Price and Rate Structures

by Hanemann, W. Michael

Chapter 5 in *Urban Water Demand Management and Planning*

edited by Duane Baumann, John Boland, and Michael Hanemann.

New York: McGraw Hill (in press)

Chapter 5

Price and rate structures

This chapter deals with the principles for pricing water delivered to municipal users by public water supply systems. In principle, there are many different rate structures that could generate the same total revenue for a public water utility. The question, therefore, is what type of rate structure to adopt. In this chapter we consider some of the criteria that have been proposed for designing rate structures , we review the economic theory of how rates should be set, and we examine some of the empirical experience with different types of rates in the U.S.[1]

[A] Components of a water rate structure

Earlier this century, most urban water agencies were financed through fixed monthly charges. While these charges did not vary directly with consumption, they did sometimes vary according to customer characteristics, which, in the absence of metering, utilities used to identify probable consumption levels. In 1907, for example, Phoenix charged $1.50 per month for domestic water supply to dwelling houses; 25¢ per month each for water for horses and cattle; $1.00 for barber shops without fixtures, $1.50 for barber shops with one basin and one chair, and 50¢ for each additional chair; and rates of $1.00-$4.00 for stores and $2.50 and up for saloons. Since these days, water rates have changed, becoming simpler in some ways, more complex in others. The common feature of most water rates today is that they are based on metered use.

There are several basic components common to most water rate structures which we briefly discuss here. We first make a distinction between rates based on *flat charges* that are entirely independent of the quantity of water used (e.g., the monthly service charge of $1.50 per dwelling unit) versus charges that *vary* in some manner with the quantity used (e.g., 50¢ per 100 cubic feet). Regarding the latter, one can further distinguish charges that vary *directly* with the quantity of water used versus those that vary only *indirectly* or approximately with water use (e.g., a monthly service charge based on meter size or based on the number and type of fixtures). During the course of the century, many urban water agencies have switched from a system of fixed charges that were partly flat charges and

[1] This is adapted from Mitchell and Hanemann (1994), which can be consulted for a fuller discussion of the issues raised here.

partly based on factors that varied roughly, but imperfectly, with usage to a system that may include a flat charge, but also will include a variable charge based on metered usage.

A second important distinction is that between variable charges that are *uniform-rate* versus *block-rate*. A uniform-rate variable charge is one where the amount paid per unit of consumption is the same over all units consumed; a block-rate is one where the unit charge varies, either decreasing with the amount consumed (*decreasing block-rate*) or increasing with the amount consumed (*increasing block-rate*). To the extent that a utility's fixed monthly charge allows some usage for free, this implies a form of increasing-block rate, even if the remaining volume charge does not increase with additional usage. In other words, the first block is priced at zero while the second block is priced at some positive amount. Until about 1980, decreasing-block rates used to be common for large, non-residential accounts, if not for residential accounts, but are now giving way to uniform and increasing-block rate structures.

Rates may also be divided between those that charge different prices in different time periods versus those that charge the same price in all time periods. In the water industry it is increasingly common to observe rates that vary by season; volume charges are higher during the peak season and lower during the off-peak season. These are referred to as *seasonally differentiated rates*, or more simply as *seasonal rates*.

In addition, water agencies may levy other charges not yet mentioned. An example is a one-time charge levied when new customers are connected to the water system, known variously as a *connection charge, facilities charge,* or *capacity charge*. This is intended to recover capital expenditures for new facilities required to meet the projected demands of new customers. Furthermore, many water agencies have some special rates for particular classes of user. For example, there may be *life-line rates* which offer low-income customers some initial quantum of usage at a reduced price. It is also common practice for water agencies to offer service to some users outside their service area at special, higher rates; or to offer water for large irrigation users on an interruptible basis at specially reduced rates.

A rate structure is built using a combination of the basic elements just discussed. The number of possible design permutations is almost endless. One finds rates that are uniform; uniform but adjusted seasonally; increasing-block; decreasing-block; increasing- or decreasing-block *and* seasonally adjusted; etc., etc. How does one choose? What type of rate design is best? In the next section we review some of the criteria for designing a rate structure.

[A] Criteria for designing water rates

Generating revenue to meet a utility's revenue requirements is the primary role of a water rate structure. But whether intended or not, any system of water rates also performs two other functions: it allocates costs among users and it provides incentives to users. To evaluate the effectiveness of a given rate structure, we must consider how well it performs these three basic functions.

1. Generate Revenue - rates generate revenue that permits a utility to cover its costs;

2. Allocate Costs - rates serve to allocate costs among different types of use and user;

3. Provide Incentives - rates provide price signals to customers which may serve as incentives for them to use water efficiently , encouraging them to modify their behavior in particular directions.

An effective rate structure, therefore, should do a good job of generating revenue, allocating costs, and providing incentives for efficiency. What is meant in each of these cases by "doing a good job"? What are the criteria for success? Below, we identify and comment on the basic criteria that are generally accepted in the literature.

[B] Criteria for revenue generation

With regard to revenue generation, a rate structure should be judged along at least three dimensions. Each represents a facet of the utility's ability to operate on a self-sustaining basis and meet its current and future financial obligations.

1. Revenue Sufficiency -- the rate structure should provide revenues sufficient to allow the utility to operate on a self-sustaining basis. The rate structure should generate revenue sufficient to cover operating costs such as salaries, chemical supplies, gas and electricity, taxes, and capital costs for system expansion, upgrades, or equipment replacement.

2. Net Revenue Stability -- the rate structure should be conducive to a stable and predictable stream of net revenue over time and as circumstances change (e.g., in a drought). Stable net revenues allow for more accurate budgeting, better planning, and can lower long-term financing costs. Conversely, unstable net revenues increase the risk of insufficient cash flow and can raise long-term financing costs. To the extent that cost does not automatically vary in line with revenue, instability in revenue can generate instability in net revenue.

3. Administration -- administration and billing costs should be balanced against the potential benefits of a more complex rate structure. The rate structure should be designed to accommodate rate modifications and updates.

[B] Criteria for cost allocation

With regard to cost allocation, the two principle considerations are equity and economic efficiency. Equity involves at least two considerations:

4. Fair Allocation of Costs -- The rate structure should apportion costs of service among the different uses and users in a manner that is fair and is not arbitrary.

5. Avoid Cross Subsidies -- The rate structure should apportion costs of service in a manner that avoids the subsidy of one group of users at the expense of another.

In the context of cost allocation, economic efficiency requires the following:

6. Fully Allocate Private and Social Costs -- The cost allocation created by a rate structure should fully reflect the private and social costs of providing service for the different users and uses that the utility serves.

[B] Criteria for providing incentives

With regard to incentives, the criteria reflect the goals of economic efficiency, system load management, and the promotion of conservation:

7. Static Efficiency -- The rate structure should encourage the efficient use of water in terms of quantity used and timing of use. To this end, at the margin, the water rate should reflect the full private and social marginal cost of supply.

8. Dynamic Efficiency -- The rate structure should encourage an efficient pattern of growth in water use and an efficient pattern of system development over time. In this regard, the marginal rate should reflect the long-run rather than the short-run marginal cost of water supply.

9. Conservation -- The rate structure should provide proper incentives for conservation, including investment by water users in cost-effective water saving appliances, fixtures, and landscaping.

10. Rate Structure Transparency -- The rate structure should be easy for water users to understand so that it provides a clear price signal to them. In order to be effective at influencing their investment decisions, it should offer a predictable price signal with minimum risk of unexpected adjustments.

[B] Applying the criteria in practice

Applying these criteria to determine the best rate structure in any particular situation is not necessarily an easy task, for at least three reasons. The first is that some of the criteria may directly conflict, forcing water agency managers to make trade-offs among them. For example, there can be a conflict between revenue stability and economic efficiency. Since a large part of a water utility's costs are fixed in the short-run, the safest way to ensure revenue stability is to raise revenues entirely through a fixed monthly service charge. This totally insulates revenues from fluctuations in the quantity of water delivered, but is counterproductive from the perspective of efficiency because it provides no incentive to use water sparingly. Conversely, some degree of revenue instability is inherent in any rate structure with volume charges.

The second reason, which builds on the first, is that more than one party is involved in the rate setting process and, typically, different parties place different weights on the alternative criteria because of their own particular interests and point of view. Beecher et al. (1990) describe the rate making process as "a continual balancing act among the divergent and often competing perspectives of utilities, consumers, and society." They characterize these perspectives through a series of questions which we reproduce here, matching them up with the criteria presented in the preceding section.

[C] The utility's perspective

Does the rate structure fully compensate the utility so that revenue requirements are met? (Criteria 1, 2)

- Does the rate structure allow the utility to earn a fair return on its investment? (Criteria 1, 2)

- Is the rate structure strategically sound for load management, competition, and long term planning? (Criteria 1, 2, 7, 8)

[C] The customer's perspective

- Are both the ratemaking process and the rate structure equitable? (Criteria 4, 5)

- Are utility rates perceived to be affordable? (Criterion 10)

- Are both the ratemaking process and the rate structure understandable? (Criteria 10)

[C] Society's perspective

- Does the rate structure promote economic efficiency? (Criteria 6, 7, 8)

- Does the rate structure promote the appropriate valuation and conservation of resources? (Criteria 6, 7, 8, 9)

- Does the ratemaking process take into account priority uses of water? (Criteria 9)

- Are both the ratemaking process and the rate structure just and reasonable? (Criteria 1, 4, 5)

Because of these differences in perspective, and because many parties have a stake in the rate setting process, one should recognize that rate design in practice is an inherently political process.

The third reason why applying the criteria can be difficult is the complicated nature of water utility costs. In many cases, the costs of service do not vary with the quantity of service in a simple or direct manner. The water industry is highly capital intensive, even compared to most other utility industries. Many of its expenses are fixed costs that hardly vary with the amount of water delivered. Moreover, the capital investment is generally long-lived, and the capital stock employed in any given water system tends to include assets of several different vintages, acquired at different times and at vastly different costs. Because capacity cannot readily be altered at short notice while demand may be highly variable -- both by day and by season -- a water utility always needs to build some amount of *excess* capacity into its system just to be sure of meeting demand with an acceptable level of reliability. Who is responsible for these costs and what price will a utility's customers be willing to pay for reliability? To complicate matters further, parts of a utility's system often serve several different functions at the same time, so that there is a problem of allocating *joint* costs among separate beneficiaries.

