Energy Policy 105 (2017) 550–561

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Price computation in electricity auctions with complex rules: An analysis of investment signals☆

Carlos Vazquez[a,b], Michelle Hallack[c,d], Miguel Vazquez[c,d,e,*]

[a] School of Mines and Energy, Universidad Politécnica de Madrid, Calle de Ríos Rosas, 21, 28003 Madrid, Spain
[b] Gas Natural Fenosa, Avenida de San Luis, 77, 28033 Madrid, Spain
[c] Faculty of Economics, Universidade Federal Fluminense (UFF), Campus do Gragoatá, Bloco F, São Domingos, Niterói, 24210-510 Rio de Janeiro, Brasil
[d] Florence School of Regulation, RSCAS, European University Institute, Via delle Fontanelle, 19, 50014 Firenze, Italy
[e] Center for Regulation and Infrastructure Studies (CERI), FGV, Barão de Itambí, 60, Botafogo, Rio de Janeiro, Brazil

ARTICLE INFO

Keywords:
Electricity auctions
Investment signals
Side payments
Integer decisions
Marginal cost

ABSTRACT

This paper discusses the problem of defining marginal costs when integer variables are present, in the context of short-term power auctions. Most of the proposals for price computation existing in the literature are concerned with short-term competitive equilibrium (generators should not be willing to change the dispatch assigned to them by the auctioneer), which implies operational-cost recovery for all of the generators accepted in the auction. However, this is in general not enough to choose between the different pricing schemes. We propose to include an additional criterion in order to discriminate among different pricing schemes: prices have to be also signals for generation expansion. Using this condition, we arrive to a single solution to the problem of defining prices, where they are computed as the shadow prices of the balance equations in a linear version of the unit commitment problem. Importantly, not every linearization of the unit commitment is valid; we develop the conditions for this linear model to provide adequate investment signals. Compared to other proposals in the literature, our results provide a strong motivation for the pricing scheme and a simple method for price computation.

1. Introduction

This paper is concerned with the study of mechanisms to coordinate long- and short-term decisions in power markets. Since often a very relevant part of this coordination happens through the price signal provided by short-run marginal costs, we will revisit the problem with the aim of showing that including start-up costs and other costs related to integer decisions in the definition of short-run marginal costs plays a relevant role in the coordination of system operation and investment.

Specifically, we will discuss the computation of marginal cost when cost functions include binary variables (those with only two possible values: 1 or 0). More precisely, we will focus on day-ahead electricity markets. Generators willing to sell in a day-ahead market face binary start-up and shut-down decisions, which are not yet fixed at the time of bidding and have to be included in the decision-making process. The existence of these binary variables makes cost functions non-convex, which in turn causes the cost derivative to be ill-defined (see Section 2 for details), so the direct application of the standard perfect competition concept of "price equal to marginal cost" is not obvious for this case. Therefore, when start-up variables are present additional criteria have to be used to define price. The existing literature shows a range of different alternatives to do so, each of them leading to different prices. This paper will try to gain insight into the reasons why marginal costs are not clearly defined, with the aim of contributing to the discussion on the choice of the criteria to be used to calculate prices.

The problem of ill-defined prices is especially apparent when the regulator has opted for a market design that is based on a complex auction. The pure complex auction is essentially a traditional unit commitment model, which is applied to clearing power markets (Hobbs, 2001). Therefore, the auctioneer receives bids from generators that include, not only their variable costs and their output capacities, but also their start-up costs, minimum stable loads, ramp rates, and other technical characteristics. The problem of the auctioneer becomes thus a non-convex optimization, so price is not anymore the cost derivative at the optimal solution point, and a number of different proposals arise for price computation.

☆ All thoughts and opinions presented in this paper are the authors' alone and do not reflect the views of their employers, clients or any other institution. Particularly, they do not reflect in any way the position of Gas Natural Fenosa.

