

REFLECTIONS ON EFFICIENCY AND GOVERNMENT REGULATION

Arnold C. Harberger

University of California, Los Angeles

May 1999

Let me begin by noting an impressive change in the way international organizations such as the World Bank, the IDB and the IMF have approached the question of privatization.

Not too long ago they practically reserved their financing for public sector enterprises and activities, and studiously avoided recommending privatization. Any step in that direction would have been considered an invasion of a country's sovereignty. But somewhere between the mid-1970s and the early 1980s the attitude of the international organizations changed drastically and they began to address privatization as a technical, economic matter rather than as a political, ideological issue.

In making this change the international organizations were recognizing certain inherent handicaps under which public-sector companies have typically operated. In a nutshell, these handicaps consist of politically-imposed constraints on such enterprises -- constraints that inhibit them from efficiency levels similar to those of private-sector operations. These constraints include:

- ⊖ The limits that keep their executive pay levels far below those prevailing in the private sector.
- ⊖ Strong political pressures against abandoning unprofitable product lines, and against downsizing in general.
- ⊖ Artificially high wages for the unskilled and the semi-skilled, as the government tries to cultivate a reputation as a "model employer".
- ⊖ An ethos of "don't rock the boat" within public enterprises discourages efforts to reduce real cost, particularly when these involve cutting the number of jobs, or restructuring the chain of command.

While we can certainly point to an occasional public enterprise that has somehow overcome them, these constraints are the common lot of public enterprises. Discovering them has represented just as serious a social-scientific inquiring as learning about the different saving habits of families or the different educational levels achieved by young people in different population groups. The changing attitude of the international organizations toward privatization stems from their appreciation of this fact, by dint of repeated, bitter lessons.

Looking at the proliferating wave of privatization over the last two decades, one might be tempted to conclude that governments throughout most of Latin America and much of the rest of the world had learned the same lesson as the international organizations. Unfortunately that would be the wrong inference, at

least as I read the evidence. A much more plausible explanation for many of the privatizations is that selling off public enterprises suddenly becomes a newly-legitimized way for governments to get their hands on more cash.

The evidence favoring this more somber interpretation is the haste with which most privatizations were carried out, once the idea was taken seriously. If the objectives are to reduce waste, to promote economic efficiency, and to advance the interests of both consumers and taxpayers at the same time, then a government should treat an act of privatization just as seriously as General Motors or General Electric would take the sale of a product line or a division of the company.

When private companies sell off a division, they worry about getting it into shape so as to get the best price for it. This is the approach governments should take when they are privatizing anything. Regretfully, governments have not done this, and their lack of due diligence has often come back to haunt them. For example, many public utilities have been sold before an adequate regulatory framework was put into place. Also, with some of the telephone companies, what was often being sold was the monopoly right to exploit the consumer for a substantial period of time. A \$420 charge that I once paid for a half hour telephone call from Buenos Aires, Argentina to San Jose, Costa Rica substantiates this kind of consumer exploitation.

The best rule in privatizations is never to use the proceeds of asset sales to cover current deficits. Instead they should be used to pay off debt. Mexico followed this rule fairly well, dedicating substantial amounts of its privatization proceeds to pay off its debt and using the rest to accumulate assets through a politically popular program called Solidaridad. However, Mexico is an exception to the rule. Few countries have followed this rule. Some, such as Argentina, followed a second-best rule, which uses the proceeds of privatizations to lower fiscal deficits while implementing major fiscal reforms -- reforms that in the long run will independently and permanently close the deficit, the hemorrhage. But a great many countries have simply used the proceeds of privatization to put a patch over a gaping hole in the fiscal budget.

Chile, in the post-Allende period, offers an interesting case and an opportunity to digress to banking and to bank regulation in particular. Chile's first wave of privatizations took place during a huge recession in 1975-76. Credit was extremely scarce. Banks had recently been made competitive, and were paying the depositors 2-3% real interest per month. They were also collecting high rates from the other side, in order to pay 2-3% to their depositors. It was into this maelstrom that the newly privatized companies of the first wave were thrown. Most were not actually being privatized in the strictest sense. They had been intervened, which meant that the government had taken over their operations while the owners

remained the owners. The great bulk of Chilean private companies that were taken under government control during the Allende period were intervened in this way. Thus, privatization meant in this case just returning them to their legal owners. The unfortunate legal owners got these hemorrhaging companies and had to deal with them somehow. At the time they were "privatized" these firms were running huge cash flow deficits. In effect, Chile had been printing money to cover these deficits. What the Chilean government essentially did was to shove these firms off on the private sector. The owners somehow patched them together for a time. But, many of these companies never really gained solvency and ultimately failed.