The result is that it can be both difficult and frustrating to answer what people might think is a simple question: "How much does it cost to supply me with the water that serves my needs?" On the one hand, it may require an enormous amount of data and extensive technical analysis. On the other hand, even when the analysis has been completed, there still may be no single, correct way to develop a rate structure that satisfies everybody. The very nature of water utility costs tends to create conflicts between the criteria of raising adequate revenues, allocating costs fairly, and providing incentives for efficiency and conservation. Indeed, it is disagreement over how to resolve those cost-driven conflicts that accounts for most of the major controversies over utility ratemaking. This is particularly true of the dispute between proponents of the traditional approach to ratemaking, based on what is known as the principle of *embedded cost*, and proponents of the newer approaches, associated with *marginal cost* pricing , that were first introduced into the electricity industry in the 1970’s, then spread to other regulated industries such as natural gas and telecommunications, and are now being introduced into the water industry. The traditional approach pools many different costs into a single average. It looks backwards to the costs that the utility has already incurred, and emphasizes the estimation and allocation of historical (i.e., embedded) average cost. In terms of the criteria presented above, it places the main emphasis on criteria 1 through 4. The newer approaches place primary emphasis on sending users the right price signal about the scarcity value of the commodity. They are generally forward looking: as a basis for setting prices, they look to marginal cost than average cost, and replacement cost rather than historical cost. In terms of the criteria, they place the most weight on criteria 6 through 9.

Given that differing notions of cost underlie the alterative approaches to ratemaking in the water utility industry, it is useful to review the economic concept of costs and the economic of pricing, to which we now turn.

[A] Water supply costs and complexities

We start off by imagining a fairy tale water utility that obtains its water supply from a magic spigot deep in a forest. The spigot is guarded by elves who also are responsible for installing and maintaining the distribution network that serves the utility’s customers. For all these services, the elves charge the water company $0.05 per hundred cubic foot of water delivered to the utility’s customers. The staff of the utility consists of one retired elf, who deals with all the customers, sends them bills for water service, collects their payments, and handles all the dealings with the elves who operate the magic spigot. Since he already has a pension, he performs these services for free.

In these circumstances, ratemaking is simple -- everybody would surely agree on $0.05/ccf as the appropriate retail charge for water. What makes the story a fairy tale is not the cheap price for water but rather the simple cost structure. The water agency only incurs costs as and when water is delivered. The costs vary directly with the quantity of water delivered. All units of water delivered cost the agency exactly the same amount of money.

The real world is different in every respect. The water industry is highly capital intensive; most of its costs are capital costs, incurred when capital assets are installed, rather than operating costs, incurred as water is delivered. Once installed, the capital is not particularly malleable: the capacity cannot be quickly expanded if you suddenly need more; nor can it be disposed of profitably if you need less. Economists make a basic distinction between *fixed* versus *variable* costs -- i.e., costs that don't vary with the quantity of service provided versus those that do. Thus, most water utility costs are fixed costs. However, there are several complications to the distinction between fixed and variable costs. One arises from differences in the time frame of the analysis. We just indicated that, because capital is relatively un-malleable once installed, its cost is to be regarded as fixed. However, when viewed from the a longer time perspective, all capital is variable -- existing capital will wear out and, with sufficient time, new capital can always be added. This corresponds to the economists' distinction between *long-run* and *short-run* -- in the short-run, the capital stock is fixed and its cost is a fixed cost; in the long-run capital is variable and its cost constitutes a variable cost. How, then, should water be priced? Should its price reflect a short-run or long-run perspective? If prices reflect short-run costs, this may encourage patterns of water use that are poorly suited to future circumstances when more expensive sources will be required. On the other hand, if prices reflect long-run costs, the utility may be in the position of charging for facilities before they exist.

Economists also make a basic distinction between *average cost* and *marginal cost*. Average cost is simply total cost divided by the quantity sold (i.e., it is the cost per unit);

marginal cost is the *change* in cost per unit change in quantity sold (i.e., it is the increment in cost per unit increase in quantity, or the decrement in cost per unit decrease in quantity). When all units of a commodity cost the same, as in the fairy tale example, there is *no* difference between average and marginal cost -- the two costs are equal. Otherwise, however, there *is* a difference between average and marginal cost, sometime a very large difference. Moreover, the distinction between average and marginal cost applies in both the short-run and the long-run. Thus, there are short-run average and marginal costs when capital is in place and its cost sunk, and long-run average and marginal costs, when capital is adjustable and all costs are variable.

Differences between average and marginal cost can arise for many reasons. An important example of divergent short-run average and marginal costs is the case of a utility with a fixed capital stock and a constant unit-operating cost. In the short run the marginal cost is simply the unit-operating cost-the cost of producing an additional unit. The average cost, however, consists of the unit-operating cost plus the cost of the fixed equipment averaged over the amount of production. The more capacity is used, the more product over which to spread the fixed cost and, therefore, the lower the average cost per unit. Hence, there is a declining short-run-cost curve lying above a short-run marginal cost curve, as shown graphically in Figure 5-1. Historically, the notion illustrated here—that with increased production one can spread fixed costs and therefore one has declining-average costs—underlies many of the arguments for declining-block prices that were used prior to the 1970s.

A more difficult case for a utility is when cost is falling in the short run but rising in the long run, as illustrated in Figure 5-2. This phenomena has to do with the fact that new supplies may be considerably more expensive to develop than past supplies. While a utility may experience economies of scale for a single project—similar to the situation shown in Figure 5-1—it may not experience economies of scale over a range of projects-meaning that cost is increasing as new capacity is added to the system. Consider a utility developing a supply of water. It secures the most convenient and least costly source available and builds a reservoir. With the investment made and the reservoir built, the cost of the water depends on the level at which the reservoir is utilized. As use of capacity increases the average cost of supplying a unit of water falls because the fixed cost is being spread over more and more units. The short run marginal cost of supplying a unit of water may also be falling because of scale economies associated with the project. When the capacity of the first reservoir is fully utilized, however, the utility will need to develop a new source of supply to meet new demand. It will likely have to develop a source not quite as conveniently located or of poorer quality than the first.

Suppose the cost of the second source is much higher than the first. How should the utility factor these two conflicting cost perspectives into its rate structure? Should rates reflect the lower unit cost of the first source, the higher unit cost of the second source, or some sort of average cost of both sources?

If we consider a utility with multiple sources of supply of varying vintages and costs, we get a situation like that shown in Figure 5-3. If the cost of water from these sources is arrayed from least to most expensive, as illustrated in the figure, we get an increasing longrun marginal cost of supply. In this case, the average cost would lie below the *marginal cost;* while rising, it is not rising as fast as marginal cost because it blends the high and the low costs together. This illustrates the complexity underlying the seemingly simple question "How much does it cost to supply me with the water that serves my needs?"

This is the fundamental dilemma of ratemaking when average and marginal costs diverge-it is, in effect, an allocation exercise with many possible solutions. One approach is to average all costs over all users. Another is to charge all users the marginal cost for at least some of their usage. The first approach may be considered simpler, but it is not necessarily less arbitrary. Economic theory supports some version of marginal cost pricing on the principle that *all* users draw on the system at the margin and should be signaled the scarcity value of the resource. Whenever average and marginal costs differ, if all units are priced at the average cost then total revenue automatically covers total cost. This is not true when all units are priced at the marginal cost-total revenue falls short of marginal cost when marginal cost lies below average cost (Figure 5-1), but exceeds total cost when marginal cost lies above average cost (Figures 5-2 and 5-3). However, one cannot conclude that utilities *always* break even with average-cost pricing or that they can *never* break even with marginal cost pricing. The reason is that we have focused so far only on variable charges while most rate structures in practice, as indicated at the beginning of this chapter, have several components, including fixed charges. The revenues from other charges can supplement the revenues from marginal cost pricing in the situation depicted in Figure 5- 1. Conversely, when the situation is like Figures 5-2 or 5-3, where marginal cost pricing would more than break even, one can hold revenues down in other ways-reducing some component of the fixed charge, for example, or charging less than marginal cost for part of a customer's usage (as with increasing block rates).

In the next section we discuss the economic case for marginal cost pricing in terms of presenting an appropriate signal to users of the commodity. We should point out, however, that determining this cost is often very complicated. The complexity of a water system will require that the utility identify and measure not just one but several marginal costs. Some of a water utility's costs vary with the number or type of customers rather than the amount of water delivered. Other costs, such as for system capacity, depend on the time as well as the *amount* of water use.

Unfortunately, the complications do not end here. Many of the facilities for a utility provide several services simultaneously, and hence the cost for those facilities may be joint or common to those services. If a facility can be expanded to increase the level of output for one service without affecting the provision of the other services connected to it, this does not raise any special issue. But, if the facility automatically benefits the provision of multiple services, then the costs of that facility or production process are said to be *joint*. For example, a distribution main jointly provides base capacity for off-peak service and extra capacity for peak service. To increase capacity by one unit for peak service necessarily increases capacity by one unit for off-peak service. In other words, producing one service implicitly results in the production of the other service, and their two costs are inextricably linked. As a result, the only economically definable cost for the distribution main-average, marginal, or otherwise—is for the composite-peak/off-peak service it provides. This does not mean, however, that it is entirely impossible to derive prices for the separate services that are both efficient and recover their joint costs of production. But, it considerably complicates the pricing of these separate services and it requires prices not based on production costs alone.

To summarize, as a matter of principle one wants water rates to reflect costs. However, there usually is not *one* cost but many costs. Which costs, then should rates reflect? While all the costs are real enough, it is how we classify them that matters for ratemaking. Any classification is a matter of judgment and, to some degree, inherently arbitrary. In the tale about the elves there was only one cost. Therefore, questions of classification did not arise and there could be total agreement on the correct rate structure. That is why it is a fairy tale.

[A] Marginal cost pricing

In this section we describe some of the approaches to pricing utility services that first emerged in the United States in the electricity industry during the 1970s and 1980s. What these approaches have in common is that they set aside the narrow focus on only meeting the utility's revenue requirements in favor of a broader set of goals that also include economic efficiency and the promotion of conservation.

Pricing to encourage more efficient use of water rests on the assumption that prices can change consumers' behavior, even for a basic commodity like water. Whether this is so is an empirical matter that certainly can vary with circumstances. A major point we wish to emphasize, however, is that *how* prices are used matters every much as *whether* they are used. Prices can be effective or ineffective as tools for influencing behavior depending on how they are deployed.

The relative importance of price and other factors as influences on water use is frequently debated among water industry professionals. Our own view is that this is an unnecessary dichotomy-price matters, but so do. other activities such as public education, advertising, and information dissemination campaigns. Indeed, the two complement one another. Raising prices without providing guidance to customers on methods whereby they might reduce their water use blunts the effectiveness of the price signal. Conversely, exhortation and cajolery unaccompanied by financial incentives may have limited impact. All too often we have heard water managers complain that customers are resisting appeals for conservation because it is not in their financial interest. Low prices at the margin are an impediment to conservation-if there is little money to be saved, this undercuts the case for changing one's behavior.

