Water rate policy: prescription and practice

Water scarcity is prone to mismanagement due to influential features that obscure efficient policy choices and establish political obstacles to the installation of efficient policy. Chief among these influences are the natural monopoly character of water service, the blended rival/nonrival nature of water consumption, and the potentials to deplete ground water stocks, reduce environmental flows and degrade water quality. Acting in concert, these features create a tangle of challenges spanning the full theoretical range of market failures – those conditions that warn us about the capabilities of decentralization. Merely "letting the market work" is not an option for processed water.

Although governments can establish transferable water permits to use the naturally occurring waters of streams and aquifers and thereafter rely on administered markets to advance allocative efficiency for raw (unprocessed, in situ, natural) water, such strategies are impractical for processed water. Consequently, if we are to coax good behavior from water consumers and avert popular pressure for uneconomic water development projects, it is necessary to get the rates right and marry them with sound regulations that can be activated during harsh seasons or drought. In this pursuit, our economic purpose is to design rates that optimize the collective rent, surplus, profit, and satisfaction derived from water enjoyment, including all environmental concerns. This is a more welfare-hungry goal than merely matching water demand with its availability.

The central objective of this chapter is to assemble the principal, theoretical advice applicable to water rate-making.[1] Much of this advice has weakly penetrated the design of rates worldwide, as competition is not available to root out unfortunate rate-making practices.

Hopefully, the advantages of efficient water rates can be achieved in the future, once the inadequacies of present policy are better revealed via poor performance. Until then, we are likely to witness growing problems related to water overuse, infrastructural shortfalls, environmental water shortage, habitat loss, too-rapid ground water depletion, and slow adaptation to climate change – all of which are with us now.

Single supplier rates

Professional thinking about water must avoid the confusion inherent to the visual similarity of raw water and processed water. Whereas furniture does not look like trees and bread does not look like wheat, processed water does resemble the raw water that is available from streams and aquifers. It is important to distinguish the policy tools for the input (raw water) from those applicable to the outputs (such as the tap water received by households). Moreover, processed water can be much more valuable than raw water because of nonwater expenditures undertaken in production.

To firm up some principles that separately address raw water policy (such as marketing) and finished water policy (especially rate-making) while also providing a platform that integrates the two, a central model setting is the sector or locale served by a single water utility. Because scale economies and the high costs of duplicate facilities infer that this utility is a natural monopoly, competition is not an economic option. Suppose that this service provider converts raw water, w, into finished water, w^f, using a leaky linear process, $w^f = kw–d$, $0<k<1$, $d>0$. The water transformation and delivery costs of this process are given by the function $C(w^f)$. The water supply is physically or legally limited to $\overline{W}$, so $w \leq \overline{W}$. Denote group benefits as $B(w^f)$, inferring among other things that the group has a pre-established policy for conducting internal allocation. This policy might be quotas, queuing, intermittent or rotating service, regulations

upon specific uses, consume-and-be-billed, or any number of instruments and combinations. Soon we shall presume that the group benefits function B is simply the sum of its members' benefits, thereby assuming that internal allocation is optimally conducted. Later, the consequences of relaxing this assumption are recognized.

If we apply the potential Pareto optimizing goal of maximizing net benefits irrespective of their distribution (aggregate economic efficiency), the consequent Lagrangian is

$$L_1 = B(w^f) - C(w^f) - \delta(w^f - kw + d) - \lambda(w - \overline{W}) \quad , \tag{1}$$

and its optimization yields first-order conditions

$$B' - C' - \delta = 0 \text{ and } \delta k - \lambda = 0$$

which can be combined to inform us that

$$B' = C' + \frac{\lambda}{k} \quad . \tag{2}$$

According to eq. (2), it is optimal to select w and w^f so that marginal benefits are balanced against marginal costs, as always. Here there are two types of marginal costs to recognize. There are the marginal accounting costs of processing water and the scarcity value of limited water supply rescaled (upwards) by k, where k is that ratio of raw water surviving a production process that necessarily involves leakages such as evaporation and conveyance losses. Failure to acknowledge this scarcity value leads to water overuse and associated social problems (such as political pressures for inefficient water supply expansions). Exercising and thinking about eq. (2), useful findings and extensions are immediately available:

- A well informed client, i, facing a volumetric rate p and having private water use benefits given by $B_i(w_i^f)$, where $\sum_i B_i$ defines B above, rationally maximizes utility or profit at

$B_i' = p$. As this is true for all knowledgeable clients, eq. (2) advises the efficiency-seeking water manager to set the optimal volumetric rate p* at marginal costs:

$$p^* = C' + \frac{\lambda}{k}, \tag{3}$$

and this is optimal pricing for all customers regardless of quantity consumed. Hence, popular policy thrusts such as increasing block rates or customized rates are inefficient.[2] The only efficiently differential rates happen across customers for which marginal costs are different, such as would occur for customers associated with different water treatment costs or different distribution distances or elevations. Otherwise, as shown by eq. (3), all customers should face an equivalent volumetric charge comprised of marginal accounting costs and the marginal opportunity cost of limited water scaled upwards in accordance with treatment and conveyance losses.

- Under fortuitous conditions the water utility may draw its water in a region where private rights support a water market, which may disclose raw water value, λ, to observers. Otherwise, the supplier must perform deeper analysis to ascertain λ, or, absent such efforts by the supplier, it may socially prudent to have a value dictated to the supplier by higher authorities. λ is a shadow price rather than an observable price for most water service suppliers, so the matter is commonly overlooked. The omitted values, which are more fully discussed in a forthcoming section, may apply to (a) limited renewable surface water which is changing seasonally or annually; (b) limited ground water which may be being mined and may have a value that is strictly rising over time; (c) infrastructural constraints restricting either processing capacity or available water; and/or (d) rationing costs stemming from nonrate policies that are switched on during shortfall events. Rate policy commonly omits these values because of public suppliers' focus on recovering accounting costs, C, whereas λ is a nonaccounting opportunity cost. Wherever raw water is legally established as state or common property, as it is in much of the world, water markets capable of transforming this opportunity cost into an accounting cost are normally infeasible until property institutions are modified.

