ABSTRACT: When the goal of water pricing is elevated from mere cost recovery to deriving the greatest value from scarce water and associated nonwater resources, conventional rate regimes are found to be deficient. To address the challenge of creating rates that are both efficient and budget-balancing, several theoretical and practical aspects of rate-making are considered. Purposeful selection of rate parameters for a specific billing system is demonstrated to serve efficiency and cost recovery objectives. Attention to non-accounting opportunity costs is an important system element, but these costs are often not fully borne by the utility or its customers. In situations where this issue is serious, state or federal pricing policy may be necessary. (KEY TERMS: water pricing; water rates; water tariffs; block rate pricing.)

INTRODUCTION

The costs of municipal water supply are outpacing inflation due to three inexorable forces: heightened water scarcity, growing infrastructure costs, and rising health and environmental regulation. One implication of these forces is water cost increases that raise water rates. As a consequence, there is progressively greater public scrutiny upon the pricing policies of water utilities, and this situation promises to intensify as water bills continue to rise and alert consumers. Simultaneously, more attention will be drawn to the policies governing the establishment of water rates.

Until now, and likely beyond judging from current inertia, both rate-setting policies and governing rules have been dominated by accounting conventions rather than by economic ones (see, for example, American Water Works Association, 1991). The difference can be substantial (Martin et al., 1984). Accounting conventions emphasize revenue neutrality, that is, rate levels that generate revenue just sufficient to cover costs. Such a goal appears sensible, even equitable. Economic conventions concentrate upon the role of price in determining the welfare of consumers. Here, price is perceived as double-edged. Price has an obvious impact on household finances and production costs for water-using businesses. Additionally, because price modifies water consumption behavior (and utilities strive to provide the quantity consumers demand), it affects supply costs. These two “edges” work in opposing directions, and the economic perspective is to select prices that make the best tradeoff in advancing consumer welfare.

Whereas some water managers embrace the fairy tale that water is unique among all commodities and does not have price-sensitive demand, decades of econometric studies falsify this claim (Espey et al., 1997). Indeed, the much celebrated potential of water marketing (Easter et al., 1998; National Research Council, 1992) to remedy water scarcity is likely to be surpassed by the promise of pricing due to the influence of price on consumption. Marketing shapes the allocation of water among water purveyors (e.g. urban utilities and irrigation districts), but its real impact on individual water users disappears once the debt of a water right purchase is retired. On the other hand, improved pricing policy affects all water users' behavior while also influencing the supply and demand forces that drive water markets.

Because of the increasing relevance of water pricing policy, the objective of this paper is to unify and extend the economic depiction of proper water pricing. While the adopted context pertains to municipal
water pricing, the full analysis is equally applicable to irrigation districts supplying metered water. The discussion begins with a short treatment of possible pricing goals and the establishment of a basic model for designing water prices. The primary elements of the model depart from those of general public utility literature in some key respects in order to concentrate on the unique attributes of water service, and these differences are observed once the basic pricing model has been presented. Following this, three distinct categories of nonaccounting opportunity costs are individually considered. In each case, the pricing model is extended to capture the pricing implications of the added concern. Because some of these opportunity costs are external to utilities, corrective policy is suggested.

A BASIC PRICING MODEL

Fundamental social questions about water pricing concern the merits of alternative rate structures in terms of social desires for efficiency, equity, and revenue neutrality. Conservation enhancement is an additional desire, but we must be careful not to undertake conservation where more valuable resources are substituted for water (Baumann et al., 1984). Consequently, economic efficiency is a sufficiently encompassing objective to include deserving conservation.

In the case of equity, public concerns pertain largely to ability-to-pay issues, but in some circumstances there are also social preferences to force “water hogs” to pay higher rates. With regard to water hogging, the economic finding is that water overuse is an artifact of water underpricing, and once prices embed the appropriate social values, the need to penalize large water users evaporates.

With respect to ability-to-pay matters involving low-income households and elderly fixed-income households, the social issue is certainly legitimate. However, care should be exercised in viewing water service as a welfare program, because government involvement in water service historically arose from the sizeable amount of capital needed to establish water services or the natural monopoly character of water delivery, not from welfare issues relating to income distribution. Water bills should be perceived as what they are: requests for payment for a valued, delivered service, not a tax invoice for funding government programs. Water rates do not have a comparative advantage in correcting income inequity, and such attempts can be damaging to both efficiency and conservation objectives. With that said, the billing system to be emphasized later does possess positive attributes in moderating the water bills of low-volume water users.

While one must be realistic about seeking rates that serve multiple objectives, interesting schemes do exist for pursuing efficiency, revenue neutrality, and some degree of equity simultaneously. To explore this topic in a purposeful manner, it is useful to initially set aside the practical matters of peak loads, reliability, seasonality, depletion, and wastewater/sewerage. Next, a basic pricing model is formulated to examine the properties of desirable water rates and to provide a platform for testing alternative rate systems.

Model Notation

Suppose that our utility serves N baseline clients or connections. Each of these clients is connected by existing infrastructure (sunk costs) to the present system. In addition to existing clients, there are \( \Delta N \) additional clients who will become connected to the system within the current fiscal period. The \( \Delta N \) prospective clients may include households and firms relocating to the area’s newly developed properties, existing clients relocating to newly developed properties, and any peripheral properties not currently receiving water service. While the usual economic convention is to treat any “\( \Delta y \)” as an incremental extension of “\( y \)," explicit attention to \( \Delta N \) underscores the important cost implications of extending water service to new customers.