Thus, we view pricing not as a substitute for a utility's existing or planned conservation programs but as something intended to work in tandem with them and enhance their impact. The hallmark of newer approaches to ratemaking, as we view it, is that they attempt to target or tailor prices in such a way as to create effective incentives to use water more efficiently.

[B] The economic argument for marginal cost pricing

The economic argument for marginal cost-based rates comes from a branch of economics known as "welfare economics" which deals with prescriptions for efficiency in the use of scarce resources." It offers rules for the efficient total production of commodities, the efficient allocation of this total among individual producers and consumers, and the efficient expansion of production capacity over time through new investment.

The central prescription of welfare economics is that commodities should be produced and allocated to the point where their marginal benefit equals their marginal cost. Whenever this is violated, there cannot be full efficiency in the economy with respect to either total level of production or the allocation of this total among individuals. Marginal benefit is defined as the extra benefit from producing and consuming on, more unit of the commodity; as explained in the previous Chapter, marginal cost is the extra cost from producing and consuming one more unit of the commodity.[2] They can equally well be defined as the benefit lost and cost saved by producing one *less* unit; seen this way, marginal cost is synonymous with *avoided cost-the* cost that would be saved (avoided) by reducing output by a small amount. Seen the first way, it is synonymous with incremental cost-the added cost of a small

[2]Economists distinguish between *private and social* marginal benefit and cost. Social marginal benefit (cost) is the extra benefit (cost) accruing to society as a whole. Private marginal benefit (cost) is the extra benefit (cost) accruing to the individuals directly involved. To the extent that there are what economists call *externalities,* there can be a divergence between private and social benefits and costs. The modern economic theory of pollution, for example, grows out of the divergence between private costs (which exclude the effects of pollution emissions on other people) and social costs (which include the effects of these emissions).

amount of additional output.

A related concept that is very important to an understanding of resource allocation is *opportunity cost.* With a given productive capacity, a decision to produce more of any *one* commodity, say water for urban use, implies a decision to produce less of all *other* commodities. Thus, the cost to society of producing anything consists, really, in the other things that must be sacrificed to produce it. In the final analysis, all cost is opportunity cost—i.e., the value of the alternatives foregone.

What is the economist's rationale for marginal cost pricing? Given a fixed productive capacity at any point in time, a decision to produce more of any one good implies a decision to produce less of all other goods as a whole. Therefore, the basic challenge for an economy is to use these resources to maximum advantage. For that to happen requires that the benefit gained from consuming one more unit of a good equal the cost to produce it. In other words, that the marginal benefit equal the marginal cost. How does one ensure this outcome? By ensuring that prices of commodities are set equal to their marginal cost. Prices provide the signals that guide people's behavior. When consumers make purchase decisions, they balance their benefit from a commodity against its price. If they are to make choices that reap the greatest possible benefit from society's limited resources, the prices that they pay must accurately reflect the opportunity costs associated with the commodities they are considering. If their judgments are correctly informed in this way, they will, by their independent decisions, guide scarce resources into those lines of production that maximize the net benefit to society *as a whole.*[3]

Could not prices equal, say, average rather than marginal costs (assuming the two are different) and do this just as well? If price had *no influence at all* on demand, it would not necessarily matter. But, in fact, the demand for all commodities is in some degree, at some point, responsive to price. As a practical matter, we know this to be the case for water, as we

[3] It is important to emphasize the "as a whole" in this statement. Economic concepts of efficiency remain quiet with regards to distributional concerns. It is quite possible to have an efficient allocation that is grossly inequitable. The question here becomes should equity concerns be addressed independent of or simultaneously with efficiency concerns. Economists generally favor that the two problems be addressed independently, though this prescription by no means enjoys unanimous consent.

noted in Chapter 2. Then, if consumers are to decide wisely from society's point of view whether to take somewhat more or somewhat less of any particular item, the price they pay must reflect the cost of supplying somewhat more or somewhat less-in short, the marginal-opportunity cost. Suppose, instead, that buyers were charged more than the marginal cost for a particular commodity; they would then buy less than the socially optimum quantity-welfare could be improved if they consumed more of the good in question and less of all other goods as a whole. Some consumers who would have consumed more of the good in question, and less of other goods, will refrain from doing so because the price exaggerates the good's opportunity cost. Conversely, if price is set *below* marginal cost, they will buy more of the commodity (and less of all other commodities taken together) than is socially optimal. Producers are diverting more resources to the production of this commodity than customers would have willingly authorized, had the price fully reflected the marginal-opportunity cost.

[C] The choice of time frame for marginal cost pricing

Although the basic economic argument for marginal cost pricing may seem compelling, we cannot end the discussion there and simply state that price should equal marginal cost. To do so is to ignore several important qualifications and practical limitations to this general rule. When it comes to moving from the theory of marginal cost pricing to its implementation, as with anything else, complications arise, and one is forced to deal with myriad details. In this and the next sections, we focus on two major issues:how one decides on the marginal cost concept to be used as a basis for setting prices, and how one ensures that the overall rate structure brings in sufficient revenue to meet the utility's requirements.

We earlier presented a brief tutorial on the economic concepts of cost and pointed out the distinction between *short-run and long-run* marginal costs, as well as the related distinction between *fixed and variable* costs. Fixed costs, i.e., the costs associated with fixed inputs, are costs that do not vary with the quantity of service provided. Variable costs, i.e., the costs associated with variable inputs, are the costs that *do* vary with the quantity of service provided. By definition, the marginal cost associated with the use of fixed inputs is zero-the quantity of output changes, but there is *no* change in the quantity of these inputs, and so there is no change in this component of total cost. Only variable inputs generate non-zero marginal costs. In the short run, the capital stock is fixed, and the marginal cost arises mainly from O&M costs. In the long run, the capital stock can be replaced and expanded, and the marginal cost includes not only O&M costs but also capital costs.

By definition, then, short-run marginal cost is always less than long-run marginal cost. But, this gap is especially huge in the water industry because of its unusually high capital intensity.[4] Thus, it matters greatly in the water industry whether prices are set on the basis of short-run marginal cost, which is only a small fraction of total expenditure, or long-run marginal cost.

What is to be done? Different opinions have been expressed in the economics literature, and there clearly is no easy answer. All agree that rice should *never* be set below short-run marginal cost. The question is whether it should be set any higher. The argument for short-run marginal cost pricing is best illustrated through the example of airline pricing. As Kahn (1988) has put it, no airplane should ever take off with empty seats as long as there exist some potential travelers who would be willing to pay the (almost negligible) short-run marginal cost associated with adding them to the flight roster. It is economically inefficient in this case to charge anything more than short-run marginal cost.[5]

This is not, however, a hard-and-fast rule. There are other considerations that may lead to a different conclusion. One consideration is price volatility. In some circumstances, short-run marginal cost could vary greatly over a short period of time as a result of

---

[4]For the water industry nationally, the asset requirement per dollar of revenue (i.e., the ratio of capital assets to annual revenue) is estimated at about $10-12; this is 3 to 4 times the capital intensity of the telephone and electric utility industries, and about 5 to 6 times that of the railroad industry.

[5]The airlines actually used this principle when they introduced cheap standby fares in 1996. But airline fares also reflect other economic considerations that are not necessarily socially efficient, such as price discrimination. The airlines reckon that the business demand for travel is far more inelastic than private individuals' demand for vacation travel. In order to discriminate between the two markets and maximize their profit from each, the airlines typically require staying over a Saturday night for their cheapest fares, since they figure this shuts out most business travelers. Then again, for standby fares to be efficient requires the airlines to be able to discriminate between different types of demand. Many airlines were compelled to discontinue standby fares because they found that standby customers were making false reservations under assumed names to ensure adequate seating.

fluctuations in either demand or supply (for example, in the case of a producer with access to many supply sources with very different costs). The variation in price that would result from short-run marginal cost pricing might be considered undesirable or counterproductive. Consequently, suppliers might prefer to abandon short-run marginal cost pricing in order to smooth out prices over time.

Another consideration is the impact setting price to short-run marginal cost may have on long-run investment. Whereas economic efficiency with respect to production from a *given* capital stock calls for setting prices equal to short-run marginal cost, economic efficiency with respect to investment and the determination of long-run capacity calls for setting prices equal to long-run marginal cost. If people are to reap the greatest possible benefit from society's limited resources and the capital stock is likely to be replaced or expanded in the foreseeable future, the prices of commodities must reflect the opportunity costs associated with not only the variable inputs but also the capital assets that are needed to produce them.

Suppose a producer will replace or expand his capital in the foreseeable future. Or, suppose that his customers will replace or expand *their* capital in the foreseeable future. In either case, charging short-run instead of long-run marginal cost provides an incorrect economic signal. If the wrong signal is given, this can lead both consumer and producer astray. If the price does not accurately reflect long-run marginal cost, the consumer may make investment decisions that are socially undesirable because they entail a long-run commitment to the use of a commodity for purposes that he would not consider justified if he had to pay the full, long-run cost of producing it. Likewise, the producer may make investment decisions that are socially undesirable because they entail a long-run commitment to the supply of a commodity whose users would not find it worthwhile if they had to pay the full, long-run marginal cost. This, in fact, is what happened on an unprecedented scale to the Washington Public Power Supply System-whose acronym WPPSS is appropriately pronounced whoops-resulting in the largest municipal bond default in U.S. history, which to this day continues to have serious financial repercussions for the Pacific Northwest.

Such questions about the correctness of investment decisions when prices are set below long-run marginal cost also have been raised throughout the history of the California water industry by critics of both urban and agricultural water policies. Gardner (1982) and others have long condemned the (wholesale) prices charged by the U.S. Bureau of Reclamation for this very reason. From the beginning, the Central Valley Project (CVP) has charged prices that were below long-run marginal cost-indeed, for most of the last two decades it charged prices that were below short-run marginal cost. As a result, the critics argue, it continued to build new reservoirs and aqueducts that irrigators would not have considered worthwhile had they been required to bear the cost. These investment projects were politically viable only because their costs were effectively invisible-they were averaged together with the costs of existing facilities and were spread over all beneficiaries of CVP water. But these phenomena-pricing below long-run marginal cost and making investment decisions for new reservoirs that would not be justified if those who used the additional supply actually had to bear the cost-are not unknown among urban water agencies, either.

