# Why Marginal Pricing?

Antonio J. Conejo

***Abstract*—Under perfect competition, marginal pricing results in short-term efficiency and the subsequent right short-term price signals. However, the main reason for the adoption of marginal pricing is not the above, but investment cost recovery, that is the fact that the profits obtained by infra-marginal technologies (technologies whose production cost is below the marginal price) allow them just to recover their investment costs. In addition, if the perfect competition assumption is removed, investment over-recovery or under-recovery generally occurs for infra-marginal technologies.**

***Index Terms*—Electricity market, marginal pricing, investment cost recovery, maximum social welfare, right spatial and temporal price signals.**

## NOMENCLATURE

### A. Constants

| | |
|---|---|
| $C^{\text{peak}}$ | Production cost of the peak technology ($/MWh) |
| $C^{\text{base}}$ | Production cost of the base technology ($/MWh) |
| $C^{\text{unser}}$ | Unserved energy cost ($/MWh) |
| $I^{\text{peak}}$ | Annualized investment cost of the peak technology ($/MW) |
| $I^{\text{base}}$ | Annualized investment cost of the base technology ($/MW) |

### B. Decision Variables

| | |
|---|---|
| $p^{\text{peak}}$ | Capacity of the peak technology (MW) |
| $p^{\text{base}}$ | Capacity of the base technology (MW) |

### C. Other Variables

| | |
|---|---|
| $e^{\text{peak}}(\cdot)$ | Annual energy produced by the peak technology (MWh) |
| $e^{\text{base}}(\cdot)$ | Annual energy produced by the base technology (MWh) |
| $e^{\text{unser}}(\cdot)$ | Annual unserved energy (MWh) |
| $h^{\text{peak}}$ | Time period of operation of the peak technology (hour) |
| $h^{\text{unser}}$ | Time period with the the unserved energy (hour) |
| $p^{\text{unser}}$ | Unserved capacity (MW) |

Manuscript received: February 4, 2023; revised: February 9, 2023; accepted: February 13, 2023. Date of CrossCheck: February 13, 2023. Date of online publication: February 22, 2023.



A. J. Conejo (corresponding author) is with the Department of Integrated Systems Engineering and the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, Ohio 43210, USA (e-mail: conejo.1@osu.edu).

DOI: 10.35833/MPCE.2023.000064

## I. INTRODUCTION

UNDER perfect competition (presence of many small investors/producers able to enter/exit the market until a null long-term profit materializes), marginal pricing results in short-term efficiency (achievement of maximum social welfare) and the subsequent right short-term price signals. However, the main reason for the adoption of marginal pricing is not the above, but investment cost recovery, that is the fact that the profits obtained by infra-marginal technologies (technologies whose production cost is below the marginal price) allow them just to recover their investment costs.

We analyze first the short-term consequences of marginal pricing, and then, with deeper detail, the long-term ones.

For the short-term analysis, we consider a competitive electricity market clearing as illustrated in Fig. 1. This figure corresponds to a given operating condition, e.g., one hour of operation (among the 8760 hours of one year). Figure 1 represents the supply curve, the demand curve, the resulting marginal price ($p^*$), and the resulting demand to be supplied ($q^*$). For simplicity, the supply curve includes two technologies, base and peak. Price $p^*$ is the marginal price because it corresponds to the increment in the supply cost as a result of a marginal (small) increment in the demand for energy (by slightly shifting right the vertical line of the demand curve). We assume that the demand is inelastic, that is, it bids a buying price large enough (horizontal line of the demand curve) for the whole demand to be supplied.

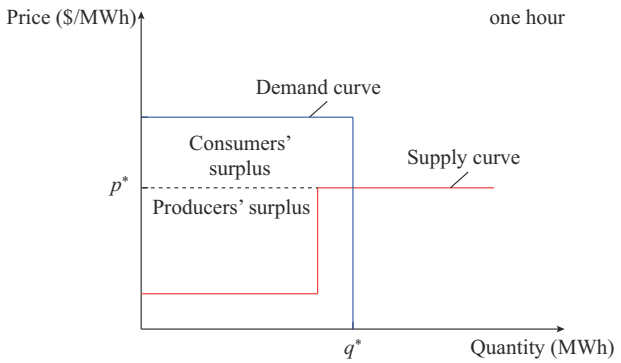


Fig. 1. Market clearing for one hour.