* Corresponding author at: Faculty of Economics, Universidade Federal Fluminense (UFF), Campus do Gragoatá, Bloco F, São Domingos, Niterói, 24210-510 Rio de Janeiro, Brasil.
E-mail addresses: vazquez.martinez@upm.es (C. Vazquez), MichelleCMHallack@gmail.com (M. Hallack), Miguel.Vazquez.Martinez@gmail.com (M. Vazquez).

http://dx.doi.org/10.1016/j.enpol.2017.02.003
Received 6 June 2016; Received in revised form 26 January 2017; Accepted 4 February 2017
Available online 28 February 2017
0301-4215/

Most of the solutions presented in the literature are derived from the algorithms used to solve the mixed-integer optimization problem. Basically, the processes used to compute the optimal solution are translated into price-setting criteria. Since there are many alternative ways to reach the optimal solution (the values of the production for each plant), marginal costs may differ greatly between the different approaches. This paper is aimed to adding some additional criteria to help discriminating between these approaches. Kahn (1970) identified two goals required from prices when fixed costs are involved: i) ensuring efficient accounting-cost recovery, and ii) defining forward-looking opportunity cost and incentives. Almost all of the solutions proposed in the literature for the day-ahead pricing problem focus on the first of them: making sure that all of the generators that are accepted in the auction receive at least their operating costs, so they are willing to produce. We will focus on the second one: providing incentives for future decisions, both for consumers and for investment in new generation plants (Vazquez, 2003). This will allow us to discard some of the proposals already on the table, narrowing the range of mechanisms to be considered.

The problem of investing in power plants can be split into several separate topics. On the one hand, there is plenty of literature regarding the adequacy issue –see for instance Vazquez et al. (2002) or Finon and Pignon (2008)– which focus on how to make sure that there will be enough installed capacity in the system to provide a reasonable level of security of supply. This includes proposals such as capacity remuneration mechanisms, the Value of Lost Load (VOLL) mechanism that was used in the original England and Wales Pool, etc. Our numerical example in Section 3.4 includes a representation of that, showing that the problems associated with lost load are not necessary the same as the ones studied in this paper. We will not address those problems in this paper. Alternatively, we will concentrate on the problem of technology choice, trying to understand how the choices are made to decide which part of the installed capacity will be baseload generators, and which part will be mid-merit, or peaking units. Those decisions are mainly driven by the prices captured at the spot market, so different ways of calculating short-run marginal costs may lead to different technology mixes.

The problem of price computation is not restricted to complex auctions. Many electricity markets have opted for a simple auction in their market design. Under this scheme, bidders just submit to the auctioneer several pairs of price and quantity for each of the hours in the market horizon (typically, one day), and prices can be computed in a clear and unequivocal way just by crossing the aggregated supply and demand curves, for each hour independently. The auctioneer does not have to consider any start-up cost nor binary decision variables when computing prices, and the problem of concern to this paper is apparently not present in simple auctions. However, in a simple auction generators have to internalize into their bids all of the technical characteristics that are not directly taken into account by the auction; for instance, they have to bid above their variable cost in order to incorporate their start-up cost in the price. When preparing such bids, sellers would typically use an optimization model to, among other things, determine how to split their start-up cost among the different hours of the following day or days. And that problem will include start-up decisions, so it will have binary variables, and ill-defined prices. The price definition issue moves from the auctioneer's problem to the bidder's problem, but it is still in place. We will concentrate hereafter in the complex auction case, but the reasoning and conclusions that we will elaborate are of application to the bidder´s problem in a simple auction. Even in other less common designs for the day-ahead market, such as clock auctions (Wilson, 1998), the issue of prices still holds, either at the auctioneer's problem or at the bidder's one.

The increasing role of renewable energy tends to stress the opportunity of this discussion, see for instance (de Sisternes et al., 2015) for numerical simulations of the effects of different pricing rules. On the one hand, more renewable energy requires a larger amount of start-ups and cycling from marginal generators, so the impact of non-convexities on prices will tend to increase and the market will benefit from a refined approach to its computation. On the other hand, we can expect a shift in the investment in new merchant generators, moving from the predominance of near-marginal technologies that we have seen in the latter decades, which are more or less isolated from the pricing problem (see Section 3.1 for details), to a larger share of base-load renewable-based capacity, which bear a much larger impact of using one pricing mechanism or another.