Each existing or new connection garners a benefit from water use, and this benefit is dependent on the amount of water consumed. Because both water use and its value may be unique to the client, let \( B_n(w_n) \) represent the \( n \)th connection’s benefit of consuming \( w_n \) units of water where \( n \in S = \{1,2,\ldots,N,N+1,\ldots,N+\Delta N\} \).

\[
C(W, N, \Delta N) \text{ represents the utility’s single period cost of serving } N + \Delta N \text{ connections that use a total of } W \text{ units of water. This cost function is defined in such a way that a new connection increases costs by } \partial C/\partial N + \partial C/\partial \Delta N \text{ even without any metered water use, meaning that } \partial C/\partial \Delta N \text{ is only the extra costs of a new connection (for added infrastructure and water supplies)} \text{ above those of an existing connection. It is also assumed that this cost function embodies a least-cost package of water sources, infrastructure, and administration, as is customary for cost functions. For consistency with the benefit functions, the cost function is expressed as a function of delivered (billed) water, rather than produced (billed plus lost) water. Hence,}
\]

\[
W = \sum_{n=1}^{N+\Delta N} w_n.
\]
By recognizing costs as functionally dependent on \( N \) and \( \Delta N \), two important facts are acknowledged. First, utilities face noteworthy operational costs that cannot be properly attributed to the amount of delivered water. For example, the maintenance of some forms of system infrastructure (e.g., storage tanks) is largely independent of water consumption. The same is true of costs due to meter reading, billing, customer service, and some aspects of administration (Martin et al., 1984). Even the water lost to distribution leakage is a consequence of maintaining a pressurized, ready-to-serve, albeit imperfect system. That is, the water cost of these leaks typically changes but slightly in response to altered water deliveries. Second, the costs of system growth for adding new connections is often laden with substantial nonwater and water costs. New infrastructure is costly to design, approve, and install. Moreover, growth commonly entails additional investments in water supply, and in water-scarce environments new water supplies can cost much more than they have in the past. In light of these observations, one must wonder if the common economic presumption of decreasing average costs in water supply is the erroneous result of associating all costs with water only, as if the cost function is well represented by \( C(W) \).

The direct incorporation of \( N \) and \( \Delta N \) as cost function arguments permits study of rate structures’ impacts on the allocation of both water and nonwater resources. This matter is rarely contemplated in water resource economics, even though it is of substantial importance in water resource accounting and rate design. Although the costs attributable to each existing connection may be functionally related to unique characteristics of the connection, in this basic model it is assumed that all existing connections are homogeneous with respect to their influences on costs. Similarly, all new connections are assumed to be have identical impacts on costs. Relaxations of these simplifications will be addressed later.

Establishing Economic Efficiency

Within this modeling framework there are three types of decisions that we wish to examine: water consumption by each customer, continuation of service by existing customers, and enrollment decisions by prospective new connections. The primary objective is to design a rate structure that encourages customers to make economically efficient consumption, continuation, and enrollment decisions. Balancing the utility’s budget is an additional objective that we wish to achieve.

To consider enrollment decisions by prospects, a multiperiod definition of economic efficiency is necessary. Here, we adopt the customary norm that efficiency is achieved when the present value of net social benefits is maximized over all current and prospective connections. Adding time \((t \, \text{superscripts})\) where needed to previous notation, the following optimization problem is obtained.

\[
\text{Max } PV(w, N, \Delta N) = \sum_{t=0}^{T} \rho^t \left[ \sum_{n=1}^{N'} B_n(w_n^t) - C^t \left( \sum_{n=1}^{N'+\Delta N'} w_n^t, \sum_{n=1}^{N'+\Delta N'} N_n^t, \Delta N_t \right) \right]
\]

where \( PV \) is the present value function employing a planning horizon of \( T \); \( w_t \) is the \( T \times (N + \Delta N) \) matrix of water consumption by all connections in all time periods; \( N \) and \( \Delta N \) are both \( T \times 1 \) vectors; and \( \rho^t \) is the discounting factor, which is the only instance in Equation (1) in which \( t \) serves as an exponent. It is assumed here that consumer and utility rates of discount are equivalent. In considering Equation (1) we employ the convention that we are only interested in determining \( \Delta N \) for the present time period, \( t = 0 \). In future time periods, Equation (1) can be reformulated and resolved with updated information, so the resolution of future decisions is not presently interesting. Any newly enrolled connections will contribute substantially to present costs, \( C^0 \), but once they are connected, their recurring costs will be different than those for existing connections.

Only the decision of \( \Delta N \) considers future periods in Equation (1). Water consumption and service continuations are static issues here, so it is practical to study these two decisions using the net benefits subproblem that is embedded in Equation (1). The static subproblem is to determine water use decisions and service continuations that maximize net benefits in the current planning period.

\[
\text{Max } NB(w, N) = \sum_{n=1}^{N+\Delta N} B_n(w_n) - C \left( \sum_{n=1}^{N+\Delta N} w_n, N, \Delta N \right)
\]

(2)

Here, \( NB \) denotes net benefits, and \( w \) is the \( 1 \times (N + \Delta N) \) vector of water consumption by all connections.

Solving the subproblem, Equation (2), we differentiate \( NB \) with respect to each \( w_n \) and set the result equal to zero. The result is
\[
\frac{dB_n}{dw_n} = \frac{\partial C}{\partial W} \quad \text{for all } n \in S. \tag{3}
\]

This is the requirement for equality between marginal benefits and marginal costs of water consumption for all connections. This is the necessary condition for allocating water efficiently.

Differentiating \(NB\) with respect to \(N\) and setting the result equal to zero, we obtain

\[
B_N(w_N) = \frac{\partial C}{\partial W} w_N + \frac{\partial C}{\partial N}. \tag{4}
\]

According to the first-order condition, Equation (4), the benefits received by connection \(N\), the system’s marginal continuing connection, should equal the marginal costs of serving connection \(N\). For every nonmarginal existing connection, benefits should exceed the marginal costs of servicing the connection.

Moving to the optimization of new connections, the necessary condition emerging from Equation (1) is

\[
\sum_{t=0}^{T} \rho^t \left[ B_{N^0+\Delta N^0}(w_{N^0+\Delta N^0}) - \frac{\partial C}{\partial W} w_{N^0+\Delta N^0} - \frac{\partial C}{\partial N^0} \right] = \frac{\partial C^0}{\partial \Delta N^0}. \tag{5}
\]

The left side of this equality is the present value of net benefits attributed to present and future water consumption by the marginal new connection, \(\Delta N^0\). The right side is the system cost of adding this user. For every nonmarginal new connection, the present value of net benefits (left side of Equation 5) should exceed the marginal costs of adding the connection (right side of Equation 5).

These necessary conditions, Equations (3) through (5), represent economically efficient targets for allocating water resources, connection resources, and new connection resources for the community. Utility managers do not directly decide these matters; however, consumers do. But consumers make their decisions in the context of policies established by utility management. The pivotal question and the focus of this paper is...is there a pricing policy which motivates consumers to make socially optimal decisions while preserving the nonprofit status of the utility?