Regarding the long-term analysis and for the sake of simplicity (and without loss of generality), we consider a trape-

zoidal annual demand curve in the form of a load duration curve, as shown in Fig. 2 and two technologies to supply the demand, base and peak. The load duration curve is a power-hour plot that represents all 8760 hourly demands of the year arranged from the largest load to the smallest one. This load duration curve embodies 8760 clearing conditions similar to the one represented in Fig. 1. We consider as well that unserved energy occurs for a number of hours of the year.

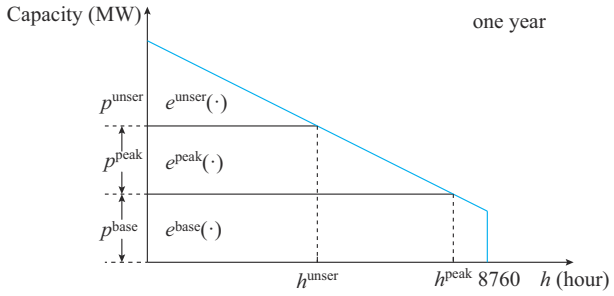


Fig. 2. Load duration curve for one year.

We first consider an investor-and-operator that has perfect (technical and economic) knowledge and invests in production facilities (base and peak) and then operates such facilities to supply the demand seeking the minimum cost (investment and operation).

Next, we consider a perfectly competitive market composed of many small investors-and-producers of the two technologies considered (base and peak) that are able to enter/exit the market until a null long-term profit materializes, which is the condition for market equilibrium.

We then compare the outcomes in terms of investment capacities (in base and in peak technologies) of the two alternatives below (omniscient investor-and-operator and perfectly competitive market).

Finally, we discuss the consequences of removing some of the assumptions inherent to the definition of a perfectly competitive market.

Regarding background information, a friendly economics manual is given in [1]. The discussion in this paper can be expanded and enriched by checking Chapter 4 in [2] and [3]. Finally, the fundamentals of power system operations can be found in [4] or [5].

## II. The Maximum Social Welfare

We note that using $p^*$ as the clearing price in Fig. 1 results in the maximum social welfare. The social welfare is the area between the supply and demand curves (upper and lower rectangles in Fig. 1) and is composed of producers' surplus and consumers' surplus.

The producers' surplus (lower rectangle in Fig. 1) is the profit of the producers as it corresponds to the energy supplied times the price difference between the marginal price and the price at which each producer is willing to sell.

Besides, the consumers' surplus (upper rectangle in Fig. 1) is the profit of the consumers as it corresponds to the energy supplied times the price difference between the per unit utility (that represents the per-unit revenue of using electricity), at which each consumer is willing to buy, and the marginal price.

Marginal prices send right signals both temporally and spatially. Temporally, a period of high prices incentivizes comparatively cheaper producers to produce during such period and disincentivizes consumers to consume during that period. Conversely, a period of low prices incentivizes consumers to consume during such period and disincentivizes producers to produce during that period.

Similarly, a location with high prices incentivizes comparatively cheaper producers to move to such location and disincentivizes consumers to move to it. Conversely, a location with low prices incentivizes consumers to move to that location and disincentivizes producers to move to it.

## III. Investment Cost Recovery

The main reason for the adoption of marginal pricing is investment cost recovery, that is, the fact that the profits obtained by infra-marginal technologies (technologies whose production cost is below the marginal price) allow them just to recover their investment costs. This is carefully analyzed below.

### A. Omniscient Investor-and-Operator

We consider in this subsection an omniscient investor-and-operator that has perfect knowledge (of all technical and economic details) and seeks the minimum total cost (investment and operation).

The target of the omniscient investor-and-operator is to identify the optimal investment in peak and base capacities ($p^{\text{peak}}$ and $p^{\text{base}}$, respectively) to supply the demand (for the whole year) at the minimum total cost.