The only reason these companies stayed around was that bankers kept renewing their credits. Now, why would a banker renew a credit when the banker knows or strongly suspects that the company in question doesn't have the wherewithal to pay back the loan? The answer lies in the regulation of the Chilean banks.

Regulations allowed banks to have loans up to 20 times their capital and surplus. In "recognizing" a bad loan in the amount of, say, a million dollars, they reduced their loans by one million dollars, and they simultaneously reduced capital and surplus by the same amount. That reduction by one million dollars of capital and surplus meant they couldn't lend \$20 million that they otherwise might have lent. So they did not want to tell anyone that this was a bad loan; they wanted everyone

to overlook it. Nor did they need to conspire with other bankers because every single banker independently came to the same conclusion. As a consequence, an overwhelmingly large number of bad loans developed in the banking system. They occupied a place on the consolidated balance sheet that could have, and should have, been occupied by good credit. So, the bad loans were taking the place of what could have been good credit. In doing this, the bad credits resulted in high interest rates on good loans. These huge interest rates on the good loans ultimately drove many good firms into bankruptcy and turned good loans into bad ones. It was a snowball effect.

Chart I illustrates the problem. In order to determine the real amount of money that people are holding or the real amount of credit available, the GDP deflator or the cost of living index is used to deflate the balance sheets. This is important because the ultimate pillar of monetary economics is the demand for real cash balances. Ultimately, it's the people who decide how much M2 there will be in real terms. M2 is the combination of currency, demand deposits, and time and savings deposits. The M2 is on the right-hand side of the balance sheet (see Chart I). Although the central bank does not dictate M2, it influences it through inflation policy. It doesn't print real money, it prints nominal money. The people then react and create enough inflation so they get the amount of real balances that they want. It is people's desired holding of real balances that determine the right-hand side of

the balance sheet in real terms. Now, the left-hand side of the balance sheet is occupied by three main components: net foreign assets, credit to the government, and credit to the private sector. Chart I distinguishes between good and bad credit. If those bad credits didn't exist, the good credit could be the entire bordered area. At the bottom of the chart is a demand curve for good credits by business firms. If these firms had access to the full supply, interest rates would be potential equilibrium real interest rate. But, these bad loans constrained the supply available to the good firms, causing the actual equilibrium real interest rate to be much higher -- and thus, the problem.

The flaw in the system was the Superintendentcy of Banks. The inspectors failed to blow the whistle on these unrecognized bad loans that were hidden in the portfolios of the banks. They failed mainly because they lacked the needed resources. They were underpaid, understaffed, and under-qualified. Good inspectors, at that time, 1975-76, tripled or quintupled their salaries by moving to the private banking sector. As these inspectors examined the books of the banks, they needn't be finding big flaws, but they needed only to go to the manager of the bank with some intelligent question that showed they really understood the situation. The bank manager would be elated and quite willing to offer the inspectors three or four times what they were earning because that was the going market price for

people of their experience and capacity. These were offers that the inspectors were very pleased to accept. The result was like a vacuum cleaner -- sucking out all the best people from the Superintendency of Banks. What was left were mainly people who were simply incapable of doing the job.

The Superintendency's challenge is to find a way to pay an adequate salary scale -- one that truly competes with the private sector. There is a solution: make the bank superintendency a joint enterprise between the government and the commercial banks, have it managed by a technical authority, and finance it by the banks themselves under the rule that the salary scale of the Superintendency must at least match that of the commercial banks. Inspectors would then earn at least as much as comparable employees of private banks. This is the only solution that has a good chance of maintaining a Bank Superintendency of adequately high quality.

Now, it is important to address a new topic -- regulation -- which sets the stage for privatization. Regulation of electricity is of special interest because the appropriate system is dictated by good economics. This is in contrast to regulation based on rates of return, simple price caps and other traditional criteria. There is an elegant economic analysis of electricity pricing that is interesting to study. It is an intellectual challenge that often opens students' eyes to facets of economics of which they have been unaware. The key element is marginal cost pricing. The spot price of energy at any given time is set equal to the marginal running cost of the

least efficient plant functioning at the moment. So, the least efficient plant just gets paid its running cost while all the more efficient plants gain an economic rent or profit because they are getting that same price while their running costs are lower. It's that economic rent or profit, accumulated over the whole year, that may eventually cover all their capital costs.