As these considerations demonstrate, whether price should reflect short-run or long-run marginal cost will vary with the circumstances. As a general principle, however, we are inclined to follow Kahn's recommendation:

> The practically achievable benchmark for efficient pricing is more likely to be a type of average long-run incremental cost computed for a large, expected incremental block of sales, instead of short-run marginal cost estimate single additional sale. This long-run incremental cost (which we shall loosely refer to as lo marginal cost) would be based on (1) the incremental variable costs of those added and (2) estimated additional capital costs per unit for the additional capacity that will have constructed if sales at that price are expected to continue over time or to grow. Both of these components would be estimated as average some period of years into the future. (Kahn (1988)[6]

Note that when Kahn speaks of average cost is referring to the average incremental

[6]This recommendation serves as a general principle rather than a hard and fast rule. It is easy to imagine situations where pricing at Silort-run marginal cost would be most appropriate. For example, if a wholesale agency had extra water available in its reservoirs that it either had to sell or release to make storage available for next year's runoff, the appropriate price would be the short-run cost of delivering the water to retail water agencies.

cost from ad a large block of new supply, not the average co supply from both old and new sources, which is the embedded-cost approach discussed in the previous chapter. When a utility prices its water service as recommends, customers are sent a signal that reflects the opportunity cost associated with their consumption level.

[C] Meeting revenue requirements with marginal cost pricing

We noted above that, when prices are set equal to average cost, total revenue by definition covers total cost exactly. If prices are set any other w such as marginal cost pricing, this does not hold. But we also noted two qualifications. One is where average cost per unit stays constant as output changes; the average and marginal cost coincide, so that marginal cost pricing is the same as average-cost pricing. The more important qualification is that there are *other charges,* and even *other ways of configuring marginal cost pricing,* through which utilities can ensure that their total revenue is adequate but not excessive to cover their total cost. Elaborating on these methods for meeting revenue targets is the focus of this section.

It has long been recognized that industries with decreasing average costs would not cover their total cost if they set prices equal to marginal cost. Two of the three main solutions have been known for over fifty years. One, involving what became known as the *two-part tariff,* was suggested by Lewis (1941). This is where a utility combines a commodity charge based on marginal cost with a fixed charge, e.g., a service charge or connection charge. Together, these make up the two parts of the tariff. The idea is that the fixed charge raises the additional revenue needed to cover total cost, but does not interfere with economic efficiency generated by having a commodity charge based on marginal cost. Since the customer pays the same fixed charge regardless of what quantity he consumes, only the commodity charge should influence his quantity decision, and this still provides the economically correct price signal. Hence, it is argued, the two-part tariff can satisfy the goals of both economic efficiency and revenue adequacy.

Subsequently, the efficiency of the two-part tariff has been challenged on two grounds. One is the argument that, even though consumers *ought to* disregard the fixed charge when deciding how much of the commodity to use, perhaps they do not. Perhaps they are impressed by the fact that, if they increase their consumption, they can spread this fixed charge over more of the commodity and so reduce the overall unit cost. If so, this would create an economically inefficient incentive to maximize consumption. When Mayor Bradley's Blue Ribbon Committee was meeting to consider LADWP water rates in 1992, it was persuaded by this argument and recommended eliminating fixed charges from the water rates. It is not clear, however, that the argument is valid.[7]

The second efficiency argument is given more weight by most economists. It arises from the fact that, while the fixed charge does not vary with the quantity consumed, it *does* vary with whether or not one consumes *anything at all if* a consumer were willing to opt out of using the commodity entirely-e.g. disconnecting from the utility's distribution network then he could avoid the fixed charge. A consumer in this position should be influenced by the fixed charge in making an economic decision. In this case, it can be shown that the fixed charge may be economically inefficient-there may be situations where the social optimum requires that the consumer *not* exit the system, but his private incentive in the face of the fixed charge *does* lead him to exit the system. In those situations, social efficiency would prescribe a smaller fixed charge than that needed to fill the gap between total costs and the revenues raised through marginal cost-based commodity charges. If so, there is *no* solution that both generates adequate revenues and is perfectly efficient from an economic point of view; there are only what economists call *second-best efficient* solutions.

Before continuing, we should emphasize that all the foregoing discussion is predicated on there being decreasing long-run average costs. In the urban water industry today, the reality is more likely to be *increasing* long-run average costs, in which case longrun marginal cost pricing generates total revenues in *excess* of total costs. There is still an efficiency issue-the utility would need to dispose of the excess revenues in a way that doesn't bias customers' decisions away from efficiency-but it may be much easier to resolve. For example, the utility

[7]Another reason for having some fixed charge is to cover costs that vary by customer rather than by quantity of commodity sold, such as the costs of administering customers' accounts. Economic efficiency suggests these be recovered not by a commodity charge but by a fixed service charge set equal to the marginal cost per customer served.

could use some type of rebate that was independent of customer use. Furthermore, the efficiency problems that arise from having a fixed charge (or rebate) would vanish if one could be sure that the charge *did not* influence customers' exit/entry decisions. For water, more than for any other utility service, this is likely to be the case. The fixed payment that most urban water agencies need to charge are unlikely to drive consumers off the distribution system and, therefore, are unlikely to cause a two-part tariff to be economically inefficient.[8]

The second main solution to the revenue problem of marginal cost pricing in the presence of non-constant average costs is known as *Ramsey pricing.* Ramsey (1927) modified the conventional economic efficiency analysis by adding an explicit constraint that commodity prices not only maximize social welfare but also break even.[9] He derived a complex formula for how one should adjust prices away from marginal cost *in inverse proportion to the elasticity of demand.* The intuition underlying Ramsey's formula is that, in order to preserve as much efficiency as possible, one wants to depart as little as possible from the pattern of consumption that would occur with unfettered marginal cost pricing, while still charging prices that secure the utility sufficient but not excessive revenue. One accomplishes this goal by imposing the *least* price adjustments on the customers whose quantity demanded is *least* sensitive to price, and the *smallest* adjustments on the customers whose demand is *most* sensitive to price. The result is a form of cross subsidization that yields a more efficient economy than if one had simply adjusted the price for all customers in the same way. There are two obvious problems with this approach, however. One is that the inverse elasticity pricing formula can be extremely complex and usually will require information on demand that simply is not available to most utilities. In order for Ramsey pricing to work, one must

[8]This is *not* the case with some other industries, such as cable TV for residential customers or long-distance telephone for large commercial and industrial customers. Those industries are more likely than water to have decreasing average costs, and they face a real prospect that their fixed charges could affect customer's entry/exit decisions. AT&T, in particular, has actively sponsored research since the late 1970s to develop a more efficient alternative to the two-part tariff known as *nonuniform or nonlinearpricing* [see Brown and Sibley (1986, Chaps 4, 5), Mitchell and Vogelsang (199 1, Chap 5), and Wilson (1993)].

[9]Because of the imposition of this constraint, his analysis was an exercise - indeed, the *first* exercise - in second-best efficiency.

have detailed information not just for the demand of the commodity in question, but for all potential complementary' and substitute goods as well. The other is that this particular solution to the revenue requirement problem relies on cross-subsidization, which, as we recall from Chapter 2, is in direct conflict with the principle of equity, and is in conflict with most regulatory codes governing the pricing of water.[10]

There are other, *ad hoc* solutions to the revenue problem created by divergent marginal and average costs. For example, Brown and Sibley (1986) note that in Europe it is sometimes accepted that public utilities should price on the basis of marginal cost with the government making up revenue shortfalls out of tax revenues. In the U.S., the tradition has generally been that utilities cover their own costs without taxpayer assistance in the case of decreasing long-run marginal cost, or that they do not earn excess revenues in the case of increasing long-run marginal cost. This has generally lead to average-cost pricing rules in the former case and *some form* of proportionally scaled-down marginal cost-based pricing in the latter.19 Either of these approaches distort the price signal to the consumer away from the true long-run marginal cost.

Another solution is to have increasing or decreasing block rates. These can be regarded as multipart extensions of the two-part tariff. They will be discussed further in Section 4.4. We mention them here only to point out their implications for revenue adequacy. Suppose there are decreasing long-run average costs. If there is a form of declining block rates with the block where most consumption is located being priced at long-run marginal cost, while the earlier blocks (called the *inframarginal blocks) are* priced at some higher amount, this-can provide many of the efficiency benefits of marginal cost pricing while still breaking even. Conversely, if there is increasing long-run marginal costs (which we consider more likely to be the case for the urban water industry) one wants a form of increasing block rates with the inframarginal blocks priced below long-run marginal cost in order that total revenues match total costs.

---

[10] In other words, by forcing those least able to adjust their pattern of consumption to bear the brunt of the effort to meet a revenue constraint, Ramsey pricing sacrifices equity at the altar of efficiency. For more on this, see Baumol.

[C] Marginal cost pricing-summary

Rate structures based on marginal cost pricing precepts are intended to provide price signals that. result in a more efficient allocation and use of a scarce supply of water. Efficient consumption requires that the benefit derived from consuming one more unit of a good equal the cost of supplying it. If the benefit is less than the cost, it is inefficient to produce additional units; the resources required to do so could be more productively employed elsewhere in the economy. By setting price equal to marginal cost, consumers are able to compare the benefit of additional consumption with its associated cost and make efficient choices. In this way, production may be guided towards more efficient levels.

Marginal cost pricing differs from average-cost pricing in important ways. Whereas, marginal cost pricing reflects the cost of producing an additional unit, average-cost pricing reflects the unit cost of producing all units. If costs are decreasing as output increases, average cost will be above marginal cost. If costs are increasing as output increases, average cost will be below marginal cost. In either case, prices based on average cost will not result in efficient consumption choices because consumers will not be matching marginal benefit with marginal cost.

Marginal cost prices are more difficult to calculate than average-cost prices. To calculate average cost prices, a utility needs some understanding of its total costs and production. Dividing one by the other yields an average price. To calculate marginal cost, on the other hand, a utility needs fairly detailed information on costs for a variety of plants and equipment, some of which may not yet be built. This data may be expensive and difficult to obtain and of uncertain reliability.[11]

Marginal cost pricing can result in over or under collection of revenue. Generating revenue insufficient to cover cost obviously is not sustainable in the to run. Collection of too much revenue may also be problem if a utility is constrained to earn zero profit. Various strategies to satisfy the zero profit constraint while retaining the efficiency properties of

[11]The latter approach avoids the undesirable cross-subsidization implicit in Ramsey pricing by applying a uniform upward or downward scaling to marginal cost-based prices so that revenue equals total cost.

margin cost pricing have been proposed. Two-part tariffs a multiple block rates, for example, can be designed that price at the margin reflects marginal cost, while prices for inframarginal consumption are set so that the utility breaks even. Another strategy, Ramsey pricing, is to set different prices for different custom or customer groupings according to the magnitude their elasticities of demand.

