Environment Canada
Conservation & Protection

Environnement Canada
Conservation et Protection

ECONOMIC THEORY OF WATER PRICING

Prepared by:

Roger McNeill

March 1989

Inland Waters
Pacific and Yukon Region
Vancouver , B.C.

Economic theory of water pricing.

TD
227
.B74
BC89-2

Economic theory of water pricing.

ENVIRONMENT CANADA
CONSERVATION AND PROTECTION

ECONOMIC THEORY OF WATER PRICING

PREPARED BY
ROGER MCNEILL

INLAND WATERS
PACIFIC AND YUKON REGION

VANCOUVER, BRITISH COLUMBIA
MARCH 1989

ABSTRACT

The purpose of this report is to review the economic theory relevant to water pricing. The theory is examined to determine how the objectives of economic efficiency and cost recovery might be achieved through water pricing. Marginal cost pricing is shown to be the most efficient system when cost recovery is not an issue. Because marginal cost pricing may not result in full cost recovery, other pricing systems including Ramsey pricing, Coase tariffs and multipart tariffs are examined. Coase tariffs would often be suitable for water pricing because they can result in maximum efficiency while at the same time ensuring full cost recovery. Peak load pricing based on long run marginal cost is also applicable to water supply utilities. Pricing is shown to be related to the issue of capacity expansion, and both should be considered simultaneously in water utility planning. Several unique features of water and their effect on optimal pricing are examined. These include storage management, natural limits to supply, non-consumptive uses and linkages between upstream and downstream demands. Pricing is shown to be effective in achieving optimal allocation of water among competing uses when natural limits on supply exist. The optimal allocation will result when price is equated to the marginal willingness to pay for all users. In a basin with a number of supply and demand nodes, more complex techniques such as mathematical programming might be necessary to determine the optimal prices for the various water uses. The study concludes with a recommendation that Inland Waters prepare a set of water pricing guidelines that can be used by water utilities and governments who wish to implement efficient pricing.

RESUME

Cette étude revoit la théorie en économie applicable à la tarification de l'eau. La théorie est examinée pour déterminer comment les objectifs d'efficacité économique et de côut de recouvrement peuvent être réalisés par la tarification de l'eau. Le côut marginal de tarification est démontré d'être le système le plus efficace quand le côut de recouvrement n'est pas important. En vue que le côut marginal de tarification ne couvrira pas possiblement le côut complet de recouvrement, d'autres systèmes de tarifications sont examinés tel que la tarification Ramsey, les tarifs Coase et les tarifs d'étapes multiples. Les tarifs Coase peuvent souvent être propices pour la tarification de l'eau parce qu'ils peuvent produire une efficacité maximum tout en assurant un côut complet de recouvrement. La tarification de débit de pointe, basée sur le côut marginal à long terme, est aussi applicable aux services publics d'aménagement des ressources hydriques. Il est demontré que la tarification est reliée à la capacité de croissance et que ces deux facteurs soient considérées simultanément dans la planification des services publics de ressources hydriques.Plusieurs aspects unique de l'eau et ces effets sur la tarification optimale sont aussi examinés tel que la gestion de l'emmagasinement, les limites naturelles des réserves, les utilisations de non-consommation et les liens de demandes d'amont et d'aval. Il est aussi demontré que la tarification est efficace à la réalisation de la répartition optimale d'eau parmi les utilisations concurrents quand il existe une limite naturelle de réserve d'eau. La répartition optimale se réalisera quand le tarif reflétera la volonté de payer de la part de tous les usagers. Dans un bassin

hydrographique avec plusieurs sites d'approvisionnements et de demandes, cela nécessitera peut-être des approches plus complexes, comme la programmation mathématique, pour déterminer les tarifs optimals pour les nombreuses utilisations d'eaux. L'étude se conclut en recommandant que les Eaux intérieures préparent des directives de tarification d'eau qui pourraient être utilisées par les services publics d'eau et les gouvernements afin d'y réaliser une tarification efficace.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

In the past few years Canadian policy makers have shown increasing interest in the use of pricing as a water management tool. This interest was sparked by a number of studies that revealed deficiencies in the pricing systems used by water supply agencies in Canada. Tate (1984) pointed out that efficient pricing would not only result in water conservation but would also promote sustainable economic development. The Inquiry on Federal Water Policy (Pearse, Bertrand and MacLaren, 1985) recommended that the federal government explicitly endorse the user pay principle for water.

The statement on federal water policy (Environment Canada, 1987) proposes water pricing as a basic strategy for achieving the goals of protecting and enhancing the quality of the water resource and promoting the wise and efficient management and use of water. The statement (p. 8) says that the federal government will:

- endorse the concept of realistic pricing as a direct means of controlling demand and generating revenues to cover costs;
- undertake, support and promote joint federal-provincial examination of the costs and pricing of water for both consumptive and non-consumptive water uses; and
- encourage the application of pricing and other strategies, such as the beneficiary/polluter pays concept, to encourage efficient water use.

Federal government staff (or anyone else involved in water pricing) should acquire a basic understanding of the economic theory of efficient pricing before promoting the concept to water supply utilities. There exists a well developed theory of optimal pricing for public utilities and natural monopolies that has been expounded in several books and journal articles. Most applications of the theory have been in the fields of power generation and telecommunications. A few examples of applications of the theory to water pricing have appeared in various journals. Most of this literature on public utility pricing is technical, and a good knowledge of mathematics and economics is required to understand it. The objective of this report is to summarize the basic theory of optimal pricing in a water use context, avoiding mathematical analysis where possible. The report is intended as a basic reference aimed at providing economists and planners with the minimum understanding necessary to develop a water pricing program.

Before determining the optimal or efficient pricing system for a water supply system, one requires some criteria by which optimality or efficiency is defined. In this report a fairly standard criterion is used which is to maximize the consumer and producer surplus from water use. The concepts of producer and consumer surplus and their relevance to water pricing are explained in chapter II. The theory of marginal cost pricing, and how it can be used to achieve efficiency, is also developed in the second chapter.

In chapter III the financial implications of marginal cost pricing for water utilities are discussed. It is generally desirable that the pricing structure result in full cost recovery by the utility or water supply agency, without a financial loss or profit. The problem of cost recovery arises in practical applications of water pricing since marginal costs of supplying water are often less than average costs. The concepts of non-uniform pricing and multipart tariffs are introduced. These concepts can be used to meet both financial and efficiency objectives.

Chapter IV outlines pricing theory relevant to a longer run context when the issue of capital expansion should be considered. It shows how pricing can be used to meet costs of future expansion and to cope with problems of peak period demands. The optimal timing of expansion of system capacity and its implications for pricing are also discussed.

Chapter V discusses the theory in the specific context of water resource management. Several of the unique features of water use such as storage, non-consumptive use, return flows, seasonal use, and constraints on natural supply must be accounted for in the pricing system. Some methods of implementing pricing in more complicated and inter-related water management systems are presented.

The final chapter summarizes the results of the discussions and makes recommendations for the promotion of efficient water pricing. Particular emphasis is placed on a recommendation that a set of practical guidelines for water pricing be prepared for use by municipalities and water supply utilities.

II. EFFICIENCY AND MARGINAL COST PRICING

This chapter outlines the basic criteria used in determining economic efficiency. The concepts of consumer surplus and producer surplus, and their use as an index of economic benefits, are explained. The theory of marginal cost pricing is introduced along with a discussion of how it can lead to the maximum amount of consumer and producer surplus from water use.

A. Economic Efficiency

Many public projects are subject to a benefit-cost analysis to determine whether the benefits of the project are greater than the cost. The project is deemed to be economically efficient if the benefits outweigh the cost. The same basic definition of economic benefits and efficiency is used in this paper as is used in standard benefit-cost analysis of public projects. Standard benefit-cost analysis has some theoretical shortcomings and is subject to a number of valid criticisms. These criticisms, however, are less critical when determining the benefits from water use. Furthermore, given the widespread use of benefit-cost analysis as an applied management tool, it seems reasonable to use its basic tenets as a starting point for determining the efficient use and pricing of water.

1. Consumer and Producer Surplus

Consumer and producer surplus are basic measures of the value accruing to the consumer and producer of a good. Consumer surplus is derived from an individual's willingness to pay for a product. Central to the concept of willingness to pay and consumer surplus is the demand curve (Figure 1).

The demand curve, which shows the relationship between the quantity demanded of a product and its price, is downward sloping in almost all cases although the steepness and exact shape of the curve will vary depending on the product and the consumer. An aggregate demand curve for a group of consumers can be obtained simply by adding up all the individual demands at various price levels. The aggregate demand curve will also be downward sloping.

The consumer's total willingness to pay is the area under the demand curve in Figure 1. Total willingness to pay can be considered as a gross measure of value to the consumers. The net willingness to pay, also known as consumer surplus, is equal to the total willingness to pay minus the actual amount paid. The area designated by the letter S in Figure 1 is the consumer surplus. Graphically it is described as the area above the price line and below the demand curve.

An analogous concept to consumer surplus is producer surplus. For a single firm supplying a product, the producer surplus can be thought of as profit to the firm. Figure 2 shows the marginal cost curve for a firm supplying one product. The marginal cost is the cost of producing each additional unit of the product. If there are no fixed (overhead) costs, the profit to the firm is equal to the area marked P, which is the area above the marginal cost curve and below the price line. If a fixed cost does exist, then the profit is equal to the same area minus the fixed costs.

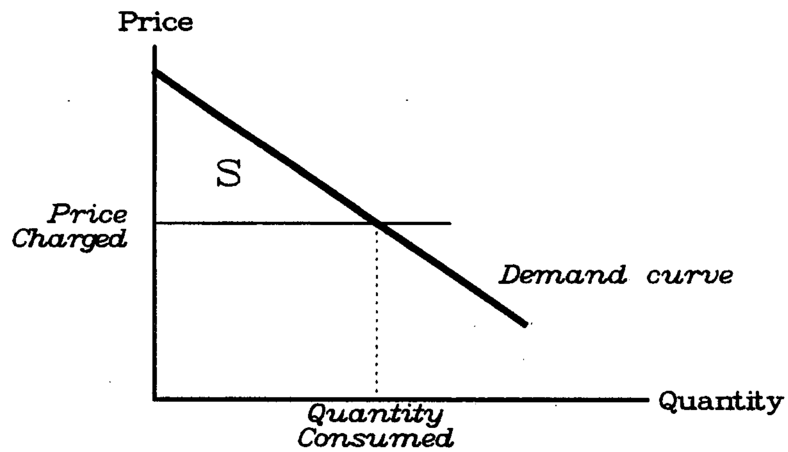


FIGURE 1 CONSUMER SURPLUS

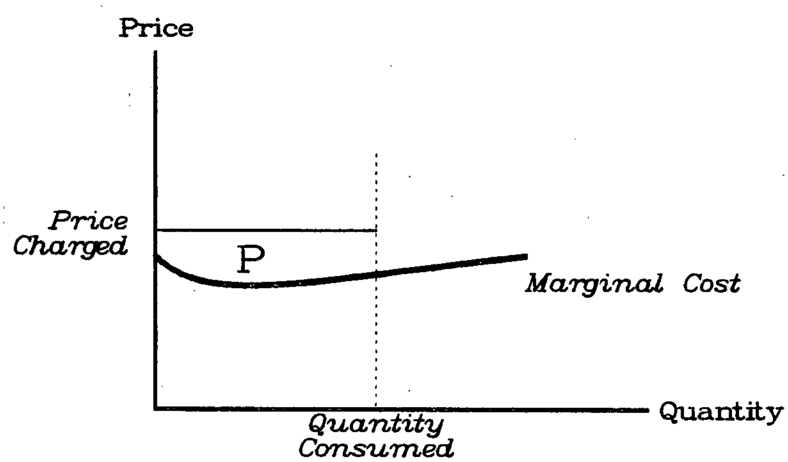


FIGURE 2 PRODUCER SURPLUS

2. Economic Benefits

In benefit-cost analysis the benefits, or economic value of a commodity, are equal to the sum of the consumer surplus and the profit (producer surplus) to the firm. Suppose for example, one were analyzing the benefits of an irrigation project that would result in increased agricultural production. The benefits from the project would be equal to the increase in consumer surplus plus the increase in the profit of the producers. The same approach can be used for analyzing the benefits of changing the price charged by a regulated firm or utility. The benefits of the price change are equal to the increase in the consumer surplus and the profit to the firm that result from the price change.

By increasing the price charged, a utility would obtain greater profit, but would at the same time decrease the consumer surplus.[1] By decreasing the price, the firm would decrease its profit and increase the consumer surplus. If the gain by either the producer or consumer is greater than the loss by the other party then a benefit results from the price change. This is illustrated in Figure 3, which shows a price change from P_1 to P_2. The firm loses profit represented by area A minus C, while the consumers gain a surplus equal to area A plus B. A net gain in total consumer plus producer surplus equal to area B plus C results from the price change. The price P_2 is thus more efficient than the price P_1 because it results in an increase in the total surplus or economic benefits.

---

1. It is assumed that the utility is large enough that it controls the market price. It should be noted that there is a certain price beyond which an increase will result in a reduction in profit because of the decrease in consumption.

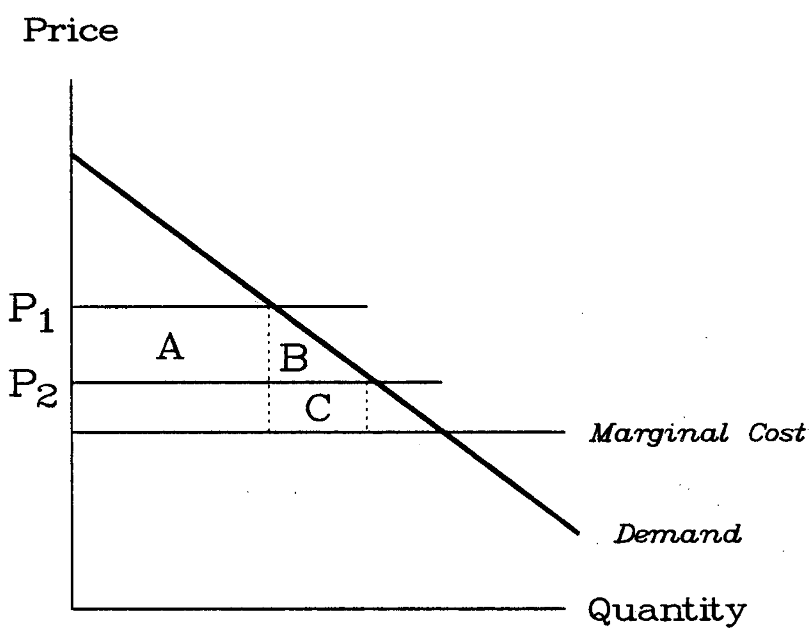


FIGURE 3 EFFICIENCY GAINS FROM A PRICE CHANGE

B. Marginal Cost Pricing

The theory of marginal cost pricing has been developed in the context of public utilities that have monopolies on the product they produce. If such a monopoly were privately owned and unregulated, the owners would choose a pricing system that would maximize profits. In the case of a monopoly that is publicly owned or regulated, a price should be chosen that maximizes the total economic benefits rather than profits to the utility. Marginal cost pricing is a method of achieving optimal output and maximum benefits from a public or regulated utility.

Figure 3 indicates that moving from price P_1 to price P_2 results in an increase in total surplus (where total surplus is the sum of consumer and producer surplus). This gives rise to the question: is there an optimal price where the total surplus is maximized? In Figure 3, moving from the price P_1, which is above marginal cost, to the price P_2, which is closer but still above marginal cost, resulted in an increase in total surplus. Lowering the price towards marginal cost will always result in a gain in total surplus because the increase in the surplus gained by consumers is always larger than the decrease in profits experienced by the firm.

Figure 4 illustrates the effect of a price change when the original price is below marginal cost. At the original price P_1, the firm makes a loss equal to the rectangle defined by MC, P_1, E_1, and M_1. By increasing the price, the loss to the firm will be reduced by the area S plus D, and consumer surplus will decrease by area S. Moving from price P_1 to price P_2, which is closer to marginal cost, will always results in a gain in total

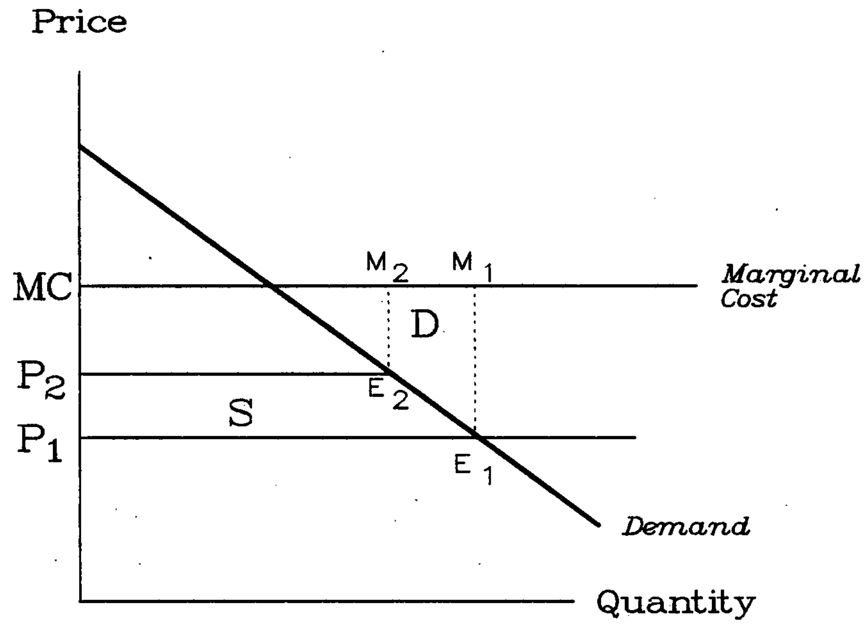


FIGURE 4 EFFICIENCY GAINS FROM ADJUSTING PRICE TOWARDS MARGINAL COST

surplus because the gains by the firm are larger than the losses by consumers.

It follows that setting the price exactly equal to marginal cost will result in the maximum total surplus. The term "marginal cost pricing" means just that: setting price equal to marginal cost in order to maximize benefits. In all the examples illustrated to this point, it has been assumed that marginal cost is constant for all levels of output. If this happens to be the case for a regulated firm, the optimal price can simply be set equal to the marginal cost without any knowledge of the consumer demand curve. In the short run, a constant marginal cost may be approximately true for many water supply agencies. For example, the marginal cost of supplying an initial unit of water to a fixed urban population may simply be the additional costs of energy for pumping. The unit cost of energy would likely be constant.