If this is an ongoing system in a growing economy, the least efficient plants will never recover their capital costs. All they will gain is their running costs. A peak-time surcharge is necessary to cover the capital cost of the *marginal peaking capacity*. Gas turbines, which have a high running cost per kilowatt-hour, but a low capital cost, are typically used for these peaking purposes. They are kept in storage much of the time, which trades off a low capital cost for a high running cost for capacity that is likely to be operating only a small number of hours a year. When some plants are operating and some are not, what they collect is the marginal running cost of those plants. When all plants are in use, a peak-time surcharge is applied. That peak-time surcharge, over the number of hours when they hypothetically are all in use, is what covers the capital cost of the peaking capacity. In turn, the peaktime surcharge is part of the price recovered by all other types of capacity in the system operating at those times.

Chart II illustrates how a peaktime surcharge works. The vertical line represents a given capacity. At the beginning, the demand is less than capacity,

which translates into spot prices that are equal to marginal running costs. Demand grows with time, so the demand curve bounces to the right. Finally, it turns into the zone where it intersects with the vertical line of capacity. The price then gets to be the crossing point of the full capacity line with the shifting demand curve (point B). That produces a rent to capacity. If demand keeps growing, that rent keeps growing. When the demand gets to the point where the price on the vertical line is equal to the full cost price of the gas turbines, then instead of letting it go up still more, more gas turbines will be acquired and used.

Most electricity enterprises or systems actually go directly to point C (see Chart II). They calculate what is the peak-time surcharge, which will justify investment in additional gas turbine capacity. They ignore the short periods of time in which they might be in situations like B. This system of "marginal cost pricing" was invented in France, and it has been applied extensively in the United Kingdom as well as in some other countries. The practice fits the theory best in predominantly thermal systems. In this kind of system, peak time is defined in terms of the hours of the day plus the season of the year.

It's trickier to deal with systems that have substantial hydro capacity. The problem is that years can go by in such a system without reaching full use of its thermal capacity. Such full use is reached only when the storage dams are nearly empty, which usually happens after two or three dry years in a row. In countries

like Chile, Mexico, Peru, and Costa Rica, strict application of marginal cost pricing would charge just marginal cost nearly all the time. After a third dry year in a row it would become necessary to use all the available thermal capacity for much of the year. This would result in very high peak-time surcharges for huge numbers of hours in the year -- charges that could be as high as 25-35 cents (of dollar) per kilowatt-hour to all users of energy. Nobody likes this.

The systems of countries like Mexico, Chile, Peru and other Andean countries really do not have full applications of marginal cost pricing. They are unwilling to go that last step of enacting huge surcharges for the second and third dry year. The end result is application of something like marginal cost pricing for the day-to-day and hour-to-hour movements. Then there are demand charges and capacity charges as well as what I call insurance policies, where a small charge is added to the hourly kwh rate to build up a surplus that will pay the cost of having the standby capacity that is needed in order to cope with a sequence of dry years. There is some loss of economic efficiency involved here, and there may be further problems in the economic pricing of electricity.

Good economics tells us that sunk costs are sunk. If you invest in a technology that becomes obsolescent, you are supposed to take the hit and write down that asset to a new lower value because it has been rendered less valuable by some new technology that is causing the market price to fall. It appears that

modern combined-cycle technology has rendered a lot of old generating equipment really obsolete in this sense. But in how many countries have the electricity enterprises taken a big financial hit as a consequence of the new dominance of combined cycle technology? They have probably made good returns on the old assets as well as on the new, and somewhere along the way, a major economic principle is not being implemented.

At this point, it is appropriate to examine other energy sector problems. Of particular interest is Chile's public sector generating entity, Endesa, which was not broken up at the time of its privatization. It was a successful enterprise with fine management. There was a concern that the process of privatization would cause a few traumas, so the emphasis was solely on a comfortable transition. To break up the enterprise would have required new management for all the various pieces. It was inertia more than anything else that kept it from being divided.

When Endesa was privatized, it was still the owner of the main transmission lines in Chile. A modern electricity system lets anyone be a competitor at the spot prices. It is essential to have a neutral referee to receive the energy from the generators and deliver it to the distributors. However, this was not the case in Chile. The referee was not only one of the participants but was and still is the biggest participant in the generating game. There have been many accusations that it uses all sorts of tricks to undercut its competitors. The division responsible for

transmission should have been separated out. It could (and probably should) have been kept as a quasi-public body together with the dispatch center that decides which plants will be turned on and in which order, working from low-cost to high-cost plants.