**Developing Efficient Rates**

Let us propose the following multipart rate system. The first component is that new connections are required to pay a one-time fee designated by \(F_A\). Once they pay this connection charge and become part of the utility’s customer base, they are regarded as existing customers. Whether this fee is paid by a property’s developer or its first occupant is inconsequential, because the developer will incorporate the fee in the property’s price.

The second component is that all connections pay for their metered water consumption according to the following billing formula

\[
\text{Bill}_n = M + p \cdot (w_n - \bar{w}). \tag{6}
\]

This billing formula should not be confused with the typical two-block rate structure which is well represented by the following equation

\[
\text{Bill} = \begin{cases} 
\text{meter charge} & \text{if } w_n \leq \omega \\
\text{meter charge} + p \cdot (w_n - \omega) & \text{if } w_n > \omega
\end{cases}
\]

The latter billing formula is included here only for comparison purposes. It includes two water prices (unless \(\omega = 0\), implying a single block), and it fails to provide efficient price signals for low-volume water users.

The billing system in Equation (6) has three parameters that must be optimized: \(w_n\) is metered water consumption; \(M\) is a fixed “meter charge” to be paid each period that is equivalent for all customers as it is unrelated to metered water consumption, and homogeneous clients have been assumed thus far; and \(p\) is the price of water. As there are no blocks in this rate structure, every connection faces the same marginal price of water (p). The “bill threshold” is \(\bar{w}\). For connections that consume exactly the bill threshold, the bill will consist only of the meter charge. For connections consuming more than the bill threshold, they will pay, in addition to the meter charge, \(p\) for every unit of water exceeding the threshold. For connections consuming less than the bill threshold, a credit of \(p\) will be generated by every unit of water consumed below \(\bar{w}\). Should this credit exceed the meter charge, the customer will receive a payment rather than a bill.

When parameterized correctly, this atypical rate structure provides economically efficient incentives for consumers, a trait not shared with conventional rates. The general purpose of each part of the rate is as follows. \(F_A\) induces efficient connection decisions by new customers and offers a means of cost recovery for system extensions. \(M\) has two purposes too. In the unlikely event that a customer’s water service benefits fall below utility-incurred costs, \(M\) encourages the customer to discontinue service. More importantly, it
Effective Water Pricing

is a means of collecting service costs functionally related to connection costs rather than water costs. Hence, \( F \) and \( M \) address the efficient allocation of nonwater resources, although \( F \) will also include the cost of water acquisitions in water-scarce regions. The purposes of \( p \) are to allocate water efficiently and to recover costs. Lastly, \( \bar{w} \) is a budget-balancing parameter, introduced as a substitute mechanism for the average-cost pricing of water, which is the traditional method of balancing utility budgets. The deployment of \( \bar{w} \) makes each connection a shareholder in the system. Should the utility generate an economic surplus, it is dispersed uniformly across all connections using this instrument. If optimal prices result in a net financial loss for the utility, \( \bar{w} \) will be negative and it serves to collect the shortfall from all connections equally. Again, these findings are specific to the model structure which includes homogeneous consumers. If the utility serves clients of different classes, e.g. residential or commercial, there may be grounds for establishing different \( \bar{w} \)'s.

To develop optimal rate parameters, \( F^*, M^*, p^*, \) and \( \bar{w}^* \), consider first the water consumption decision of an existing customer. The \( n \)th connection rationally selects a consumption level that maximizes own net benefits. This presents the following problem:

\[
\text{Max } NB_n(w_n) = B_n(w_n) - M - p \cdot (w_n - \bar{w}).
\]  

(7)

Solving \( N+\Delta N \) of these problems gives rise to the following set of first-order conditions

\[
\frac{dB_n}{dw_n} = p \quad \text{for all } n \in S.
\]  

(8)

Comparing Equation (8) to Equation (3) we find that economic efficiency is advanced by marginal-cost pricing. That is, in an efficient rate structure, we must have

\[
p^* = \frac{\partial C}{\partial W}.
\]  

(9)

Every connection must face the same price of water, and that price must be identical to the system's marginal cost of supplying water. Thus, we have determined one parameter of the three that specify the billing system given by Equation (6). According to this result, there are three general aspects of contemporary rate structures which are deficient. First, low-volume water users face a zero price of water in the common system of entitling each connection to a certain quantity of free water. Second, multiblock systems contain more than one price of water. Third, rates are commonly founded on average costs rather than marginal costs. These deficiencies may advance revenue neutrality goals, but they are clearly inefficient with respect to water allocation.

With respect to continuing connections, the marginal customer, client \( N \), will be indifferent to maintaining service when benefits received are precisely offset by the bill. Expressed mathematically,

\[
B_N(w_N) = M + p \cdot (w_N - \bar{w}).
\]

Substituting from the optimality condition of Equation (4) and the optimal water pricing rule of Equation (9),

\[
\frac{\partial C}{\partial W} w_N + \frac{\partial C}{\partial N} = M + \frac{\partial C}{\partial W} (w_N - \bar{w})
\]

and solving for the meter charge, we obtain

\[
M = \frac{\partial C}{\partial W} + \frac{\partial C}{\partial W} \bar{w}.
\]  

(10)

Intuition readily suggests the first term of the righthand side of this equality but not the second, which constitutes a precise offset of the budget-balancing mechanism introduced by \( \bar{w} \). The second term emerges in the above derivation in order to prevent consumers from maintaining service mainly to receive their "rent share". On the other hand, the combined result of Equations (6), (9), and (10) is to eliminate the budget-balancing term, \( \bar{w} \), from the billing structure given by Equation (6).

It is difficult to design a rate structure that serves multiple masters, in this case efficiency and revenue neutrality. The billing system in Equation (6) does not distribute revenue excesses or shortfalls in a fully lump-sum manner (they only go to active connections), so something has to be sacrificed. Either we abort the pursuit of efficiency or we abort balanced budgets. In the present case, it seems very sensible to select balanced budgets over allocative efficiency in the continuation of connections. The lowered meter charge given by Equation (11) is arguably a proper tradeoff for this issue.

\[
M^* = \frac{\partial C}{\partial N}.
\]  

(11)

Relative to other factors, capturing rent share should be a very small enticement to continue service, so little allocative efficiency should be lost specifying the
meter charge with Equation (11) rather than Equation (10). If a property is idled or vacant and therefore consumes no water, its owner may wish to continue service if rent share exceeds the meter charge (thereby causing an allocative inefficiency), but this problem is virtually negated by the opportunity cost of land and buildings, which make it costly for property owners to maintain idle connections.