1. Increasing or Decreasing Marginal Costs

If the marginal cost curve is increasing or decreasing, the same basic rule of optimal pricing applies, but one must also take into account the consumer demand curve. To obtain maximum benefits, a price must be chosen where the amount supplied by the firm is equal to the amount demanded. In Figure 5, the equilibrium point between the marginal cost curve and the demand curve results in the maximum benefits. At this point, the price P^* is equal to marginal cost, and the amount supplied by the utility is

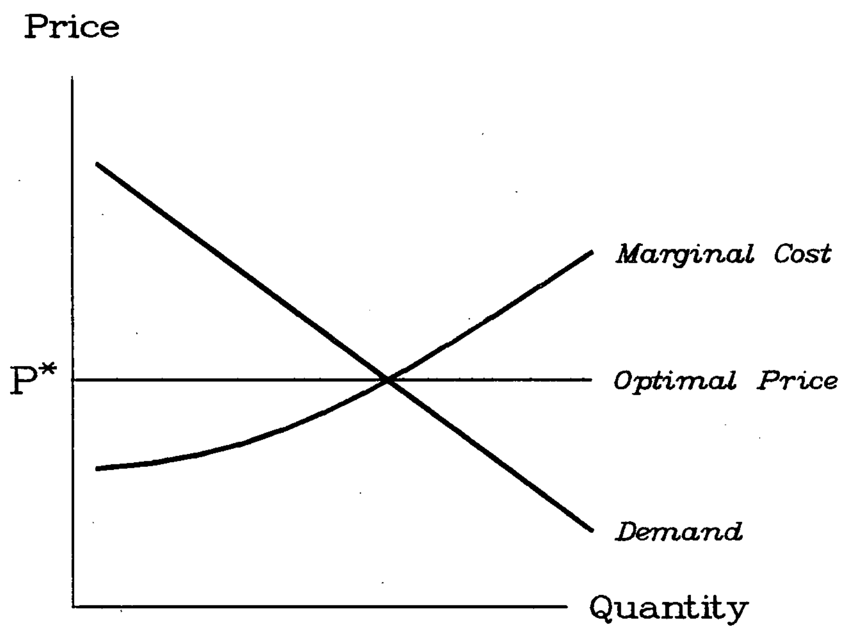


FIGURE 5 OPTIMAL PRICE WITH INCREASING MARGINAL COSTS

exactly equal to the amount demanded by consumers. The marginal pricing rule can be stated formally as:

(1) Price = Marginal Cost = Marginal Willingness to Pay.

The marginal willingness to pay is represented by the consumer demand curve. Each point on the demand curve represents the marginal willingness to pay or price that consumers will pay for an extra unit of consumption. At the point where the marginal cost curve crosses the demand curve, marginal willingness to pay is equal to marginal cost. To find the price that results in this equilibrium, it would be necessary to have a graphical or mathematical estimate of both the demand curve and the marginal cost curve.

2. Short Run and Long Run Marginal Cost

The marginal cost pricing rule has both short run and long run interpretations. In the short run, capital costs cannot be varied and marginal costs include only the variable costs of production or delivery. As long as there is adequate physical capacity for the foreseeable future, setting the price equal to marginal cost will maximize benefits. In the long run, capital expansion must be planned for since all inputs, including capital are variable. A long run marginal cost function (see Chapter IV) can be defined and used as the basis for long run pricing.

Water utilities should, in general, consider the long run marginal cost when setting price. Because population and industry are increasing in most regions, the issue of system expansion will have to be addressed at some point. Therefore, to ensure efficient water use and optimal capacity of future expansions, the long run marginal cost should be used as the basis for pricing.

The marginal cost pricing rule does not state anything about the distribution of benefits. In its basic form it does not even require that the regulated firm break even. If the short run marginal cost curve is used as the basis for price setting, then the utility will lose money because fixed costs are not covered. Even if the long run marginal cost curve is used, the utility may still experience a net financial loss and possibly a financial profit. These problems and their policy implications are discussed in greater detail in Chapter III.

3. Marginal Costs for Different Classes of Consumers

It is possible that some classes of consumers will have lower marginal costs than other classes for a given level of consumption. In other words, the utility faces a separate marginal cost curve for each class of consumer. This situation would be encountered, for example, if energy costs for distributing water were related to the distance from the main storage reservoir. The marginal costs of supplying distant customers would then be greater than the marginal costs of supplying close-by customers. Long run marginal costs might also be less for large consumers because greater amounts of water can be supplied through a single connection, resulting in lower distribution costs.

The optimal pricing rule when different classes of customers have different marginal cost curves is basically the same as equation (1). The price for each class should be based on their particular marginal cost curve. Thus, some classes of users with higher marginal costs would be charged higher prices than other classes with equivalent usage. In practice the difference in marginal costs between users may be slight and the increase in efficiency might not warrant the extra billing expenses.

4. Losses from Flat Charge Pricing

It is quite common to observe flat charges for water use by utilities in Canada. These utilities charge the consumer with a connection price, but do not charge according to how much the consumer actually uses. The marginal price to the consumer is therefore nil, and he will continue consuming water until his marginal willingness to pay is also zero. In Figure 6, water will be consumed until the point Q on the horizontal axis. This results in an over-consumption equal to the distance Q minus O, where O is the optimum consumption that would occur if price were set equal to marginal cost. The area under the marginal cost line (C plus T) represents the total cost (excluding any fixed costs) of supplying the water and the total area under the demand curve represents the consumer surplus. The triangle labeled T, to the right of the demand curve represents the total loss in benefits that result from using a flat charge instead of marginal cost pricing. This area is often called a deadweight loss.

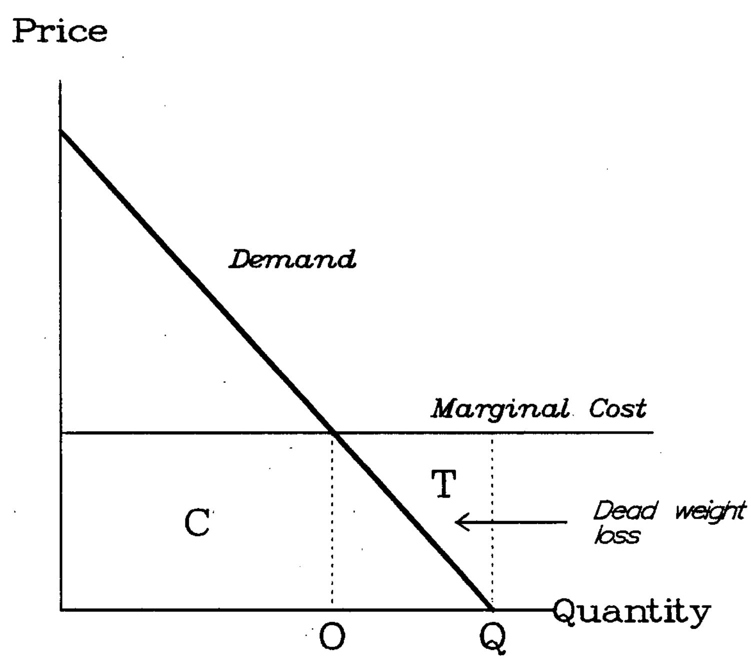


FIGURE 6 LOSSES FROM FLAT CHARGE PRICING

C. Summary

In this chapter the concepts of consumer and producer surplus, efficiency gains from pricing, and marginal cost pricing have been introduced. The standard definition of economic benefits, represented by the sum of consumer and producer surplus, has been used as a criterion for determining the efficient pricing system. Using this criterion, it was found that setting price equal to marginal cost resulted in the maximum benefits. Flat charge pricing, which is often used by water supply utilities in Canada, was shown to cause excess consumption of water and a loss of benefits from its use.

For utilities that have constant marginal costs, price can be set equal to marginal cost without any knowledge of the consumer demand curve. If marginal costs are increasing or decreasing, then an estimate of the demand curve is required so that the optimal price and quantity can be determined. The optimal price is obtained at the point where the demand curve and marginal cost curve intersect.

The marginal cost pricing rule, as presented in this chapter, is concerned only with efficiency and does not consider any financial constraints on the utility. Financial constraints would usually require the utility to break even by recovering all costs without making excess profits. In the short run, it is possible, and even likely, that marginal cost pricing will result in a financial loss for the firm even though total benefits are maximized. The loss will result from the fixed (overhead) costs that the firm is faced with in order to supply the product. If the long run marginal cost is used as a basis for price setting, the utility might still experience a loss or possibly an excess profit. Losses will have to be covered from some other source such as general tax revenues or from fixed user charges that can be charged in addition to the per use price based on marginal cost. Excess profits may also have to be redistributed back to customers. In the following chapters, pricing systems are discussed that ensure that the regulated firm breaks even while at the same time retaining maximum efficiency.

III. EFFICIENT PRICING WITH FINANCIAL CONSTRAINTS

Marginal cost pricing will achieve efficiency objectives but may not allow the utility or regulated firm to meet its financial objectives. These financial objectives will usually require that the utility recover all costs but not make excessive profits. Ideally it should just break even on its operations. Such financial objectives have a strong tradition in public policy and can also be supported on economic grounds. The purpose of this chapter is to examine pricing systems that ensure full cost recovery while providing the maximum possible economic benefits to consumers.

If the utility experiences a financial loss, it will have to be subsidized from some other source of revenue to stay in business. Subsidization through taxation and income transfers can cause distortions in the economy that result in a loss of benefits in other areas of the economy. Another criticism of subsidization is that it will lead the system planners to incorporate excess capacity into their operations knowing that the utility will not be responsible for recovering its costs. The costs of subsidization thus have to be weighed against the loss in total benefits that would result if the utility became fully self supporting by deviating from marginal cost pricing.

From a public policy perspective, subsidization of public utilities, crown corporations and other regulated firms has not been popular in North America. In Canada, deficits run by crown corporations often become a great source of embarrassment for the government. Hanke and Fortin (1985), in a survey of provincial regulations, found that recovery of costs was a common requirement for municipal water utilities. In the United States there is a general tradition that utilities are expected to cover their full costs. This tradition is probably due to public concerns over equity of taxation and subsidization. For example, the public in one area of the country is unlikely to support subsidization of power production in another region. For these reasons, it would be naive of economists to advocate pure marginal cost pricing without looking at the issue of cost recovery. Interest in this issue has spawned significant theoretical development aimed at maximizing total benefits subject to full cost recovery.

Social policy may dictate that water utilities not make a consistent profit from their operations. Hanke and Fortin (1985) point out that profits by water utilities are usually against provincial regulatory tradition in Canada, and an excess profit can be explicitly prohibited or subject to regular review. Many of the principles developed in the following sections can be used to deal with excess profits as well as financial losses.

A. Relationship between Marginal Cost and Average Cost

To just break even, a water utility would have to set its price equal to the average cost of supplying the water. Depending on the amount of water delivered, the average cost could be either above or below the marginal cost. As shown in Figure 7, marginal cost is below average cost when average costs are declining. It is above average cost when average costs are increasing. Under marginal cost pricing, a deficit will arise if

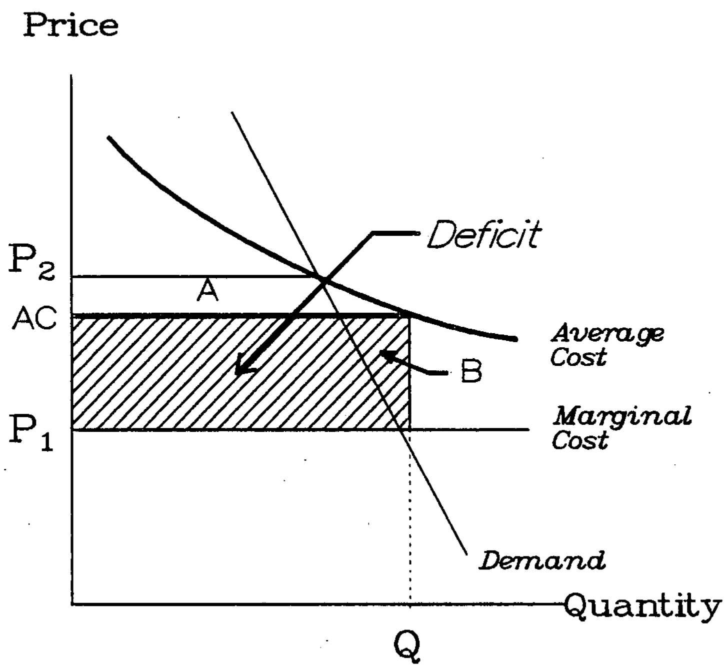


FIGURE 7 DEFICITS FROM MARGINAL COST PRICING

average cost is above the marginal cost. Figure 7 illustrates the situation where the utility charges a price P_1, equal to the marginal cost. At this price, the amount demanded would be equal to Q, and the average cost of production would be AC. The utility would experience a deficit indicated by the shaded area. If the price were set at P_2, where demand is equal to average cost, the utility would break even. However, the resulting losses to consumers would be greater than the gains to the utility. The net loss in benefits would be equal to the area A plus B.

It is also possible that the optimal price, which occurs at the intersection of marginal cost and the demand curve, could be higher than the average cost. This price would result in the utility making a profit on its sale of water. If the price were lowered to the same level as average cost, the profit would disappear and consumers would benefit. Despite the benefits to consumers, a loss in efficiency would result because the gain in consumer surplus is less than the profit lost by the utility.

In the short run, the average cost will always be greater than marginal cost if there is a fixed (overhead) cost that is born by the utility. Basing the price on short run marginal cost will thus result in a net loss to the firm. Water supply utilities will almost certainly have overhead costs, so basing the price on short run marginal cost will result in a financial loss.

In the long run the average cost curve may be declining in the vicinity of the optimal price and the subsequent demand level. Average costs decrease when there are economies of scale that make it more efficient to deliver large quantities of water than small quantities. Companies in this category are known as natural monopolies, since it is more efficient to have one firm acting as a monopoly producing all of the output rather than having several smaller firms undertake the same function. In the case of water utilities, there is not much empirical evidence to assess whether average costs are declining in the long run. In many instances it would be reasonable to assume that there are economies of scale in water collection and distribution, meaning that it would be more efficient to have one central utility undertaking this function.

The problem of cost recovery for regulated natural monopolies has been examined extensively in the literature as is discussed in subsequent sections. Methods of obtaining full cost recovery with both uniform (single price) price systems and non-uniform (multiple price) systems are presented. The principles discussed can also be used for distributing revenues back to consumers if marginal cost pricing results in a profit for the water utility.

B. Ramsey Pricing

Ramsey pricing is one method of trying to maximize total benefits while at the same time ensuring that the regulated firm does not incur a loss. In its basic form it uses uniform prices to achieve this objective. A uniform price is a unit price that does not vary with the amount consumed; it will not include any quantity surcharges or discounts. The amount that a consumer pays is a straight multiple of this price times consumption. Prices that vary between individuals or groups of consumers, but are not based on the amount consumed, are also classified as uniform prices.

If a firm is restricted to imposing a single uniform price that does not vary between consumers, then the only way it can recover costs is to charge a price equal to average cost. If the firm is given the option of charging different prices among different groups of consumers, then it is possible to achieve cost recovery and increase the total benefits compared to average cost pricing. This is achieved by charging a greater mark-up for the groups of consumers with the most inelastic demand, and a lesser mark-up for those consumers with elastic demands.[1] When demands are inelastic, the response in quantity demanded is less and by concentrating the price increases in these markets there will be less deviation from the optimal marginal price solution.

The basic problem is to determine the optimal mark-up in each market that will maximize total benefits while ensuring full cost recovery. A formula for achieving this objective has been derived as shown in equation (2).

$$\text{(2)} \qquad \text{mark-up}_i = \frac{P_i - C_i}{P_i} = \frac{\lambda}{\varepsilon_i}$$

1. The elasticity of demand is a measure of the response in quantity demanded to a change in price. For a small change in price it is defined as:

$$\frac{\text{percentage change in quantity demanded}}{\text{percentage change in price}}$$

where:

P_i = price charged in each user category

C_i = marginal cost for each user category

λ = constant

ε_i = elasticity of demand for each user category

In equation (2) the subscript i refers to the particular market served. For example, suppose there are two markets for water supplied by the utility; agricultural and industrial. The percentage mark-up over marginal cost for each market will be determined by dividing a constant λ by the elasticity of demand for the respective market. If, for example, industrial water demand is less elastic than agricultural water demand, then industrial users will be faced with a higher mark-up because their elasticity coefficient ε is lower. The actual value for λ is found by trying different values until total cost of supplying water is just covered. The rule shown in equation (2) is known as the inverse elasticity rule (IER), as the mark-up is inversely proportional to the elasticity. It is a well known result in the literature on optimal pricing, and the concept will be encountered again in some of the more complex pricing systems discussed in a subsequent section.

Ramsey pricing becomes more complicated if the demand by one segment of the market is related to the demand by other segments. This problem does not usually occur in water markets, as the demand for water by one group of users is unlikely to be related to the water demand in other markets. If the marginal cost is not constant, the Ramsey price solution will become more complex, although the general principle will still apply.

The Ramsey principle could also be used as a method for distributing excess profits when they they occur. In such cases, the users with the inelastic demands would be charged lower prices than the other users in order to reduce the revenues to the utility. By lowering prices for the inelastic users, the deviation from the optimum marginal cost solution will be minimized.

Two problems are apparent with Ramsey pricing. First, the elasticities of demand will have to be known for each market that the utility serves. This type of information is not easy for utilities to obtain. The demand elasticities typically have to be estimated by econometric studies requiring extensive data. The second problem is one of perceived equity or fairness. The consumer groups with the highest mark-up are not likely to consider Ramsey pricing as a fair system even though it maximizes total benefits subject to the utility breaking even. However, it may be possible to compensate the high mark-up users through a system of lump sum transfers that do not affect their decision on the amount consumed.[2]

C. Two Part Tariffs

Another approach to ensuring that regulated firms break even, is to charge an entry fee (connection charge) to consumers in addition to a uniform price per amount consumed. The extra revenue gained from the connection

---

2. The lump sum transfers may result in a slight increase in consumption because of an income effect. Because consumers' incomes effectively increase because of the transfers, they may choose to dispose of some of the extra income by increasing consumption of water.

charge can help offset the loss to the utility if prices are set at or near marginal cost. The connection charge, because it is not related to the amount consumed, may not affect the consumption decision of an individual. It seems likely then that a two part tariff may allow the utility to break even while still maximizing total benefits.