Mexico's handling of electrical generation also is interesting and quite different from Chile's. Mexican generating capacity was badly de-optimized in the wake of the 1980s debt crisis. It is a story that shows how different pieces of economic policy work against each other -- one of them fouls up another. Mexico had practically no money for investment. The electricity enterprise, the Comisión Federal de Electricidad, realized that the country was in the trough of a big depression, that demand for electricity would be growing fast in future years, and that it was not being allowed to build to meet the future demand. So it played a trick. It used very expensive energy, gas turbine energy, to feed the current demand during this depression. It did not use energy from its dams, but instead allowed the dams to fill all the way up to the top. In this way it stored power now for future use. It was a very ingenious way of dealing with the constraint it faced of not being given funds to expand capacity "right now". In the process of all of this, it introduced peak-time surcharges for large users of energy. These charges were unlike Chile's, where the charges were meant to cover just the running costs plus the capital costs of the marginal capacity during the hours of full use. Mexico raised

the peak-time charges to levels that would stop peak-time demand. Some enterprises were paying 30 and 35 cents (of dollar) per kwh for peak-time energy. It's amazing that once this system was in place, it stayed.

A look at the privatization of highways in Mexico is also of interest. The process has been referred to as the Mexican Highway Disaster. The people who constructed the expressways (*autopistas*) became the owners of them. The government guaranteed certain minimum traffic volumes. The constructors then set tolls that in a practical sense were prohibitive. As a result, it was not unusual to find a stretch of tollway like that from Mexico Cith to Puebla practically empty. The main reason was that there was a free parallel road, while the toll road charged 25 cents (of dollar) per kilometer. The toll road simply did not have an advantage sufficient to support such a high toll. This resulted in traffic levels that were far below the guaranteed minimum. In the end the government caved in and began buying back most of these roads from the constructors. It was a scandal, a mess.

There is absolutely no advantage to private sector ownership of roads. There is an advantage to the private sector building and managing roads, but not to its owning them. A private owner does not have the incentive to widen or upgrade a road at the time it is economically warranted. The private incentive is to raise the price in such circumstances. Also, the private owner has little incentive to

straighten out dangerous curves or otherwise follow where good cost-benefit analysis leads.

The Double Auction approach is a solution for governments that need roads but do not have money to construct them. In the first auction, each constructor or bidder submits two envelopes. One of these envelopes contains a cash price and the other contains a credit price with payments spread over 10, 15, or 20 years, whatever the government asks for in the bid. The faces head-on the main reason why governments ever thought about privatizing roads: they didn't have the money to construct them.

Governments were simply short of money. They badly wanted roads to be built, and they found build/own/operate (B,O,O) to be a very tempting alternative. However, the auction with two envelopes solves the problem because governments can always pick the credit envelope and not have to lay out any cash. The second auction in the double auction system is for the administration and maintenance of the road once it is built. There is no reason why the constructor should be any good at this. Such a contract should be for a 3-5 year period and no longer. At the end of the contract, the government can make necessary adjustments such as widening the road, straightening out the curves, etc. During the period when the road is being adjusted, management can be handled through a patch contract. A new full auction for another 3-5 year period would take place once the needed

adjustments had been completed. The Double Auction system accomplishes what a B,O,O system does, but without its disadvantages.

Proper highway pricing should be mainly based on congestion tolls. This is where economic theory comes into the picture. Congestion tolls are meant to make drivers recognize the costs that they impose on others as they add themselves to the list of traffic on the road. Congestion tolls are first-best road pricing. There is no reason why congestion tolls should cover costs, at least not in the first instance. They just compensate for whatever congestion there may be. There are, of course, other reasons to consider a set of charges that extract from users the full economic cost of the road. The users may be asked to bear the full costs of the road rather than have the taxpayers paying all or part of these costs. If such charges are implemented via tolls, they are distinctly second best, but they can certainly be considered as the government auctions the maintenance and administration of the roads. But full cost pricing of roads is a second best and not a first best alternative. What does not make much sense is giving private road owners a free or almost free hand in setting tolls. Roads need economic safeguards.

My conclusion as well as a unifying theme of this presentation is that in designing economic policy, there is nothing more important than thinking in terms of costs and benefits. It leads to new ways of looking at certain issues and

inevitably impacts decisions, helping guide the authorities toward sound economic policy.