[B] Seasonal rates

The simple of emulation of marginal cost pricing requires that all customers purchasing a good should be charged the same price, namely the good's marginal cost. But suppose that it costs more to provide the good during certain periods than others. Suppose, for instance, that a water agency does not have one but several sources of supply, and that its most expensive sources are required only when demand is high, whereas its cheaper sources are sufficient when demand is low. How then should the utility price its water? Under marginal cost pricing, the price should reflect the cost of the last unit produced. But since this cost may differ over different production periods, we get different prices at different times. This can be considered an example of peak-load pricing: the service is priced higher during the peak demand period, when cost is high, than during the off-peak period, when cost is low. For some utilities the major variation in demand occurs during different parts of the day (e.g., early evening, middle of the day, night) or during different days of the week (e.g. weekday versus weekend). For water, while there are daily and weekly cycles, the main variation is seasonal-summer versus winter-because outdoor water use varies with seasonal changes in temperature, precipitation, and plant evapotranspiration. For this reason, the terms seasonal pricing and seasonal rates are commonly used in the water industry when discussing peak-load pricing. Below, we discuss both the theory and the application of peak-load pricing for water service in terms of economic efficiency, cost allocation, and incentives for conservation.

[C] Principles of peak load pricing

It is safe to say that the economics profession is fairly unanimous in its prescriptions for pricing capacity required to meet peaks in demand. In general, there are two. Both follow from marginal cost theory, and both assign capacity costs to those that are causally responsible for them. One is appropriate in instances where variation in demand is random, and therefore unpredictable; the other is appropriate when the variation in demand is systematic, and therefore predictable, at least to a reasonable degree of accuracy. As it turns out, both situations occur with urban water service. Fire service is perhaps the best example of an unpredictable demand variation that requires that excess capacity be on hand. Summer sprinkling is an example of a much more predictable demand variation that also requires that excess capacity be on hand. Before discussing in more detail each pricing prescription, however, additional insight into the nature of the service provided by a water utility may be useful.

A fundamental condition of urban water service is that the utility must be ready to deliver at any time. In this sense, as Hirshleifer, et al., aptly state, the utility must stand "ready to enter into a contract for delivery at the option of the buyer." It is the customer rather than the utility that is in the proverbial driver's seat determining the quantity of service the utility must provide at each given moment. To oblige the demands of its customers with a given level of reliability, the utility requires an excess capacity above *the average* level of demand it can expect. All urban water systems are designed with this in mind.

Essentially, then, a customer holds an option to utilize this reserve capacity, and can exercise this option at any time. The cost of the option to the utility is the cost of providing the extra capacity necessary to meet maximum expected demand. Efficiency requires that the customer's willingness to pay for the option equal the cost of the option. In other words, the customer gaining access to additional capacity must be willing to pay the incremental cost of providing it. If the price for additional capacity is less than its marginal cost, over time too much capacity will be demanded. If customers were to face the true cost of the additional capacity, they would scale back their demand and apply the savings to more valued alternatives. Conversely, if the price is above its marginal cost, too little will be demanded. In this case, if customers were to face the true cost of additional capacity, they would scale back their demands for other goods and services and apply the savings to the purchase of more capacity.

With this in mind, Kahn states the economist's rule for pricing capacity as follows:

> If the same type of capacity serves all users, capacity costs as such should be levied only on utilization at the peak. Every purchase at that time makes its proportionate contribution in the long-run to the incidence of those capacity costs and should therefore have that responsibility reflected in its price. No part of those costs should be levied on off-peak users. (Kahn 1988)

This is the rule of "Peak Responsibility." Those that create the peak should pay for the peak. However, as we alluded above, identifying those responsible depends on whether the capacity is required to satisfy random or systematic spikes in demand. If it is random and can occur at any time (e.g., fire service), the procedure is to allocate the cost of the necessary capacity according to the maximum instantaneous demand a customer or class of customers can be expected to place on the system (Hirshleifer, Dehaven, et al. 1960). This can be accomplished with a "capacity' charge for the extra capacity that a utility holds in readiness for its customers unexpected demands. It is important to note that this should not be a volume charge but -rather a fixed charge that reflects the cost of providing capacity sufficient to ensure a reasonable degree of reliability in service.

If, on the other hand, the variation in demand is systematic, occurring during a certain period or periods (e.g., sprinkling demand), then the price for service during the peak period(s) should include the cost of capacity that makes consumption at the peak level possible.[12] This results in a very different pricing formulation than when the variation in demand is random. In this case, a volume charge is not only appropriate, it is necessary to obtain efficiency. When peak demand occurs during a specific period, then any consumption during that period contributes to the peak and thus to the need for capacity. In other words, any consumption during the peak period is, in part, responsible for the capacity required to satisfy it. Whether a customer's consumption is small or large during that period is of no

[12]Note that the timing of the peak is different for water than, say, electricity. With the latter, the peak tends to occur at certain hours of the day (e.g., late afternoon) and on weekdays instead of weekends. While daily and hourly variation in water use also occurs, the ability to store large amounts of water and the need for fire-flow capacity make it of much less significance than for electricity. For water, the seasonal peak is the focal point of peakpricing issues.

relevance, since it is the sum of all demands that creates the peak. Therefore, the price of a unit of water during the peak period for all customers should reflect the cost of providing this additional amount of water.

It is also true, however, that capacity available for service during the summer peak is also available during the winter off-peak. Indeed, if summer ceased to exist, part of the capacity would still be needed to serve winter. Should not, then, winter be responsible for the share of capacity it requires? Should, then, summer only be charged for the *extra* capacity above and beyond that needed to serve winter? The answer to both questions, in general, is no.[13] As Kahn states:

> Any attempt to shift capacity costs to the offpeak demands, by raising prices for that service above its own separate, incremental cost..., will cause available production capacity at that time... to be wasted, and would cut off purchasers willing to pay the additional cost of serving them. Any reduction of the peak... price below the full joint cost... would stimulate additional purchases at the peak, requiring additions to capacity that would not be made if buyers had to pay the full opportunity costs of the additional resources required to supply them. (Kahn 1988)

In the case of water service, it is the second instance that is of most importance. The underpricing of service during the peak period results in over consumption, and, in the long run, will encourage over development of water resources. From a practical standpoint, would adjusting the peak period price to reflect the full cost of water service have much impact on consumption? The answer depends on the responsiveness of demand to price. The empirical evidence suggests that residential demand for water service during the summer (the peak period for water service) is markedly more elastic than for the winter, and water demand for outdoor uses is more elastic still. Since outdoor use tends to drive the peak, this would suggest that underpricing service during the peak period could indeed have a significant

[13]Actually, a better general answer to the two questions posed would be "it depends." However, answering the questions in this Way would lead us too far astray. The reader interested in a very lucid and detailed discussion of capacity costing from the economist’s point of view is referred to (Hirshleifer, Dehaven, et al. 1960; Kahn 1988).

impact on consumption.

[C] Implementing seasonal rates

Before a utility can implement seasonal rates, however, it must address several additional considerations. These include the stability of the peak period with respect to price and time, measurement, and administrative feasibility. We address each of these in turn.

A legitimate concern for a utility is that the institution of peak-load pricing may simply shift the peak to another period. Faced with a higher price in the peak period, customers shift, en mass, to the offpeak period where price is lower.[14] Whether or not this is a significant possibility depends on the cross-price. elasticity between the two periods (i.e., the degree to which a change in price in one period affects the demand for the good in another period). In the case of seasonal rates for water, it is very unlikely that this would pose a problem. To a great extent, the difference in demand between the peak and off-peak periods is determined by factors other than price, such as climate. While an increase in the price of water during the peak period is likely to induce a decrease in consumption during that period, it is not likely to induce an appreciable increase in consumption during the off-peak period. Overall, the level of demand is likely to fall.

Another equally counterproductive possibility exists, however. Suppose that upon instituting a seasonal rate schedule the utility discovers that while the average level of demand during the peak period has declined, the actual peak day or hour demand, which determines the capacity requirement, has remained unchanged. This phenomenon has been labeled "needle peaking' (Beecher, Mann, et al. 1990). If we 'were to graph the pattern of consumption the peak would appear as a needle. Overall, the demand for water during the peak period has declined, which is good from the point of view of conservation, but the level of capacity required to meet demand has remained unchanged, which may be very bad from the point of view of utility finance. It means that a greater amount of capacity must stand idle for a longer period of time, thus deteriorating annual load factors and possibly eroding

[14]There are many examples, such as telecommunications (where customers switch some calls and fixes to the evening), airlines (where people switch vacations to off-season months), and electricity (where, for example, electric storage heaters are designed to use electricity during the evening hours).

revenue.[15] Whether this is a significant risk is largely an empirical question that must be answered with the hard hand of experience.

A final consideration is that the peak may shift through time for reasons unrelated to price. Changes in technology, preference, or social policy each have the possibility of shifting the peak period. The relevance of this is that consumers may base long-term investments in water conservation, in part, on the cost of water during the peak period, which ideally would reflect the long-run marginal cost of additional supply. As the peak shifts, so too will the long-run marginal cost, and thus what was a wise investment before the shift may no longer be so wise after the shift. Again, this does not appear to be a considerable risk in the case of water service, where hydrologic conditions maintain a very stable peak.[16]

Before a utility can institute peak pricing it must have some way to differentiate between peak and nonpeak consumption and to measure system costs that are associated with peak-period consumption versus those that are associated with off-peak consumption. In the case of urban water service, identifying the peak season of demand is straightforward. Low winter use and high summer use are fairly typical for urban water agencies in the west. Urban water utilities also experience daily peaks and valleys in demand in a manner similar to those seen in the electricity supply industry. To measure these demands, however, requires meters that can mark both the amount of consumption and the timing of consumption, which currently is beyond the grasp of almost all water utilities. Therefore, the practical application of peak-load pricing for retail water service is restricted to monthly demands.

Measuring system costs associated with peak consumption often is less straightforward, though not impossible. Marginal costs associated with a particular service or

[15]However, if the decline in demand is purely price motivated, and demand is price inelastic (as empirical studies suggest), then a positive change in utility revenue would result from a price increase.