Next we consider the decision calculus of a prospective connection. Once the enrolling connection pays $F_\Delta$, it is eligible to receive the present and future benefits of water deliveries from the utility. This is sensible to the prospect if the present value of net benefits exceed $F_\Delta$. For the indifferent (marginal) prospect designated as $N$,

$$F_\Delta = \sum_{t=0}^{T} \rho^t \left[ B_t^{N+\Delta N} (w_{N+\Delta N}^t) - M - p \cdot (w_{N+\Delta N}^t - \bar{w}) \right].$$

When the prior results embodied in the pricing findings of Equations (9) and (10) and the social optimality condition of Equation (5) are substituted into this expression, we obtain the optimal fee for new connections

$$F_\Delta^* = \frac{\partial C^0}{\partial \Delta N}.$$ (12)

This rule is the intuitively expected result. New additions to the system must pay the marginal costs of their new connections. To do otherwise encourages nonoptimal expansion and detracts from the social net benefits of water service in the community.

The final billing parameter to be resolved is the billing threshold, $\bar{w}$. Rather than being efficiency-based, the billing threshold acts to balance the utility's budget. Hence, it is obtained by finding that level of $\bar{w}$ for which utility revenues are equivalent to costs. In the following expression, revenues are assembled on the left side of the equation; costs are on the right

$$F_\Delta^* \Delta N + M^* \bullet (N + \Delta N) + p^* W - p^* \bar{w} \bullet (N + \Delta N) = C(W, N, \Delta N).$$

Solving for $\bar{w}$ yields the optimal billing threshold,

$$\bar{w}^* = \frac{F_\Delta^* \Delta N + M^* \bullet (N + \Delta N) + p^* W - C(W, N, \Delta N)}{p^* \bullet (N + \Delta N)}.$$ (13)

Looking beyond the complex appearance of Equation (13), $\bar{w}^*$, is the level of the billing threshold necessary to assign all revenue shortfalls or to disperse all rents/profits accruing to the utility. While there are certainly other candidate mechanisms for achieving revenue neutrality, this one is straightforward, and it introduces a measure of fairness by treating all connections as equal partners in the utility. Most importantly, unlike average-cost pricing, the billing threshold does not impede the efficient allocation of water resources.

Depending on cost function properties relating to returns to scale, it is possible for the numerator of Equation (13) to be negative, in which case $\bar{w}$ is negative and its purpose is to assign responsibilities for revenue shortfalls. In such cases, $\bar{w}$ performs the same service as the second part of the two-part tariff idea that was promulgated long ago for utilities experiencing decreasing average costs (Kahn, 1988, Chapter 4).

**GENERAL PUBLIC UTILITY PRICING THEORY**

Economic literature pertaining to public utilities has had a longstanding interest in efficient pricing, so it is useful to consider both the overlap and departures of the above water pricing results in relation to the more general literature. Public utility economics has provided many contributions (reviewed by Berg and Tschirhart, 1995), most of which are not directly germane for the objectives of the pricing model set forth above.

One relevant facet of public utility theory focuses upon the specification of optimal prices for goods supplied by natural monopolies. A natural monopoly is a "firm" operating in a situation in which the introduction of any competition (other firms) would actually lower consumer welfare due to the nature of the cost function for this industry. Loosely speaking, when average costs per unit of output are decreasing across the relevant range of demand, we have a natural monopoly (Baumol, 1977). Public utility literature has long scrutinized and espoused the notion of a "two-part tariff" for achieving an economic efficient level of production in natural monopoly situations (Lewis, 1941; Brown et al., 1992). The first part involves a marginal cost based price for each unit of the good consumed. The second part involves a flat "access fee" for each consumer, so that the sum of these fees cover revenue shortfalls implied by the fact that marginal costs are lower than average costs for natural monopolies.
The multipart rates of the water pricing model developed here emerge for largely these reasons but not entirely. Optimal water and new connection rates \((p \text{ and } F_{A})\) are found to be marginal cost prices required to achieve efficiency in the allocation of water and nonwater resources. The “optimal” meter charge \((M)\) deviates from the efficient level in order to preserve revenue neutrality, but it is the marginal cost of connections, and the efficiency sacrifice should be small. The fourth part of the rate structure, \(w\), serves to balance the utility’s budget, so that it neither suffers a loss nor captures a profit. However, the derivation of this system of water prices did not presume that the cost structure is one of decreasing average costs. Indeed, no structure was assumed for the cost function except to say that it may be functionally dependent upon three primary items, \(W, N, \text{ and } \Delta N\). Average costs may be decreasing, constant, or increasing in these three arguments, and the water pricing results will be unchanged.

Viewed in this way the water price findings are only slightly more general than an extended two-part tariff, but they are also more applicable in some key respects. The sole per unit price of the two-part tariff is a water charge, so nonwater resource costs are implicitly portrayed as being functionally dependent on the amount of supplied water. However, water utilities incur large costs which are functionally unrelated to water deliveries, and consumers make important choices influencing these costs. Additional price signals are needed so that (1) efficiency can be attained for the allocation of nonwater resources, and (2) water price signals are not overstatement. The separability of water distribution functions from water production functions are well observed within contemporary utility accounting procedures (Raftelis, 1993). Therefore, the multipart water tariff developed here serves to bring water pricing theory into compliance with accounting practice as well as the economic implications of separable costs.

It is noteworthy that the public utility literature’s assumption of a decreasing cost industry may have declining relevance for water utilities. Water demand growth has generally allowed decreasing-cost ranges of average cost functions to be more fully exploited. Also, when cost functions include the opportunity costs of scarce resources such as water and infrastructure, marginal-cost prices are more likely to generate adequate revenue for the utility. The inclusion of resource opportunity costs is an important matter to be considered later.

SOME REFINEMENTS

The billing system specified by Equations (6), (9), and (11) through (13) provides socially attractive price signals because the derivation of these prices emphasized efficiency while preserving revenue neutrality. These signals are also water conserving in that any reduction in water use from that motivated here would be socially wasteful, as their benefits would be exceeded by their costs. Of course, the application of this system for any utility hinges upon the elaboration of the cost function, \(C(W,N,\Delta N)\), but such cost information is regularly collected as part of recommended accounting practices for contemporary utilities.

