The role of prices in the management of computing resources

Permalink
https://escholarship.org/uc/item/76n7n8f4

Author
Feldman, Julian

Publication Date
1976

Peer reviewed
The Role of Prices in the Management of Computing Resources

Julian Feldman

1 December 1976

Technical Report #96

This is a preliminary draft of material to be included in a forthcoming book on management of computing resources. Please do not reproduce, quote, or mention without the permission of the author.
The Role of Prices

in the Management of Computing Resources

Many computing centers in large organizations charge the computer users in their organizations for computing services. Why do organizations go thru the trouble of accounting for computer usage by internal users? and of shifting funds from one internal account to another?

Organizations have computer centers charge internal users for two reasons: (1) The organization does not know how much money to spend on computing. So it lets the subunits decide how much computing they want to buy. The charging mechanism is used to decide how to allocate organizational resources between computing and other activities. (2) The organization uses prices to allocate computing resources, both physical and temporal, between contending users.

The first section of this paper contains a discussion of these two reasons for pricing computing services. The second section contains a series of criticisms of the practice of pricing computing services and responses to these criticisms.
Question 1. How much money should an organization spend for computing services?

Every introductory economics text provides an answer to this question: A decision making unit, e.g., consumer, organization, should allocate resources between alternatives so the utility received from the last dollar spent on each alternative is the same. If the marginal utilities are not the same, the decision making unit can increase total utility by shifting resources away from commodities which have lower marginal utility to commodities which have higher marginal utilities.

The organization can be viewed as making choices between alternatives like production equipment, raw material, research and development, computing, etc. To maximize its utility, the organization should use the basic strategy in making decisions about how much money to spend for computing, i.e., it should increase its expenditures on computing as long as the return from the computing expenditures is greater than the return from alternative expenditures.

But the organization often does not know what the marginal return is from spending a dollar on computing. This is not surprising because the organization is in the same predicament on many other decisions (cf. the extensive literature on how organizations set prices). So the organization looks for other guidelines, heuristics, or rules of thumb to determine how much to spend on computing.

One place to look for guidance is to other organizations. How much are similar organizations spending on computing? as a percentage of sales? per employee? per student? If an organization is spending a lot more or a lot less than its peers, the situation is worth looking at.
In some industries, this information is exchanged or otherwise disseminated. In some industries, particularly in the public sector, some effort is made to establish guidelines or standards for expenditures. For example, the per capita expenditures of cities on law enforcement, fire protection, library services, etc. are widely disseminated and tend to be used to compute de facto standards. There was an effort to establish a standard for expenditures for instructional uses of computers in higher education. But that standard has not been widely accepted. (See C. Mosmann, Evaluating Instructional Computing, 1976, for a recent discussion of the problems involved in setting standards for instructional uses of computers in higher education.)

Another way of deciding how much to allocate for computing is to look at individual applications, e.g., payroll, accounts receivable, registration of students, teaching, and compare the cost of accomplishing these tasks with alternative systems. If the cost of a system which utilizes computers is less than the cost of a system which doesn't, then the use of computing resources is justified. This is an example of the dictum which enjoins the decision maker to select the cheapest means for accomplishing a goal. The decision maker using this method needs information on the costs and benefits of using computing resources.

Many economists and management scientists advocate decentralizing decisions on computing expenditures. Let each organizational unit decide how much it wants to spend for computing in the same fashion each unit makes other decisions. The organization only has to allocate resources between major organizational programs, e.g., between products in a multi-product firm, between campuses in a multi-campus university. Then each program
can allocate resources between its alternatives.

Each program or organization unit can buy computing services from outside organizations or provide their own services. If the organization decides to create a central computing center for any of several reasons, the organization can still decentralize the decision about how much money to spend on computing by establishing the center as a cost center which charges internal users for computing services according to some price schedule. The use of prices then permits the organization to decentralize the decision about how much to spend and to maintain a central computing center if it so desires.

Summary. If the organization does not know how much money it should spend for computing services, it can turn the decision over to its major program units. In a large, multi-product firm, each product line could decide how much it wants to spend for computing. In a university, each department or school could decide how much it wants to spend.