[16]Technology, too, has been shown to exert its influence on water consumption patterns, sometimes in rather amusing ways. For example, starting in the late 1940s and early 50s, water agencies started to detect small evening peaks in demand recurring at fifteen and thirty-minute intervals. Eventually, it was determined that these peaks were the result of concentrated use of bathroom fixtures during TV commercial breaks. Perhaps with the adoption of ultra-low-flush toilets, evening demand will again resemble its pre-TV pattern.

class of customer should be assigned accordingly. This corresponds to the pricing rule that a good must cover its variable operating cost. Many system costs, however, are joint costs, and it is sometimes difficult to determine which use or combination of uses is the driving factor. This is particularly true for assigning causal responsibility for capacity. Fire service and peak summer service are the two principle candidates. As we discussed above, the pricing rules for the two types of service are very different-one being a fixed charge, the other being a volume charge. Thus, it is important that one not be assigned cost responsibility for the other. In general, fire-flow requirements will determine sizing for the distribution and storage network in the immediate vicinity of customers, while seasonal peak-day demand will determine sizing for most transmission lines, pumping stations, and treatment plants.

Finally, before adopting a seasonal rate structure-or any price structure for that matter -a utility should be reasonably confident that the gains from doing so will exceed the costs. Marginal cost principles, in addition to guiding the pricing of a good such as water, also can be applied to the feasibility and cost-effectiveness of a particular price structure itself. It is always possible that the cost of designing and administering a peak-load price structure will exceed its benefits. Whether this is a likely occurrence depends on numerous factors, including: the cost differential implicit in serving peak and off-peak uses; the price responsiveness of demand during the peak; the cost of measuring and assigning system costs to peak and off-peak uses; and, the cost of metering and billing peak and off-peak uses. For example, given the existing stock of metering equipment in place, it is very unlikely that potential water savings and load factor improvement derived from time-of-day rates would exceed the added cost of metering and billing necessary to implement them. On the other hand, it is very possible that seasonal pricing, which could utilize existing metering devices and would not require drastic changes in billing procedure, would be efficient; though this is not guaranteed to always be

[C] Seasonal rates summary

Seasonal rates are a potentially effective means realizing more efficient use of scarce water resources when demands on a water utility's system v systematically across seasons. Their primary advantage is that they provide to consumers an accurate signal of the *cost of* consumption, including the cost of capacity, in a given period. In this regard, seasonal rates have several advantages over more traditional approaches to pricing capacity, including:

- Consumption within periods responsible for capacity costs pay for those costs. Traditional approaches typically spread these costs over all periods. This can increase inefficiency and decrease equity. It increases inefficiency by underpricing water service in the peak period, thus encouraging too much consumption, and by overpricing water service in the offpeak period, thus encouraging too little consumption. It decreases equity because offpeak users, by paying a share of cost that they are not causally responsible for, implicitly subsidize the consumption of peak users.

- All uses during the peak period are recognized as contributing to and are charged for the cost of meeting the peak.

- Ideally, seasonal rates will also reflect the full cost of capacity required to meet the peak rather than just that portion in excess of average demand. Traditional approaches, on the other hand, go to great lengths to identify whether capacity is meeting average demand or peak demand requirements. In fact, the capacity is jointly meeting both, and causal responsibility for costs depends on the relative magnitudes of peak and offpeak demands. In cases where the differential between the two demands is large, the peak period will bear responsibility for the costs.[17]

The practical considerations of developing and applying a seasonal rate also must be taken into account. Seasonal rates may present a somewhat more complicated, and hence more expensive pricing formulation than one that does not use different rates for different periods. Therefore, the costs of design and administration must be carefully weighed against

[17] The exception being the capacity required for fire flow.

the potential gains in efficiency. Still, the increasingly common occurrence of a seasonal rate differential would suggest that these costs are not prohibitive for many urban water utilities.

[B] Increasing block rates

Increasing block rates are rates where the volume charge increases for successive quantum of water. For example, a rate that charges $1 per unit for the first 100 units and $1.50 per unit for all subsequent units is an increasing block rate. Many utilities now employ increasing block rates of one variety or another.

An increasing block rate is often proposed to meet two objectives: (1) to promote customer conservation; and (2) to satisfy revenue sufficiency constraints. Its potential to promote customer conservation is well described by the following passage:

> The impact [on conservation] of an increasing block rate design is best illustrated by comparing it to the simplest alternatives uniform design where all water is sold at the same price. Because of revenue sufficiency and other constraints, it can be assumed that both rates are initially designed to recover the same total revenue. The increasing block design, then, must contain one or more prices which are higher than the uniform design, and one or more which are lower. Customers facing the higher prices at the margin will, in theory, use less water than they would under the uniform design; customers facing lower prices at the margin will use more. The increasing block design will conserve water if the sum of decreases in use exceeds the sum of increases. (Metropolitan Water District of Southern California, 1991)

The expectation is that demand in the high blocks will be more elastic than demand in the low blocks, resulting in a net decrease in water use. The empirical evidence on elasticity suggests that this is not an unreasonable expectation; but neither is it a guaranteed result. The schedule's design, customer attributes, and regional climate will be important factors to any outcome.

Increasing block rates also are advanced as a means to preserve revenue neutrality. For example, when marginal costs are increasing, a marginal cost based rate will collect more than the utility's revenue requirement. A commonly proposed solution to this is an increasing block rate that sets the marginal price for only the last block and sets the price for prior blocks so that the utility breaks even. If all customers face the marginal cost for at least some of their consumption, the efficiency properties of a marginal cost rate design will be preserved. The problem, of course, is that it is exceedingly difficult to design a rate schedule that results in every customer facing the marginal rate for at least some of their consumption. More typical is a rate where some customers pay a high price and some pay a low price. Unless this is cost-justified, questions of equity will arise.

[C] Case studies of increasing-block rates

We noted above that increasing-block rates have been used by a number of water agencies at various times. In 1990, for example, sixteen of the sixty plus water agencies served by The Metropolitan Water District of Southern California had increasing block rates, as shown in Table 5-1 (some other agencies within Metropolitan's service area have adopted increasing-block rates since then, including Los Angeles whose experience is discussed in this section). Those rates were not necessarily adopted as conservation rates, but it may be instructive to analyze them from that perspective. In most cases, there are only two blocks, as opposed to having many, finely graduated blocks. There is considerable variety with regard to the magnitude of the price differentials between the blocks: in nine cases, the differentials are in the range of five to fifteen percent; in three cases, they are in the range of thirty-five to sixty-five percent; and in five cases, they exceed one hundred percent. What is most striking is the location of the switch points. In ten of the sixteen rates, the switch point is at 125-250 gallons per account per day. One thinks of a typical single family residential account as having three or four people in the household with each having an indoor use of between 60-80 gallons per day. Thus the total indoor use for the typical account may range between 210-280 gallons per day. Added to this amount is outdoor use, which will vary considerably by account, but that could easily average an additional 100-200 gallons per day, bringing the total to between 310-480. What these numbers suggest is that the majority of the utilities-ten of sixteen-employing increasing-block rates in Metropolitan's service area are locating switch points at levels too low for most singlefamily households to attain without a very substantial change in the pattern of household water use. A more appropriate location to induce households to switch from above the switch point to below might be somewhere in the range of 300-500 gallons per day.

Table 5-1. Retail agencies in MWD service area with block rates in 1990

Agency	Population Served	Rate Structure (Marginal Price in $/AF) (g/d - gallons/day/account)
LOS ANGELES COUNTY		
Beverly Hills	34,300	$0 up to 250 g/d $433 up to 374 g/d $469 up to 748 g/d $503 up to 1500 g/d $525 above 1500 g/d
Glendale	166,100	$159 up to 125 g/d $182 above 125 g/d
Inglewood	103,000	$601 up to 199 g/d $806 above 199 g/d
Las Virgenes	54,400	Zone:1 $283 up to 150 g/d $828 up to 300 g/d $558 above 300 g/d Zone 2: $923 up to 150 g/d $1032 up to 300 g/d $1198 above 300 g/d
Long Beach	419,800	$177 up to 125 g/d $362 above 125 g/d
Pasadena	132,200	$148 up to 374 g/d $317 above 374 g/d
Pomona	118,000	$136 up to 150 g/d $227 above 150 g/d
Santa Monica	96,500	$266 up to 250 g/d $301 above 250 g/d
ORANGE COUNTY		
Fountain Valley	55,600	$331 up to 250 g/d $392 above 250 g/d
San Clemente	35,100	$0 up to 125 g/d $540 up to 648 g/d $810 above 648 g/d
RIVERSIDE COUNTY		
Riverside	199,400	$65 up to 150 g/d $261 above 150 g/d

Table 5-1. Retail agencies in MWD service area with block rates in 1990 (Cont.)

Agency	Population Served	Rate Structure (Marginal Price in $/AF) (g/d - gallons/day/account)
SAN DIEGO COUNTY		
El Cajon	83,200	$365 up to 250 g/d $421 above 250 g/d
La Mesa	52,200	$365 up to 250 g/d $421 above 250 g/d
San Diego	1,027,360	$409 up to 125 g/d $468 above 125 g/d
VENTURA COUNTY		
Oxnard	124,000	$0 up to 125 g/d $457 up to 250 g/d $482 above 250 g/d
Simi Valley	94,500	$283 up to 685 g/d $414 above 685 g/d

Source: Metropolitan Water District of Southern California, Water Conservation Pricing, Approaches of The Metropolitan Water District, Staff Report, October 1990.

There are some cases where water districts have adopted increasing-block rates with the promotion of conservation as an explicit consideration in the design of ’the rate structure. We consider three of the more informative examples. The first involves a small water district serving irrigators in the San Joaquin Valley rather than M&I users. While the agricultural context is very different from what we have been discussing, in this instance, the basic principles of rate design remain the same and are particularly well illustrated by this example. The second example is Tucson Water, which has employed increasing-block rates, in part to promote conservation, for well over a decade. The third is the LADWP, which adopted increasing block rates in 1993.

What these three utilities have in common is that they attempted to use the rate structure to change the distribution of water use by targeting consumption at the high end of the range of consumption. By way of explanation, when one sees data on distribution of water usage by individual residential accounts there is almost always a distinctive *long-right tail-a* small fraction of the users at the high end account for quite a substantial fraction of total use. Rather than the symmetric, bell-shaped distribution of the normal probability distribution, one finds something closer to the distribution in Figure 5-2. For instance, the top 24 percent of single-family residential accounts in the LADWP service area in 1988 accounted for 47 percent of the total use, and the top 1 0 percent accounted for almost 27 percent of the total use. Overall, usage per residential account varied from as little as 25 gallons per day to as high as 22,400 gallons per day (there were 5 accounts using this amount). Some of this variation is due to differences in household size or lot size-but not all. Some of it is due to life-style, habit, and preference (e.g., whether or not people bother to fix leaky faucets, how much attention they pay when watering the yard, etc.).[18]

The existence of this distinctive pattern of household water use, with a long right tail, raises a fundamental question about the design of conservation oriented water rates. If one wants to reduce overall water use by *x percent,* should one aim to *shift the entire distribution* of water use to the left by *x percent, or* should one seek to *change the shape of the distribution,* pulling in the right tail in such a way as to reduce overall use by *x percent* The first strategy aims to change *every* customer's water use. The second *targets* customers with substantially above-average consumption; it rests on the notion that, if one could just attract their attention and induce them to change their pattern of use, this could lower total use without having to impact the behavior of the entire customer base. Which is better is an empirical matter that will certainly vary with circumstances. If targeting a large incentive at the fraction of users in the right tail is more effective than offering a smaller incentive for everybody, this strengthens the case for the strategy of changing the shape of the distribution. That was the objective in each of the three case studies, to which we now turn.