There exist certain complexities of water and nonwater allocation that may require more thorough consideration than that undertaken above, but several such complexities are readily entertained through slight extensions of the above theory. Examples include different water types or different connection characteristics having distinguishable effects on the utility’s costs. When the costs attributable to water change seasonally, the \(W\) term of our cost function should be partitioned appropriately, and we should pursue seasonal prices. The same is true if the utility delivers any water types with different cost implications (e.g., treated water, unprocessed water, and treated wastewater). Likewise, if different customer classes impact costs differently, as with service capacities (pipe sizes), peaking characters, or connection elevations for example, then the cost function should be refined to acknowledge the differential impacts for both continuing connections and, if appropriate, new connections. The result is that the correct cost function may be of the form \(C(W_a,W_b,...,N_i,N_{ii},...;\Delta N_i,\Delta N_{ii},...),\) and the rate structure will possess \(p_a,p_b,...;\) \(M_i,M_{ij},...\) and \(F_{aij}, F_{aiii},...\). Readily obtained enhancements of Equations (6), (9), and (11) through (13) specify this pricing system completely.

EXPLORING MARGINAL COSTS

Having established a general framework that demonstrates the sensibility of volumetric water pricing working in concert with nonwater fees, the informational requirements of proper pricing no longer differ much from those underlying normal pricing procedures. Conventional accounting and pricing practices involve the separation of utility-incurred costs into various categories. An intended purpose of these partitions is to enable the computation of separate prices for water usage and for connections.
Consequently, the basis for establishing efficient prices is in place within recommended accounting procedures.

In addition to the inefficiencies of block rate water pricing, conventional pricing practices are deficient in two broader ways. In the first, the correct cost information is misprocessed by calculating average prices rather than marginal prices. Average-cost pricing is only justified when there are legitimate arguments indicating that marginal costs are well approximated by average costs.

The second area of deficiency concerns nonaccounting opportunity costs (NOCs, pronounced “knocks”). Opportunity costs are a fundamental element of economic decision making. An opportunity cost is the value of a foregone option resulting from a decision. In many instances, opportunity costs are accounting costs too (e.g., labor, energy, and land), so there is no difference between the accounting and economic perspectives on what counts. But some opportunity costs do not become accounting costs. These we label NOCs, and when they are significant, accounting-based prices can seriously misvalue either water or nonwater aspects of the rate structure. Due to the great importance of NOCs in certain situations, the following three sections discuss three distinct NOCs and their implications for resolving socially attractive rate structures.

**Marginal Value of Raw Water**

Interestingly, the greatest share of the rates we customarily pay for finished (tap) water are based on the nonwater resources used as inputs. For the most part, finished water prices are the consequence of value added by the utility for the administration, conveyance, storage, pressurization, and treatment of water. If clients are paying 2¢ per cubic meter at the tap, it cannot be said that added raw water supplies are worth 2¢/m³. To determine the implicit unit worth of additional raw water, all the value-adding cost items must be subtracted from the 2¢. Such calculations often indicate a zero or near zero value for raw water, because the finished water price includes no economic rent component (in the case of a nonprofit utility) and the utility possesses a debt-free raw water entitlement. Under these conditions, there is no value assigned to the raw water resource. This is not necessarily bad, because in some times or regions water may not be scarce – meaning that water use involves no forfeitures in sacrificed alternative uses. But when and where water is scarce, finished water prices should incorporate the marginal value of raw water, MVW. This is readily demonstrated.

When the optimization subproblem of Equation (2) is augmented to observe water scarcity, the most general approach is to add net benefit functions for the other water users in the watershed and a raw water constraint for the watershed. Alternatively, we can assume that \( Z^* \) is the optimal raw water allocation to our utility and we must not use more water than that. If \( L \) denotes the ratio of raw water lost in conveyance, the utility’s constraint is

\[
\sum_{n=1}^{N+\Delta N} w_n \leq (1-L) \cdot Z^*.
\]

Employing the Lagrangian method of optimization in the presence of a constraint, the revised, current period subproblem is

\[
\max \sum_{n=1}^{N+\Delta N} B_n(w_n) - C \sum_{n=1}^{N+\Delta N} w_n \cdot N, \Delta N
\]

\[+ \lambda \cdot (1-L) \cdot Z^* - \sum_{n=1}^{N+\Delta N} w_n \]

where \( \lambda \) is the introduced Lagrange multiplier. When the constraint is binding, \( \lambda \) is positive, and it connotes the marginal value of raw water in the community. Following the prior procedure to develop an efficient water price, we obtain

\[
p^* = \frac{\partial C}{\partial W} + \lambda.
\]

That is, it is easily demonstrated that price should include \( \lambda \), the MVW. Though not an accounting cost for the utility, \( \lambda \) is an opportunity cost. Failure to include it in water price detracts from net benefits in the watershed. Failure to include it must either lead to inefficient nonprice rationing in the community (thereby lowering net benefits for the utility’s consumers) or lead to overallocations of water to the utility to the detriment of the watershed’s other water users (lowering net benefits for the region). To avoid double counting, any raw water costs appearing as accounting costs in \( C(W, N, \Delta N) \) must be dropped from \( C \) before applying the pricing rule given by Equation (14).

As a consequence of the revised water pricing rule in Equation (14), the utility receives revenue that is unmatched by accounting costs. Hence, the billing threshold must be increased in order to maintain a balanced budget, but the adjustment to Equation (13)
Effective Water Pricing

is easily made. Similar adjustments are necessitated by the incorporation of all NOCs.

Operationalizing the pricing rule indicated by Equation (14) does not always require solution of the constrained subproblem. In regions where evidence for MVW is available, this value need only be converted into the proper units. For example, if permanent surface water rights are presently trading at $1/m^3, and the real rate of discount (excluding inflation) is 4 percent, then 3.85¢/m^3 is the implied value of raw water for the present period ($1 • (0.04/(1+0.04))). This procedure assumes a constant value for raw water in all periods, thereby overvaluing MVW in periods of plenty and undervaluing it in times of drought. If short-term (rental) water market values are observable and market participation is robust, then these values will be a preferred alternative.