The total annual cost $C^{\text{total}}(\cdot)$ includes annualized investment and yearly operation costs. The annualized investment cost $C^{\text{investment}}$ is:

$$C^{\text{investment}} = I^{\text{peak}} p^{\text{peak}} + I^{\text{base}} p^{\text{base}} \tag{1}$$

While the yearly operation cost $C^{\text{operation}}$ is:

$$C^{\text{operation}} = C^{\text{peak}} e^{\text{peak}}(p^{\text{peak}}) + C^{\text{base}} e^{\text{base}}(p^{\text{base}}) + C^{\text{unser}} e^{\text{unser}}(p^{\text{peak}}, p^{\text{base}}) \tag{2}$$

Pursuing optimality, the omniscient investor-and-operator solves the problem below.

$$\min_{p^{\text{peak}}, p^{\text{base}}} C^{\text{total}}(\cdot) = C^{\text{investment}}(\cdot) + C^{\text{operation}}(\cdot) \tag{3}$$

The criterion to get the optimal solution of problem (3) is $\partial C^{\text{total}}(\cdot)/\partial p^{\text{peak}} = 0$ and $\partial C^{\text{total}}(\cdot)/\partial p^{\text{base}} = 0$.

Regarding the peak technology and considering the annualized investment cost (1) and the yearly operation cost (2), we obtain:

$$\frac{\partial C^{\text{total}}(\cdot)}{\partial p^{\text{peak}}} = I^{\text{peak}} + \frac{\partial C^{\text{operation}}(\cdot)}{\partial p^{\text{peak}}} = 0 \tag{4}$$

To compute $\partial C^{\text{operation}}(\cdot)/\partial p^{\text{peak}}$, we consider a marginal increment in $p^{\text{peak}}$, i. e., $\mathrm{d}p^{\text{peak}}$, and compute the corresponding increment in the yearly operation cost, i. e., $\mathrm{d}C^{\text{operation}}$. For this, we consider Fig. 2. The resulting yearly operation cost change is $\mathrm{d}C^{\text{operation}} = \mathrm{d}p^{\text{peak}} h^{\text{unser}} (C^{\text{peak}} - C^{\text{unser}})$.

This is because an increase in the available peak production capacity $dp^{peak}$ reduces the unserved power by $dp^{peak}$ during unserved energy hours, $h^{unser}$ in see Fig. 2. The increase in peak power during unserved energy hours results in an extra cost of $dp^{peak}h^{unser}C^{peak}$, while the decrease in unserved power results in a cost reduction of $dp^{peak}h^{unser}C^{unser}$. Then, $\partial C^{operation}(\cdot)/\partial p^{peak} = h^{unser}(C^{peak} - C^{unser})$.

Recalling (4), the first optimality condition is:

$$I^{peak} - h^{unser}(C^{unser} - C^{peak}) = 0 \tag{5}$$

Regarding the base technology, we proceed similarly:

$$\frac{\partial C^{total}(\cdot)}{\partial p^{base}} = I^{base} + \frac{\partial C^{operation}(\cdot)}{\partial p^{base}} = 0 \tag{6}$$

Again, to compute $\partial C^{operation}(\cdot)/\partial p^{base}$, we consider a marginal increment in $p^{base}$, i.e., $dp^{base}$, and compute the corresponding change in the yearly operation cost, i.e., $dC^{operation}$. For this, we consider Fig. 2. The resulting operation cost change is $dC^{operation} = dp^{base}h^{peak}(C^{base} - C^{peak}) + dp^{base}h^{unser}(C^{peak} - C^{unser})$.

This is because an increase in the available base capacity $dp^{base}$ displaces up the available peak production capacity during peak hours $h^{peak}$, and reduces the unserved power by $dp^{base}$ during unserved energy hours $h^{unser}$.

On the one hand, the displacement of the peak power during peak hours results in an extra cost by the base technology of $dp^{base}h^{peak}C^{base}$, and a cost reduction by the peak technology of $dp^{base}h^{peak}C^{peak}$. On the other hand, the increase in peak production during unserved energy hours results in an extra cost of $dp^{base}h^{unser}C^{peak}$, while the decrease in unserved power results in a cost reduction of $dp^{base}h^{unser}C^{unser}$. Then, $\partial C^{operation}(\cdot)/\partial p^{base} = h^{peak}(C^{base} - C^{peak}) + h^{unser}(C^{peak} - C^{unser})$.