Because of the advantages of sharing people, software, and hardware, the organization might create a central computing center. If this center is administered as a cost center charging internal users for computing services in accordance with a price schedule, the organization can have both a central center and decentralization of the decision on the amount of computing activity.
Question 2. How much of the available computing services should be allocated to each unit of the organization?

Even if an organization is prepared to decide how much it should spend on computing without decentralizing the decision and using prices, the organization may well decide to use prices to allocate the services provided by the central computing center to the organizational units.

Many organizations with computing installations have adopted pricing as an allocation mechanism only after they have gone thru one or more of the following allocation procedures:

First-come-first-serve. When a computing system is first installed, there is usually plenty of capacity. Application programs have not yet been completed, and most installations are oversized to allow for growth. So a first-come-first-served allocation scheme is usually quite adequate. (This was especially true prior to the widespread use of multi-user operating systems.) When the usage increases, the arguments begin, and alternatives to FCFS are considered.

Allocation by administrative rules. The first step in resolving the conflicts resulting from increased usage under a FCFS allocation policy is often to establish a set of administrative rules to modify the FCFS policy. A common set of modifications might consist of giving high priority to short jobs which do not require extensive use of peripherals (comparable to the express check out line in a supermarket—six items or less and cash) and to payroll and/or other critical organizational activities.
Average cost pricing. Any allocation system which does not require
the user or the user's patron to pay for computing will eventually lead
to uneconomical uses of computing. If a manager has to decide whether
to use a computing based system or another method to solve a problem
and the services of the computing system are free, the manager's decision
will be biased in favor of the computing system. A frequent first
solution to this problem is the installation of average cost pricing.
The average cost of computing is calculated by dividing the total cost
of the computing operation by the expected number of units of usage
(e.g., cpu hours, wall clock hours).

The advantages of average cost pricing are two: (1) The organization
recovers all of its computer costs, and (2) the user has some number to
use in making a decision about the cost of computing services.

The disadvantages of average cost pricing are that the price
charged to the user varies inversely with usage. Because there is
a large fixed cost component in computing, average cost will be high
when usage is low; and average cost will be low when usage is high.
But a high average cost (price) will discourage usage when usage is
low and could (should) be increased. And a low average cost (price) will
encourage usage when usage is high and it should be leveled off.

The problems of average cost pricing can be exacerbated by using
a short accounting period, e.g., a month, and ameliorated by using a
long accounting period, e.g., three-four years. A short period could
lead to frequent price changes and resulting instability.
Flexible pricing. While average cost pricing can be used as an allocation tool, it does not really provide the flexibility required to allocate computing resources. Average cost pricing is primarily concerned with short-term cost recovery. While a pricing system can be used to recover costs, it should be concerned with long-term cost recovery. In the short term, the pricing system should be concerned with maximizing the usage of the system. Assuming that more computing is preferred to less and that much of the computing costs are fixed, the organization should adopt a pricing schedule which will make maximum use of the computing system subject to the constraint of providing the desired quality of service.

One of the hallmarks of a flexible pricing schedule is that short-term prices may have little relation to costs. Short-term prices are market clearing devices. If the demand is small, prices may be lowered below average cost to clear the market. If the demand is large, prices may be raised above costs to reduce the quantity demanded (i.e., encourage users to seek alternative products or services, e.g., use tape storage when disc storage becomes overcrowded).

A second hallmark of flexible pricing is that response time or turnaround time is considered a commodity just like CPU time and can be (should be) priced. So if the demand for first-shift services is greater than the demand for off-shift services, first-shift users are charged more. If users want high priority service, they pay for it. The charges for time as well as physical services can be rationalized in a welfare sense. Those people who want to work days and get good service should compensate those whom they dispossess (assuming all are equal to begin with). Or the first-shift services can be viewed as a separate commodity from off-shift services. If the demand for a commodity is greater, than the user will have to pay more.
Pricing with fiat money. If the organization is prepared to decide how much money should be spent on computing without decentralization and the use of prices, the organization can still use flexible prices to allocate computing resources. The organization has only to allocate a budget of fiat money (tokens, yet, etc.) to each organizational unit in the same way that the organization makes any budget allocation. The organization sets prices, and the users use fiat money to buy services.