Marginal User Cost

In the prior section we considered the value of raw water that arises when its use involves a sacrifice to alternative uses. In situations relating to unrenewed ground water supplies, water use causes a sacrifice in alternative future uses. That is, when scarce ground water is used now, those units of water will be unavailable for future use, begetting another type of NOC. The technical name for this opportunity cost is marginal user cost, defined as the value of sacrificed future uses, discounted to present value (Tietenberg, 2000:90). Moncur and Pollack (1988) simply refer to these opportunity costs as a “scarcity rent” in their investigation of its magnitude for a specific setting. To obtain the proper pricing rule, the optimization problem in Equation (1) must be augmented by linking future water costs to water usage in the present period, thereby making the problem dynamic.

We have already determined that it is optimal to price water so that everyone faces the same rates, so we can omit client-specific aspects of the optimization model and aggregate all benefits into the terms $B'(W';N',\Delta N')$. To acknowledge that the utility’s pumping costs are affected by water table elevation or, loosely, hydraulic head in the pumping period, $H^t$, the cost functions are revised to $C'(W^t,N^t,\Delta N^t,H^0 + \sum_{\tau=0}^{t-1} h(W^\tau;R^\tau))$. Again, we have no immediate interest in affecting $N$ or $\Delta N$ beyond period 0. Higher head lowers pumping costs, so $\partial C/\partial H^t < 0$. Pumping in the current period tends to lower the hydraulic head experienced in all future periods. Likewise, natural ground water recharge, $R^t$, increases the head. The dynamic connection between head in consecutive periods is given by

$$H^{t+1} = H^t + h(W^t; R^t) \quad \text{for all } t > 0.$$  \hfill (15)

where $h$ is the head gain function translating pumping and recharge into changes in hydraulic head. The current period’s hydraulic head is $H^0$. Successive substitutions of Equation (15) result in

$$H^t = H^0 + \sum_{\tau=0}^{t-1} h(W^\tau;R^\tau) \quad \text{for all } t > 0.$$  \hfill (16)

Substitution of the refined cost function and Equation (16) into Equation (1) provides an optimization problem appropriate for utilities mining ground water:

$$\begin{align*}
\text{Max } PV(W,N,\Delta N) &= \sum_{t=0}^{T} \rho^t \\
&= \left[ B'(W^t) - C'(W^t,N^t,\Delta N^t,H^0 + \sum_{\tau=0}^{t-1} h(W^\tau;R^\tau)) \right].
\end{align*}$$  \hfill (17)

Optimization results pertaining to the selection of $N^0$ and $\Delta N^0$ are unaltered in Equation (17). The necessary condition for resolving water use in the immediate period is

$$\frac{dB^0}{dW^0} - \frac{\partial C^0}{\partial W^0} - \frac{\partial h}{\partial W^0} \sum_{t=1}^{T} \rho^t \frac{\partial C^t}{\partial H^t} = 0.$$  

Comparing this finding with Equation (8), the optimal pricing rule is obtained:

$$p^* = \frac{\partial C^0}{\partial W^0} + \frac{\partial h}{\partial W^0} \sum_{t=1}^{T} \rho^t \frac{\partial C^t}{\partial H^t}.$$  \hfill (18)

When Equation (18) is contrasted to Equation (9), it is seen that optimal pricing involves a new component — marginal user cost (MUC)

$$\text{MUC} = \frac{\partial h}{\partial W^0} \sum_{t=1}^{T} \rho^t \frac{\partial C^t}{\partial H^t}.$$  \hfill (19)

MUC is positive because both derivatives of Equation (19) are negative. The most challenging element of this formula to operationalize is $\partial C/\partial H^t$, the impact of head changes on costs. In circumstances where present and future mining is slight because of the contributions of natural recharge, $\partial C/\partial H^t$ will be dominated by the energy costs of greater pumping lifts. If ground water mining is more severe,
decreased head may periodically induce the utility to install new wells together with connecting infrastructure to offset lowered pumping capacities. In more extreme cases, the utility may need to bear the costs of switching to surface water as ground water is depleted. All of these investments become part of \( \partial C/\partial H^t \) during the periods when pumping capacity or surface water sources are to be expanded.

Although the fixed costs of induced ground and surface water investments may easily dominate MUC in cases involving moderate to severe mining, it can be instructive to bound MUC by focusing on the energy elements of \( \partial C/\partial H^t \). In general, modified head has a linear effect on pumping costs, implying that this aspect of \( \partial C/\partial H^t \) is relatively constant across periods. Denote the energy cost implications of lowered head by \( \partial C/\partial H^t \). Employing these assumptions, we obtain the following lower bound for MUC:

\[
MUC \geq \frac{\partial h}{\partial W} \frac{\partial C}{\partial H^t} \sum_{t=1}^{\infty} \rho^t = -\frac{\partial h}{\partial W} \frac{\partial C}{\partial H^t} \rho^{t-1} \,\rho - 1.
\]

It is relatively simple to operationalize the latter expression. The first derivative on the right hand side can be estimated from general aquifer parameters. The value of the second derivative can be determined through statistical analysis of pumping records containing energy usage and water table elevation or through modeling of ground water physics. The result constitutes a minimum MUC, and the omitted values are possibly large, so there is strong motivation for both including the minimum MUC in price and striving for the additional information needed to achieve Equation (18) more completely.

**Marginal Capacity Cost**

An interesting facet of water planning concerns capacity expansion. Here lie crucial decisions relating to the choice of new facilities as well as the timing and sizing of these facilities. Such decisions often have large cost implications for the utility, so their efficient resolution can be important. A noteworthy aspect of new investments in water supply capital is their lumpiness. That is, in many instances it is efficient to undertake projects of greater scale than that necessary to satisfy the current quantity of water demanded. Hence, utilities tend to expand in spurts. As system capacity becomes limiting, utilities undertake the next project. Upon project completion, excess capacity exists for accommodating more demand growth, but growth eventually consumes this capacity and the next project is engaged.

The economic view is that optimal project timing maximizes the present value of net benefits. This is a more demanding criterion than requiring projects to pass a cost-benefit test. Pursuit of this goal typically implies that it is not rational to build projects in advance of demand. Moreover, it is often economically efficient to not build projects when demand for their capacity is slight. Premature construction is costly due to the time value of money and capital depreciation. Following an efficient path for the timing of lumpy projects therefore means that there are periods during which water supply capacity is less than the quantity of water demanded. During such periods, there is a third NOC to contemplate – marginal capacity cost (MCC). The existence of this NOC can be demonstrated to have implications for efficient pricing.