Interregional rates and allocation

Expanding the model to examine two water service providers, we can index the prior notation to distinguish a potential exporter, x, of raw water and a potential importer, m. If x and m lie on the same river, export may involve nothing more than reducing withdrawals, leaving the water instream, and allowing the importer to withdraw it. Or it may involve the use of existing infrastructure or the original construction of new conduits. The model's results generate additional insights about intersectoral and interregional raw water values and optimal processed water rates.

The exchange of raw water can involve two sacrifices: some water may be lost to the environment during transport, and there may be financial costs incurred for transport and transaction costs. We shall assume that when x foregoes taking w_t units of water so that M can have more, m receives $rw_t - d_t$ units ($0<r\leq 1$, $d_t \geq 0$). Reallocation costs are $ew_t + E$ with $e,E \geq 0$. Additional costs such as the exchange of money from m to x to pay for water rights are treated as zero-sum and ignored here. In the present model (to be modified later), water use is assumed to be rival, implying that return flow from one of these service areas is not available for reuse in the other area.

Instead of applying the potential Pareto criterion of maximizing summed benefits, unequal social weighting of the two regions can be tested by employing the Pareto criterion. Here, we shall maximize m's net benefits subject to the constraint that x receives net benefits of a minimum arbitrary amount, N_x. Combining all elements in Lagrangian form yields

$$\begin{aligned} L_2 = {} & B_m - C_m - \alpha(N_x - B_x + C_x) - ew_t - E \\ & - \delta_m(w_m^f - k_m w_m + d_m) - \delta_x(w_x^f - k_x w_x + d_x) \\ & - \lambda_m(w_m - \overline{W}_m - rw_t + d_t) - \lambda_x(w_x - \overline{W}_x + w_t) \end{aligned} \quad . \qquad (4)$$

It is harmless to assume interior (nonzero) values for all decision variables within this problem because positive water consumption is to be expected for both regions, and because the two regions become fully independent when $w_t = 0$ (inferring two problems of the eq. (1) variety above are sufficient to isolate rate recommendations). Focusing on rate-making implications, the several first-order conditions derivable from (4) can be algebraically processed into the following relations:

$$B'_m = C'_m + \frac{\lambda_m}{k_m} = p^*_m \quad \text{and} \quad B'_x = C'_x + \frac{\lambda_x}{\alpha k_x} = p^*_x \; ; \tag{5}$$

$$\text{and} \quad \lambda_x = r\lambda_m - e \; . \tag{6}$$

Primary implications are:

- According to eqs. (5) the spirit of pricing rule (3) is preserved for the two-region case, and it is now seen that it may be optimal to maintain different prices across regions even when their marginal accounting costs (C') are equivalent.
- Normatively, the two regions are equally weighted only when $\alpha=1$. When $\alpha=1$ is regarded as socially appropriate, the Pareto problem given by eq. (4) is reduced to a potential Pareto problem maximizing summed rewards, and the resulting pricing rules (5) are slightly simplified.
- Whereas water reallocation between the two regions is only motivated by the discrepancy in raw water's shadow value between the two regions, reallocation succeeds in bringing these values closer together but not enough to equilibrate them.[3] Eq. (6) states that equivalence between the two shadow prices occurs only in the polar case where (i) there are no delivery losses in transporting water from x to m and (ii) there are no variable costs such as occur for pumping. For example, if m and x are coexisting sectors such as residences and commercial enterprises being served by the same utility on the same distribution network, then the opportunity costs of water are equal for both sectors, and it is also necessary that $\alpha=1$ for $p^*_m = p^*_x$.

Necessary extensions

Detailing the opportunity costs

Throughout the public utility literature there is ample discussion about the advantages and disadvantages of marginal cost pricing (Hotelling 1938; Ruggles 1949; Kahn 1988). These writings strongly emphasize the application of rates to recover accounting costs, as opposed to nonaccounting opportunity costs (NOCs), focusing on industries such as electricity and telecommunications. The tendency in these industries is for all inputs to be reasonably valued as true accounting costs, unlike the circumstances for water. The public utility literature recognizes the desire to ration limited infrastructural capacity, extending to peak loading issues (Steiner 1957) as well as weather-driven uncertainty. Deterministic peak loads can be well managed through the design of time-dependent rates (Hirshleifer 1958). However, once uncertainty of demand and/or supply enter the picture, rates become an incomplete policy solution that must be supplemented by other public policies (Crew, Fernando, and Kleindorfer 1995).

At a formative level, the NOCs (λ, λ_m, and λ_x above) of water use are readily understood. If the utility is expecting a supply shortfall in the sense that quantity demanded exceeds the utility's ability to supply, then the basic short-run opportunity cost can be computed as that extra, per-unit value required to reduce quantity demanded to the available supply. Empirically, the economist only needs to possess some demand information, such as demand elasticity, and a quantitative estimate of the shortfall in order to estimate this opportunity cost. Theoretically, this value has different origins and names, and it may only represent part of the needed revisions to rates if we are to achieve economic efficiency (Griffin 2001). One additional matter is the longer run objective of supporting an optimal supply (composed of infrastructure and water rights), recognizing that the currently developed supply may under- or over-shoot its efficient level. In a

changing economy with or without population growth, there is an efficient schedule of supply expansions[4] to pursue in light of the social costs and benefits of such projects. Good rates support this schedule by seeking to align quantity demanded with supply under expected weather conditions (Dandy, McBean, and Hutchinson 1985; Riordan 1971). Variable weather will of course produce periods of excess demand or excess supply if rates do not rise and fall in response. The generally low price elasticity of water demand limits the ability of rates alone to equilibrate demand and supply, because rates would have to have large variability across seasons.