1. The Coase Two Part Tariff

The Coase two part tariff, first suggested by Coase (1946), is the simplest form of two part tariff. It charges a fixed connection charge that is the same for all consumers and a single price on consumption equal to marginal cost. The size of the connection charge is chosen so that it exactly offsets the loss to the utility that would occur with a uniform price equal to marginal cost. The connection charge can be calculated simply by dividing the fixed costs by the number of connections. Equations (3) and (4) define the Coase tariff.

(3) $P = MC$

(4) $E = FC/K$

where:

P = price based on usage

MC = marginal cost of supply

E = fixed entry fee or connection charge

FC = fixed or overhead costs to the firm

K = number of connections

At first glance it appears that the Coase tariff should result in the maximum benefits that would be obtained with straight marginal cost pricing, while at the same time ensuring that the utility breaks even. However, maximum benefits will not occur if marginal consumers decide not to buy the product from the utility because of the fixed connection charge. Under straight marginal cost pricing, these consumers would be part of the utility's market and if they cease purchasing from the utility, a loss in benefits results.

It is unlikely that customers of a water utility would drop out of the market due to the imposition of a fixed connection charge, especially if there are no other sources of water for the consumer. If there are no consumers, or if there are a negligible amount of consumers who would drop out, then the Coase tariff will maximize total benefits while ensuring full cost recovery for the utility. Thus the Coase tariff should warrant serious consideration for a great many water utilities in Canada.

In some instances, water utilities may be supplying customers who could find other sources of supply at a reasonable cost. For example, some major industrial users might find it more feasible to supply their own water from wells if the fixed connection cost to the utility were too high. In these cases a Coase tariff might not be the most appropriate pricing structure.

The Coase two part tariff could be used by a water utility that was making a profit on pure marginal cost pricing and wished to distribute the extra funds back to its customers. The fixed connection charge would take the form of a refund instead of a charge to the consumers. Note that the refund would be a fixed amount, not related to the amount of water consumed by each customer. Under such a system it is theoretically possible, although unlikely, that some low volume users would actually receive more from the refund than they paid through the usage charge.

2. The Optimal Two Part Tariff

The optimal two part tariff is the particular combination of connection charge and price that will maximize total benefits subject to the constraint that the utility break even. The Coase tariff will be the optimal two part tariff if no consumers leave the market because of the connection charge. If some consumers do drop out of the market because of the connection charge, the problem of finding the optimal connection charge and price becomes more difficult.

Determining the optimal two part tariff is considerably more complex than determining the optimal uniform (Ramsey) price. The problem, however, can still be solved using the same general principle that is used in Ramsey pricing. To recover full costs, the price setter has the choice of either increasing the connection charge, increasing the usage price, or increasing both charges. The Ramsey principle of increasing the charge in the most inelastic markets where the least distortion on quantity demanded can be used. For example, if consumer response is extremely inelastic to the connection charge, then the connection charge should be raised to recover costs, and the usage charge can be left near marginal cost. This process follows a similar logic to the Ramsey inverse elasticity rule.

The complication with two part tariffs and the Ramsey principle is that cross-elasticities between the connection charge and the usage price may be encountered. A cross-elasticity occurs, for example, when an increase in the usage price causes a decrease in the number of connections. The Ramsey solution must take into account the interaction between the two variables and as a result becomes quite complex. Two conditions that must hold for an optimal Ramsey two part tariff are shown in equations (5) and (6). Equation (5) defines the optimal mark-up on the connection charge.

$$(5) \qquad \frac{E - v}{E} = \frac{M}{\varepsilon} - \frac{(P-c) \cdot Q_m}{E}$$

where:

E = the connection fee (entry fee)

v = the cost of adding an additional customer

P = the usage price

c = the marginal cost of the product

Q_m = the quantity demanded by the marginal consumer

M = constant set high enough for the firm to cover all costs

ε = elasticity of participation with respect to the connection charge

The quantity Q_m is the amount demanded by the marginal consumer who would leave the market if there were only a marginal increase in the entry fee. Equation (5) is fairly similar to the basic inverse elasticity rule shown in equation (2), except that it has the additional term on the right hand side to account for the cross-elasticity effect.

Equation (6) sets out the condition for the optimal mark-up on the usage price, again accounting for the cross-elasticity effect.

$$(6) \qquad \frac{P - c}{P} = \frac{M}{T} \cdot \left[1 - \frac{Q_m}{Q_a}\right]$$

where:

P = the usage charge

c = the marginal cost of supply

M = constant set at sufficient rate to ensure cost recovery

T = elasticity with respect to the usage charge

Q_m = Quantity demanded by the marginal consumer

Q_a = average consumption

Equation (6) again follows the basic Ramsey concept, marking up the usage charge in inverse proportion to the elasticity. The final term in square brackets accounts for the cross-elasticity with the entry fee.

When calculating the optimal two part tariff, information is needed on both the elasticity of the connection charge and the usage price. At each price the quantity demanded by the marginal consumer must be known so that the cross-elasticity terms can be calculated. This detailed knowledge of individual demands for the service makes the actual calculation of the optimal two part tariff very difficult. However, if we can be reasonably certain that the imposition of a connection charge will have a negligible effect on the number of subscribers, then the optimal two part tariff will be represented by the basic Coase tariff described earlier.

D. Multipart Tariffs

A multipart tariff is a price schedule that has different rates for different levels of consumption. It is sometimes referred to as a non-uniform price. The most common such tariff is the declining block rate tariff as shown in Figure 8. This type of price schedule is quite commonly found among Canadian water utilities. Because high volume users pay a lower unit price than low volume users, declining block rate pricing is sometimes considered inefficient and wasteful. However, recent work on utility pricing (Brown and Sibley, 1986) has shown that, under some circumstances, declining block rate pricing can be more efficient than single or two part tariffs for obtaining full cost recovery by the utility. This section summarizes the work of Brown and Sibley, showing why declining block rate pricing can be an efficient price system in some circumstances.

The efficiency of declining block rate structures arises because they allow consumers to sort themselves onto various sections of the price schedule. Consumers have a range of choice; from low quantities at a high price to higher quantities at a lower price. This might seem to cause inefficiency by rewarding higher consumption, but a correctly designed block rate system can make all consumers and the utility better off than a uniform tariff.

The efficiencies can best be illustrated by starting with a single price tariff and comparing it to a two part tariff. To ensure cost recovery, the single uniform price is set equal to average cost. The two part tariff is composed of an entry fee and a usage charge that is less than average cost.

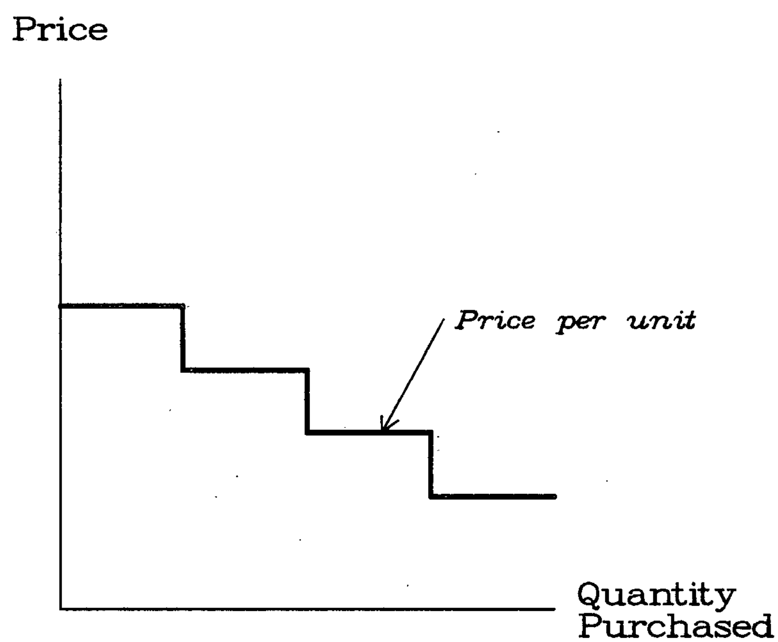


FIGURE 8 DECLINING BLOCK RATE TARIFF

Suppose there are two customers as shown in Figure 9, and initially both are charged the single average cost price equal to P_1. What would happen if they were offered a choice between the average price tariff and the two part tariff consisting of an entry fee equal to the shaded area and a usage price equal to P_2? By moving from price P1 to price P_2, Consumer One increases his consumer surplus by the area P_1,E_1,P_2,E_2. This increase is not enough to offset the entry fee, so he will opt to stay with the original price, P_1. Consumer Two would increase his consumer surplus by the area P_1,F_1,P_2,F_2, which is more than enough to offset the entry fee so he would choose the two part tariff with an entry fee plus the P_2 usage charge. The utility would also increase its revenues as a result of the entry fee and the increased consumption (from point C to point D) by Consumer Two. Thus, as a result of offering the optional two part tariff, Consumer Two is better off, the firm increases its profit and Consumer One remains the same. The net result is a gain in efficiency (total of consumer plus producer surplus).

Figure 10 shows the expenditures by consumers under the two optional price schemes. The shaded portion of each price schedule represents the expenditure function. At any consumption level less than the critical level Q_c, the consumer will elect to be on the shaded portion of price option 1. If his consumption is greater than Q_c, he will be on the shaded portion of price option 2. Figure 11 shows the portion of the price structure that generates the expenditure function. As can be seen, it is equivalent to a two part declining block price schedule with a change in rates once consumption reaches the level Q_c.

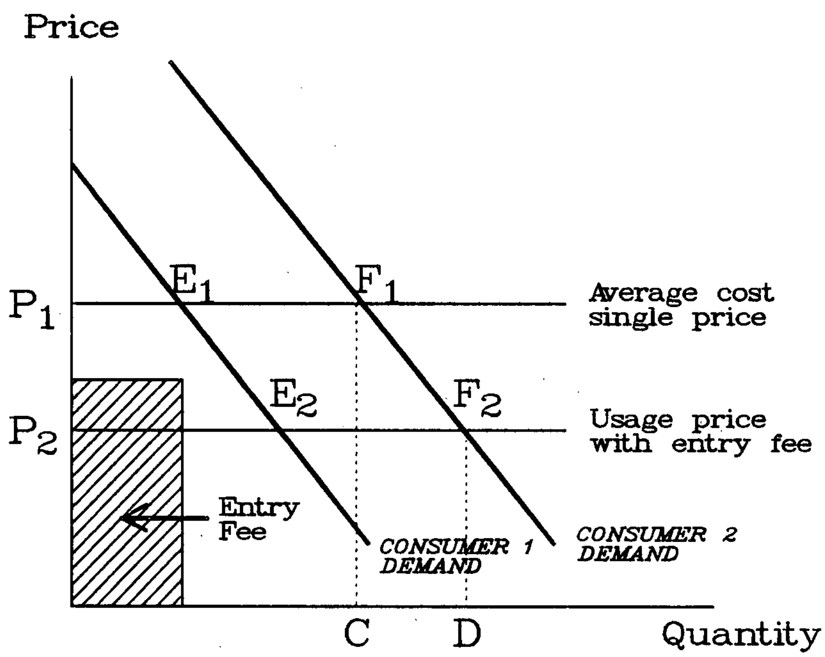


FIGURE 9 OPTIONAL TWO PART TARIFF

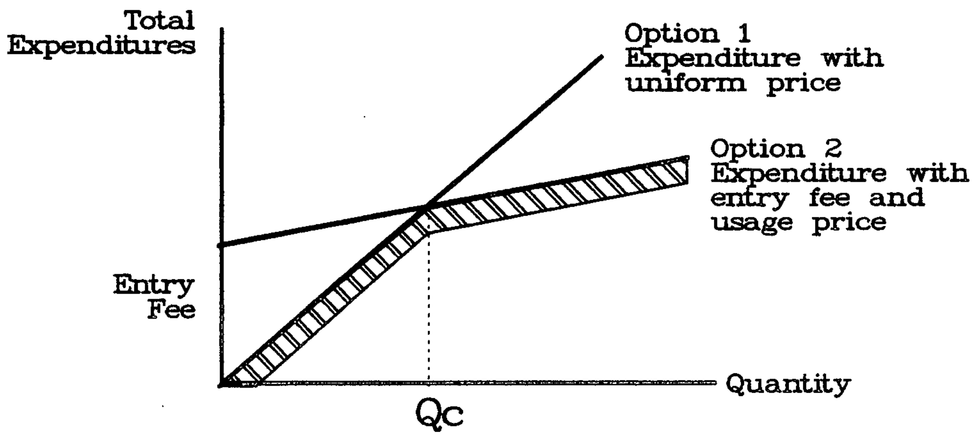


FIGURE 10 EXPENDITURE WITH OPTIONAL TARIFFS

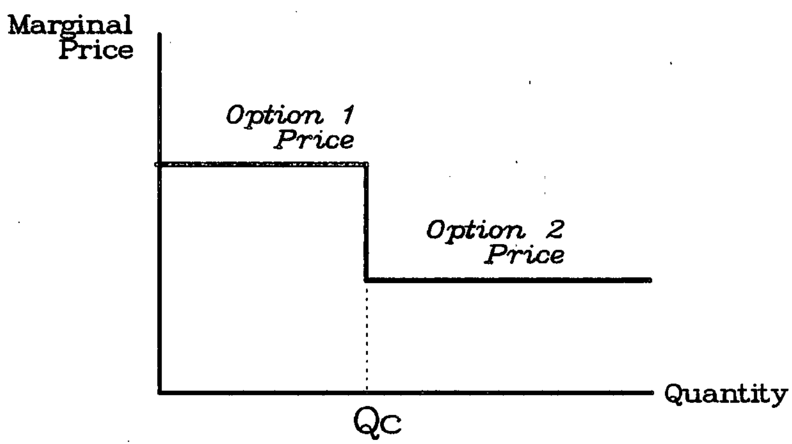


FIGURE 11 MARGINAL PRICE WITH OPTIONAL TARIFFS

More sections on the declining block rate schedule would be generated by adding further optional tariffs. As long as the usage charge is above marginal cost, the effect of adding more options is to increase the efficiency of the system, allowing consumers more choice of where they want to be on the price system.

Willig (1978) showed that it is always possible to construct a non-uniform price schedule that will improve efficiency over any uniform price system charging a price greater than marginal cost. He also demonstrates that no consumers will be made worse off than they were under the uniform price. The fact that nobody will be made worse off makes the adoption of such a price schedule easier for policy makers. Economists also consider this a desirable outcome of a project or policy since the issue of compensation and fairness will not have to be addressed. A situation where one or more individuals is made better off without anyone else being made worse off is known as a Pareto improvement in the economy. Brown and Sibley (1986) show that it is often possible to obtain a Pareto improvement through adding more blocks to an existing multipart tariff. The additional blocks should charge a usage price less than previous blocks but greater than marginal cost. The addition of new blocks can be done with relatively little information about the consumer demand functions.

The improvement in efficiency from adding more block rates to a multipart tariff suggests that maximum efficiency could be obtained by reducing the size of the blocks and adding more of them until the price schedule becomes a smooth function as shown in Figure 12. A question remains as to the

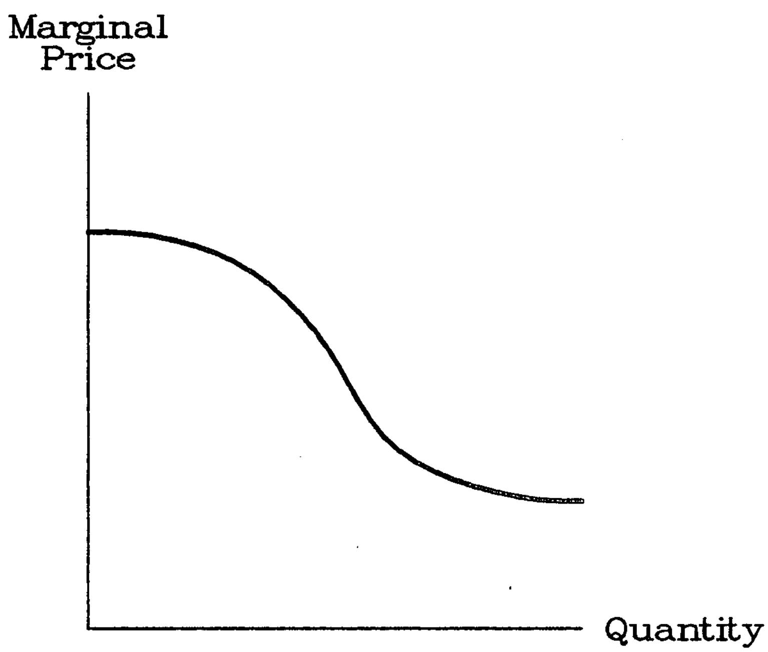


FIGURE 12 SMOOTH DECLINING RATE SCHEDULE

shape of this smooth function. For example, should it be steeply sloped or gently declining, or should it be convex or concave? It is possible to determine the optimum function using the Ramsey principle of increasing prices where demands are more inelastic. At each point on the quantity axis of the price schedule there is a corresponding price elasticity. The size of this elasticity depends on the relative distribution of demands by individual consumers. Using the Ramsey principle, a utility would charge higher prices at quantities where demand is inelastic. At each point on the optimum price schedule equation (7) should apply.

$$\text{(7)} \qquad \text{mark-up} = \frac{P - MC}{P} = \frac{m}{\varepsilon}$$

where:

P = price charged

MC = marginal cost of supply at Quantity Q

m = constant that is set high enough so that firm breaks even

ε = elasticity with respect to price at quantity Q

The major difficulty in determining the optimum smooth price schedule is finding the elasticity at each point on the quantity axis. To do this requires knowledge of the demand curve for each individual consumer. The problem can be simplified somewhat by assuming that each consumer has a similar sloped demand curve, but varies in the absolute amount consumed at a given price. However, the relative distribution of consumers according to quantity consumed would have to be known. An extremely good data set on individual (or small groupings of individuals) would be required before an analyst could construct an optimal smooth price schedule with reasonable accuracy.