Thus, recalling (6), the second optimality condition is:

$$I^{base} - h^{peak}(C^{peak} - C^{base}) - h^{unser}(C^{unser} - C^{peak}) = 0 \tag{7}$$

Therefore, considering both optimality conditions (5) and (7) together renders:

$$\begin{cases} I^{peak} - h^{unser}(C^{unser} - C^{peak}) = 0 \\ I^{base} - h^{peak}(C^{peak} - C^{base}) - h^{unser}(C^{unser} - C^{peak}) = 0 \end{cases} \tag{8}$$

or

$$\begin{cases} I^{peak} - h^{unser}(C^{unser} - C^{peak}) = 0 \\ I^{base} - h^{peak}(C^{peak} - C^{base}) = I^{peak} \end{cases} \tag{9}$$

The system of equations in (8) allows deriving optimal values for $h^{unser}$ and $h^{peak}$. That is:

$$\begin{cases} h^{unser*} = \dfrac{I^{peak}}{C^{unser} - C^{peak}} \\ h^{peak*} = \dfrac{I^{base} - I^{peak}}{C^{peak} - C^{base}} \end{cases} \tag{10}$$

Considering the load duration curve in Fig. 2, $h^{unser*}$ and $h^{peak*}$ allow computing the optimal values of the peak and base power to be built, i.e., $p^{peak*}$ and $p^{base*}$, respectively.

Additionally, we note as well that the system of equations in (8) can also be expressed as:

$$\begin{cases} I^{peak} + C^{peak}h^{unser} = C^{unser}h^{unser} \\ I^{base} + C^{base}h^{peak} = I^{peak} + C^{peak}h^{peak} \end{cases} \tag{11}$$

This allows the graphical interpretation provided in Fig. 3.

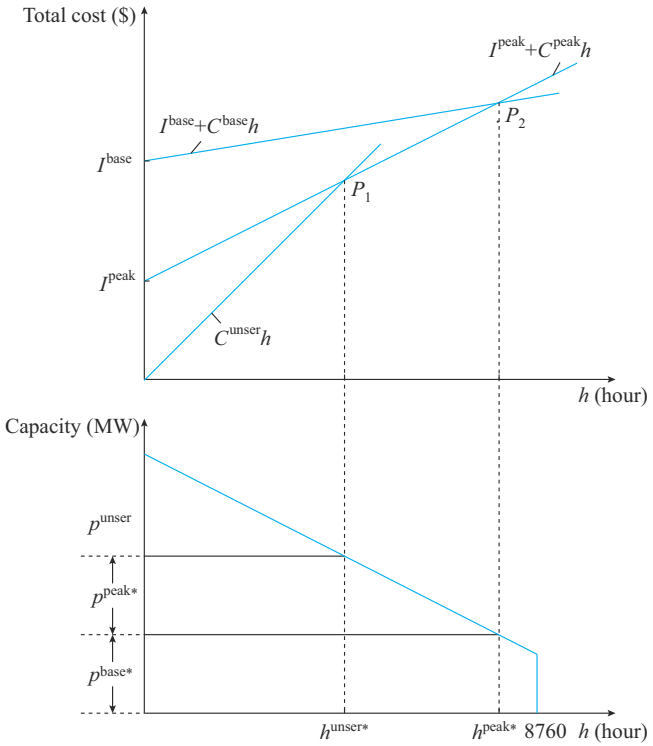


Fig. 3. Optimal investment and operations.

The upper plot of Fig. 3 includes three straight lines. Each of these lines expresses the total cost (annulized investment and yearly operation) as a function of the production hours $h$. The investment cost corresponds to the $y$-axis intercept while the operation cost increases linearly with the production hours (in the $x$-axis). These lines are:

1) $Cost^{unser}(h) = C^{unser}h$.
2) $Cost^{peak}(h) = I^{peak} + C^{peak}h$.
3) $Cost^{base}(h) = I^{base} + C^{base}h$.

The first line corresponds to the total unserved energy cost, the second one to the total peak cost, and the third one to the total base cost.

Point $P_1$ in Fig. 3 corresponds to the intersection of the unserved energy cost line and the peak cost line, which is equivalent to the first condition in (9), and in turn corresponds to the first condition in (8).