In the standard case of a variable surface water supply that is renewed by precipitation, λ may be called the *marginal opportunity costs of raw water*. This value is likely to vary seasonally as well as year-to-year. It acknowledges the regional worth of naturally occurring water, not merely worth to the clientele of the utility.

In the case of depletable ground water supplies, the applicable economic term for λ is *marginal user cost*, and this value may grow in a structured Hotelling fashion over time. This value is reflective of the tradeoff between using marginal units now versus using them in the future, so it is explicitly designed to accompany an economically efficient rate of depletion. Computation in this case must be forward looking and dynamic (as in Pitafi and Roumasset 2009).

In the case of restricted infrastructure rather than restricted water, the operational term is *marginal capacity costs* (Turvey 1976). Because of the essential character of both raw water and infrastructural capital in the delivery of treated water, in some settings it may difficult to ascertain which portion of λ is attributable to capacity constraints and which portion is based on water value. Regardless of what we call them however, their rate-making implications are

identical. Because increments to water supply and water delivery infrastructure should be timed so as maximize rewards, marginal capacity costs also have a strong dynamic character. For example, they fall to zero in periods of excess supply.

When water service planning involves uncertainty and part of the solution (yes, solution!) is periodic water supply shortfalls, the concept of *marginal rationing costs* arises. As explained in the next section, when temporary policies are applied during shortfall events and these policies lack the ability to ration water on the basis of willingness-to-pay (as can rates), there are losses accompanying these policies that can be efficiently reduced with an appropriate markup to water rates (marginal rationing costs).

Taken as a group, these opportunity costs constitute an analytical burden for well-intentioned economists and practitioners. Yet, they are also complementary in that the proper incorporation of one may reduce or even nullify another's relevance, and they are all working to recommend reduced consumption in the interest of economic efficiency. Moreover, there may be other opportunity costs that point in the same direction, such as when water use degrades a water body's quality in a thermal or chemical manner.

Uncertainty and reliability

Variable weather pushes water demand and supply in opposing directions, exacerbating both general scarcity and peak loads. When summer peaks in temperatures and deficient precipitation worsen during hotter/drier years, the opportunity costs of water surge, and the reliabilities of utility systems are tested. As general water scarcity rises in a region and the effects of uncertainty become more acute, neither the economist's nor the engineer's prime policy measures are efficient tools in isolation.

Although scarcity-sensitive rates are underutilized within contemporary policy, excess demand cannot be trimmed adequately solely using higher rates. One limitation is public rejection of rates sufficiently varying to do the job, acknowledging the low price elasticity of demand. Another is the impracticality of short-run water metering. Rates cannot change with sufficient velocity to reflect the moving "spot market" value of water, because customer water meters are not continuously read. With monthly readings at best, shorter term water shortfalls are incompletely manageable with rates. Nor can excess demand be controlled efficiently by "overbuilding" the supply system, given the ultimate physical scarcity of water. Recognizing the great variability in weather, perfectly reliable water supply systems are too expensive to support. As scarcity and the inherent costs of reliability rise, the optimal reliability of water supply systems tends to be reduced.

When seasonal or annual patterns of demand and supply are stochastic, there continues to be a mutually supportive balance to strike with rates and infrastructure. Moreover, non-rate and non-infrastructural strategies must be introduced to ration water during periods of excess demand. Examples are "drought management programs," temporary policies such as lawn watering bans or alternate-day watering, conservation mandates, water use audits, and educational efforts. In more extreme situations uncommon to developed countries, rationing may involve intermittent water service, public taps, and private arrangements such as trucked water and point-of-use storage (e.g. household cisterns). Together, all of these policies – rates, infrastructure, and nonrate rationing – are mutually related and interdependent, and it is constructive to think of them as portfolio elements.

During extreme periods when rates are incompletely rationing and other rationing mechanisms are in play, it must be acknowledged that water is not being denied to its lowest-

valued applications (Visscher 1973). For example, banning or restricting a particular use of water because it is allegedly low-valued overlooks the heterogeneities of personal preferences and water productivities. Although marginal uses such as outdoor water applications in urban areas and agricultural irrigation tend to be pivotal control behaviors during shortfall events, it should be conceded that some instances of irrigation by either sector are high valued relative to other uses. The same concession applies for all allegedly low-valued uses targeted by nonrate rationing strategies. Consequently, rationing costs occur when a rationing policy incorrectly limits some not-so-low-valued uses. Other rationing costs arise from the administration requirements of all non-rate policies. Nonrate policies can also be "disruptive" in ways that impose additional costs (Crew, Fernando, and Kleindorfer 1995).

Nonrivalness and return flow

A fortunate though perplexing facet of water use is that consumption can be nonrival to a degree. For example, whereas households mostly use water rivally within their utility's service area, this same consumption may be imperfectly nonrival vis-à-vis downstream households due to surface water return flow or ground water recharge. Such situations lower the opportunity costs of water use and have rate-making implications.