In this paper we will not address the discussion of whether the regulator should adopt a complex or a simple auction, which we assume that depends on the conditions at each market. Also, we will restrict ourselves to a perfect competition situation, ignoring market power, in order to concentrate on the pricing issues linked to non-convexities.

The paper will first present the pricing problem in Section 2, using some simple examples to illustrate why price may not be defined when the only criterion considered is ensuring that generators agree with the centralized dispatch, while reviewing in light of this description the different proposals presented in the literature. Then, we will incorporate into the discussion the criterion of providing incentives for the investment decisions of other generators (Section 3), identifying some additional requirements for the prices. Section 4 will be devoted to discuss the implications of the results obtained, while Section 5 will sum up the conclusions.

2. Statement of the problem and literature review

2.1. A simple example with binary variables

Let us assume a single-hour problem, where demand is d and there are three generators $i=\{i_1, i_2, i_3\}$, with maximum output g_i^{max} and a cost function that only involves a start-up cost ca_i and a variable cost cv_i for each of the generators, being $ca_i < ca_{i+1}$ and $cv_i < cv_{i+1}$. This is a very simple example, but it keeps the essential feature of including binary variables, which are the source of the pricing problems under study. The optimization of the centralized problem is rather easy in this case: if demand is smaller than the capacity of the cheapest generator, then only this one should produce; if demand is higher than the maximum output of the cheapest generator, but smaller than the aggregate capacity of the first and second generators, then the first one should produce at its maximum and the rest of the demand should be provided by the second unit; if demand is larger than the capacity of the first two generators, then both of them should operate at their maximum and the rest of the demand should be produced by the third generator. It results in the curve, presented in Fig. 1, of total production cost as generation increases. The optimal solution is the point where generation is equal to demand.

In a perfect competition context, each generator decides its output by maximizing its income from market sales minus its operating cost. If

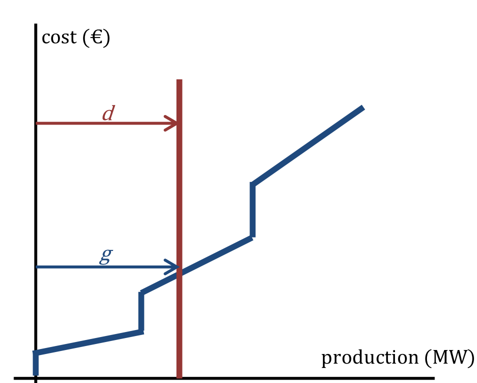

Fig. 1. Production costs as a function of production.

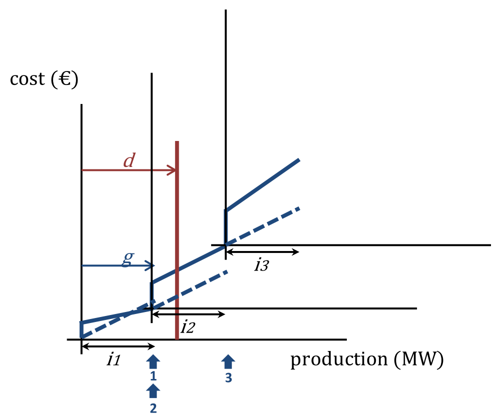

Fig. 2. Production decisions for price equal to the variable cost of the marginal unit (plant by plant view).

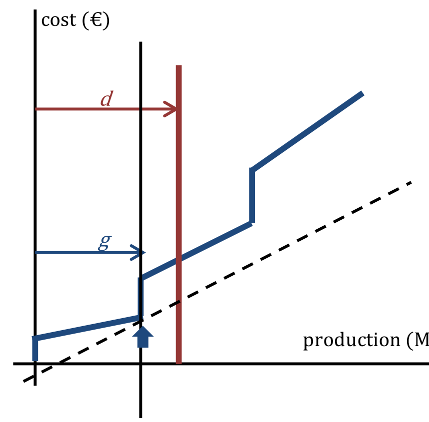

Fig. 3. Production decisions for price equal to the variable cost of the marginal unit (aggregated view).