In practice, declining block pricing can cause inefficiencies if peak period usage is aggregated with non-peak period usage for billing calculations. As a result the heavy peak period users would pay a lesser price than the lower non-peak users. This would be inefficient because the peak period users are responsible for the capacity requirements of the system and thus have higher marginal costs than the off-peak users (see Chapter IV).

E. Summary

In this chapter it has been shown that marginal cost pricing, although efficient in terms of maximizing consumer and producer surplus, may not result in the regulated firm breaking even and a loss or profit might result. Subsidization of the regulated firm may not always be feasible or even desirable from an economic and social point of view. Therefore, other pricing schemes were examined to determine if losses to the firm could be prevented while at the same time obtaining the maximum benefits from water use. These pricing schemes can also be used as methods of distributing excess revenues back to consumers if the utility is in a net profit situation.

The principle of Ramsey pricing allows a mark-up in certain inelastic segments of the market, so that the firm can cover fixed costs and break even. By charging higher prices in inelastic portions of the market, less deviation from the benefit maximizing marginal cost price solution occurs. The Ramsey principle could be applied to uniform price schedules if the market could be segmented into different consumer groups with different elasticities of demand. It could also be applied to non-uniform tariffs (two part and multipart tariffs) in order to achieve full cost recovery while obtaining the maximum benefits possible. A considerable amount of data about consumer demand and the distribution of consumption among consumers would be necessary before optimal non-uniform tariffs could be derived.

It was also pointed out that the commonly occurring block rate structures often used in Canadian utilities represent a set of optional two part tariffs. This type of price schedule, if designed correctly, can be very efficient; ensuring full cost recovery while benefiting both low consumption and high consumption consumers. The potential efficiency increases as more rate blocks are added to the price schedule. The most efficient such schedule would be one that is completely smooth. To design the most efficient smooth function, considerable information about the slope and distribution of individual consumer demands is required.

One encouraging result from the theory reviewed in this chapter is that a simple two part Coase tariff can often lead to the same total benefits from water use as would occur with marginal cost pricing, as well as ensuring complete cost recovery by the regulated firm. For the Coase tariff to result in the maximum benefits, it is required that the connection charge or entry fee not cause any marginal consumers to leave the market. For most water utilities, this would not be a problem, given the universal requirement for water and the lack of alternatives in many municipal markets. Furthermore, Coase pricing usually will not require detailed information about demand in various markets and at various prices. Coase pricing would, therefore, warrant serious consideration in applications of pricing to water demand management.

The principles discussed in this chapter have been based on static models, without consideration of changes over time. The next chapter examines the implications of increasing demands over time and the resulting needs for expansion of system capacity. The effects of pricing on the timing of system expansion and in managing peak loads are discussed.

IV. WATER PRICING AND SYSTEM CAPACITY

The correct pricing policy for a water supply system is dependent on both the current capacity and future needs for expansion. In fact, pricing and capacity expansion should be considered simultaneously since the type of pricing will effect the needs for future capital investment. Because of the high costs of capacity construction for most utilities, there has been considerable emphasis in the literature on the joint problem of capacity expansion and optimal pricing.

This chapter outlines several issues related to system capacity and marginal cost pricing. A discussion is presented comparing short run and long run pricing policies, and the concept of long run marginal cost is presented in more detail. It is shown how the short and long run marginal cost curves can be used to derive an optimum pricing system when peak loads cause a strain on system capacity. Finally, the issues of system expansion and price adjustments over time are addressed.

A. Short Run and Long Run Marginal Costs

The difference between short run and long run marginal costs can be illustrated by the case of a utility with a fixed maximum capacity. In the short run, output cannot be increased beyond the maximum capacity. In the long run, the system can be expanded so that output can be increased beyond the short run maximum. An example of a fixed maximum capacity would be a water distribution system whose maximum delivery capacity is limited by the intake pipe diameter. In the short run, the water supply can be increased to maximum capacity with only small incremental labour and pumping costs. Once maximum capacity has been reached, the diameter of the intake pipe will have to be increased in order to increase output. The long run marginal costs of increasing output will therefore have to include the capital costs of increasing the intake pipe capacity.

Figures 13 and 14 illustrate the case where capacity is fixed in the short run. Figure 13 shows the total short and long run costs for the output produced. The short run total cost curve rises vertically after the maximum capacity point Q^m, indicating that output cannot be increased beyond this point. the short run curve also begins at a point FC on the vertical axis that represents the short run fixed costs. The long run total cost curve begins at the origin since all costs are variable in the long run. After point Q^M, the costs of increasing output are based on the long run cost curve.

Figure 14 shows the short and long run marginal cost curves derived from the total cost curves in the previous figure. The short run marginal cost curve is horizontal beginning at level V, which is the slope of the short run total cost curve, until the maximum capacity point is reached at which point it becomes vertical. The long run marginal cost in this particular instance is horizontal starting at point LMC, which is the slope of the long run total cost curve, although it could be increasing or decreasing depending on the production process.

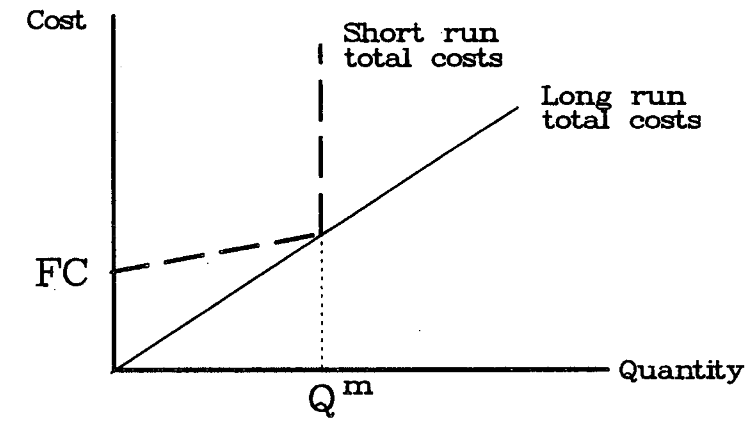


FIGURE 13 TOTAL COSTS WITH CAPACITY CONSTRAINTS

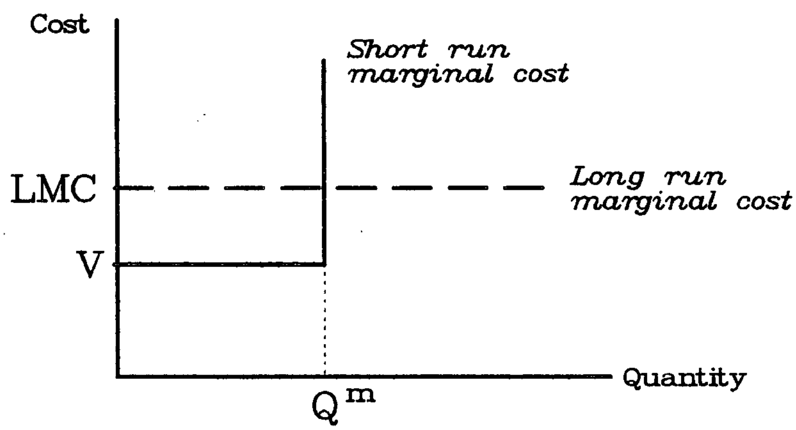


FIGURE 14 MARGINAL COSTS WITH CAPACITY CONSTRAINTS

B. Short and Long Run Marginal Cost Pricing

It becomes necessary to distinguish between short and long run marginal cost pricing when demands are increasing over time. If demands were stable, there would be no need to address the issue of future expansion and price could be based on short run marginal cost. Any time there is a shift in demand, represented by a shifting of the whole demand curve, system managers must consider the issues of price adjustment and capacity expansion. The following sections discuss implications for pricing when demands are increasing over time. Initially, only two time periods are considered: the short run when capacity is fixed, and the long run when capacity can be expanded and demand has increased.

The short and long run demand curves are shown in Figure 15. Using these demand curves, specific short and long run marginal cost pricing rules can be derived. The optimum short run price is the intersection of the short run demand and marginal cost curves. The resulting quantity, Q_s is feasible because it is below the maximum capacity of the plant shown by Q_m. In the longer run the optimum price, Q_L is given by the intersection of the long run demand curve and the long run marginal cost curves. Note that the long run demand curve exceeds the short run capacity, so that increased capacity must be constructed.

In the long run situation shown in Figure 15, the firm will completely cover its costs. The price, which is set at long run marginal cost, is also equal to long run average cost, so the firm just breaks even. The equality between long run marginal and average costs occurs because long

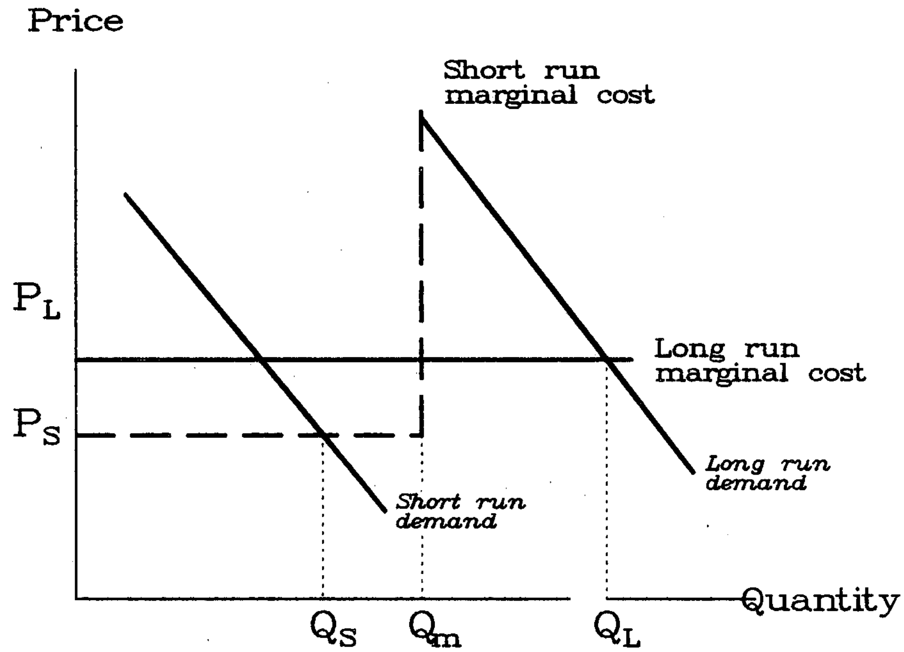


FIGURE 15 OPTIMAL SHORT RUN AND LONG RUN PRICES

run marginal costs are assumed to be constant (constant returns to scale). It is possible, however, that the long run marginal cost could be declining. In this case setting the price equal to long run marginal costs would mean that the price charged would be less than average costs and the firm would lose money. Conversely, if long run marginal costs are increasing, setting the price equal to marginal cost will result in the firm making a profit. It is often assumed that utilities are subject to constant returns to scale, so the marginal cost pricing rule will result in the firm just breaking even.

The deviations from the breakeven point will likely be less in the long run than in the short run even if the firm is not subject to constant returns to scale. In the short run, the losses usually occur because the fixed costs of producing the output are ignored in the optimal pricing rule. In the long run these costs are variable and included in the long run marginal cost. Only if the firm were experiencing dramatically increasing or decreasing long run marginal costs would a large deviation from the breakeven point result.

When pricing is considered in a planning context, long run costs and demands must be projected. It is likely that long run costs can be estimated with greater accuracy than demands can be forecast. Some error in demand forecasting is probable, so that the plant may again end up with either excess or under capacity. If it turns out that the plant does have excess capacity, as shown in Figure 16, then the optimum price to charge would be P^* which, although less than long run marginal cost, ensures that

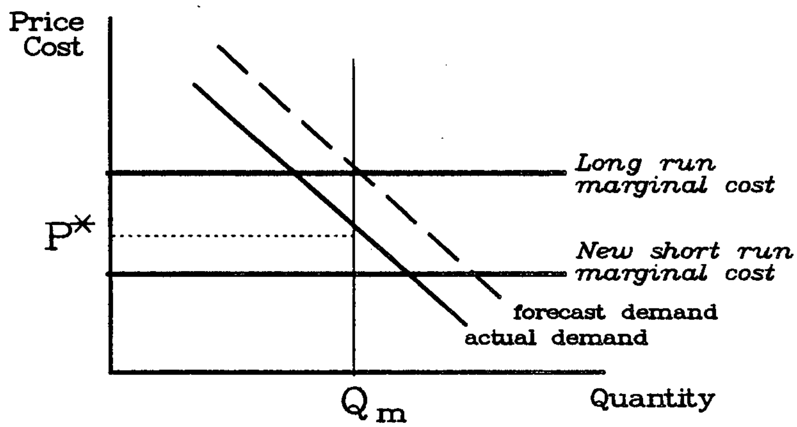


FIGURE 16 EXCESS CAPACITY
DUE TO OVER-ESTIMATED DEMANDS

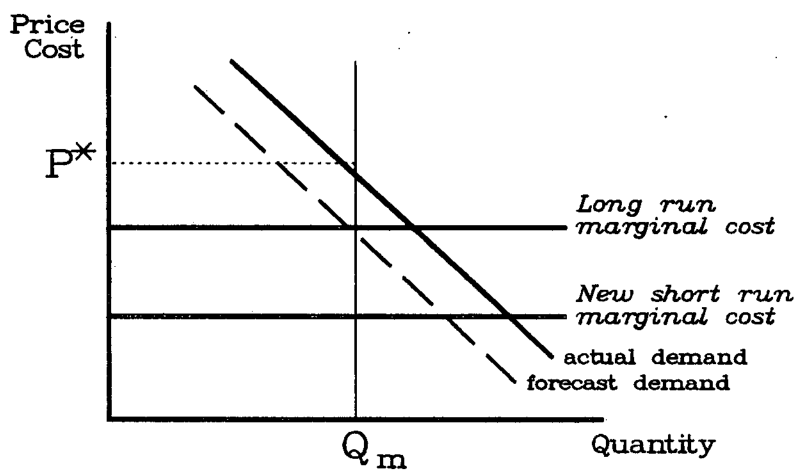


FIGURE 17 CAPACITY SHORTAGE
DUE TO UNDERESTIMATED DEMANDS

maximum capacity, Q_m is used. This will result in a short term loss to the firm since the full costs are not covered by a price less than long run average cost. It is not feasible to charge a price equal to short run marginal cost, since this would result in demand being greater than the maximum capacity.

If demand is underestimated when capacity is expanded, then a price higher than long run average cost will have to be charged in order that plant capacity not be exceeded, as illustrated in Figure 17. A price equal to P^* would have to be charged in order to constrain demand to plant capacity. Since P^* is above the long run marginal cost, the utility will make a profit on its sales. The size of the profit (or loss) realized be the utility is a function of the accuracy of the demand forecasts. If demand is forecast accurately then there will be neither a profit or loss and the firm will just break even. This emphasizes the importance of obtaining accurate demand function estimates and growth forecasts.

C. Peak Load Pricing

The relationship between long run marginal cost and price, shown previously in Figure 15, also provides a conceptual basis for the problem of peak demands. Peak demands often create a problem in water supply systems since the amount of water demanded is likely to be highly dependent on the time of day or season of the year. Peak demands may be high enough that most of the existing or planned delivery capacity is constructed for the purpose of supplying these peak periods. Intuitively, it would seem reasonable to charge the peak period users with more of the costs since their demands have resulted in the increased need for delivery capacity. In contrast, the off-peak users do not require the use of the extra capacity and so should be charged a lower price.

Figure 18 can be used to represent a two period peak demand model. Two demand curves are shown; the high peak demand and the lower off-peak demand. This is a similar to the short and long run model that was shown in Figure 15. The off-peak period demand is analogous to the short run demand and the peak period demand is analogous to the long run demand. Using this model, the off-peak users are charged a price of P_l and the peak users are charged the higher price of P_h, and the respective quantities demanded are Q_l and Q_h. The high price paid by the peak period consumers is based on the long run marginal cost and therefore includes the marginal cost of the necessary capacity as well as operational costs. When planning prices and plant capacity in a peak load situation, the pricing rule for peak periods can be restated as in equation (8).

(8) Price = MC + MCE

where:

MC = short run marginal cost

MCE = marginal cost of capacity expansion

For the off peak periods, price is set equal to marginal cost since no extra capacity is required to meet demands.

The peak period pricing model shown in Figure 18 will result in the maximum benefits from water use. It will also result in conservation of water and

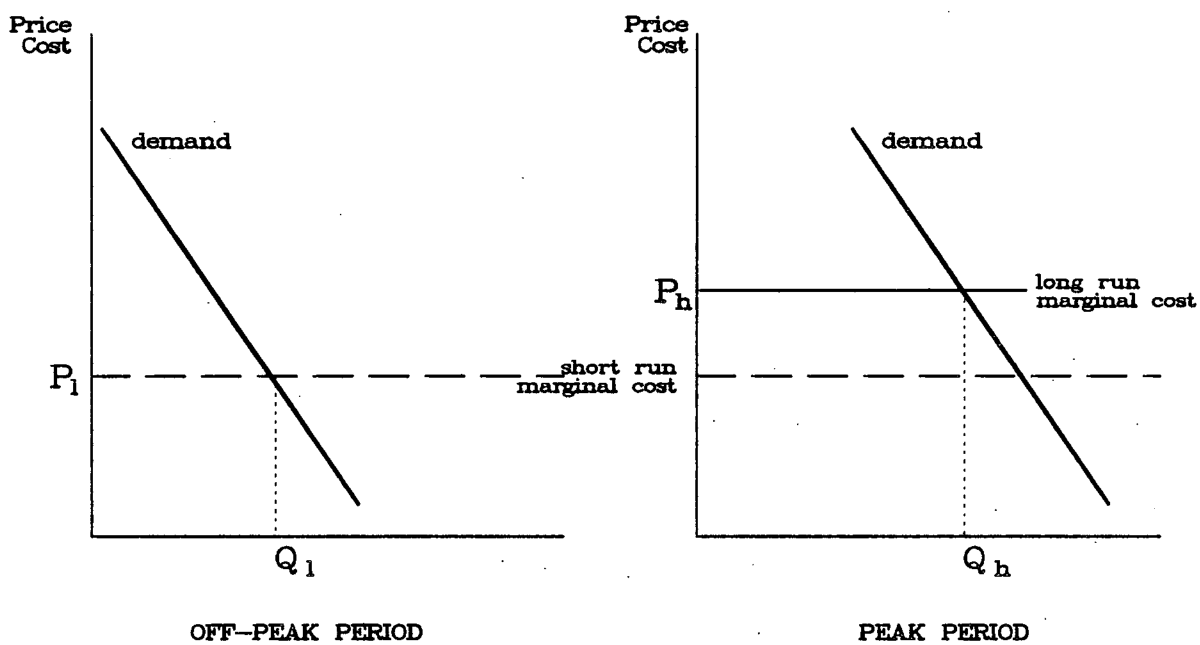


FIGURE 18 PEAK LOAD PRICING

reduce the amount of capacity needed to service demands. If a different pricing approach were used, for example a single price levied on users in all seasons, a loss in efficiency would result. A single price system would benefit peak users at the expense of off-peak users and would result in a loss in total benefits.