Suppose that a ratio u of the exporting region's use of finished water returns to the raw water supply so as to be reusable by the importing region. Similarly, suppose that ratio v of the importing region's finished water use becomes available to region x. The last two constraints within the Lagrangian (4) above are then altered to obtain this revision:

$$\begin{aligned} L_3 = {} & B_m - C_m - \alpha(N_x - B_x + C_x) - ew_t - E \\ & - \delta_m(w_m^f - k_m w_m + d_m) - \delta_x(w_x^f - k_x w_x + d_x) \\ & - \lambda_m(w_m - \overline{W}_m - rw_t + d_t - uw_x^f) - \lambda_x(w_x - \overline{W}_x + w_t - vw_m^f) \end{aligned} \quad . \qquad (7)$$

Typical hydrological conditions may not support both u>0 and v>0, but both are allowed for generality here. The recommended water rates become

$$B'_m = C'_m + \frac{\lambda_m}{k_m} - v\lambda_x = p^*_m \quad \text{and} \quad B'_x = C'_x + \frac{\lambda_x}{\alpha k_x} - \frac{u}{\alpha}\lambda_m = p^*_x \tag{8}$$

with no modification to condition (6). A reusability credit now appears in each rate-making rule of eqs. (8). Therefore, rates are lowered in accordance with scarcity in the other region.

These phenomena can become more complex as other water-using entities of a watershed are considered and multiple uses become practical for individual water molecules. Complexity is generated by the linked-but-different local scarcity circumstances of a watershed. Environmental values stemming from streamflow maintenance (for habitat and recreation for example) and end-of-river estuary inflows can also affect optimal rates. For example, if x lies down-river from m, consumption by m may have an opportunity cost that does not apply to x: the reduction in streamflow along the river segment separating them (Griffin and Hsu 1993). These varied hydrologic circumstances give rise to distinct spatial and dynamic relationships among the raw water values in a basin (Chakravorty and Umetsu 2003). Although we know that these optimality relationships exist, they are difficult to estimate well. This is dismaying in the sense of making optimal rate making more difficult to achieve, but it is not an excuse for ignoring shadow prices entirely as is common today.

Given the narrow accounting stances of local water suppliers, how would they quantify the out-of-jurisdiction values within eqs. (8)? What of the out-of-jurisdiction values that appear in similar guidance from more complex hydrological circumstances? Given their self-interests, why would utilities bother to try? If a water utility is economically vigilant insofar as pursuing the best interests of its clients, it will be motivated to assess own-λ but not the opportunity costs of nearby jurisdictions utilizing shared waters. Only under idealized water market conditions

might the full spatial character of optimal shadow prices be revealed in the absence of careful study. Therefore, from a regional, provincial, state, or watershed perspective, there are oversights to be expected when local authorities dominate the design of rates (Griffin 2001). Accounting stance matters in determining applicable opportunity costs. A compelling conclusion is that stronger policies are needed to guide rate-making. Recently, a similar idea became a highlighted element of Europe's "Water Framework Directive" which devotes considerable attention to the ideal of "full cost water pricing" (Unnerstall 2007), but the Directive appears to emphasize accounting costs and the removal of popular subsidies. Whether or not this policy development can be successfully extended to incorporate NOCs remains unclear.

Revenue sufficiency

The preceding theory slights utility managers' desires to have a balanced budget. Indeed, achieving revenues sufficient to cover all costs is crucial in most U.S. water utility settings whereas economic efficiency is lightly appreciated. Simultaneous pursuit of efficiency and balanced budgets can be accomplished in two ways. First, pricing rules such as (5) can be optimally perturbed so as to achieve a balanced budget as efficiently as possible. This is the strategy behind Ramsey pricing and its second-best, inverse-elasticity pricing findings for public utilities in general (Baumol and Bradford 1970). Such findings are not particularly compelling for water ratemaking because water utilities do not normally rely on a unidimensional volumetric price for billing customers.

The second avenue is to apply all of the pricing instruments at the utility's disposal, thereby tapping into a realistic "two-part tariff" opportunity (Martin et al. 1984). Utilities that meter each customer's water usage will normally charge customers both a volumetric fee and a flat fee. The most simple scheme is for customer i's bill to be calculated as $pw_i^f + M$. As p* defined by either

eqs. (3) or (5) can promote allocative efficiency, the flat fee (M) can be tasked to balance the utility's budget. Because water utilities typically experience average and marginal accounting costs that decline with rising production across a large quantity range (implying that marginal costs are less than average costs), conventional circumstances are that summed volumetric payments, $p * \sum_i w_i^f$ from a marginal cost pricing rule will be inadequate to cover costs. Using the second pricing instrument, the shortfall can be collected as a payment of M from every client without disturbing water efficiency. Moreover, as nonaccounting scarcity values – the λ's of eqs. (3) and (5) – come to affect and raise marginal-cost prices, decreases in M can preserve a utility's balanced budget. In this way, customers are treated as shareholders in the system, and the efficiency rewards of marginal-cost pricing are partially distributed to customers as M-declines (Griffin and Mjelde 2011). A consequence of this strategy is that equity can be improved, perhaps even beyond that offered by increasing block rates.

Moving people to the water in the long run

Water utility traditions are to welcome new members to the community. Local policies tend to be very supportive of growth, to the point that political preferences are often to subsidize growth rather than to pursue "user pays" or economic efficiency. Growth does, however, exact important costs in the water sphere. In regions where water scarcity impacts are enlarged by growth, past depletion, habitat injuries and other costs, it becomes appropriate to attack conventions by casting off subsidies.

In addition to the recurring flat fee (M above) and the volumetric water price (p) commonly applied by water service providers, urban utilities usually charge a variously labeled "new connection charge." This is a one-time fee for connecting a newly built home or business to the distribution network. Residential developers routinely pay this fee and embed it in the cost of

new homes. Progressive utilities in water-scarce areas have begun to incorporate the cost of expanding their water supply, not just the cost of new infrastructure, in the fee (Hanak 2008). Some utilities have requested that developers acquire water rights sufficient to serve new connections, and then transfer those rights to the utility as a condition for service. Yet, in most of the U.S. utilities are still applying very low connection charges – essentially sufficient to cover the costs of installing the necessary water meter.

