Row, London, 1973, pp. 52-63.

131. van der Veer,G.D., "The Role of Rail Transport in the Total Transport Scene", paper delivered at the Annual Transport Convention, August 1982, Pretoria.

132. van Gent,H.A. and Kuyvenhoven,R.A., "Professor J.P.B. Tissot van Patot: His Approach to the Theories of Transport Policy and Spatial Transport Economics", in Polak,J.B., and van der Kamp,J., op.cit., pp. 26-37.

133. Verburgh,C., South African Transportation Policy, Bureau for Economic Research, University of Stellenbosch, 1961.

134. Vickery,W., "Some Objections to Marginal Cost Pricing", Journal of Political Economy, June 1948.

135. ________, "Some Implications of Marginal Cost Pricing for Public Utilities", American Economic Review, vol. 45, 1955, pp. 605-20.

136. Voigt,F., and Witte,H., "A General Concept of Transportation Theory", International Journal of Transport Economics, vol. IX., No.1, April 1982, pp. 15-23.

137. Warford, J.J., "Water Requirements: The Investment Decision in the Water Supply Industry", The Manchester School, vol. 34, 1966, pp. 87-112.

138. Whitcomb, D.K., Externalities and Welfare, Columbia University Press, New York, 1972.

139. Wiles, P.J.D., Prices, Costs and Output, B.Blackwell, Oxford, 1961.

140. Williamson, O.E., "Peak-Load Pricing", in American Economic Review, vol. 56, 1966, pp. 810-27.

141. Wiseman, J., "The Theory of Public Utility Price - An Empty Box", Oxford Economic Papers, vol. 9, 1957, pp. 56-74.