Example 1: broadview water district

Broadview water district is located near Firebaugh, California. It serves an area of about 10,000 acres of farmland, making it one of the smaller agricultural water districts in the state. Broadview is one of several irrigation districts on the west-side of the San Joaquin Valley implicated in the pollution of Kesterson Wildlife Refuge and the nearby San Joaquin River with selenium laden drainage water. Interim water quality standards for reaches of the San Joaquin River that serve as outfall for Broadview drainage were announced in 1987-88. To comply with the standards, Broadview would have to reduce its drainage into the river by

[18]The variation in use does *not* reflect difference in price, since these households were all paying the same, flat-rate price in 1988.

about 15 percent. It was determined that this could probably be done by reducing the amount of irrigation water applied to crops by approximately 10 percent, since over-irrigation was identified as a major cause of high drain flows.

In October 1988, the start of the 1989 crop year, the district announced a new rate structure. Previously, the district had charged a uniform commodity charge of $16 per acre-foot of delivered water, together with a fixed assessment of $42 per acre served by the district. Part of these charges covered the cost of the district-wide drainage collection and conveyance system. The annual cost to operate and maintain this system amounted to about $21 per acre-foot of collected drainage water or, when averaged over the roughly 25,000 acre-feet of water delivered annually by the district, about $3.08 per acre-foot. The district felt, however, that a surcharge of $3.08 per acre-foot would be too insubstantial to elicit a significant reduction in on-farm water use. Instead, it adopted a two tier rate structure; water in the first block continued to be priced at $16 per acre-foot; water in the second block would be priced at $40 per acre-foot. The idea was to use the revenues from the sales in the second block to cover the drainage system costs. Separate switchpoints were set for each crop grown in the district; these were set at 90 percent of the district-wide average water usage for the crops over the period 1986 through 1988, as shown in Table 5-2.[19] If irrigators, behaved towards the incentive as anticipated, applied water within the district would decline by about 10 percent as desired.

The results of the new rates in 1989 and 1990 are shown in the last two columns of Table 5-2. In four cases, there was already some reduction in water use in the 1989 growing season (a drought year), but not for cotton and wheat. With those two crops, efforts at conservation were offset by unusually hot temperatures during key parts of the 1989 growing season. For cotton and wheat, the most pronounced responses to the new rates occured in 1990. Cotton is the most important crop grown in the district, planted on about half the district acreage. Field level data for cotton indicate the effect the rate had on water use decision, as shown in Figure 5-5. Cotton was grown on 32 fields in 1989 and 33 fields in 1990. In 1989, the modal application rate (i.e., the rate that was applied on the greatest number of fields) was 3.6 acre-feet per acre. On eight fields, the application was higher than

---

This corresponds to the notion that switchpoints, to be effective, must be set at levels that are within the reach of customers being asked to conserve. If the district had simply set a single switchpoint for all crops, irrespective of individual crop water requirements, the effectiveness of the rate would have been greatly reduced.

this amount. In 1990, the modal application rate fell to 3 acre-feet per acre, very close to the switchpoint of 2.9 acre-feet per acre, and only one field had an application rate above the 1989 modal rate. Since 1990, the trend in water use for both cotton and other crops has continued downward, though the drought has been a confounding factor that makes it difficult to determine the extent to which the new rates have encouraged this trend. It is clear, nevertheless, that there has been a substantial reduction in water use in Broadview and that this is likely to be permanent. Numerous improvements in irrigation practice since 1989 have now become fixtures in the district. There is little doubt that these changes were greatly encouraged by the new rate structure.

Table 5-2. Crop specific water use in Broadview Water District, 1986-1990

Crop	Average Use 1986-1988 (af/acre)	Switchpoint (af/acre)	Average Use 1989 (af/acre)	Average Use 1990 (af/acre)	% Change 1986-88 to 1990
Cotton	3.20	2.9	3.34	2.84	-11
Tomatoes	3.22	2.9	2.73	3.03	-6
Melons	2.11	1.9	1.93	1.79	-15
Wheat	2.30	1.9	3.02	2.18	-5
Sugarbeets	4.58	3.9	3.73	2.54	-45
Alfalfa Seed	2.06	1.9	1.84	1.88	-9

SOURCE: Dennis Wichelns and David Cone. “Irrigation district programs motivate farmers to improve water management and reduce drain volume.” Presented at U.S. Committee on Irrigation and Drainage, 12th Technical Conference on Irrigation, Drainage and Flood Control, San Francisco, CA, November 13-16, 1991

There are several features of this story, some not yet mentioned, that act to make it unique, but also highly illustrative of the factors that contribute to the effectiveness of a conservation rate structure. First, the district was able to tailor its block rates in a way that enabled it to motivate all of its customers. Rather than apply a single rate to a heterogenous group of users, it differentiated rates by crop, and thus was able to apply individual rates to more or less homogenous groups of users. This allowed it to tailor the price signal to increase its effectiveness. Secondly, farmers within the district had detailed knowledge of their water use, monitored it fairly closely, and faced relatively drastic financial consequences if their water use was excessive. In other words, the farmers were very cognizant of the relationship between price and quantity, which is fundamental to customer responsiveness to price. Thirdly, farmers in the district had direct access to information on possible conservation strategies, knew how well their neighbors were doing, and understood the normative expectations of the district. Indeed, water usage rates were posted by the district each month, and anyone deviating from the norm could expect a good razzing from his fellow growers. Thus, the rate setting environment satisfied each of the factors discussed above that influence price responsiveness and make conservation rates effective: rates were tailored to target homogenous groups of users; the rate differentials presented users with significant incentive to curtail use; users themselves had a firm understanding of their use and how they could curb it; and normative expectations were clearly established and adhered to.

In an urban water district the situation is different in each respect. Urban utilities seldom face a homogeneous group of users for which to tailor rates. Typically, a utility's customers are differentiated only up to the point of primary class (e.g., residential, commercial, industrial, and institutional). Within any class, however, the degree of customer heterogeneity is still apt to be very great, which will make it much more difficult to develop a rate with price differentials and switchpoints relevant to a wide variety of customers. It is also unrealistic to expect an urban utility's customers, particularly residential customers, to understand and be aware of their water use to the same degree as the farmers of Broadview. While water is central to our day-to-day tasks, we do not face the same degree of risk to income if we use a little more than we should or fail to conserve as much as we should, and thus it is not too surprising that we are fairly ignorant of our water uses. For example, when asked how much water they consumed in a day, surveyed customers in Southern California were frequently off by a factor of ten. Thus, a key factor contributing to the effectiveness of conservation rates -- knowledge of use -- may be missing or muted in the urban setting. Finally, normative expectations become increasingly difficult to convey as the size of the group increases. In Broadview, the small number of growers made it relatively easy to impose a group standard. The same will seldom be the case for an urban utility.

Given these differences, is it reasonable to expect the same type of conservation potential from an inverted block rate in an urban setting? Probably not to the same degree as experienced in Broadview, but the experiences of at least one utility -- Tucson Water -- indicate that an inverted block rate can measurably decrease high-end consumption and result in significant water savings.

Example 2: Tucson water department

Tucson Water first instituted an inclining block rate in the mid-1970's. In 1977, it combined a four block rate for summer months with a flat rate for winter months. By 1986, it had switched to a six block rate for both winter and summer, though rates in the summer were set higher than in the winter. Water rates were increased each year between 1977 and 1986, though in inflation adjusted dollars water bills either remained constant or declined until 1982. Starting in 1982, rates were adjusted in a way to make it more expensive in real as well as in nominal dollars to consume above average amounts of water but left water bills for customers using average or below average amounts mostly unaffected. The notion was to motivate without coercing customers using considerably above average amounts of water to scale back their consumption.

Between 1982 and 1986, water bills calculated with inflation adjusted dollars for single family residential customers with usage three times 1978-79 average usage increased by 26 percent. Over the same period, bills for customers with usage equal to 1978-79 average usage increased about 7 percent, while bills for customers with usage equal to one-half this amount increased about 9 percent. The impact of this rate adjustment on consumption, particularly with respect to usage in the upper blocks, was analyzed by Cutherbert and Nichols (1987). Their analysis indicates that the rates were relatively successful in shifting consumption out of the upper blocks. Normalizing for weather and other factors, they found that for the period 1982 to 1986, annual usage in the upper three blocks as a share of total residential usage declined from 8 percent to 6.6 percent. More importantly, however, weather normalized average monthly usage in the peak usage months of June-July was 11 percent lower in 1986 than in 1982, but was left unaffected during the non-peak winter months, clearly suggesting that the rate structure was effectively curtailing discretionary outdoor use during the peak season when the utility is most at risk for shortage. Annual water savings were estimated to be 550,000 ccf in 1983, 1,000,000 ccf in 1984, 1,500,000 ccf in 1985, and 1,300,000 ccf in 1986. By 1986, then, estimated annual savings from shifting the distribution of consumption away from the high-end range equaled an amount of water sufficient to supply more than 9,000 customers.

Example 3: Los Angeles department of water and power

In the summer of 1991, in the face of a serious drought that had required rationing and emergency water rates to cope with a 15 percent shortfall in supply, Mayor Bradley appointed a Blue Ribbon Committee to consider LADWP’s water rates for the future. The committee made an extensive analysis of LADWP’s costs and then proposed an increasing-block rate structure designed specifically to target higher use customers.[20] The committee was

[20]One of the present authors (MH) served as a technical adviser to the committee.

committed to the basic principle of marginal cost pricing, but wanted to ensure revenue sufficiency in the face of a rising marginal cost of new supply sources. It saw the increasing-block rate structure coupled with carefully designed automatic rate adjustment mechanisms as a means of providing revenue sufficiency while at the same time promoting conservation through a targeted incentive. The committee also was concerned that the adoption of the new rate structure might be misconstrued by some simply as a revenue-generating device rather than as a means to promote more efficient use of water. Therefore, the structure was designed to be revenue neutral in comparison to its predecessor for the first year following implementation. [21] Another advantage of the two-tier structure was that it protected more conserving users, whom the committee felt should be rewarded for their efforts, not penalized. After some analysis, the committee decided that reclaimed water should be taken as the marginal source of supply for the purpose of estimaing off-peak marginal cost.[22] In addition, it propsoed seasonally varying rates, based on the change in marginal cost between off-peak and peak periods, with the peak users covering certain of the capacity costs of treatment, transmission and distribution.[23] The committee felt that the incentive offered by two-tier rates could still be effective for consumers who were below the switch point, as long as it was sufficiently close that the higher price loomed in their consciousness and could influence their behavior. Finally, the committee felt strongly that there should be the same rate

[21]After the first year, revenue would be allowed to vary from that generated under the old rate schedule. The rate ordinance guarantees LADWP a base revenue of $277 million and includes a Water Revenue Adjustment Factor to adjust rates if revenues fall short of targets.