For utilities operating in the presence of water and infrastructural scarcities, low connection charges infer that the costs of growth are born by all customers, not merely the new connections. In addition to the equity issue involved in such a "nonuser pays" convention, the costs of growth are being mis-signaled and subsidized, with negative implications for overpopulating water scarce regions. It is interesting that pro-growth policy is combined with the rhetoric of "need" to argue first that economic growth is necessary for community welfare improvements and second that growing communities need assistance to address their water problems. Alternative policy thrusts can contribute here.

The actual costs imposed by a new connection on the utility can be assessed by comparing costs without the connection to costs with it. This cannot be accomplished by examining a single-period effect. Even a nongrowing, constant-demand utility faces changes in its supply costs. For example, the pace of ground water depletions and infrastructural depreciation for the nongrowing utility implies that there is a schedule of future costs to be paid for construction programs and supply enhancements. Growth quickens that schedule and may add other costs. Whereas the impact of truly marginal, single-customer growth may be difficult to discern, we can lump together all expected new connections for the coming year and consider their "average marginal" effect. Denote the number of new connections as Δn. Let $\mathbf{C}^{wo} = \left\{C_1^{wo}, C_2^{wo}, ..., C_T^{wo}\right\}$ be

the without-growth schedule of relevant costs over the next T years. Relevant costs may exclude variable costs ordinarily offset by water price p while including capital costs and all water acquisition costs. Hence, this vector is not a compilation of all utility costs, just those affected or shifted by growth. Higher values of T enlarge the scope of the calculation and reduce growth subsidies.[5] Let $\mathbf{C}^{wh} = \left\{C_1^{wh}, C_2^{wh}, ..., C_T^{wh}\right\}$ be with-growth costs. Because of the forward shifting of costs, it is not necessarily true that every element of $\mathbf{C}^{wh}$ is higher than the corresponding element of $\mathbf{C}^{wo}$. Under these circumstances, an appropriate new connection fee, F, is computable as

$$F = \frac{1}{\Delta n} \sum_{t=1}^{T} \frac{C_t^{wh} - C_t^{wo}}{(1+d)^t} \tag{9}$$

using a discount rate of d.[6] Such a new connection fee compensates the utility for water supply expansions caused by growth. It also raises the costs of new homes (and businesses). The resulting shift in the supply function for new residences interacts with housing demand to capitalize a portion of this cost into the value of existing homes. Hence, a policy change from traditional new connection charges, which are limited to meter installation costs, to new connection charges defined by eq. (9) increases the value of existing residences. But the real intent of computing F appropriately is to induce people to consider water scarcity in their location decisions. F-differentials across regions are not the only factor in such decisions, but contemporary policy tends to make them a nonfactor, unfortunately. Optimal F policy also preserves welfare for existing customers by (i) maintaining their water supply, (ii) exempting them from paying for water supply improvements wanted by new customers, and (iii) capitalizing the value of their utility connection into the value of their residence or business.

Actual practice and utility conventions

The theory above, as lodged by economists, has had limited impact. This becomes more burdensome as scarcity advances, in that society is incompletely capturing water's available rewards just as the stakes are getting high. To better understand the slow maturation of policy, it is helpful to consider the traditional methods of water rate making. Professional pricing principles are anchored in accounting-based "Hopkinson" conventions (1892) that have arisen in support of balanced books; not only is this fiscally prudent and normatively palatable (the "users pay" principle), but utilities' bond ratings are sensitive to their ability to generate sufficient revenue (Hewitt 2000), so revenue matters in multiple ways. Yet, from an economic, let's-signal-consumers-well perspective, there are strong obstacles established by Hopkinson procedures. [John Hopkinson was a brilliant engineer.]

The accounting principles of conventional U.S. rate making focus on the revenue requirements of an average fiscal year representative of the forthcoming year. These requirements include operation and maintenance expenses and debt service (principle and interest payments), capital costs, and/or depreciation costs.[7] For-profit utilities are also entitled to a fair rate of return, so this too is part of revenue requirements. As commonly conducted, the process of converting the various fixed and variable costs into rates is essentially a weighted average cost calculation. The confounding issue is the joint nature of many of the important costs (e.g. pumps, land, tanks and pipes). These joint costs must be allocated in some fashion across revenue instruments, and this is inherently subjective. By definition, joint costs are collaboratively caused, so "There is no unique correct method" of allocating joint costs across responsible parties (James and Lee 1971, p. 538).

Standard water utility practice has been to base rates on the different peak-loading characteristics of different sectors. Because utility systems are scaled to meet peaks, much of the system may be underutilized much of the time, and the residential sector is argued to be responsible for a higher ratio of peak use to base use. Hence, traditional practice is for the residential sector to bear a larger portion of joint costs. Whereas the fixed nature of such costs makes them candidates for recovery via the recurring M charge noted above, they are generally recovered using volumetric charges (which is commendable from a long-run signaling perspective). Recommended practice is for M to emphasize "customer charges" associated with meter reading and billing tasks, yet it is also regarded as acceptable for M to include a portion of joint costs (American Water Works Association 2012, p. 138). Calculation of each rate element is completed by dividing assigned costs by the number of units served. So the volumetric water rate for residential users is their collective, expected, caused costs divided by their expected water use, and the flat rate is total caused customer costs divided by the number of served units (e.g. number of connections times 12 months). Clearly, these procedures are satisfied with average cost pricing. Economic disappointments with this procedure have long been asserted:

> "From an economic point of view, this principle [cost distribution to customer classes] is defective for several reasons. First, fully distributed costs must imply that prices are based upon average cost rather than marginal cost. Second, where several classes of service exist, the allocation of all costs will certainly mean that costs that are shared between all classes of service (joint costs) will be improperly assigned. Third, the principle requires the allocation of historical and sunk costs that are not relevant for current decisions" (Milliman 1964, p. 129).