Summary. A flexible pricing scheme, which provides users with a variety of services at prices set to balance demand and supply for those services, appears to be the best of the methods which have been proposed for allocating computing resources. If the prices are set in real dollars, the price system can also be used to determine the amount of money which should be spent on computing. If the prices are set in fiat money, the price system is still the most flexible method for allocating computing services.
Criticisms of the use of prices to make internal allocations of computing services

**Criticism 1.** Computing is not a class of service which should be allocated on the basis of price. Most societies and most organizations have two classes of goods and services. One class is allocated on the basis of price. The consuming unit gets as much of these commodities as it is willing to buy at the prevailing price. The other class is allocated to consumers on the basis of needs or requirements independent of their individual willingness or ability to pay. In this second class are commodities like free libraries, roads, public schools, etc. Computing in an organization should be treated like this second class of commodities. The organization should decide how much computing it needs, and then the services should be allocated without charge.

**Response.** While it is true that many societies and organizations provide the services mentioned above to their members for little or no charge, there are many who argue that the method of distribution is inappropriate or in need of modification. Experts in higher education argue that support should not be provided to schools but provided to students so that private and "public" schools can then compete for students. Recent scandals in medical distribution programs suggest that efforts to improve health care for the indigent and the aged may have succeeded largely in increasing the income of medical practitioners and hospitals. Neo-classical economists argue that rent controls only succeed in depressing housing starts and maintenance, i.e., prevent long-run solutions to the problems of inadequate housing.
If computing is made a "free good", i.e., consumers are not charged, the criteria for deciding between alternatives will disappear. The organization won't know how much or how little to spend on computing. The user will do more computing than he/she "should" because of the artificially low price. Some other mechanism(s) (perhaps fiat money) will have to be introduced to allocate the available computing services between contending users. At zero price, demand will exceed supply unless supply is relatively very large.

Criticism 2. Because computing services come in large chunks for which relatively long-term commitments have to be made, the organization has to make a commitment to pay the fixed cost of the computing system. Because the fixed cost is the largest part of the cost, the user should only be charged the marginal cost which is either zero or very small.

Response. In the short run, the organization should provide services for marginal cost, i.e., services should be provided as long as the value (return) to the organization exceeds the marginal cost. So the computing center should set price at marginal cost and provide services up to capacity.

If capacity is not reached at marginal cost pricing, then capacity is clearly too great and should be reduced with a corresponding increase in prices. Once capacity is reached, prices will have to go up to ration capacity. If prices exceed average cost (long run marginal cost), then capacity should be increased. In the long run, prices should approximate average cost.
The pricing model does not recommend that capacity be withheld from users if the value to the users, in their organizational roles, is greater than the cost of providing the services.

However, strong arguments have been raised about the advantages of price stability. Relatively few users of computing services want to operate in an environment in which prices change direction like a roller coaster. For example, a new system comes in with excess capacity, so prices are very low. Usage builds up, prices rise, capacity is exceeded. Another new system is brought in ... . Most computer usage is not yet so portable that one is willing to move to the lowest bidder on a daily, weekly, or even monthly basis.

The price oscillations can be damped without grossly reducing the value of the computing system to the organization by slightly reducing prices when demand exceeds capacity and slightly increasing prices as demand increases. Excess capacity can be used to ease conversion costs, to help reduce development costs for new applications, i.e., for short-run, oneshot deals. Because it is practically difficult to conduct sales of computing services, short-run reductions in price will reduce income but won't increase usage very much because of the inelasticity of the short-run demand curve.

Basing prices on long-run average costs has the advantage of stability. It is the price that the user, the center, and the organization can use in comparing alternatives.