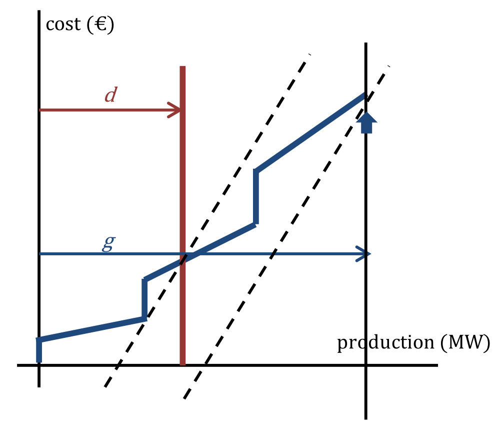

Fig. 4. Production decisions for price equal to the average cost of the marginal unit.

the prices were calculated from the direct application of the results of the convex case, see for instance (Bohn et al., 1984), then the price would be equal to the pure marginal cost: in this example, to the variable cost of unit i_2. The decision process of market players for this price is represented in Fig. 2, where the solid lines show their operating costs and the dotted ones represent their market income.

For generator i_1, operating margin is maximized when producing at its full capacity, since the price is higher than its marginal cost, and the higher the production the higher the net income. For generators i_2 and i_3, on the contrary, the best decision is not to produce at all, since the price is not enough to compensate for their full operating costs. Total generation under this price is lower than demand, and different from the centralized solution. Since the price is calculated as the pure marginal cost, any cost that is related to the discrete decisions is not incorporated into the prices, so the marginal unit does not recover its full operating costs, and therefore it is not willing to produce. This price does not attract enough generation to fulfill demand. Therefore, we can conclude that this price does not support the global optimum to be a competitive equilibrium.[1]

The same result could have been obtained by noticing that the net profits associated with any plant can be written as $R = \pi g - C$, where C represents the production costs, g is the plant output and π is the market price. Thus, the net profits can be recast as $C = -R + \pi g$, so they constitute a set of straight lines with slope π and intercept $-R$. The individual decision of any certain market player, given the market price, is to choose the straight line with a slope equal to the market price and with the lowest possible intercept (i.e., maximum possible operating profit), taking into account that only the lines with at least one point in common with the cost curve are feasible. Therefore, a straight line with slope equal to the market price and tangent to the cost curve represents the optimal response of the system generation portfolio to a certain price. The optimal production is given by the tangent point. In Fig. 3 this reasoning is applied to the previous case, market price equal to the pure marginal cost, with the dotted line being the lowest line which is tangent to the cost curve and has the same slope than the variable costs of generator i_2, and it yields the same results as in Fig. 2.

An alternative candidate for the market price is the average cost of the marginal plant, which should ensure that the price pays for all of the operating costs of this generator, and therefore it will be willing to come on line. This is represented in Fig. 4.

This price is in fact high enough to make generator i_2 willing to produce. However, i_2 does not maximize its profits by operating at the partial level determined by the global optimization, but rather by producing at its full capacity. Furthermore, the price makes plant i_3 willing to produce at its maximum too. Then, the total generation induced by this price is much larger than demand. The equilibrium implied is not the optimal dispatch either.

A possible compromise solution is represented in Fig. 5. In this case, the market price is defined as the average cost at its maximum output of the marginal plant.

Under this solution, all the units that were not dispatched in the optimal solution are not willing to produce, and no infra-marginal unit has incentives to stop producing. This means that, for all units except the marginal one, this price supports the global optimum as an equilibrium. However, the optimal response of the marginal plant is either to stop producing or to produce at its maximum output. Any generation level different than its maximum production would reduce the income of unit i_2, making it lower than its operating costs, and thus leading the plant to stop producing. The average-cost-at-maximum-output price does not create incentives for generator i_2 to produce at the level required to fulfill demand as in the optimal dispatch.