With less sensitivity, but with lucid implications for the merits of increasing block rates, Lewis (1941) assails the Hopkinson-based traditions of partitioning costs on the basis of loads:

> "The maximum rate at which the individual consumer takes is irrelevant; what matters is how much he is taking at the time of the station's peak. This point is now generally accepted among the better writers on the subject, but the persons actually engaged in framing tariffs (they are usually engineers) do not seem to have mastered it yet" (p. 252).

Whereas the usual design of rates is led by these "cost causation" principles, the accounting theory becomes relatively arbitrary in practice due to the heavy presence of joint costs. The inherent subjectivity of assigning joint costs combines with the large number of available revenue-generating instruments to make highly disparate rate outcomes feasible. Not only are there separate volumetric and flat rates to decide as in the economic theory above, but by using cost causation theory, different rates for different sectors can be rationalized as can differing rates for inside-city-limits and outside-city-limits customers. Rates can (and often should) be seasonal, adding further dimensionality to the joint cost partitioning exercise. Block rates of both increasing and decreasing varieties add further dimensionality since the number of blocks, their ranges and their rates require selection. These block elements are particularly arbitrary in practice. Overall, the large number of available pricing instruments combine with the intrinsic nature of joint costs to infer that cost causation underidentifies the overall rate package, especially when block rates are selected. A consequence is that rate designers have considerable latitudes. Interestingly, part of the process of revising rates usually involves a survey of rates being applied in the region or by like-sized water utilities. The evident purpose of this step is to improve the defensibility of the selection, with the result that actual rates tend to be everywhere errant (from an efficiency perspective).

Hence, whereas the economic goal of optimizing net benefits produces the ideal of a uniform rate based on short-run marginal costs (Boland 1987), complemented by a second type of rate (M) to balance revenues, the accounting-led traditions of utilities is to utilize average cost

prices that balance the budget. Moreover, economics asks that intrinsic water value and infrastructural opportunity costs be part of rates, but shadow values are invisible to accounting cost-causation directives.

Cost-causation also points in a different direction regarding peak management. If sector A is more responsible than is sector B for peak month demand, say in July, cost causation theory is applied to argue that sector A's water rate should be higher (all year). In the same situation but in the spirit of Lewis's quote above, economic theory argues that all demand in July, regardless of sector, contributes to the utility's monthly peak loading problem and that the appropriate strategy is higher July water rates for all sectors so as to enlist water-conserving behavior from all parties in accordance with their marginal benefits (Hanke and Davis 1971).

It is difficult to inventory all of the discrepancies between purist, net-benefit-maximizing water rates and those rates ordinarily applied by utilities. Departures from economic recommendations include average-cost pricing, absence of time-of-year prices, block pricing, customized (agent-specific) pricing, omissions of particular opportunity costs, and growth subsidies.

Imperfectly informed consumers

There are important, and unmet, challenges that arise for the economic theory of water ratemaking. The transaction costs of monitoring consumption, both by clients and utilities, is at the heart of the most severe problems. As a compounding problem in many countries, water bills are low portions of household budgets and do not inspire much attention. Utilities read meters no more frequently than monthly. This approach is practical in light of meter reading and billing costs, but it does limit consumer information. Also, although the utility industry is making improvements, bill formats are often uninformative in the sense that water units and billing

formula are not readily accessible or understandable to many customers (Gaudin 2006). Water bills are often part of a more comprehensive bill that may include wastewater service, trash collection, and/or electricity consumption, thereby adding complexity for consumers. Block rates further complicate things. A previously discussed difficulty for economic recommendations is that monthly billing prevents utilities from using rates to signal particular peaking problems, such as the opportunity costs of consuming water on the peak day of the year.

Consuming tap water is not accompanied with the same clarity that accompanies the purchase of typical commodities. When a consumer loads a loaf of bread into the grocery cart, the weight of the bread and its price is well labeled. When a consumer turns on a faucet or operates a water-using appliance, the quantity of use is difficult to know. [In some countries, consumers regularly read and self-report their consumption data utilizing easily accessed meters, thereby improving these conditions.] Monthly bills provide the user with delayed information about total consumption, but not individual uses. There is a negative feedback injected here; imperfect information about quantities reduces the rational consumer's motivation to fully understand rates. Although these features of water metering and billing are largely sensible in that they are reasonable responses to the transaction costs of the situation, they limit the applicability and penetration of marginalist pricing ideals.

Full realization of efficiency via marginal cost prices complemented by budget-balancing flat fees is dependent on the ability of customers to understand billing practices. This is an evolving condition which should improve as (a) consumers pay greater attention to water costs as water's share of production costs or household expenditures rise, (b) new billing formats continue to advance the transparency of rates, and (c) new instrumentation increases the ability of consumers to comprehend the billing implications of their water-using activities. Here lie

important research opportunities for integrating the opportunity cost theories of information and water, so as to better understand the tradeoffs and to better prescribe rate policy.