The continuous long run marginal cost curve, as was shown in Figure 15, is difficult to estimate by econometric methods so some means of approximating the long run marginal cost will have to be used. The particular difficulty will be in estimating the marginal cost of capacity expansion, since expansion typically occurs only in large increments rather than by marginal amounts. Some methods of approximating marginal capacity cost have been developed that take into account large increments in planned expansion of capacity.

Herrington (1987) summarizes a number of approximation methods for marginal capacity cost based on planned expansion. Any section of the curve representing marginal cost of capacity expansion can be approximated by the increased capacity costs resulting from a planned increase in water supply. The basic formula can be stated as in equation (9).

$$(9) \quad MCC = \frac{PWSC_1 - PWSC_2}{PWQD}$$

where:

MCC = Marginal Capacity Cost

$PWSC_1$ = Present worth of system costs with one planned expansion

$PWSC_2$ = Present worth of system costs with a different planned expansion

PWQD = Present worth of difference in quantity of water delivered under the different capacities.

The selection of the PWSC1 and PWSC2 variables may be somewhat arbitrary in practice, depending on the availability of cost data on planned expansions. A fairly common method is to consider the effects of reducing growth in demand for a selected single year to zero, with the result that the next planned expansion can be delayed for one year. The delayed expansion costs are then used in the calculation of the PWSC2 variable, while the original planned expansion costs are used for the PWSC1 variable.

An alternative method is to calculate PWSC2 on the basis of zero growth in demand and no expansion costs for a selected planning horizon. The PWSC2 variable thus reduces to zero in equation (9). PWSC1 is based on the current forecast of demand growth over the same planning horizon. Using this procedure, the marginal capacity cost is simply the costs of the next planned expansion, PWSC1, averaged over the expected increase in water delivered above current levels.

A final note of caution is necessary before the peak load pricing rule in equation (8) can be applied. Some public utilities have experienced a shifting peak phenomenon where the peak demand shifts from the high price period to the low price period. This may occur if the price differential is high relative to the demand differential in the original peak and off-peak periods. Referring back to Figure 18, if the peak demand curve were lower and the long run marginal cost curve higher, the peak pricing rule would cause a shift in peak load between the periods. In this case the pricing rule would have to be modified to decrease the price differential between the two periods so that demands would be leveled out in both periods.

Water supply utilities may or may not experience shifting peaks under marginal cost pricing. When peak prices are charged on a seasonal basis, they are unlikely to cause shifting peaks, since seasonal uses such as lawn watering cannot be shifted to the off-peak winter period. The peak load pricing rule, if applied to daily peaks in water use, might cause shifting peaks, since there is more leeway for consumers to shift their water use to off-peak times of the day.

D. Capacity Expansion

In the long run, the questions of optimal pricing, optimal capacity, and optimal timing of expansion must be considered simultaneously. The issue of pricing cannot be separated from the issue of when and by how much capacity should be expanded. As was shown in Figure 15, the optimal output and price over the long run is determined by the intersection of the demand curve and the long run marginal cost curve. In practice, the problem of optimal capacity selection is more difficult because expansion can take place only on an incremental basis. In other words, there is a certain minimum economic level by which capacity can be expanded. Expanding capacity by less than this level may be unfeasible or extremely costly over the longer run. This problem is often referred to as the indivisible or lumpy investment problem.

The lumpy investment situation will lead to problems with both the timing of capacity expansion and the correct pricing regime. Given that expansion can take place only by minimum fixed increments, the timing of the expansion becomes important. If carried out too soon before demands have grown enough to warrant it, the capacity will be unused for an excessive period. If expansion is put off too long a shortage will persist resulting in a loss of benefits. Even if capacity is carried out at the correct time, there will still be a period where capacity will not be utilized fully unless the price is set lower than long run marginal cost. This will be followed by a period where the price is set higher than marginal cost to ensure that capacity is not exceeded.

The decision when to expand capacity is really a benefit-cost decision as pointed out by Rees (1984). Figure 19 illustrates the costs and benefits of expanding capacity by the minimum fixed increment from M_1 to M_2 in a given year. The benefits of the capacity expansion are represented by the area under the demand curve to the right of M_1, which is the original capacity constraint. This area is equal to the triangle M_1,e,g. The costs of expansion include the increase in variable costs equal to the rectangle M_1,M_2,a,d plus the capital costs of capacity expansion equal to the rectangle a,b,c,d. If in a given year the benefits of expansion are greater than the costs, then capacity should be expanded to the new level.

In the example illustrated in Figure 19, it is apparent that the costs of expansion are greater than the benefits. As demands grow over time, the

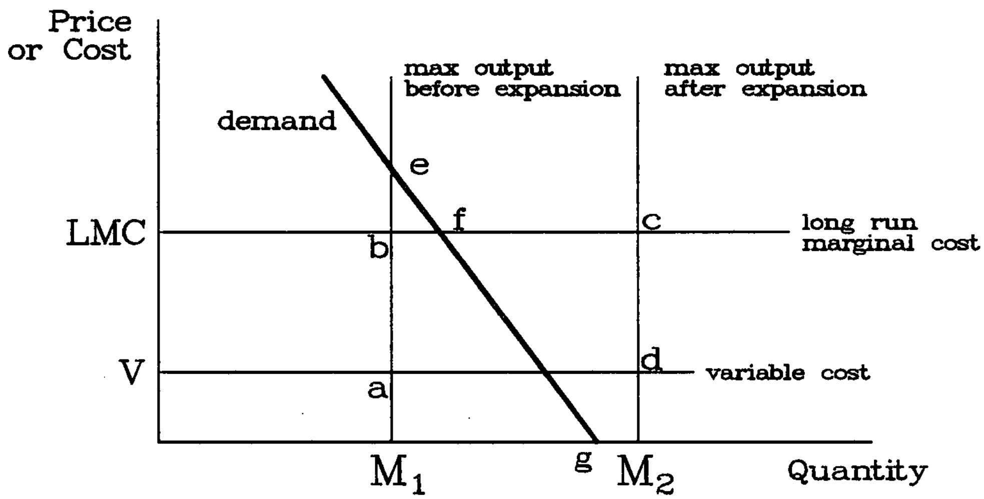


FIGURE 19 COSTS AND BENEFITS OF EXPANDING CAPACITY

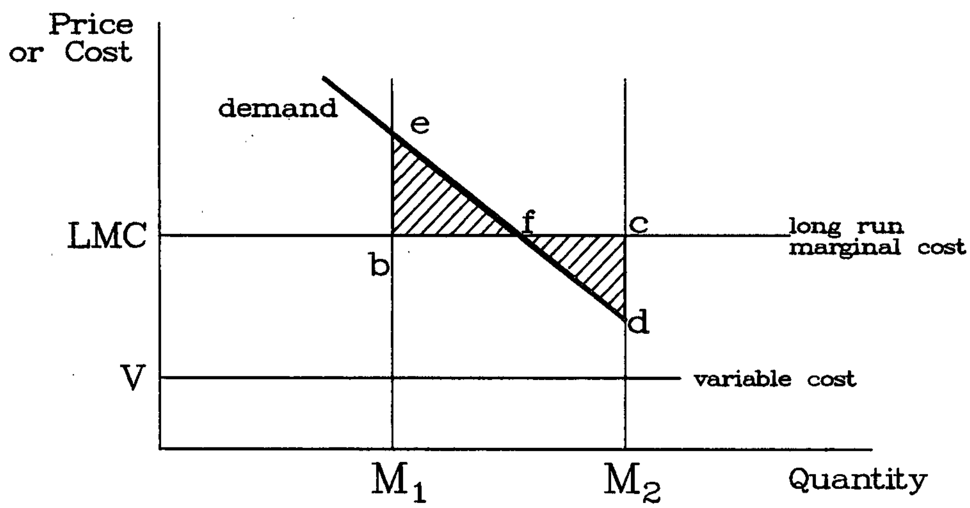


FIGURE 20 OPTIMAL TIMING OF CAPACITY EXPANSION

demand curve will shift outward and the benefits will become larger while the costs remain the same. At some point in time, a critical year will be reached where the benefits of expansion exactly equal the costs. This will be the optimum time to expand capacity. This situation is illustrated in Figure 20. As shown in this figure, benefits are equal to costs when the two triangles b,e,f and d,c,f are equal in area.

The optimal price will follow a definite cycle during the time between capacity expansions. At any time before capacity expansion, the price should be sufficiently high to ensure that capacity is not exceeded. As demands increase over time, this price will have to be raised. Eventually a point will be reached where it becomes economic to expand capacity to the next incremental level. Once this occurs, the price can be lowered to the point where the new capacity is completely utilized. As demands increase, the price will again have to be increased and the cycle will repeat.

Figure 21 illustrates the price cycle during periods of demand increase. Initially, there is a maximum capacity equal to M_1. In year zero, price should be set at P_0 in order not to exceed this capacity. In year one, when demands have increased the demand curve has shifted out to the level D_1, and the price must be increased to P_1 for capacity not to be exceeded. In year two, demands again increase to the level D_2. At this level of demand, it becomes economical to expand to the next capacity level designated by the vertical line M_2. The price can then be set at the lower level P_2, where the new capacity would be fully utilized. As demands increase in

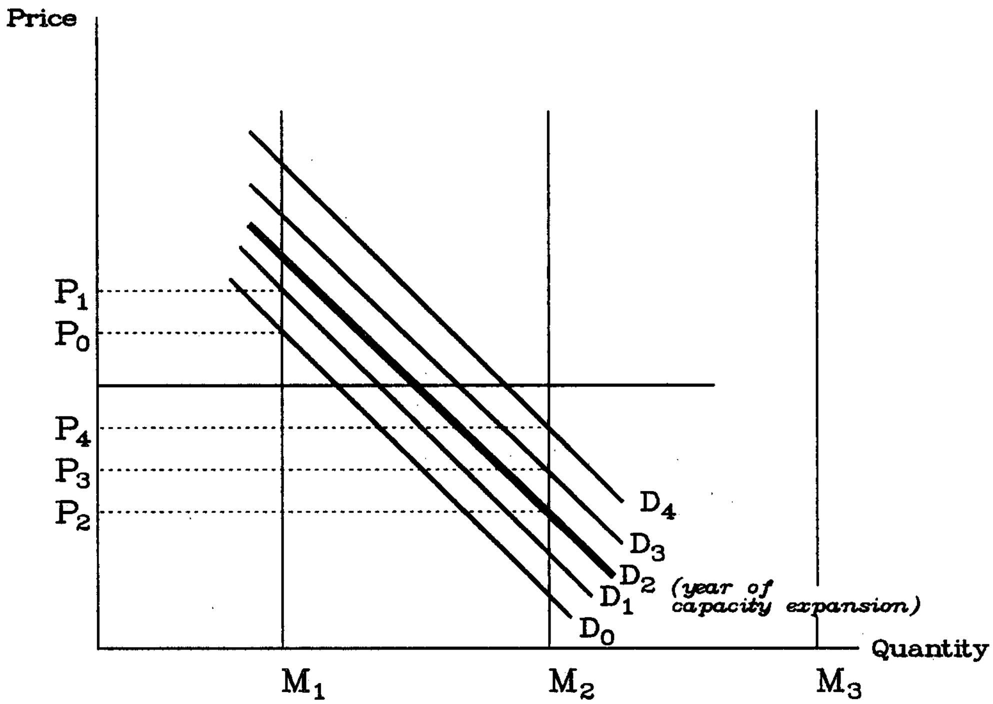


FIGURE 21 PRICE FLUCTUATIONS BEFORE AND AFTER CAPACITY EXPANSION

periods three and four, the price would again have to be augmented. Eventually a point would be reached where it is again economic to increase capacity, this time to the level M_3, and the price could be lowered again.

The price cycle illustrated in Figure 21 is also applicable to peak load pricing. The cycle of increasing prices, followed by system expansion and lower prices would be used in the peak period when capacity constraints are encountered. In the off-peak period, there are no demands made on system capacity, so the short run marginal price would still be used as the basis of the price.

The variability of prices over time may be considered undesirable by the policy maker, but there are possibilities for mitigating the cyclical effect. In the first place, the actual cycle of optimal prices would not likely have the same amplitude as shown in Figure 21. Demands, on average, would probably only grow by one or two percent a year, so the resulting increase in prices due to demand growth would not be dramatic on a year to year basis. A two part Coase pricing tariff might be used to average out the total amount that consumers pay over the years. This would be accomplished by increasing or decreasing the fixed block part of the price over time to balance out fluctuations in the usage charge. A final alternative is to institute some limited quantity rationing and settling for a lesser price increase when capacity becomes constrained. This alternative will result in a less efficient allocation of the available water than a pure price rationing scheme.

Demand and supply fluctuations for water may also show substantial variations from year to year because of weather and run-off conditions. These fluctuations may in fact be much greater than increases in demand caused by long-term population and economic growth. Using price rationing in such cases may not be feasible because of the difficulty in calculating and administering a highly variable price. The problem would be further complicated because of the time it takes consumers to reduce consumption of water in response to a price increase. Thus, it may be necessary to accept some degree of quantity rationing in times of extreme fluctuations in run-off and in weather related demands, despite the loss in efficiency that would result.

The price cycle that was illustrated in Figure 21 will not necessarily result in the utility breaking even. There will be periods when the price charged is below the long run marginal cost, but in other periods the price will be above long run marginal cost. On average, the utilities might be expected to approximately cover costs, but their exact level of profits or losses will depend on a variety of factors such as the slope of the demand curves, the minimum capacity increment and rate by which demands are increasing. In this situation, A Coase two part tariff could be used for adjusting revenues to ensure that the utility breaks even while maintaining efficient allocation of the available water.

E. SUMMARY

This chapter introduced the relationship between system capacity and water pricing. The efficient price level was shown to be a function of existing system capacity. As demands increase over the long term, correct pricing would lead to optimal timing of capacity expansion and maximum benefits from water use.

Much of the theory in this chapter rests on the concept of long run marginal cost. Long run marginal cost is defined as the marginal cost of an additional unit of output when all factors of production (including capital) can be varied. It becomes relevant to pricing when demands are increasing over time and the short run plant capacity is exceeded, thus requiring expansion of system capacity. The general pricing rule derived was to choose the price that would result in an equilibrium between long run demand and long run marginal cost.

Long run marginal cost pricing provides a framework for peak load pricing. Peak load use often accounts for most of the system capacity required in a utility. The peak pricing rule should then be applied with the peak users paying the costs of the extra plant capacity, with the non-peak users paying the short run marginal cost. This pricing rule should be applicable to water pricing utilities, since peak period usage is a common phenomenon in water demands.

The fact that system expansion can only be carried out by minimum increments will lead to problems in strict application of long run marginal cost pricing. Because utility capacity can typically only be expanded by "lumpy" increments, there will be periods where there is either over capacity or under capacity. During periods of under capacity, the price will have to be set higher than the long run marginal cost. Conversely, during periods of over capacity, a price lower than long run marginal cost would have to be charged so that the capacity would be completely utilized. This will lead to a price cycle where prices increase over time until capacity is expanded, at which point the price will be lowered again. Possibilities for mitigating this price cycle exist through the use of two part tariffs.

The long run marginal cost pricing models discussed in this chapter do not ensure full cost recovery by the utilities. However, they are more effective in cost recovery than the short run models discussed in earlier chapters because some consideration is given to capital costs. Whether costs are fully recovered or not depends on a number of factors including the slope of the long run marginal cost curve, the accuracy of demand forecasts and timing of capacity expansion. Approaches for recovering costs discussed in the previous chapters such as Ramsey pricing and multipart tariffs can be applied to long run pricing situations. Again the possibility of using Coase two part tariffs to ensure full cost recovery while maintaining efficient pricing would seem especially applicable to water utilities.

V. WATER PRICING - SPECIFIC PROBLEMS

The previous chapters presented selected economic theory that could be generally applied to water pricing. Water, when considered as an economic commodity, has many unique qualities that must be considered before water pricing can be implemented. In the first place, water is not manufactured. The supply is for a large part dependent on nature, although man-made storage can mitigate seasonal variations in supply. In many areas there are absolute restrictions on the amount available and quantity restrictions have to be considered. On the demand side, the picture is also complex. Some demands are consumptive, while other uses such as fish habitat and hydro-power generation only use the water temporarily and then return it to the system. To take into account all the linkages between supply and demand, water pricing should be considered in the context of a complete water system or river basin. The objective of this chapter is to discuss how the theory of efficient pricing can be used in specific cases that are unique to water supply and demand.

A. Quantity Constraints

In some instances the water supplied by a utility may be affected by the absolute amount of water available. For example, reservoirs fed by snow-melt are limited by the physical amount of snow available. In some seasons, a river basin may not have enough natural run-off or water storage to meet all requirements. Some alternative sources of supply may be available through such mechanisms as interbasin transfer or through bulk shipment, but these can be prohibitively expensive. A reasonable alternative for coping with with shortages would be to use pricing as a means of reducing the amount of water demanded.

If there are physical constraints to the amount of water available, do the same principles of efficient pricing discussed in previous chapters still apply? To examine this question, first consider the case where the marginal cost of supplying water is zero or negligible. Figure 22 presents the case where there are two consumers of water and an absolute maximum on the water supply represented by Q_m. Total water use by the two consumers must be less than or equal to water available. The problem is to find the pricing scheme that maximizes total benefits of water use while satisfying this constraint. It turns out that the optimal price is P^*, a single price charged to both users, as shown in Figure 22. This price is chosen such that the sum of Q_1 and Q_2 is equal to Q_m. The reason why a single price is optimal goes back to the original pricing rule presented in Chapter I, which stated that price should be equal to marginal willingness to pay. By charging the same price for both users, their marginal willingness to pay will be equal at their equilibrium consumption point. The optimal pricing rule when marginal costs are zero is expressed by equations (10) and (11).