[22]It was understood that this could change over time as new developments occurred, such as increased supplies becoming available from demand-side management or water markets sales by agricultural users.

[23]Having adopted flat rates in place of declining block rates in the previous drought (1977), LADWP introduced seasonally differentiated flat rates in 1985. The summer rate was initially set at 15 percent higher than the winter rate; by 1992, the differential had risen to 25 percent. The committee felt, however, that this differential was too small to reflect the real differences in correctly calculated peak and off-peak marginal costs, as well as too small to attract much notice from water users.

structure throughout the city, as opposed to having different structures (e.g., different switch points or different prices) in different areas.

The rates that the Blue Ribbon Committee recommended are shown in Table 5-3 and Table 5-4. The rates are structured differently for single-family residences compared to other users. For single-family residences, the switch point is located at 525 gallons/account/day. The other customer classes-multifamily residential, commercial and industrial- are considerably more heterogenous than the single families and it was felt that, for them, the switch point should be based not on some absolute level of use that would be the same for all users within the class but rather on a *relative* level of use, namely usage in the winter. Thus, in the winter there is a single block rate; in the summer, the first block applies to consumption up to 125 percent of winter consumption, and the second block rate applies to consumption beyond this level. [24] The second-block rate is the same for *all* users and is based on the estimate of LADWP's marginal cost of supply; it differs between summer and winter because of the peak-load pricing design. The rate for the first block varies among customer classes and is set so as to meet the revenue targets for that class.

In addition, there is a *separate* set of rates for *drought* years. The idea is to set down ahead of time the principles that will be followed when it comes to adjusting water rates in the course of a drought. The same type of block-rate structure still is applied in a drought year, but it is modified in two ways to adjust to the shortage. First, the switch point that which the second block commences is reduced, roughly in proportion to the severity of the shortfall. This means that the higher price will be triggered sooner during a drought than during normal supply years. Second, the rate charged in this second block is raised to equal what the committee estimated to be the rationing price that would equilibrate demand to

[24]It should be noted that sewer charges in the LADWP service area are based on the volume of water used during the winter period. This provides a complementary incentive against artificially boosting wintertime use for the sake of lowering summer water bills.

supply, given the shortfall.[25]

In January 1993, the L.A. City Council adopted a rate ordinance which closely followed the committee's recommendations. The main change was to raise the switch point for single-family residential accounts and differentiate it by season, placing it at 575 gallons/account/day in the winter and 725 gallons/account/day in summer. This was done mainly to accommodate the interests of residents living in the San Fernando Valley who face a warmer climate and tend to have larger lots than residents in the downtown and costal areas.

[25]These equilibrium prices were based on an analysis of LADWP's experience in the summer of 1991 when increasing-block rates with punitive upper rates were introduced temporarily to cope with the drought.

Table 5-3. Water rates proposed by LADWP Blue Ribbon Committee

Normal Year Rates				
	Price in Low Block ($/ccf)	Switch Point	Price in High Block ($/ccf)	
			Winter	Summer
Residential				
Single-Family	$1.71	525 gallons/day	$2.27	$2.92
Multifamily	$1.71	125% of winter use	NA	$2.92
Nonresidential	$1.78	125% of winter use	NA	$2.92

Drought Year Rates			
	Price in Low Blcok ($/ccf)	Switch Point	Price in High Block ($/ccf)
Residential			
10% Shortage	$1.71	475 gallons/day	$3.70
15% Shortage	$1.71	450 gallons/day	$4.44
20% Shortage	$1.71	425 gallons/day	$5.18
25% Shortage	$1.71	400 gallons/day	$6.05
Multifamily			
10% Shortage	$1.71	115% of adjusted winter use	$3.70
15% Shortage	$1.71	115% of adjusted winter use	$4.44
20% Shortage	$1.71	110% of adjusted winter use	$5.18
25% Shortage	$1.71	110% of adjusted winter use	$6.05
Nonresidential			
10% Shortage	$1.78	115% of adjusted winter use	$3.70
15% Shortage	$1.78	115% of adjusted winter use	$4.44
20% Shortage	$1.78	110% of adjusted winter use	$5.18
25% Shortage	$1.70	110% of adjusted winter use	$6.05

Table 5-4. Normal year rates adopted by L.A. City Council

Normal Year Rates				
	Price in Low Blcok ($/ccf)	Switch Point	Price in High Block ($/ccf)	
			Winter	Summer
Residential				
Single-Family	$1.14	575 gallons/day 725 gallons/day	$2.23	 $2.98
Multifamily	$1.14	125% of winter use	NA	$2.92
Nonresidential	$1.21	125% of winter use	NA	$2.98

LADWP staff had estimated that 71 percent of the single-family residences would have a lower water bill in a normal year with the new rate structure it replaced. They could have a higher bill in some of the summer months when their usage spilled over into the higher priced block, but not of the year as a whole. Residents of the San Fernando Valley would be somewhat more adversely affected—only 61.4 percent would have a lower bill over the course of the year. By contrast, in the downtown and on the west side, respectively, 84 percent and 91 percent of the single-family accounts would have a lower bill with the new rate structure.

In fact, these predictions have borne out well during the first twelve months with the rate experiment. Bye the end of the fall of 1993, however, it had run into strong political opposition from the San Fernando Valley. Several factors were responsible. In addition to the change in the water bill, there had been an increase in sewer charges, which are included on the water bill along with various other city charges. Customers attributed all of the change to the water rates. This coincided with the first hot months, September and October, when many households experienced a higher bill for the first time, after several months of lower bills. Moreover, a new mayor of Los Angeles had been elected with considerable support from the San Fernando Valley. As a result, the Mayor reconvened the Blue Ribbon Committee November 1993, and directed it to reconsider the rate structure to make it more equitable for residents of the San Fernando Valley. The new commitee decided to preserve the two-block rate structure and the rates associated with the block, but to make the switchpoint vary among different groups of residential users in a manner that would fit their individual circumstances and be more equitable. The committee decided that users should be grouped according to climate and lot size. It identified three seperate climatie zones with the LADWP above 85 F. It identified four ranges of lot size — below 7500 square feet, 7500 - 10,999, 11,000 - 17,499, and above 17,499. Because each group was now more homogeneous, the switchpoint was set at about 120% of the existing median use of customers within the group, with some minor adjustment in the lowest groups. The resulting switchpoints are shown in Table 5-5. This change had two advantages: compared to the earlier reate design with a uniform switchpoint for all residential users, the new structure increased the number of users actually facing the marginal cost incentive to conserve, and it distributed the benifits and costs of rate reform more evenly among all of LADWP's residential customers. These rates were adopted by the Los Angeles City Counsel with some small increases in the switchpoint and the subdivision of the largest lot size category into two categories, to provide a more generous allowance for very large lots, as shown in Table 5-6.

Table 5-5. Switchpoints recommended by the 1996 Blue Ribbon Committee

Lot Size (sq. Ft.)	Summer Average Daily High	Usage Charges at Low Initial Block Rate (gallons/day)	
		Winter	Summer
< 7,500	< 75°	325	400
	75-85°	325	425
	> 85°	325	425
7,500 - 10,999	< 75°	400	575
	75-85°	400	625
	> 85°	400	650
11,000 - 17,499	< 75°	575	900
	75-85°	600	975
	> 85°	600	1000
> 17,499	< 75°	725	1125
	75-85°	750	1200
	> 85°	750	1225

Table 5-6. Switchpoints adopted by city

Lot Size (sq. Ft.)	Temperature Zone	Usage Billed at Low Initial Block Rate (gallons/day)	
		Winter	Summer
< 7,500	Cool	325	400
	Moderate	350	450
	Hot	350	475
7,500 - 10,999	Cool	400	575
	Moderate	425	650
	Hot	425	675
11,000 - 17,499	Cool	600	900
	Moderate	625	1000
	Hot	625	1050
127,500 - 43,559	Cool	700	1125
	Moderate	725	1275
	Hot	725	1325
> 1 Acre	Cool	900	1375
	Moderate	950	1550
	Hot	950	1625

[C] Increasing block rates-summary

An increasing-block rate is often proposed to promote customer conservation through higher rates without violating revenue sufficiency constraints. For example, increasing block rates have been used by several utilities to reshape the distribution of consumption by discouraging high-end uses. They can be formulated based on average or marginal cost data. If designed so that every customer pays the marginal cost of service for at least some of their consumption, they can mimic the efficiency properties of marginal cost pricing. The heterogeneity of customer demands, however, makes this exceedingly difficult to accomplish. More typical is a block rate structure that results in some people paying a higher price for service than others. Unless there are cost-based reasons to do so , concerns about equity will emerge.

References

Beecher, J.A., P.C. Mann, er al. 1990. *Cost Allocation and Rate Design for Water Utilities*. The Natioanl Regulatory Research Institute.

Brown, S. And D. Sibley. *1986.The Theory of Public Utility Pricing*. Cambridge University Press.

Cuthbert, R.W. 1989. "Effectiveness of Conservation - Oriented Water Rates in Tuscon." *American Water Works Association Journal*. 81(33) 67-69.

Hall, Darwin C., and W. Michael Hanemann, "Urban Water Rate Design Based on marginal Cost," in Darwin C Hall (ed.), *Advances in the Economics of Environmental Resources Volume I: Marginal Cost Rate Design and Wholesale Water Markets.* JAI Press Inc., Greenwich CT, 1996.

Hirshleifer, J., J.C. Dehaven, and J.W. Milliman. 1960. *Water Supply: Economics, Technology, and Policy.* Chicago, IL, University of Chicago Press.

Kahn, A. 1988. *The Economics of Regulation: Principles and Institutions.* Cambridge, MA, The MIT Press.

Metropolitan Water District of Southern California. 1991. Water Conservation Pricing Approaches of the Metropolitan Water District.