Similarly, point $P_2$ in Fig. 3 corresponds to the intersection of the peak cost line and the base cost line, which is equivalent to the second condition in (9), and in turn corresponds to the second condition in (8).

Finally, since the $x$-axes for both upper and lower plots in Fig. 3 are production hours, both plots can be connected regarding: ① optimal production hours (unserved and peak); and ② optimal total costs (unserved, peak, and base).

### *B. Perfectly Competitive Market*

We consider in this subsection a perfectly competitive market composed of many small investors-and-producers, or investors-and-producers that are price takers. Each investor-and-producer has the capability of entering or exiting the market at will until it achieves zero long-term profit (including investment and operation), which is the condition for market equilibrium. We assume that this market relies on

marginal pricing.

We consider first a peak producer, compute its long-term profit, and make it equal to zero.

A peak producer does not operate at base hours, and its profit is null while operating during peak hours. Therefore, it only makes a profit during unserved energy hours.

Considering marginal pricing, the revenue of such producer during unserved energy hours is $C^{\text{unser}} p^{\text{peak}} h^{\text{unser}}$ (the marginal pricing during unserved energy hours is $C^{\text{unser}}$), while its total cost (investment and operation) is $I^{\text{peak}} p^{\text{peak}} + C^{\text{peak}} p^{\text{peak}} h^{\text{unser}}$.

The long-term profit is thus $C^{\text{unser}} p^{\text{peak}} h^{\text{unser}} - I^{\text{peak}} p^{\text{peak}} - C^{\text{peak}} p^{\text{peak}} h^{\text{unser}}$.

Making this long-term profit zero renders:

$$I^{\text{peak}} + C^{\text{peak}} h^{\text{unser}} = C^{\text{unser}} h^{\text{unser}} \tag{10}$$

Next, we consider producers that use the base technology.

A base producer experiences profit different from zero during peak hours ($h^{\text{peak}} - h^{\text{unser}}$) and unserved energy hours ($h^{\text{unser}}$). We note that the profit during base hours is zero. Thus, the revenue of such producer during unserved energy hours ($h^{\text{unser}}$) with marginal price $C^{\text{unser}}$ and strictly peak hours ($h^{\text{peak}} - h^{\text{unser}}$) with marginal price $C^{\text{peak}}$ is $C^{\text{unser}} p^{\text{base}} h^{\text{unser}} + C^{\text{peak}} p^{\text{base}} (h^{\text{peak}} - h^{\text{unser}})$, while the total cost (investment and operation) is $I^{\text{base}} p^{\text{base}} + C^{\text{base}} p^{\text{base}} h^{\text{peak}}$.

The long-term profit is thus $C^{\text{unser}} p^{\text{base}} h^{\text{unser}} + C^{\text{peak}} p^{\text{base}} \cdot (h^{\text{peak}} - h^{\text{unser}}) - I^{\text{base}} p^{\text{base}} - C^{\text{base}} p^{\text{base}} h^{\text{peak}}$.

Making this long-term profit zero leads to $(h^{\text{peak}} - h^{\text{unser}}) C^{\text{peak}} + h^{\text{unser}} C^{\text{unser}} - h^{\text{peak}} C^{\text{base}} - I^{\text{base}} = 0$, and considering (10) renders:

$$I^{\text{base}} - h^{\text{peak}} (C^{\text{peak}} - C^{\text{base}}) = I^{\text{peak}} \tag{11}$$

Considering together the peak and the base no-profit conditions ((10) and (11)), we can obtain:

$$\begin{cases} I^{\text{peak}} - h^{\text{unser}} (C^{\text{unser}} - C^{\text{peak}}) = 0 \\ I^{\text{base}} - h^{\text{peak}} (C^{\text{peak}} - C^{\text{base}}) = I^{\text{peak}} \end{cases} \tag{12}$$

Observing conditions in (8) for the omniscient planner-and-operator and conditions in (12) for the perfectly competitive market, we conclude that both approaches result in exactly the same optimality conditions.

The above is a relevant fact indicating that a perfectly competitive market will produce, in terms of both investment and operation outcomes, identical results to those produced by an omniscient investor-and-operator. Thus, instead of an omniscient investor-and-operator, a perfectly competitive market can be used.