Suppose that the value of the utility’s existing capital is given by \( K^0 \). A “yield” function, denoted by \( Y(K) \), translates capital into available water supply. In future periods, the capital base will be enhanced by investments and decreased by depreciation. Let investment in period \( t, I^t \), contribute to capital in period \( t + 1 \), and suppose the rate of depreciation is constant over time and given by \( 1 - \alpha \) where \( 0 < \alpha < 1 \). Hence,

\[
K^t = \alpha K^{t-1} + I^{t-1} \quad \text{for all } t > 0.
\]

Successive substitutions of this expression result in

\[
K^t = \alpha^t K^0 + \sum_{\tau=0}^{t-1} \alpha^\tau I^{t-\tau - 1} \quad \text{(20)}
\]

where only the superscripts to \( \alpha \) are exponents. In all periods, water supply is constrained by capacity, so

\[
W^t \leq Y(K^t) \quad \text{for all } t \geq 0. \quad \text{(21)}
\]

When Equation (1) is augmented by investment costs and the constraints of Equations (20) and (21), the resulting Lagrangian is

\[
\text{Max} \sum_{t=0}^{T} \rho^t \left[ B^t(W^t) - C^t(W^t, N, \Delta N) - rI^t \right] \]

\[
+ \sum_{t=0}^{T} \delta^t \left[ Y\left( \alpha^t K^0 + \sum_{\tau=0}^{t-1} \alpha^\tau I^{t-\tau - 1} \right) - W^t \right] \quad \text{(22)}
\]

where \( r \) is the price of investment, and \( \delta^t \) are the introduced Lagrange multipliers for the supply capacity constraints of each period. For periods during
which water supply capacity is binding, \( \delta^t > 0 \). The revised efficiency problem of Equation (22) has no new implications for the selection or pricing of \( N \) or \( \Delta N \), so let us focus upon the immediate decisions \( W^0 \) and \( \delta^0 \). Differentiating with respect to these two variables and setting them equal to zero yields

\[
\frac{dB^0}{dW^0} - \frac{dC^0}{dW^0} - \delta^0 = 0
\]  

(23)

and

\[
-r + \sum_{t=1}^{T} \delta^t \frac{dY}{dK^t} \alpha^{t-1} = 0.
\]  

(24)

Equation (24) indicates the efficient level of investment for the current period. In concert with Equation (8), Equation (23) tells us how to price current water

\[
p^* = \frac{\partial C^0}{\partial W^0} + \delta^0.
\]  

(25)

Price should include \( \delta^0 \) which is marginal capacity cost in the current period.

The effect of including MCC in price is to efficiently ration available capacity during capacity-constrained periods. If price omits MCC during these periods, the quantity demanded will exceed supply. Two inefficient consequences can then occur. First, the shortfall will have to be accommodated through some nonprice allocation policy, implying that the marginal value of finished water will not be equivalent across all clients. Second, deficient pricing will create a prevailing opinion, among both clients and responsive utility managers, that projects should be initiated to rectify the perceived shortage. If such action is taken, the investment regime will be accelerated beyond an efficient pace (failing Equation 24), and the present value of net benefits will be decreased.

The problematic aspect of incorporating MCC in price is its changing level over time. As supply capacity becomes more constraining over time due to growth in demand, MCC increases, but upon completion of a supply-enhancing project, MCC commonly falls to zero and remains there until further demand growth eliminates the excess capacity (Turvey, 1976). Hence, MCC rises and falls over time. Its incorporation in rates brings about a long-term cycling in rates which may be objectionable to customers. Recognition of this issue has led to the creation of smoothing substitutes for MCC (Mann et al., 1980), with the acknowledgement that some sacrifice in economic efficiency is made whenever rates do not embed true MCC. Still, use of an MCC substitute in water pricing is a likely improvement over abandonment of the matter.

Whereas the MCC derived from Equation (22) would appear difficult to obtain in light of the dynamic extent of the problem, the character of the solution illuminated by Equations (23) and (24) reveals some separability between the optimal investment strategy and MCC. Current MCC can be observed without calculating optimal current and future investments. Since only current MCC should be part of current water price, there is no need to solve an elaborate problem to acquire the future. Assuming that price is being established for the current period using Equation (25) and that excess demand of an estimatible amount is anticipated in the absence of an MCC-inclusive price, we need only know demand elasticity in order to estimate a price increase sufficient to assuage the excess demand. This rate hike is MCC.

COLLECTED PRICING IMPLICATIONS

The previous sections develop water pricing recommendations pertaining to three distinct NOCs. If all apply at the same time in an additive manner, the pricing advice would be to include all three along with \( \partial C^0/\partial W^0 \), but this is an unlikely scenario. The MVW pertains primarily to surface water sources, but it is also applicable to fully renewable ground water sources. In either case raw water must be socially scarce for the MVW to be nonzero. MUC arises from the future opportunity costs of present water consumption, so it is primarily relevant to mined ground water sources. Hence, the marginal value of raw water and marginal user cost tend to arise in different circumstances, depending on the raw water source. For utilities employing both renewable and nonrenewable water sources, economic efficiency dictates a blending that achieves equivalent marginal costs (inclusive of NOCs) or a scheduling that first employs sources having the lowest marginal costs (inclusive of NOCs). Considering all possibilities, the emerging price proposal is

\[
p^* = \frac{\partial C^0}{\partial W^0} + \text{MVW} \quad \text{or} \quad p^* = \frac{\partial C^0}{\partial W^0} + \text{MUC}
\]  

(26)

depending, respectively, on whether renewable or nonrenewable water resources are being employed at the margin, \( W^0 \). In the case of blending at the margin, the prices of Equation (26) are equal. In hydrologic situations where surface water use lowers ground water recharge or where ground water use induces surface water flows into ground water bodies, MVW
and MUC values become interdependent, but Equation (26) still applies.