Ratemaking for irrigation organizations

An interesting set of questions pertain to the depth to which efficient water rate-making prescriptions extend to the various water authorities that deliver water to agricultural irrigators. All of the issues discussed above are applicable, so the full slate of marginal pricing ideals would seem to apply. Irrigation organizations (IOs) are natural monopolies too, and they have at their disposal various means of collecting revenue through bills upon their irrigator clients, as well as through more broadly based taxes on the direct and indirect beneficiaries of irrigation. In particular, IOs may exercise area-based charges (parallel to M above) where producers pay on the basis of the amount of land they are irrigating, and IOs can apply volumetric prices where water deliveries are metered. It is common for both instruments to be conjunctively applied in the western U.S., and rates there may also employ blocks. As in urban settings, block pricing is strictly inefficient in IOs (Bar-Shira, Finkelshtain, and Simhon 2006). In many parts of the world, the costs of irrigation are subsidized at the national level, so it is often the case that full-cost pricing has been politically rejected.

Molle and Berkoff argue that the principles of efficient pricing have weak applicability for the IOs of developing countries (2007, pp. 31-2). Part of their argument is that volumetric pricing is less practical due to metering requirements for large numbers of small fields and the opportunities for tampering. Metering is certainly difficult for low-pressure flows as typify canal delivery of surface water. Other arguments include the role of subsidized water in redistributional policy intended to assist the rural poor and the often promoted, though economically odd idea (from an invisible hand perspective) that the entire population benefits

from food production, thereby justifying subsidy. Still more arguments against the notion of marginal cost or full cost pricing have been observed and even compiled (Johansson et al. 2002; Tsur et al. 2004). Yet, much of this literature emphasizes the pricing of water in a developing nation context. Accounting costs are typically underscored for these scenarios, and intrasectoral water allocation is a lesser concern. Based on this literature, one may conclude that the merits of efficient pricing ideals are dulled in the case of IOs.

Yet, the troublesome matter of intrasectoral allocation is sharpening arguments for stronger incentives. In some regions of scarcity (e.g. southwestern U.S.), there are elevating differentials between the marginal value of water used in irrigation versus water's marginal value in other sectors. These differentials recommend reallocation from agriculture to urban and environmental applications. Where this occurs and especially where irrigation is the dominant consumer of water, there is a strong need for policies that allow irrigators, not merely IOs, to directly witness the external value of water in some manner (Griffin 2012). If water rates are not acceptable instruments for addressing these problems, then other strategies will have to be entertained.

Conclusions

Recognizing that all resources are scarce to some degree, it cannot be scarcity that is socially bad, it can only be our policy responses to it. Self-correction mechanisms, as may exist for marketable commodities, are extremely weak in the case of processed water, thereby providing considerable longevity for an unfortunate water rate-making doctrine. Rate styles have been evolving, but modern incarnations of conservation-oriented water rates, such as increasing block rates, do not satisfy economic efficiency objectives. Perhaps more efficient rate regimes will disperse across the industry once a forthcoming wave of innovation is initiated (Teodoro 2010).

Yet, pressures for these improvements will likely require participation by professionals having exposure to a technical literature.

The water rate-making advice produced by economics is decidedly forward-looking; its goals are to establish signals that motivate socially appropriate behaviors by private agents. The targeted behaviors are several: how much to consume, when, where (location choices), what conservation activities to deploy and to what degree. A crucial dimension of these objectives is a desire to reshape public sentiment and political pressure pertaining to water supply enhancement projects, many of which are costly relative to their potential benefits. Viewed from the economist's perspective, constant hue and cry about scarcity and crisis are clear evidence of deficient rate making. Likewise, political clamor for expensive structural solutions is potentially indicative of unimproved rates.

There are intricacies encountered in designing better rates. Better rates infer better welfare, by definition, though this ideal is slow to be grasped in debate. Whereas economically commended water rate-making principles arise from a more general public utility literature that has deeply studied many things, including the objectives of rates, marginal cost pricing, Ramsey pricing, two-part tariffs, and optimal rates in the presence of uncertainty, there is more than accounting costs to be addressed in the case of water. There are several conceivable opportunity costs that may be variably active. Some of these arise from the scarcity of raw water inputs. Others pertain to the scarcity of infrastructure used to process or convey water. These opportunity costs directly influence volumetric water rates, and they indirectly affect nonvolumetric rates customarily applied by water service providers. These nonvolumetric components have important, supporting roles to play.

Cited References

American Water Works Association (2012), *Principles of Water Rates, Fees, and Charges*, Denver: American Water Works Association.

Bar-Shira, Z., I. Finkelshtain and A. Simhon (2006), 'Block-rate versus uniform water pricing in agriculture: An empirical analysis', *American Journal of Agricultural Economics,* **88** (4), 986-999.

Baumol, W.J. and D.F. Bradford (1970), 'Optimal departures from marginal cost pricing', *American Economic Review,* **60** (3), 265-283.

Boland, J. J. (1987), 'Marginal Cost Pricing: Is Water Different?', in D.D. Baumann and Y.Y. Haimes (eds), *The Role of Social and Behavioral Sciences in Water Resources Planning and Management,* Santa Barbara, CA: American Society of Civil Engineers, pp. 126-137.

Chakravorty, U. and C. Umetsu (2003), 'Basinwide water management: A spatial model', *Journal of Environmental Economics and Management,* **45** (1), 1-23.

Crew, M.A., C.S. Fernando and P.R. Kleindorfer (1995), 'The theory of peak-load pricing: A survey', *Journal of Regulatory Economics,* **8**, 215-248.

Dandy, G.C., E.A. McBean and B.G. Hutchinson (1985), 'Pricing and expansion of a water supply system', *Journal of Water Resources Planning and Management,* ***111*** (1), 24-42.

Gaudin, S. (2006), 'Effect of price information on residential water demand', *Applied Economics,* **38**, 383-393.