## VI. Market Imperfection

Removing some of the strong assumptions of perfect competition generally results in undesirable consequences [6].

Particularly if, for example, the peak technology suddenly increases its operation cost by a factor of 10 and, against perfect competition assumptions, cannot be expelled from the market, the clearing marginal price will increase by a factor of 10. As a consequence, the base technology (infra-marginal) will increase its revenue by a factor of 10, and will be able to recover its investment cost quickly, generally at the cost of the consumers. Conversely, if the operation cost of the peak technology decreases by a factor of 10 as a result of a technology breakthrough and the base technology cannot leave the market (against perfect competition assumptions), it may not be able to recover its cost, and an under-capacity situation may arise as a result of lack of investment.

Needless to say, removing the perfect competition assumption of the presence of many small producers, leading to an oligopolistic market with few strategic producers, will generally result in important price distortions, again at the cost of the consumers.

Producers and consumers need to be independent agents, so that the supply and demand curves form separately. An agent that is (implicitly of explicitly) both a producer and a consumer (against perfect competition assumptions), and that is large enough, may generate important market distortions.

Besides, it is most important that the market rules are designed and implemented to induce any producer to offer its true marginal cost. Likewise, the market design should induce any consumer to bid its true marginal utility. Additionally, the market design should ensure short-term cost recovery (by producers) and revenue adequacy (the payment by consumers should be equal to or higher than the revenue of the producers), that is, in addition to perfect competition, the market rules need to be correctly designed and implemented.

Other market agent features or market clearing rules not explicitly represented in the analysis provided in this paper may alter market outcomes significantly, but do not generally modify the conclusions reported. These features include:

1) The minimum power output, and the minimum up- and down-time of production facilities (leading to lumpiness and non-convexity in market clearing algorithms).

2) Inter-temporal constraints due to ramping limits of production facilities and demand inflexibility.

3) Network bottlenecks.

4) The presence of a significant number of producers using technologies with near zero marginal cost.

5) The availability of a future market and contracts to mitigate the risk of price variability.

6) Increasing uncertainty in the market due to a demand that increasingly incorporates behind-the-meter devices, and the presence of an increasing number of weather-dependent production facilities.

7) Temporal or spatial aggregation of locational marginal prices (common in EU electricity markets).

## V. Conclusion

Marginal pricing is universally used in electricity markets because under perfect competition such pricing ensures short-term efficiency and, what is more important, investment cost recovery, that is the fact that the profits obtained by infra-marginal technologies allow them just to recover their investment costs. Using a stylized model, this tutorial paper formally shows the cost recovery property.

However, if some of the components of the perfect competition assumption are removed, the cost recovery property does not generally hold true. This paper briefly reviews marker imperfections that alter the cost recovery property.

## REFERENCES

[1] H.-J. Chang. *Economics. The User's Guide.* New York: Bloomsbury Press, 2014.
[2] M. Tanaka, A. J. Conejo, and A. S. Siddiqui. *Economics of Power Systems.* New York: Springer, 2022.
[3] M. Ventosa, P. Linares, and I. J. Perez-Arriaga. "*Power System Economics*," in *Regulation of the Power Sector.* New York: Springer, 2014.
[4] A. J. Conejo and L. Baringo. *Power System Operations.* New York: Springer, 2018.
[5] A. Gómez-Expósito, A. J. Conejo, and C. A. Cañizares. *Electric Energy Systems: Analysis and Operation*. Boca Raton: CRC Press, 2018.
[6] A. J. Conejo and R. Sioshansi, "Rethinking restructured electricity market design: lessons learned and future needs," *International Journal of Electrical Power & Energy Systems*, vol. 98, pp. 520-530, Jun. 2018.

**Antonio J. Conejo** received the M.S. degree from the Massachusetts Institute of Technology, Cambridge, USA, in 1987, and the Ph.D. degree from the Royal Institute of Technology, Stockholm, Sweden, in 1990. He is currently a Professor at the Department of Integrated Systems Engineering and the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, USA. His research interests include control, operations, planning, economics and regulation of electric energy systems as well as statistics and optimization theory and its applications.