Whereas the MVW and MUC arise from opportunity costs relating to water sources, the MCC is the consequence of constrained capital, so the prices of Equation (26) must be augmented by MCC to allocate water efficiently. Of course, if the MVW or MUC has been incorporated in price, water quantity demanded is moderated and the necessary MCC will be lower. The complete pricing recommendation is

\[
p^* = \frac{\partial C}{\partial W^0} + \text{MVW} + \text{MCC} \]

or

\[
p^* = \frac{\partial C}{\partial W^0} + \text{MUC} + \text{MCC}.
\]

**FLAWED LOCAL PRICING POLICY**

One of the most noteworthy features of the NOCs just discussed is the accounting stance over which each is relevant. To the extent that each possesses a local accounting stance, one may be confident that well informed local utility management will design efficient rates. But where these opportunity costs fall outside the utility's jurisdiction, local pricing policy may be suspect.

The MCC has to do with capital-constrained water supply. In many cases these limiting facilities will be controlled and owned by the utility, so the MCC calculated by the utility will be a correct one. Should the utility be receiving water supply from a water purveyor that serves as a wholesaler to the utility and other buyers, and should the facilities of this purveyor be the constraining ones, then the achievement of efficiency hinges upon the allocative policies of the water purveyor. If the purveyor establishes quotas for each of its clients as a means of partitioning its capacity, the buyer's own quota will be a sound basis for determining its specific MCC. Alternatively, the purveyor can compute an overall MCC and make this part of its wholesale price, but such purveyors commonly serve few clients. The paucity of buyers will likely confuse the application of this approach, because MCC payments generate a rent for the wholesaler. Assuming that the wholesaler is obliged to be nonprofit, this rent must be dispersed in a lump sum, efficiency-serving manner – a challenge for a wholesaler serving few buyers. However, assuming the wholesaler calculates MCC well and the wholesaler's buyers include this MCC in finished water prices for their customers using Equation (27), economic efficiency is achievable.

The prospects for sound local pricing policy are weaker for the MVW and MUC. Ideally, the MVW should be the regional value of renewable water. In areas where water marketing is not robust, clues regarding the value of raw water may be hard to find. In such cases, utilities are unlikely to commission proper regional studies of raw water value so that they may enhance regional efficiency by improving local pricing. Often, the efficiency gains of improved pricing lie substantially exterior to the utility or its customers. For example, by virtue of history or aggressive planning, some utilities in water scarce regions possess water entitlements in excess of current demand quantities even when MVW is omitted from price. For such utilities, the adoption of MVW-inclusive pricing will lower quantity demanded, potentially revealing substantial excess supply. Such a surplus may dismay customers, who may have footed the bill for this accumulation and are now being asked to pay MVW, and it may persuade other water users in the region to undertake political or legal actions to expropriate the exposed water surplus.

Thus, depending on circumstances and institutions, the prime beneficiaries of MVW-inclusive pricing may be external to some utilities, and such pricing may be locally regarded as unattractive. Of course, if water marketing is effective in the region, these problems may be circumvented through mutually beneficial trades, but water marketing is not yet practiced in many jurisdictions.

The remaining NOC, marginal user cost, was demonstrated earlier to be based upon the future costs consequent to current water consumption. Thus, it pertains primarily to nonrenewable ground water. As shown by Equation (19), the pivotal elements of MUC are the future costs of lowered hydraulic head. Unless the utility is the sole user of the aquifer – an unlikely situation, some of the costs of lowered head fall to other pumpers. Hence, the normal circumstance is that a local accounting stance for assessing MUC would cause MUC to be undervalued. In the extreme, the utility may be a minor pumper of the aquifer in which case a locally computed MUC might drastically underestimate the social MUC. At the heart of this issue is the common property nature of ground water and the importance of collective action for achieving economic efficiency in mining ground water.

Overall, the extrajurisdictional elements of these NOCs suggest that local pricing policy will not achieve economically efficient resource allocations. This problem implies a need for well crafted policies to guide all utilities in their development of water tariffs. Many such policies are conceivable, but the most obvious is for state or national authorities to formulate, update, and require specific MVW and MUC tariffs.
amounts pertaining to each major watershed and aquifer. If this approach is taken, NOCs can be collected at the local level and dispersed across clients employing an appropriately revised bill threshold (Equation 13), thus raising the role of this instrument.

CONCLUSIONS

When the goal of water pricing is elevated from mere cost recovery to deriving the greatest value from water and associated nonwater resources, conventional rate regimes are found to be deficient. This is important because revenue neutrality need not be sacrificed to achieve economic efficiency. A simple billing system is demonstrated to be both efficient and revenue neutral, providing that its parameters are purposefully chosen. Other efficient billing systems may also exist, and the economic theory assembled here serves as a benchmark for gauging alternative rate proposals.

It is noteworthy that revenue neutrality can be achieved by a simple and practical device, the billing threshold (\( \bar{w} \)), thereby allowing abandonment of wasteful average-cost pricing considering only accounting costs. The effect of \( \bar{w} \) is to take the financial surpluses or deficits of water supply that are experienced by the utility and assign them equally across connections. The result is efficient, and it is arguably equitable. In addition, low-income households can utilize the threshold to their advantage.

From the perspective of this investigation, it is redundant to seek rates that advance water conservation. Because reduced water employment must be accompanied by reduced profits, reduced consumer satisfaction, and/or the substitution of other valuable resources for water, economically efficient water use embeds the proper degree of water conservation. Contemporary calls for water conservation in public policy are artifacts of deficient pricing schemes. Conventional rates encourage inefficient levels of water use because they omit important social values pertaining to water scarcity. This scarcity originates from intersectoral competition (MVC), depletion (MUC), and limited infrastructure (MCC). Once these values are reflected in rates and once marginal cost pricing is adopted, the need to advance water conservation vanishes. Moreover, nonprice conservation policy is likely to be a poor substitute for proper pricing, in the sense that it is price guides that most easily induce consumers to maximize the value derived from water.

Whereas it is demonstrated that water price should include applicable nonaccounting opportunity costs, it is also found that these costs are not fully borne by the utility and its clients. Hence, even if utilities choose to rely on a pricing doctrine that attends to NOCs, their own accounting stances may cause the NOCs to be understated. In some cases, this issue is likely to be serious, and state or federal policy regarding pricing practices may be a necessary remedy. Alternatively, water markets have some capability to address this issue by allowing transactions that signal surface water or ground water scarcity. Indeed, pricing and marketing policy appear to be complementary institutions insofar as the efficiency of water markets, which are fueled by excess water demand and excess water supply, is dependent on the establishment of efficient rates. Where rates are wrong, marketing will be misled.

LITERATURE CITED