Griffin, R.C. (2001), 'Effective water pricing', *Journal of the American Water Resources Association,* **37** (5), 1335-47.

Griffin, R.C. (2012), 'Engaging irrigation organizations in water reallocation', *Natural Resources Journal,* **52** (2), 277-313.

Griffin, R.C. and S.-H. Hsu (1993), 'The potential for water market efficiency when instream flows have value', *American Journal of Agricultural Economics,* **75** (2), 292-303.

Griffin, R.C. and J.W. Mjelde (2011), 'Distributing water's bounty', *Ecological Economics,* **72**, 116-128.

Hanak, E. (2008), 'Is water policy limiting residential growth? Evidence from California', *Land Economics,* **84** (1), 31-50.

Hanke, S.H. and Davis, R.K. (1971), 'Demand management through responsive pricing', *Journal of the American Water Works Association,* **63**, 555-560.

Hewitt, J.A. (2000), 'An Investigation into the Reasons Why Water Utilities Choose Particular Residential Rate Structures', in A. Dinar (ed), *The Political Economy of Water Pricing Reforms*, New York: Oxford University Press.

Hirshleifer, J. (1958), 'Peak loads and efficient pricing: Comment', *Quarterly Journal of Economics,* **72**, 451-62.

Hopkinson, J. (1901), 'On the Cost of Electric Supply (1892 Presidential Address to the Junior Engineering Society)', in B. Hopkinson (ed), *Original Papers by the Late John Hopkinson*, vol. 1, Technical Papers. Cambridge: Cambridge University Press.

Hotelling, H. (1938), 'The general welfare in relation to problems of taxation and of railway and utility rates', *Econometrica,* **6** (3), 242-269.

James, L.D. and R.R. Lee (1971), *Economics of Water Resources Planning*, New York: McGraw-Hill Book Company.

Johansson, R.C., Y. Tsur, T.L. Roe, R. Doukkali and A. Dinar (2002), 'Pricing irrigation water: A review of theory and practice', *Water Policy,* **4**, 173-199.

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

Lewis, W.A. (1941), 'The two-part tariff', *Economica,* **8** (31), 249-70.

Martin, W.E., H.M. Ingram, N.K. Laney and A.H. Griffin (1984), *Saving Water in a Desert City*, Washington, D.C.: Resources for the Future.

Milliman, J.W. (1964), 'New price policies for municipal water service', *Journal of the American Water Works Association,* **56** (2), 125-131.

Molle, F. and J. Berkoff (2007), 'Water Pricing in Irrigation: Mapping the Debate in Light of the Experience', in F. Molle and J. Berkoff (eds), *Irrigation Water Pricing: The Gap between Theory and Practice*, Oxfordshire, UK: CAB International.

Pitafi, B.A. and J.A. Roumasset (2009), 'Pareto-improving water management over space and time: The Honolulu case', *American Journal of Agricultural Economics,* **91** (1), 138-153.

Riordan, C. (1971), 'Multistage marginal cost model of investment-pricing decisions: Application to urban water supply treatment facilities', *Water Resources Research,* **7** (3), 463-478.

Ruggles, N. (1949), 'The welfare basis of the marginal cost pricing principle', *Review of Economic Studies,* **17**, 29-46.

Steiner, P.O. (1957), 'Peak loads and efficient pricing', *Quarterly Journal of Economics,* **71** (4), 585-610.

Teodoro, M.P. (2010), 'Contingent professionalism: bureaucratic mobility and the adoption of water conservation rates', *Journal of Public Administration Research & Theory,* **20** (2), 437-459.

Tsur, Y. (2009), 'On the economics of water allocation and pricing', *Annual Review of Resource Economics,* **1**, 513-535.

Tsur, Y., T. Roe, R. Doukkali and A. Dinar (2004), *Pricing Irrigation Water: Principles and Cases from Developing Countries*, Washington, DC: Resources for the Future.

Turvey, R. (1976), 'Analyzing the marginal cost of water supply', *Land Economics,* **52** (2), 158-68.

Unnerstall, H. (2007), 'The principle of full cost recovery in the EU-water framework directive – genesis and content', *Journal of Environmental Law,* **19** (1), 29-42.

Visscher, M.L. (1973), 'Welfare-maximizing price and output with stochastic demand: Comment', *American Economic Review,* **63** (1), 224-229.

Endnotes

[1] For a differently oriented and more formal economic network model of these issues, see Tsur (2009).

[2] It is intriguing that economists might champion nonuniform rates on purely allocative grounds. When marginal costs rise with quantity supplied, as is common for all goods, we do not recommend prices that discriminate on the basis of consumption levels. When aggregate consumption causes marginal costs to be high, everyone's consumption is equally relevant at this high margin.

[3] The truth of this claim is not fully evident from the first-order conditions set forth here. If conveyance losses and/or marginal transfer costs are sufficiently large, it may be optimal to undertake no reallocation even when $\lambda_m > \lambda_x$. Intuitively, however, if r and e are sufficiently small and if $\lambda_m - \lambda_x$ is sufficiently large, it will be optimal to reallocate water, and this reallocation will close the gap between λ_m and λ_x to that indicated by eq. (6).

[4] The same can be said of supply contractions for sectors or regions experiencing decline or falling in relation to growing sectors/regions, but expansion is the usual condition.

[5] If it is reasonable to presume that the ΔC pattern exhibited in the numerator of eq. (9) is repeated beyond an easily projected period, say for T=5, then it becomes possible to approximate F for a larger T.

[6] A reasonable candidate for d is the utility's cost of borrowing via the issue of bonds.

[7] There is some double counting in this listing as utilities posting depreciation to their revenue requirements would not ordinarily also post debt service and capital costs.