This example illustrates that often there is no price that supports the short-term competitive equilibrium in the market. Any price higher than the average-cost-at-maximum-output price will result in a total generation that is above demand; any price lower than that one will yield less generation than demand; and, for the average-cost-at-maximum-output price, generation can be either lower or higher than required, but not equal to demand.

2.2. Duality gap

The lack of short-term equilibrium is not restricted to this example, it is rather a general result of the non-convexities of the cost functions.

[1] We understand that a certain generation dispatch is a competitive short-term equilibrium under some market prices if, once the prices are known, generators are not willing to change their productions.

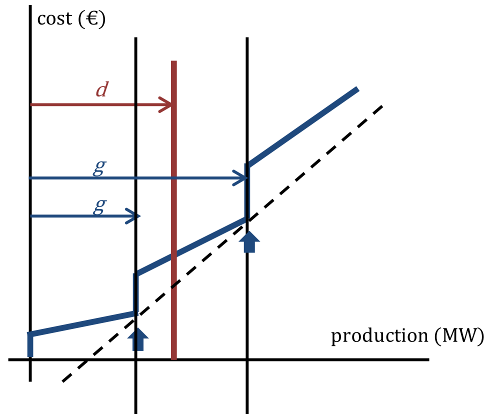

Fig. 5. Production decisions for price equal to the average cost at maximum output of the marginal unit.

One could describe the market as a process where a central problem (the day-ahead auction) sets the price and each producer resolves his own subproblem to determine his production. In other words, the centralized unit commitment problem is solved through a decomposition technique, allowing for a distributed optimization (each generator decides his own production) and leaving the price as a control variable that creates economic signals to ensure that the generation-demand balance is maintained. This kind of decomposition, where a certain constraint is replaced by a price signal, is known as Lagrangian relaxation, see for instance Minoux and Vajda (1986).

A common phenomenon when applying Lagrangian relaxation to non-convex problems is the existence of a duality gap. This is a difference between the solution of the primal problem (in our case, the centralized optimization) and the dual problem (in our case, the market), which is zero when the problem is convex but not in the presence of non-convexities. More precisely, in the areas where the problem is not convex, small changes in the price signal leads to large leaps in the solution of the dual problem, making it impossible to satisfy the relaxed equation (the supply-demand balance, in our case). As a result, the solution of the primal problem, which satisfy exactly all of the constraints, differs from the solution of the dual problem.

This is the case with the day-ahead electricity market. There is no price that creates incentives for the generators to produce the same amounts that were dispatched in the centralized model: in the market, they will produce either at their full capacity or zero. Thus, the solution of the market is different from the global optimum. In the presence of non-convexities, there is no price that supports a short-term competitive equilibrium.

It is often mentioned, e.g. (Ferreira, 1993), that the duality gap decreases with the size of the problem, being negligible for large-scale optimizations. This is true, in relative terms, when total supply cost is considered: the value of the duality gap for the marginal unit remains unchanged but, as long as the system is larger, the number of generators at full load, with no duality gap, increases and, when compared with total cost, the value of the total duality gap is relatively smaller. However, it does not hold for the revenue streams of generators in a single-price market: the duality gap for the marginal generator remains unchanged, and it is translated into an error in the price computation, which is in turn paid to all producers. In a market, the problems linked to duality gap do not reduce with the size of the problem.

2.3. Side payments

If there is no price that supports the competitive equilibrium, then the only option for identifying a reasonable price-setting mechanism is to move away from the single-price scheme and try to find an alternative approach to deal with price computation. Most of the proposals in the literature assume, with more or less emphasis, that generators receive their short-term income from two sources: on the one hand, an energy price, paid uniformly to all of the generators that are producing at a certain hour, and calculated as some form of marginal cost (details vary depending on the solution); on the other hand, a side-payment, different for each generator, and often restricted only to a few of them, or even only to the marginal one.