(10) $P_1 = P_2 = P_3 = P_n$

(11) $Q_1 + Q_2 + Q_3 + Q_n \leq Q_{max}$

where:

P_1 to P_n = prices paid by customers 1 to n

Q_1 to Q_n = quantities consumed by consumers 1 to n

Q_{max} = maximum quantity available

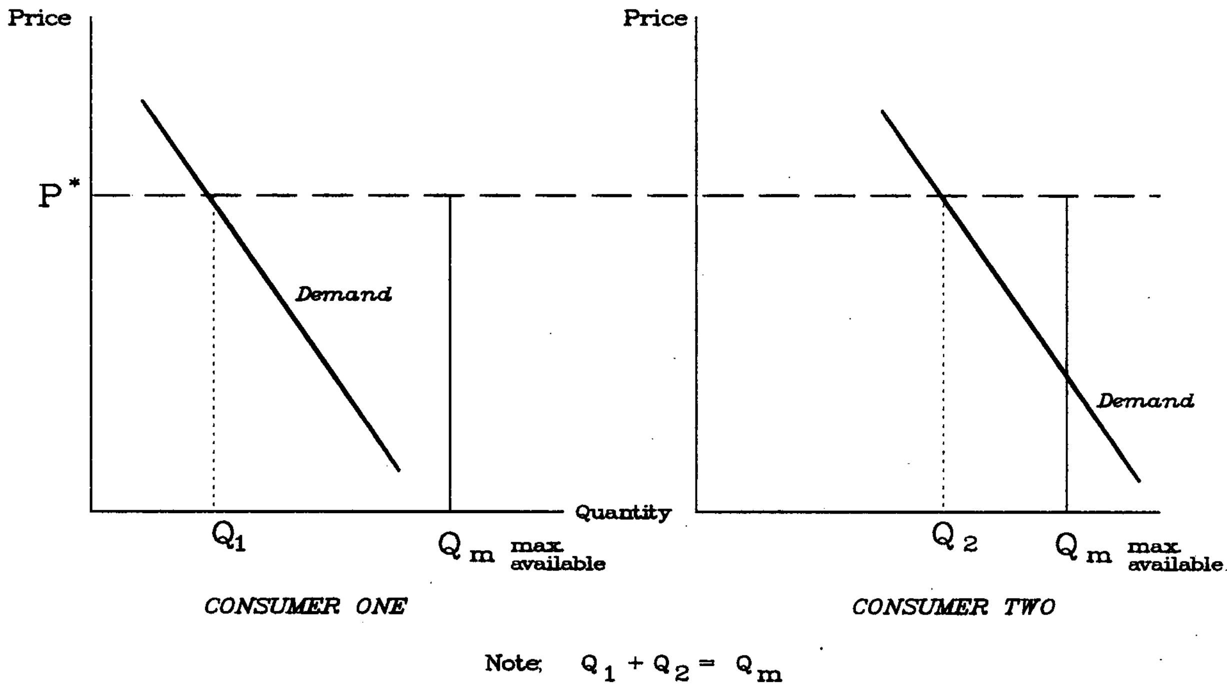
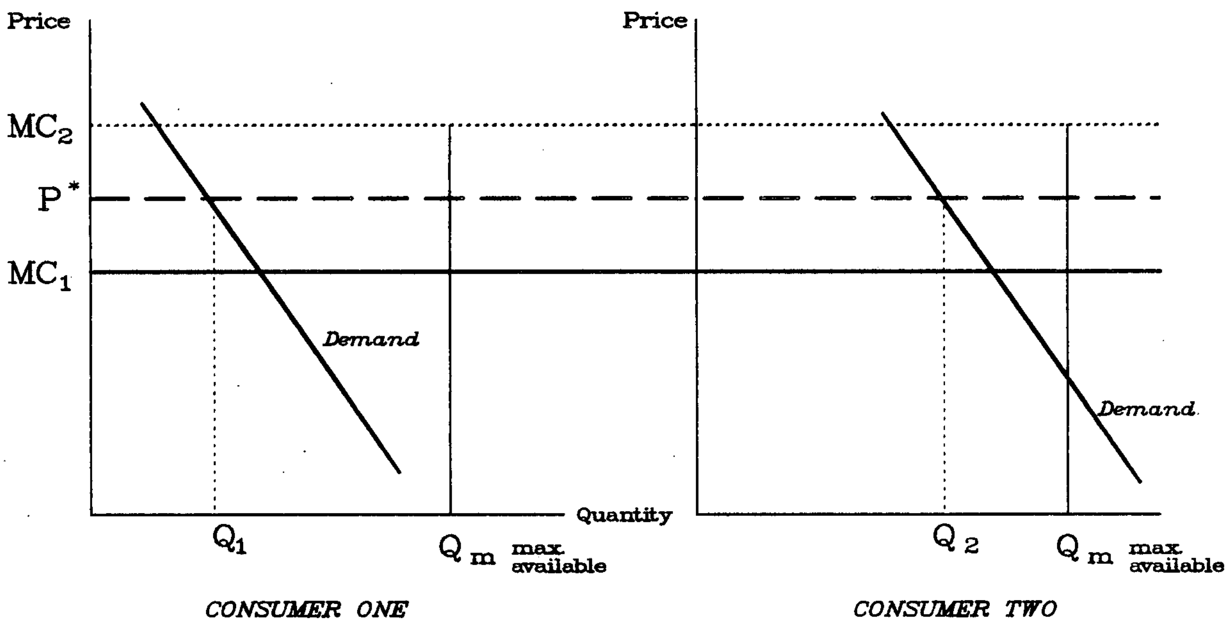


FIGURE 22 OPTIMAL PRICE WITH QUANTITY CONSTRAINT AND ZERO MARGINAL COST

What happens if there is a non-zero marginal cost as well as a maximum quantity constraint on the amount of water available? The answer depends on the magnitude of the marginal cost relative to the price that would just satisfy the quantity constraint. This situation is illustrated in Figure 23, which shows two possible levels of marginal cost, MC_1 and MC_2. MC_1 is below the critical price P^*, and cannot be used as a basis for setting price, because the total quantity demanded would be greater than the maximum amount available. Therefore, one would have to revert to the optimal pricing rule as expressed in equations (10) and (11) and find a higher price that would just result in all of the available water being used. MC_2 is above the critical price P^* and can therefore be used as the basis for setting the price. At price MC_2, the quantity demanded is less than the quantity available and total benefits from water use would be at their maximum.

We can more generally state our pricing rules for maximum quantity constraints as follows:

1. Price should be set so that the marginal willingness to pay is the same for all users.
2. If the marginal cost is greater than the critical price P^* that would result in all of the water being consumed, then price should be set at marginal cost.
3. If the marginal cost is less than the critical price P^* that would result in all of the water being consumed, then price should be set at P^*.



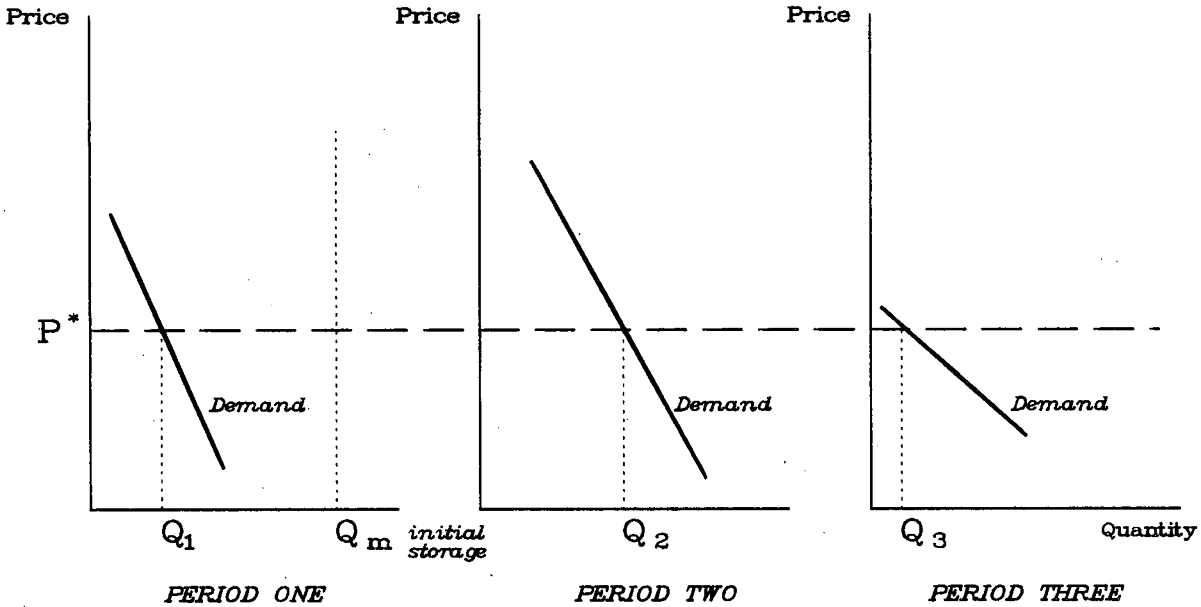
Note; $Q_1 + Q_2 = Q_m$

FIGURE 23 OPTIMAL PRICE WITH QUANTITY CONSTRAINT AND NON–ZERO MARGINAL COST

B. Water Storage

The fact that water can be collected and stored in a reservoir for use in a later time period has ramifications for pricing of water use. Selection of efficient prices becomes a dynamic management problem, where pricing decisions in one time period are not independent of pricing decisions in subsequent periods. Therefore, an optimal price schedule covering all the time periods must be selected. In some cases this can be done in a fairly straightforward manner using the marginal cost pricing principles. In other cases a detailed analysis of the storage capacity, water available for recharge and consumer demand for water must be simultaneously considered.

The simplest case is a reservoir with a limited storage capacity that is only filled once per year, with no intra-year recharge. An example of this would be a reservoir in a dry area that is charged in the spring by snowmelt, but receives no further water during the rest of the year. All yearly demands must be supplied by the initial amount of stored water available in the spring. The storage will allow the water to be allocated over the course of the year. Figure 24 illustrates such a case where there are three time periods in the year each with a unique demand curve based on seasonal requirements. It can be seen from Figure 24 that such a scenario is analogous to the case of three separate users facing a maximum quantity constraint. The beginning storage level Q_m, is equivalent to an absolute maximum of water available. The three seasonal demand curves are equivalent to three separate demand curves in a single period. The sum of the quantities consumed in each period, Q1, Q2, and Q3, must be less than the beginning storage level. Therefore, equations (10) and (11) can be



Note: $Q_1 + Q_2 + Q_3 = Q_m$ *(initial storage)*

FIGURE 24 OPTIMAL PRICE OVER THREE TIME PERIODS used to determine the optimal price. The solution would involve setting the same price in all three time periods such that the total live storage of the reservoir is depleted by the end of the last period. By setting a common price, the marginal willingness to pay will be the same in all time periods and total benefits will be maximized.

Optimal pricing with storage becomes more difficult when there is some recharge of the reservoirs during the year. It is still possible to use the basic strategy of trying to ensure that the marginal willingness to pay is the same in all periods while utilizing the maximum amount of water possible. This strategy, however, may not result in maximum benefits from water use if there are large variations in the natural supply between periods. For example, if there is a large recharge that occurs only in the last period, it might be desirable to lower the price in the final period to ensure that all of the water supplied by the recharge can be used. Alternatively, one could operate the storage in order to utilize as much water as possible in the previous periods and then rely on the recharge to meet the needs in the final period. By doing so, it might be possible to ensure that the marginal willingness to pay is equal in all periods.

The optimal pricing problem described in the previous paragraphs is complex because there are several variables and options to consider. Both demands for water and natural supply may vary from period to period. The manager or policy maker has to consider both the price charged and the quantity released from storage in each period. Furthermore, he is faced with a restriction on the maximum capacity of the reservoir. It may be helpful to take such a problem and divide it into two sub-problems, each of which is more easily solved than the complete problem. The two sub-problems are:

1. Determine the storage operation (the amount of water released for use in each period) that will maximize total benefits,
2. Determine the price in each period that will result in full consumption of the water released in each period.

Determination of the optimal storage regime is a unique problem for each system. The total storage capacity, recharge rates and seasonal demands all enter into the solution. Given this information, it is possible to set the problem up as a mathematical programming problem, where the object is to maximize total benefits subject to a number of constraints. These constraints would represent the physical parameters of the problem such as storage balance, run-off in each period. Of course the demand curves in each period will have to be known in order to determine the total consumer surplus. There are many examples of such maximization problems in the literature, and computer solutions can usually be obtained quite easily if all the data are available.

Once the optimal amount of water released for consumption in each period is computed, the optimal prices can be obtained quite easily. The seasonal demand curves can be examined to see which seasonal prices will exactly result in full consumption of the water released. These will be the optimal prices.

C. Non-consumptive Demands

Because water can often serve a useful purpose without actually being consumed, the ability of pricing to serve as a demand management tool for non-consumptive uses deserves some discussion. It may seem that pricing is not a relevant issue, since it will not effect the amount of water consumed. As an example, consider the use of water for fish habitat. Charging a price to fishermen may reduce the amount of fishing, but will not affect the amount of water available for habitat. If one takes the example of a hydropower generation station, which is a non-consumptive water user, it also seems unnecessary to impose a price, since all the water used will be returned to the system. Nevertheless, it can be shown that pricing may still be an appropriate demand management tool, even for non-consumptive uses.

Although it is not possible to put a direct price on instream uses such as fish habitat, such uses often occur in competition with consumptive uses. If the correct price is applied to these other consumptive uses, the optimum amount of water can be allocated towards each use. Figure 25 illustrates a case where there is one consumptive user and one instream non-consumptive user and a constraint on the total amount of water available. Note that it is still possible to define a demand curve for the non-consumptive use even though price will not effect the quantity of water utilized. The demand curve can be thought of as representing the marginal value product of water for the non-consumptive use. For example, if the instream use is for fish habitat, the demand curve reflects the net value of the extra fish per unit of water available for habitat.

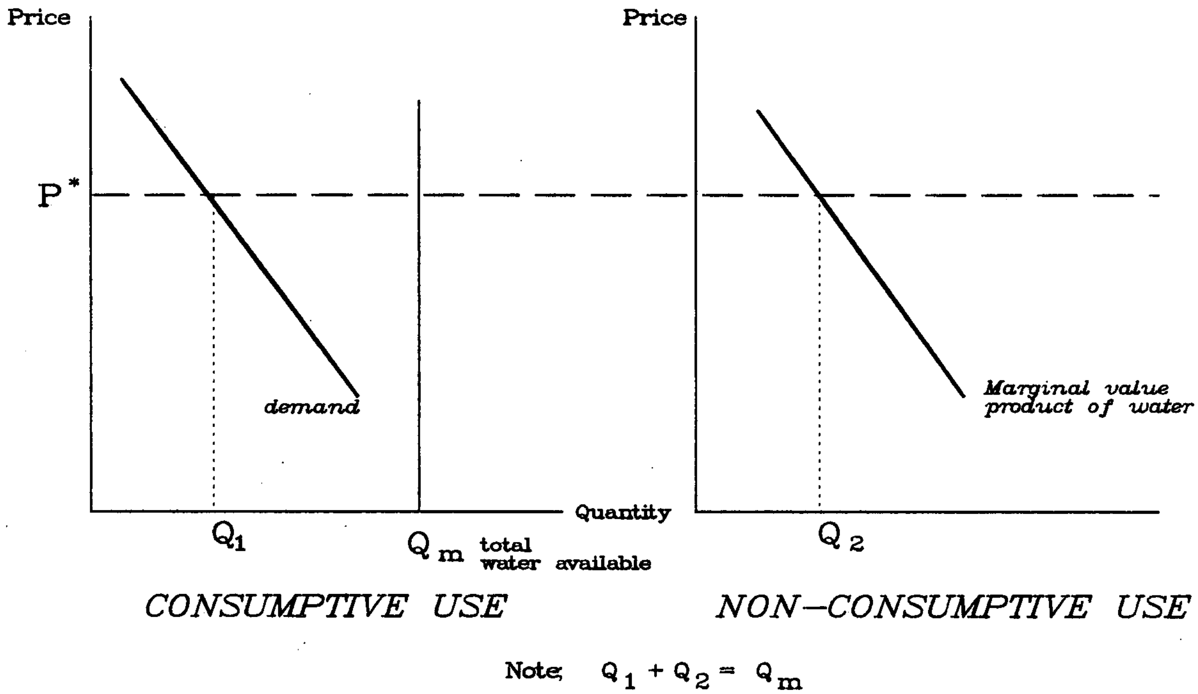


FIGURE 25 COMPETING CONSUMPTIVE AND NON–CONSUMPTIVE USES

In the example shown in Figure 25, the optimal pricing rule with a quantity constraint is still applicable if the two demands occur in a single reach. In this case the demands are competitive in that water diverted for one use will not be available for the other use. The basic strategy will be to select the price, P^* for the consumptive user such that his marginal willingness to pay is equal to the marginal value product for the non-consumptive use. Although the price does not directly affect the quantity demanded by the non-consumptive user, it does so indirectly through its effect on the consumptive user. The price causes the consumptive user to select his own optimum quantity, Q_1 and the remaining water, Q_2 will go towards the non-consumptive user. If a price is selected where the marginal willingness to pay by the consumptive user is equal to the marginal value product of the instream use and all water is allocated, then total benefits will be maximized.

Occasionally a situation is encountered where there are two or more non-consumptive uses competing with each other. Although these uses do not actually consume water, they can be in competition with each other because the timing of their demands is different. For example, instream flows for fisheries occurring in the spring can compete for the same stored water that is required for recreational lake levels in the summer. Since both uses are non-consumptive and non-market, pricing will not have any effect on the allocation of the water and alternative methods of resolving the conflicts will have to be found.