When average cost pricing leads to large amounts of unused capacity, the organization must consider reducing capacity or stimulating demand. One way of stimulating demand is a lowering of prices in the
short run. Because price instability can cause other problems, the
organization may want to consider alternative ways of stimulating demand.

**Criticism 3.** Charging internal users for computing services is
fine in theory, but in practice it doesn't work. Pricing gets you into
a situation where you have unused capacity and unmet needs.

**Response.** What appears to be unused capacity may only be the
capacity necessary to give good service. There is considerable empirical
and theoretical evidence that response times will increase sharply at
certain levels of utilization. So some apparently unused capacity
is required to provide good service.

The organization can adopt a pricing schedule which would produce
unused capacity other than that required to provide the desired level
of service. If this happens, then the price schedule may be wrong.
The organization is not getting all it can from the computing center
if the value which can be obtained from the unused capacity is less
than the cost of providing these services. But poor selection of prices
does not mean that the use of prices to allocate computing services is
wrong.

The issue of "unmet needs" is a difficult one. Nobody gets all
of the capacity they want in the same sense that nobody gets all of
anything they want. But if the user is free to allocate resources
between desired goods and services, then the user will select a bundle
of goods and services (including computing services) which will
maximize the user's utility. If the market is open and there are no
impediments, then the user will move towards a bundle of goods and
services which will be optimal. If the user had more money, he/she
might buy more computing services. But the allocation of resources
between users is another problem.
Criticism 4. A major problem with using prices for allocating computing services is that the computing center can fail to earn sufficient income to cover its costs.

Response. It is quite true that the computer center's income might not cover its costs. Under the model presented here, the computing operation should be evaluated on the basis of whether the value of its services exceeds the costs of providing these services. And the value will in general exceed the revenue. In a market where everybody pays the same price for the same service, the price paid is the least value. Everybody pays the price paid by the person who values the services least. Therefore, some value the services more than what they pay for them.

At the beginning of the life cycle of a piece of equipment, the center will generally have excess capacity and should provide services for less than average cost. The center could run a deficit in such a situation.

In the long run, the center should charge internal users average cost; and should therefore recover its costs. If not, prices may be set incorrectly or the center has too much capacity. But it is possible to have the center run at a financial loss and still contribute more value to the organization than the cost of operating the center.
Criticism 5. Computing is an infant industry. The organization should encourage its units to utilize computing because the benefits will eventually exceed the costs. However, if a strict cost-benefit accounting is required too early, units will be discouraged from using computing.

Response. There is some validity in this criticism. However, the organization must be careful to assess the cost and benefits of computing in some way if it does not use the price mechanism. And the organization may well find it worthwhile to allocate the furnished services with fiat money and a price system so that the services will be well used.
Summary. A computing center within an organization should function as a cost/profit center (i.e., charging users for services) to help the organization decide how much of its resources should be allocated to computing by decentralizing the decision. But even if the organization is prepared to make the decision about the level of computing activity centrally, the computing center should use a price system as the basis of allocating services to users.

The prices charged by the computing center should be those which maximize the net value of the organization. Whether the cash profit accumulates to the user or producer of computing services matters little to the organization as long as the prices selected maximize the net value to the organization.

For the center charged with cost recovery, I recommend that the prices be set to average cost at capacity. If revenue is less than cost, there can be several possible causes:

(1) Capacity is too great.

(2) The center and/or organization is not providing the non-computer services required to make effective use of the capacity.

(3) This is a short run phenomenon happening at the beginning of the life cycle of the machine or at some other major change, e.g., the introduction of a major new piece of equipment.

Because the price list of a multi-user system consists of a whole set of prices for the different services provided, a price must be assigned to each class of services. The basic rule should be
applied to each class of service, i.e., the base price for each service should be the average cost at capacity recognizing the problem of calculating average cost includes the distribution of overhead.

The appropriate shift differentials and other charges for temporal (as opposed to physical) services can be calculated if the several demand curves and cost curves are known. In practice they may not be known and a reasonable strategy is to set shift differentials so that each shift is used to capacity.

For the center charged only with making the best allocation, prices should be set to maximize usage subject to constraints on providing certain levels of service.