Making use of side payments, the average-cost-at-maximum-output price solution described above could be transformed into a short-term competitive equilibrium solution just by giving to the marginal unit an extra payment. This additional income would be equal to the portion of its fixed costs corresponding to the capacity that is actually not producing. Being δ_i the side payment received by producer i, in the example of Section 2.1, $\delta_i = ca_i \cdot \left(1 - \frac{g_i}{g_i^{max}}\right)$ for $i = i_2$ and $\delta_i = 0$ for $i = i_1, i_3$. Note that δ_i is not fixed, but changes with g_i. For this new pricing scheme with side payments, the income received by generator i_2 is equal to its production costs regardless of its output, so the generator is indifferent to which is its production, being willing to sell any amount of generation, including the mid-range level that is required to meet demand in the most efficient way. Therefore, these set of prices are compatible with the competitive equilibrium.

Nevertheless, one could construct an equivalent argument using the variable-cost prices described in Section 2.1. In this case, if the marginal generator received as side payments its full fixed costs, it will be willing to produce at any output level, including the one that is required to meet demand. Additional side payments would be required to make sure that the rest of the infra-marginal generators receive their fixed cost and are willing to produce. In the example of Section 2.1, $\delta_i = ca_i$ for $i = i_2$ and $\delta_i = \max\{0, (ca_i - (\pi - cv_i) \cdot g_i)\}$ for $i = i_1, i_3$. This combination of price and side payments would also allow for a competitive equilibrium.

Furthermore, even the average-cost price solution could support a competitive equilibrium with the correct side payments. If the non-dispatched generators received negative side payments that discourage them from producing, the incentives for all of the players would be compatible with the optimal dispatch. In the example in Section 2.1, $\delta_i = 0$ for $i = i_1$ and $\delta_i = (cv_i \cdot g_i^{max} + ca_i - \pi \cdot g_i^{max})$ for $i = i_2, i_3$. Side payments for generator i_2 are in principle null, but could become negative if its production were increased.

Taking this argument to the extreme, if the price were zero, one could devise side payments equal to the full production costs of each of the generators accepted in the auction, and all of them would be willing, under this pricing scheme, to produce the generation required to meet demand, so the overall pricing scheme will support the competitive equilibrium too. This is often known as pay-as-bid auction.

For any possible market price there is a set of side payments for the different generators that makes it compatible with a competitive equilibrium. In effect, under any given price π, it only takes one condition for each generator to make sure that the result of the centralized unit commitment is a competitive equilibrium.

$$\begin{aligned} &\pi \cdot g_i - cv_i \cdot g_i - ca_i + \delta_i \geq 0 && \text{if } g_i = g_i^{max} \\ &\pi \cdot g_i - cv_i \cdot g_i - ca_i + \delta_i = 0 && \text{if } 0 < g_i < g_i^{max} \\ &\pi \cdot g_i - cv_i \cdot g_i - ca_i + \delta_i \leq 0 && \text{if } g_i = 0 \end{aligned} \tag{1}$$

Being the costs parameters, the productions and the price known when calculating the side payment, there is one side payment for each generator, which can be freely adjusted to satisfy the conditions in (1). This is a system of equations with the same number of equations than unknowns. It can be solved and side payments that comply with the conditions in (1) can be computed. Thus, a set of side payments exists that makes any price support a competitive equilibrium for the cost-minimizing productions.

A reasoning based just on the short-term competitive equilibrium criterion yields mixed results. Without side payments there were no price that allowed for the productions of the unit commitment to be a competitive equilibrium, but when side payments are considered then any price fulfills the competitive equilibrium condition and apparently any pricing computation would be correct. Although it helps in achieving an equilibrium, it provides no information regarding which is the most suitable mechanism to compute prices. In Section 3 we would incorporate an additional criterion, with the aim of discriminating among the different solutions.