In some instances, a non-consumptive use may not conflict with consumptive uses. For example, if a power station is located upstream of a consumptive use such as irrigation, then the water used for power generation can also be used for irrigation. There would be no sense in charging the power station for water, unless there were some marginal cost involved in delivering it to the station. If there is a non-zero marginal cost of delivering water to the power station, then its price should be set at marginal cost. A price higher than marginal cost would be charged for the downstream consumptive use if quantity constraints were a factor.

D. Private Withdrawals

In several areas of Canada private withdrawals of water are significant and the question of pricing for these uses is important. Most of the concepts developed in previous sections and chapters for public utilities should also apply to private abstraction of water. The major difference that must be considered is that self-supplied water users incur the capital and operating costs of supply.

1. Pricing of Self-Supplied Users

In theory, self-supplied users should be optimizing their consumption by abstracting water until their marginal willingness to pay for it is equal to their marginal cost of withdrawing it. Therefore, no price should be charged to these users if there is no alternative use for the water they use. If there is an opportunity cost to the water because of restricted natural quantities, then it would be necessary to levy an additional usage fee on these users in order to obtain an optimal allocation of water. This additional charge has the effect of rationing the total water available.

The amount that should be charged is shown in Figure 26. It is the difference in the marginal cost of supply, MC, borne by the private user and the optimal price, P^* at which all water is allocated. A similar principle would apply if the alternative uses were non-consumptive. An extra price would be charged such that the marginal willingness to pay by the private user was equal to the marginal value product of water for the non-consumptive use.

When a private user and a public utility withdraw water from the same source, the pricing problem becomes more complex, especially if there is a seasonal pattern to supply and demand for water. An example is a private user who withdraws water in the summer, thereby reducing the amount of water available to the public utility. The utility might have to respond by increasing the storage capacity of its reservoir system to capture more of the natural supply in spring and winter. Thus, there is a marginal capacity cost associated with the private user's consumption of water. This suggests that the peak load pricing rule can be applied to the private user who should then be charged the marginal capacity cost resulting from his water withdrawal in the critical water-short periods.

Even if there are no current alternative uses for the water that is withdrawn by a private user, it may still be desirable to charge a fee for the right to withdraw the water. This would be the case if there were possible future alternative uses for the water licensed to the current user. Although water is a renewable resource and the current consumption is usually not in conflict with future consumption, the rights to future

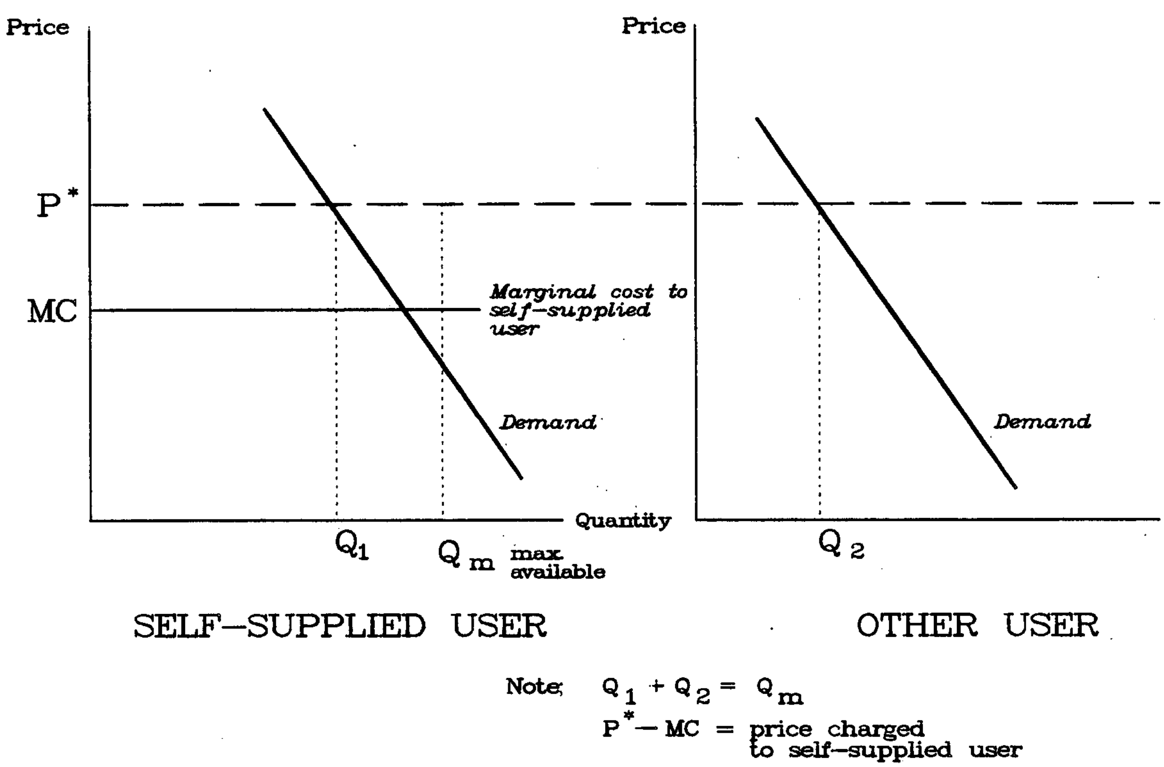


FIGURE 26 QUANTITY CONSTRAINT WITH SELF-SUPPLIED USER

water may be dependent on the current amount consumed or licensed. Under the doctrine of prior appropriation, it is advantageous for a water user to apply for a license for the largest quantity of water possible if there is no charge related to the size of the license. Future inefficiencies would result if new users, with higher marginal willingness to pay, wish to use the same water source, but cannot because of the prior appropriation. The solution to this problem would be to charge a fee to the current user based on the present value of the future value of water to other users.

2. Market Exchange of Water Rights

Self-supplied water users usually have rights to or licenses for the water they withdraw based on prior appropriation or riparian rights. This system can often lead to inefficient allocations of water because the value of the water may be higher for future unlicensed users than for the current licensed users. Water pricing could be imposed on all users to reduce demand and free up supplies for future users, thereby increasing the benefits from water use. Legal problems would still remain concerning the transfer of licenses for the freed up supplies. An alternative to government pricing on private withdrawals would be to allow the buying and selling of water rights on an open market. This would effectively transfer the pricing function from government to the market place.

Under certain conditions, a market system could result in the same benefits from water use as would occur under government regulated pricing. These conditions would require that no single current user be large enough to control or effect the market price. In other words, the market for water rights would have to be competitive. If there were only one licensed owner who was selling rights to users, he could maximize profits by restricting the amount sold, thereby increasing the price. A competitive market would result in the complete use of all available water, a constant price to all users and the maximization of economic benefits.

A market for water rights also allows non-consumptive users such as sport fishermen to have direct influence on the allocation of water between users. If water is supplied by a government regulated public utility, non-consumptive users must rely on the utility to allocate water for non-consumptive uses even though the utility does not receive revenue from such uses. In a market system, the non-consumptive users could levy a tax on their own water based activities and use the proceeds to buy water rights from other users.

In practice, a market system for water rights could be less efficient than a system that uses a single regulated utility to supply water. A market system would be inefficient when the costs of withdrawal and distribution of water are less for a single agency than for many small individual users. This is the natural monopoly situation referred to earlier in Chapter II. In the natural monopoly situation, it is more efficient to have a single utility supplying all the water at a price based on marginal cost than to have many smaller licensees abstracting water and selling it on the open market.

E. Non-Classical Demands

Classical economic theory deals with smooth and continuous downward sloping demand curves. The discussion to this point has used these type of demand curves for convenience. These classical demand curves often will not give a realistic depiction of the demand for water especially in the short run. There may be upper limits on the total water demanded by various sectors due to such factors as limitations on available land for irrigation. The demand curve may not be continuous or smooth due to the underlying nature of the industry that requires the resource. In other cases the demand curve will be practically inelastic within a wide price range.

It is quite possible that certain sectors will have practical upper limits to the amount of water they will consume. For example, the agricultural sector will have an upper limit dependent upon the irrigable land available and the manufacturing sector's demand will be limited by the physical capacity of their plants. These upper limits are represented by a vertical or near vertical section of the demand curve as shown in Figure 27. In the example illustrated, the marginal pricing rule will still result in the maximum benefits from water use. However, a higher price up to the level of P_m could be charged by the utility and maximum benefits would still be obtained, since the quantity demanded would not change. The distribution of the benefits would be effected by this price increase with some of the consumer surplus being transferred to the utility in the form of increased revenues.

If there are a number of users, each of which has an upper bound constraint

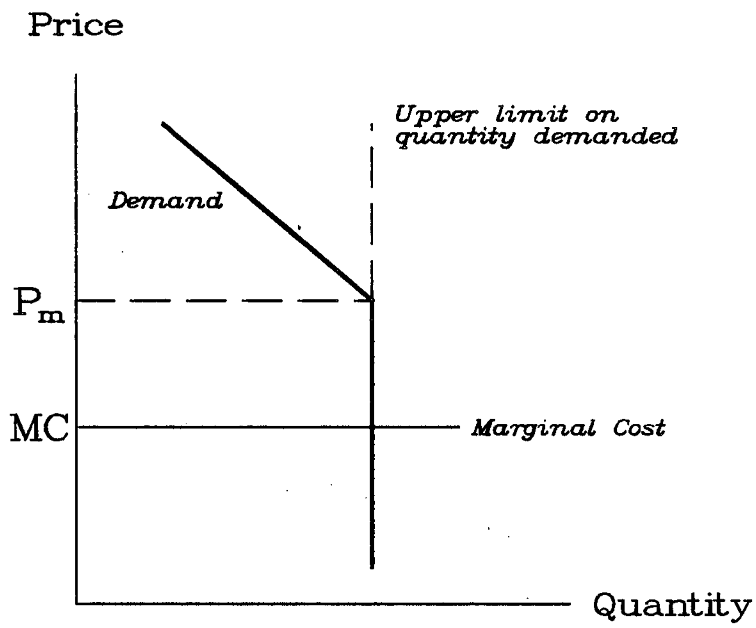


FIGURE 27 CONSUMER WITH UPPER LIMIT ON DEMAND

on his demand for water, the marginal cost pricing rule will still result in the maximum benefits from water use. As in the case of a single constrained user, the utility will have the option to increase prices for some users without affecting the optimal allocation of water. If there is an absolute limit on the amount of water available to these users, then maximum benefits will be obtained when a common price is charged that is just high enough to ensure that all the water is used. This is essentially the same price rationing approach discussed earlier that can be used when there is an absolute limit on water available.

The demand curve for water by certain sectors may be composed of a series of vertical and horizontal sections with a general downward trend. This is likely to happen if there only a few major users in the sector who use water in fixed proportions to certain processes. It could also occur in the agricultural sector if there are distinct blocks of irrigable land, each with different productivity. At very high prices of water, only the most productive block of land would be irrigated. As the price was lowered, the less productive blocks would be added, resulting in a stepwise declining demand curve composed of vertical and near horizontal segments. The marginal cost pricing rule and price rationing approach discussed earlier would generally be appropriate for demands curves of this type.

F. Upstream and Downstream Allocation of Water

In most of the cases discussed in this chapter, the supply of water was assumed to be a single reach or reservoir. Although this may be appropriate for many water supply utilities, it will only provide a partial solution to the optimum pricing of water use in a connected system. Many water supply systems are best considered part of a linked network of water storage, use and supply. In such cases a pricing system must consider the opportunity cost of water at each point in the network, and attempt to maximize total benefits from water use in the whole system.

The same basic principles of efficient pricing will apply in a complete system such as a river basin. The difficulty will be in the actual determination of these prices over the network. If the network is relatively simple, then the prices can be determined in a straightforward manner based on the concepts discussed earlier. Figure 28 presents a simple case where there is a limited water supply that occurs only at the head of the system serving three sequential downstream users. Since the three users are competing for water from a single source of supply, the situation is exactly analogous to multiple use under a quantity constraint. The same rule illustrated in Figure 22 and outlined in equations (10) and (11) can be used to determine the best price. That is, equate marginal willingness to pay by charging each user the same price, P^* such that the total water supply, Qm is completely utilized.

It may well be possible that the network has several supply nodes as well as demand nodes as shown in Figure 29. This case is analogous to the single node multiperiod storage situation where some recharge occurs during the year. The best strategy is to attempt to allocate water to each user such that the marginal willingness to pay is the same and that all the available supply is used. As for the multiperiod case, this will not

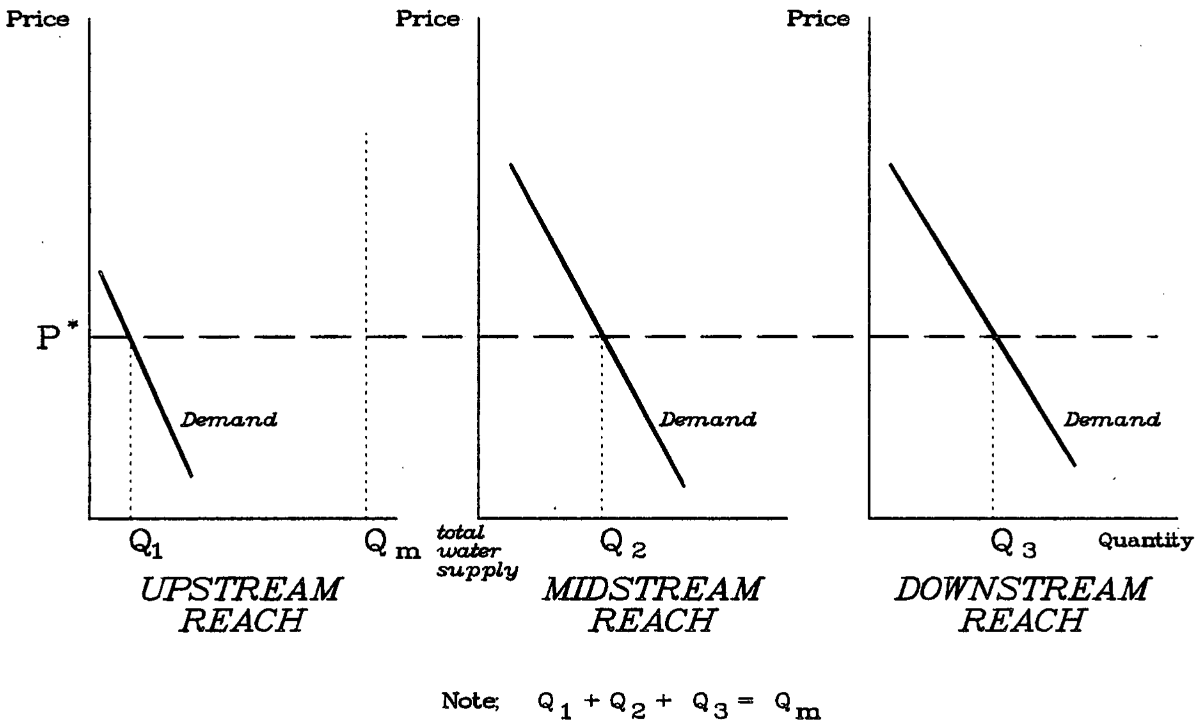


FIGURE 28 DOWNSTREAM COMPETITION FOR WATER

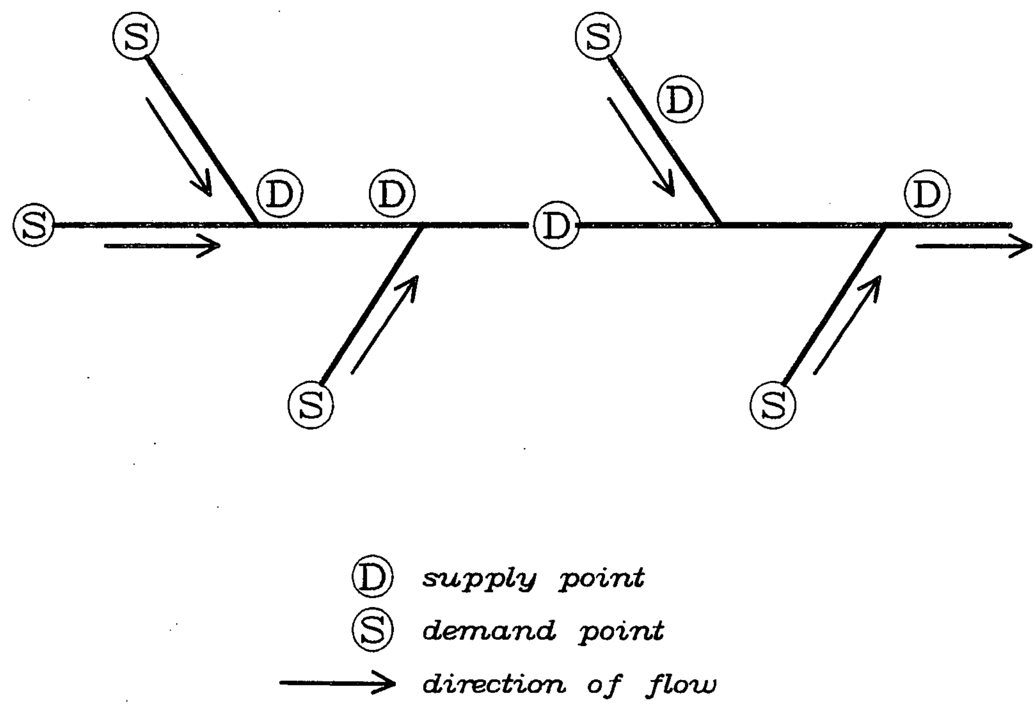


FIGURE 29 SUPPLY AND DEMAND POINTS ON A RIVER SYSTEM

always be possible. For example, if there is a large supply that is only available to a downstream user, it would probably be desirable to lower the price for this user so that he would make use of the extra water.

The pricing problem becomes more complicated when multiperiod use and storage is considered. The separate demand curves in each reach and time period would have to be inspected in order to find a pricing system that maximizes total benefits. At this level of complexity, the set of optimal prices may be too difficult to find by simple inspection of the demand curves and water supplies. It would then be necessary to use the mathematical programming approach described earlier that accounts for linkages between reaches and time periods.

G. Summary

Water has several unique features that must be considered when determining optimal prices. Because it is a naturally occurring, non-manufactured commodity, its supply cannot be perfectly controlled. The allocation and pricing of water must be subject to the natural occurring variance and restrictions of supply. Optimal pricing systems must consider the timing and location of this supply as well as storage management. On the demand side, water is also unique in that value can be obtained from it without actually consuming it. The consumptive and non-consumptive demands have to be considered in arriving at optimal prices.