2.4. Related literature

A relevant piece of related work is (O'Neill et al., 2005), which proposes solutions that are similar to the scheme presented previously of a variable-cost-based price plus a side payment. The motivation of their work is to address the fact that no linear price supports an equilibrium in presence of non-convexities. They use the format of their mixed-integer optimization algorithm, along the lines of (Gomory and Baumol, 1960), to define a pure marginal price -which by construction is the variable cost of the marginal unit- and a set of side payments, derived from the conditions used by the algorithm to attain integrality. Those side payments support the equilibrium, ensuring that all dispatched generators recover their full operating costs. Other proposal relying on the same motivation, i.e. imposing conditions to ensure operatinal-cost recovery, is (Bjørndal and Jörnsten, 2008), where Benders cuts are used to define more stable prices.

On the other hand, (Hogan and Ring, 2003) propose to compute prices with the objective of minimizing side payments (uplifts, in their naming). The main motivation for this proposal is an auction-design argument: since pay-as-bid auctions are not desirable for power markets, because of their bad bidding incentives, the price adopted should be as close as possible to the uniform-price auction solution, and that implies minimizing uplifts. Gribik et al. (2007) shows that this uplift minimization is equivalent to computing prices as the dual variables of a Lagrangian relaxation. A recent analysis of the properties of this proposal can be found in Schiro et al. (2015). In the context of the examples of Section 2.1, this result is similar to the scheme of an average-cost-at-maximum-output price plus a side payment. Additionally, both Hogan and Ring (2003) and Gribik et al. (2007) discuss an alternative approach motivated by actual operations at the New York ISO, which they name dispatchable model, where binary variables are relaxed in order to allow for the costs associated with them to be reflected in marginal prices. While Hogan (2014) considers this solution as a workable approximation of the Lagrangian relaxation results, (Gribik et al., 2007) show some differences for a small example.

Several alternative proposals argue along the lines of uplift minimization, or some similar kind of optimization that explicitly considers side payments into their objective function. Alternative formulations of this kind can be found in Galiana et al. (2003), Bouffard and Galiana (2005), Toczylowski and Zoltowska (2009), Van Vyve (2011), Andrianesis et al. (2013), and Liberopoulos and Andrianesis (2016). Some of them add additional criteria, such as zero-sum side payments, that possibly move them away from the Lagrangian relaxation results. For instance, (Van Vyve, 2011) assumes that only already committed units are relevant for price calculation, which may distort prices at some cases. All of these proposals use different specifications and algorithms but they all share the characteristic of explicitly manipulating side payments under a general goal of reducing them.

Finally, there are a number of works that face the pricing issue from a numerical simulation perspective. de Sisternes et al. (2015) compare a single-price scenario (no side payments) with a centralized unit commitment with side payments. They find out that total costs for consumers in the single-price scenario are higher than in the centralized one. In the context of Section 2.1, this can be interpreted as average-cost-at-maximum-output price being lower than average-cost price: when there are no side payments, producers often calculate their bids as average costs at their estimated production point, while the centralized auctioneer in de Sisternes et al. (2015) calculates them as average costs at maximum output. Herrero et al. (2015) simulate investment decisions for several pricing models. They conclude that none of the pricing mechanisms under their scrutiny is able to induce optimal investment decisions at their simulations. However, since their approach is basically numerical, it is not easy to explain or further analyze which are the reasons for these results.

Summing up, the vast majority of the proposals in the literature are concerned with ensuring that generators are willing to produce according to the instructions of the auctioneer. There are several alternatives that are well suited to this goal, but the arguments presented to choose any of them among the others are axiomatic solutions to address the cost-recovery problem. Our paper focus on including additional criteria for this decision, stressing the role of prices as long-term signals.

3. Investment in new generation facilities

3.1. Solution concept

Looking into the example in Section 2.1, we can observe that the income received by the marginal generator does not change across the different pricing methods. Both the proposal in O'Neil et al., (2005) and the convex hull pricing in Gribik et al. (2007) provide the same total remuneration for the marginal plant, even though the prices and side payments of both methods are different. In fact, this is necessary if the pricing scheme has to be a competitive equilibrium: if the remuneration for the marginal generator were lower than its operation costs, then the generator would decide to stop producing, deviating from the equilibrium production; if its remuneration were higher, then the generator would decide to produce at its maximum output, which is not the equilibrium production either. Thus, all of the pricing schemes are equivalent from the point of view of the marginal unit, and the main difference among them is how much they pay to the infra-marginal units (i.e., those generators with bids that are cheaper than the marginal plant, which are producing at their maximum output).