Some basic strategies were introduced in this chapter. In particular, it is desirable, whenever possible, to charge users the same price in order that their marginal willingness to pay be equal. Where absolute quantity constraints on the amount of water exist, a price would have to be chosen that results in an equilibrium between the amount of water available and the amount demanded. This strategy could be altered to ensure full utilization of water supply that only occurs locally or at specific times and is not generally available to all users.

It is also desirable to consider pricing in the larger context that includes linkages between downstream and upstream users. If there are significant linkages within the system, the optimal price regime for the whole system under study should be determined. In the context of river basin planning and management, prices for each time period and reach should be determined simultaneously, since the optimal use of water at each point or time period is dependent on the opportunities for its use elsewhere. In such situations it will often be difficult to calculate the optimal set of prices using any of the simple rules discussed in this chapter. Mathematical programming models or other optimization techniques will be required to determine both optimal allocations and prices. Despite these difficulties, the optimal prices, if correctly administrated, will still accomplish the objectives of achieving the maximum benefits from water use.

VI. SUMMARY AND CONCLUSIONS

In this chapter the important concepts of the theory of pricing are summarized. Some conclusions are made about the theory most relevant to water pricing that warrants further investigation. Environment Canada is actively promoting efficient water pricing, and some steps in this process are described. A major function in this process is to generate and disseminate the technical information necessary for efficient water pricing. As part of this function, it is recommended that Inland Waters Directorate prepare guidelines on water pricing suitable for use by municipalities and water management agencies.

A. Summary

Optimal pricing can be considered in both a short run and a long run context. In the short run, when adequate system capacity exists to supply all needs, price can be set equal to short run marginal costs. System capacity is fixed and output can only be varied by changing the variable inputs. In the long run, as demands increase or are projected to increase, the quantity demanded will become higher than system capacity if price is set equal to the short run marginal cost. When demands are increasing over time, the long run marginal cost should be used as the basis for setting prices, since the planning process must at some point incorporate an increase in system capacity.

The theory of public utility pricing includes several concepts relevant to water pricing. Foremost among these is the concept of marginal cost pricing, which will theoretically result in the maximum benefits from a commodity produced by a regulated utility. Marginal cost pricing means setting the price equal to the marginal cost of delivering or producing a commodity. At this price the total producer and consumer surplus is maximized. In the short run, marginal cost pricing does not ensure that the regulated firm will break even. Many water utilities can be classified as natural monopolies with marginal costs generally below average costs. For these utilities marginal cost pricing will result in a financial loss.

There are a number of reasons why policy makers might desire that public utilities break even on their operation, and there has been a general regulatory tradition in North America that they do so. Economists have responded to this tradition by developing a fairly extensive body of theory on achieving maximum efficiency subject to a breakeven constraint for the regulated firm. The major concepts in this theory, which can be applied both in the short run and long run, are summarized below.

Pricing systems are considered as either uniform or non-uniform. A uniform price is invariable over the quantity demanded. In other words, high quantity users are not necessarily charged a higher or lower price than the low quantity users, although the price may still vary between different classes of users. For example, agricultural users might be charged less for water than industrial users under a uniform pricing scheme. A non-uniform price varies with the quantity demanded. An example of a non-uniform price system is the common declining block rate price. Both uniform and non-uniform price systems have been developed that can theoretically result in full cost recovery.

Ramsey pricing is a uniform price method of ensuring that the regulated firm breaks even while obtaining the maximum benefits possible. It works by marking up the price charged to various classes of users at different rates depending on their elasticity of demand. Groups with high demand elasticities are charged more than the low elasticity groups. This results in a minimal distortion from the optimal quantity that would be demanded under pure marginal cost pricing. Although there is extensive use of this concept in the theoretical literature, it is difficult to apply in practice because of the detailed knowledge required about the demand elasticities of the various users.

Non-uniform prices present some scope for obtaining full cost recovery with a minimum loss of efficiency. In general, they take the form of multipart tariffs, with different block rates depending on the amount demanded. It was shown that declining block rate tariffs could be a highly efficient method of ensuring cost recovery and retaining the maximum benefits possible. Such pricing schemes are efficient because they allow consumers to sort themselves into the most beneficial pricing block. Increasing the number of blocks on the schedule generally increases the efficiency of the pricing system.

A simple Coase two-part tariff, when used for water pricing, usually results in the same efficiency as pure marginal cost pricing and at the same time ensures full cost recovery. This type of tariff consists of a fixed connection charge and a price related to usage. The principle is to set the price near the marginal cost and to use the fixed connection charge as a means to covering the remaining costs. Coase pricing will be an efficient pricing scheme as long as the fixed charge does not cause significant numbers of consumers to leave the market for the commodity. In most cases it would be an appropriate pricing system for water utilities because consumers are not likely to leave the market because of the fixed connection charge.

In a longer run planning horizon when system expansion must be considered, all inputs, including capital, are variable. In this context the long run marginal cost becomes the theoretical basis for optimal pricing. The long run marginal cost is the marginal cost of producing an additional unit of output when all inputs are variable. Setting the long run demand equal to the long run marginal cost will result in the maximum benefits possible.

The concept of long run marginal cost pricing can also be used as a framework for optimal peak load pricing. In many utilities the existing capacity is required because of peak load demands. The basic peak load pricing principle states that the peak load users should be charged the long run marginal cost while the off peak users should be charged the lower short run marginal cost. Peak load pricing should be considered by water utilities because seasonal demands are often responsible for most of the system capacity requirements.

In practice the use of long run marginal cost pricing runs into some difficulties because water supply systems cannot be continuously expanded from year to year. Usually, expansion is only feasible by minimum fixed increments. The timing of expansion depends on the size of the increment and the rate by which demands are growing, and there may often be several years between capacity expansions. During the intervening years demands will continue to grow so setting the price equal to long run marginal cost will result in excess demand in some years and under-utilized capacity in other years. To ensure that supply and demands are kept in equilibrium during all years, prices will have to be continually increased until an expansion takes place, and then decreased immediately following the expansion. The cycle will be repeated as demands again increase. This cycling of prices is efficient for maximizing consumer and producer surplus, but may be undesirable because of the instability it causes. Policy makers might wish to resort to some means of price stabilization utilizing connection charges or non-price rationing.

In long run pricing models, the problem of cost recovery may still exist although deficits may not be as serious as those encountered in the short run. Many of the same principles used in the short run pricing context can also be used in the long run to ensure cost recovery. For example, multipart tariffs could still be applied on a long run basis. In this role they might serve the dual purpose of both price stabilization and full cost recovery.

B. Conclusions

The theory discussed in this report is primarily drawn from a body of economic literature concerned with public enterprises and the regulation of utilities. In general this theory is relevant to most public or cooperative water supply systems, although there are some specific aspects that are especially relevant to water pricing. There are also some particular features of water that require some special theoretical treatment. In this section the particular theoretical aspects relevant to water are highlighted. These aspects of the theory deserve further study and development by professionals in the field of water resources.

Water pricing should be based on a standard definition of efficiency. In this report, and in the vast majority of applied welfare studies, total consumer and producer surplus is used as an indicator of economic efficiency. Given its widespread use in benefit-cost analysis, it is probably the measure of choice for analysis of water pricing. However, analysts should be open to new methods of measuring or specifying social welfare and economic efficiency as they arise.

There are some economic arguments in favour of pricing systems that achieve full cost recovery. These include the avoidance of possible tax distortions that could result from subsidization and the tendency of planners to produce over-capacity when their agency does not bear full responsibility for costs. A further policy argument for cost recovery is that it places the responsibility for finances on the users and avoids the issue of cross-subsidization where one group of users may feel that they are unfairly financing other users. The aspects of price theory that relate to cost recovery deserve particular study and understanding by federal agencies.

One of the most appropriate price systems for full cost recovery is the Coase two part tariff. This tariff consists of a fixed connection charge plus a charge based on usage. When used for water pricing, this system will probably result in the complete benefits of pure marginal cost pricing while at the same time ensuring full cost recovery. Water utilities are particularly appropriate for this type of pricing because the connection charge is unlikely to have any significant effect on the number of users using the service. Furthermore, the Coase tariff is a fairly simple pricing scheme and should be easier to implement than multipart tariffs.

There may be a few water supply utilities for which multipart tariffs, as represented by declining block rate schedules, may be the most appropriate system. In cases where water users have alternative sources for their water, a multipart tariff could be constructed that would be more efficient than the simple two-part Coase tariff described above. In these situations, a declining block tariff can increase the revenue to the water supply utility while increasing the benefits of many of its customers.

The theory of peak load pricing is directly applicable to many water demand situations in Canada. In many regions the greater part of system capacity is constructed to meet peak summer demands. Peak load pricing would result in better use of system capacity and better planning of future requirements. Because of the widespread nature of peak load demands, this body of theory also deserves considerable focus in efforts to implement and encourage efficient water pricing.

System expansion is an important issue in optimal water pricing. Correct pricing policies will prevent premature expansion and maximize the benefits that occur during the system expansion cycle. Because there is a minimum economic increment by which expansions can be carried out, there are likely to be periods where marginal cost pricing would lead to an under or over demand on a system's capacity. An efficient price rationing scheme would exhibit a cyclical pattern over time, increasing before system expansions and then decreasing immediately afterwards. Methods of stabilizing this price cycle, while maintaining economic efficiency, warrant further examination.

C. Recommendations for Further Research

As stated in the Federal Water Policy document (Environment Canada, 1987), Environment Canada has made a commitment to encourage the application of water pricing. There are a number of roles that Environment Canada could undertake in promoting the application of pricing to encourage efficient water use. Some current activities by Inland Waters Directorate include the following: increasing the public awareness of the efficiency of water pricing; documenting current pricing practices; and, increasing the awareness of water managers of the benefits of demand management. An additional role that would be effective would be that of information broker, providing technical information to the numerous water supplying utilities and price setting agencies in the country. As can be seen from the theoretical material summarized in this report, the issue of water pricing is complex, and individual water utilities will likely not have the technical expertise to determine the pricing system most suited to their situation.

A recent survey of municipal water rates in Canada (Environment Canada, 1989) showed considerable inconsistency in pricing systems. The study concluded that water rates failed to meet the criteria of economic efficiency and cost recovery. This failure suggests a lack of clear and accessible information on correct water pricing that can be used by municipal rate setting authorities. One fairly widespread information source is the manual on water supply practices by the American Water Works Association (1983) which provides some guidelines for rate setting. These guidelines stress cost recovery and equity rather than economic efficiency. There does not appear to be any other standard and understandable reference on water pricing that is based on efficient marginal cost pricing principles.

Inland Waters Directorate is in a position to supply considerable information and expertise to agencies wishing to implement efficient pricing strategies. When providing this information, it would be useful to consider two categories of water pricing: (1) pricing for public water utilities; and, (2) pricing of private withdrawals of water. In the first case, the information would be provided to municipalities or local governments that operate water utilities. In the second case, the information on water pricing would be of use to provincial and territorial governments that have the responsibility of licensing and controlling private withdrawals of water. Information could be provided to these agencies on the most appropriate pricing systems, data requirements and the kinds of analysis required. Empirical information from our current databases and our own research would also be extremely useful.

An important step in carrying out the role of information broker would be to prepare a report or reports on federal guidelines to water pricing. An initial report should be prepared for municipal utilities wishing to implement efficient pricing systems. Subsequently, a report could be prepared with guidelines for provincial and local governments on pricing of private withdrawals. These reports would contain a summary of the relevant theory, the available data, the demand and cost functions required and the basic process to be followed in determining the best pricing system. Preparation of such reports would represent a considerable commitment by Inland Waters Directorate, and would require input and review from other agencies and the academic community. These reports would represent an effective method of assembling the current knowledge and expertise into practical and accessible tools for use by water managers and planners. The contents of the proposed reports are explained in more detail below.

1. Theoretical Contents

The theoretical discussion would be minimal, outlining only the basic concepts required, and the relevant type of pricing systems. Emphasis would be placed on the most relevant theory discussed in the previous sections.

2. Description of Process for Setting Prices

The general process would involve choosing the correct theoretical pricing model, then estimating the demand and cost functions necessary to implement it. The correct theoretical model would depend on the particular features of the water utility or the kind and location of private withdrawals. In general, the pricing models for municipal utilities would emphasize marginal cost pricing and would incorporate theory related to system expansion. For private withdrawals, where individual water users bear their own direct costs of supply, the problem would be one of choosing correct prices to ration and allocate water between different uses. Estimation of the cost and demand functions may turn out to be unfeasible for individual utilities without the help of outside expertise. The report should clearly set out these information requirements.

3. Data and Information Requirements

As can be seen from previous chapters of this report, information requirements to determine an optimum price schedule are significant. The report should outline these information requirements and where possible, supply the information that has already been assembled in previous studies. Knowledge of the consumer demand functions for water will sometimes be required. The demand functions will vary according to many geographical variables, and may be quite specific to the particular water supply utility. In certain situations, demand estimates for both consumptive and non-consumptive water uses will be required. Demand forecasts are also necessary for determining optimal timing of system expansion and future price levels. Econometric estimation of the demand relationships will sometimes be necessary although, in other cases, existing estimates of price elasticities may be adequate.

A range of cost factors must be known to determine optimal prices for municipal utilities. Some knowledge of costs for private water withdrawals will also be necessary to set optimal prices for self supplied firms or individuals. Foremost among these are the long and short run marginal cost curves. Knowledge of total and average costs will also be required to ensure cost recovery. Again the cost functions are likely to very specific to the particular utility. Engineering cost data are often available but methods for estimating marginal operating and marginal capacity costs from these data will have to determined.

As part of the program activities of Inland Waters Directorate, a number of econometric water demand studies have been completed or are currently underway including Renzetti (1988), Shaw (1988) and Gai (1989). In many instances water management agencies will be able to use the demand relationships from these studies, rather than having to undertake new demand studies specific to their situation. The report on water pricing guidelines will present a summary of the available demand information and will discuss situations where supplementary studies should be carried out.

Little work has been done in Canada on cost function estimation for water utilities except by Renzetti (1989), who used an econometric approach to estimate marginal costs of the Greater Vancouver water supply system. Given the importance of cost functions for determining optimal water pricing, this area deserves much more emphasis in research programs in both government and academic communities.

REFERENCES

American Water Works Association. Water Rates Manual, AWWA Manual M1, New York, 1972

Brown, S., and D. Sibley. The Theory of Public Utility Pricing, Cambridge University Press, New York, 1986

Clark, R. and R. Stevie. "A Water Supply Cost Model Incorporating Spatial Variables", Land Economics, Vol. 57, p. 18-32, February, 1981

Coase, R. "The Marginal Cost Controversy", Economica, Volume 13, p. 169-189, 1946

Dandy, G., E. McBean and B. Hutchinson. "A Model for Constrained Optimum Water Pricing and Capacity Expansion", Water Resources Research, Vol. 20, No. 5, p. 511-520, 1984

Environment Canada. Federal Water Policy, 1987

Environment Canada. Municipal Water Rates in Canada, 1986 - Current Practices and Prices, Social Sciences Series No 21, Inland Waters Directorate, Ottawa, 1989

Gai, S. Estimation of Residential Water Demand Functions for Greater Victoria, Environment Canada, Inland Waters, Vancouver, 1989

Gai, S. Estimation of Residential Water Demand Functions for Greater Victoria: An Expansion on the Theory and Testing of the Econometric Model, Environment Canada, Inland Waters, Vancouver, 1989

Grima, A. "Empirical Basis for Municipal Water Rates Adjustment", Canadian Water Resources Journal, Vol. 9, No. 3, p. 31-52, 1984

Gysi, M. "The Cost of Peak Capacity Water", Water Resources Bulletin, Vol. 17, No. 6, p. 956-961, 1981

Hanke, S., and M. Fortin. The Economics of Municipal Water Supply: Applying the User-Pay Principle, Inquiry on Federal Water Policy, Research Paper #21, Ottawa, 1985

Hayes, K. "Cost Structure of the Water Utility Industry", Applied Economics, Vol. 19, No. 3, p. 417-425, 1987

Herrington, P. Pricing of Water Services, Organization for Economic Cooperation and Development, Paris, 1987

Hines. L. "The Long-Run Cost Function of Water Production for Selected Wisconsin Communities", Land Economics, Vol. 45, p. 133-140, 1969

Hirschleifer, J., J. DeHaven and J. Milliman. Water Supply: Economics, Technology and Policy, University of Chicago Press, Chigago 1960

Mann, P., R. Saunders and J. Warford. "A Note on Capital Indivisibility and the Definition of Marginal Cost", Water Resources Review, Vol. 16, No. 3, p. 602-604, June 1980

Millerd, F. "The Role of Pricing in Managing the Demand for Water", Canadian Water Resources Journal, Vol. 9, No. 3, p. 7-16, November, 1984

Pearse, P., Bertrand, F., and MacLaren, J. Currents of Change, Final Report, Inquiry on Federal Water Policy, Ottawa, 1985

Randall, A. Resource Economics, An Economic Approach to Natural Resource and Environmental Policy, John Wiley and Sons, New York, 1981

Rees, R. Public Enterprise Economics, Weidenfeld and Nicolson, London, 1985

Renzetti, S. An Econometric Study of Optimal Water Pricing, Unpublished Ph.d Thesis, Department of Economics, University of British Columbia, Vancouver, 1989 (forthcoming)

Renzetti, S. The Economic Aspects of Industrial Water Use, Environment Canada, Inland Waters Directorate, Ottawa, 1987

Shaw, A. Sensitivity of Municipal Water Use to Water Price, Environment Canada, Inland Waters Directorate, Ottawa, 1988

Tate, D. "Canadian Water Management; A One Armed Giant" Canadian Water Resources Journal, Vol. 9, No. 3, p. 1-6, November, 1984

Teeples, R. and N. Glyer. "Production Functions for Water Delivery Systems: Analysis and Estimation Using Dual cost Function and Implicit Price Specifications", Water Resources Research, Vol. 25, No. 5, p. 765-773. 1987

Turvey, R. "Analyzing the Marginal Cost of Water Supply", Land Economics, Vol. 52, No. 2, p. 158-168, 1976

Willig, R. "Pareto-Superior Nonlinear Outlay Schedule", Bell Journal of Economics, Vol. 9, p. 56-69, 1978