Presented in these terms, the pricing problem is reduced to determining which part of its operational costs is paid to the marginal generator in the form of price, which all other generators receive, and which part of that cost is paid to it in the form of side payments, which are discriminatory and typically most of the infra-marginal producers will not receive. Thinking just in terms of minimizing consumer's payments at the short-term market, one could argue that the larger the side payment, the lower the price and the lower the costs for consumers, but this reasoning would not be correct, since it ignores the impact of price on investment decisions, and hence on future costs. The standard line of argumentation, see for instance (Caramanis, 1982), to explain why infra-marginal generators should receive a price that is higher than its bids and higher than its operational costs is that the net income captured by the generators in the market is required to remunerate for their investment costs. In other words, the difference between the price and the operational costs of the producers represents the market income for each plant, which is in general required to pay for the capital costs of constructing the facility. If prices were close to the actual operational costs for all of the generators, then no plants with high investment costs -typically base-load equipment- could be financed and new capacity additions would be based on low-investment-cost generators, such as peaking plants. If prices were substantially higher than operation costs, then new additions would be predominantly base-load generators.

The need to induce the correct investment decisions in generation equipment is thus an additional criterion that can be used to choose among different pricing schemes (Vazquez, 2003). We will reformulate the question of which is the pricing mechanism that should be used as which is the pricing mechanism that provides better long-term signals. For this analysis, we will use the approach of comparing the results of the market with the results of an ideal centralized optimization with perfect information, see for instance Bohn et al. (1984) or Caramanis (1982). We will describe the ideal aggregated investment-plus-operation optimization problem, and then we will extrapolate to the market environment. The Appendix provides further details on the formal derivation of these results.

3.2. Investment-plus-operation model

The starting point for this kind of analysis (see for instance Caramanis (1982) or Joskow and Schmalensee (1983)) is the ideal optimization model. A perfectly-adapted generation mix is defined to this end, which consists in the set of production facilities that provides least-cost supply, computed as if there were no existing generation plants; i.e., generation capacity in the system is planned from scratch, ignoring the present situation of the generation equipment. If the installed capacity of any certain technology were lower than its capacity at the ideal model, then that technology would receive more income from the spot prices than required to pay for its capital costs. Therefore, it would become very attractive for investors and more capacity of this kind would be constructed. Reversely, if actual capacity for some technology were higher than the ideal capacity, it would not recover its full costs. Thus, no more capacity would be built of this technology, until demand grows and the capacity of this kind results to be adapted again. Through this mechanism, prices act as long-term signals for generation investment.

In order to compute the ideal mix, the optimization problem assumes that no size constraints apply to the generation facilities; i.e., the model decides to build plants of any size, as long as they are required to attain the least-cost solution. It does not mean that the centralized planning algorithm is unaware of the potential size limits for some equipment, or of the fact that many plants are already installed, with their investment costs sunk, but means that the optimization model is not computing a realistic expansion plan. Instead, the centralized algorithm is defining a reference ideal mix that the planner would like to have, and prices are constructed to provide incentives for the players to approach to this ideal mix. In this context, competitive market means that generators are paid according to their value to the system on an hour-by-hour basis. For instance, a certain plant that becomes un-adequate when a new technology appears will not be paid its full investment costs by the market, even if it was optimal when it was built. Equivalently, a production unit that has to build more megawatts than optimal due to the lumpy nature of its capacity additions will receive less remuneration than some other more flexible technology that is better adapted in size to the demand.

The assumption of no size limits, which has a minor impact for a convex model, becomes of capital importance when the unit commitment case is considered. We will assume that, for the perfectly-adapted generation mix, there is enough flexibility to build production equipment of any size. Again, not because it is realistic, but because it is the aspirational scenario that is being promoted. We will further discuss the implications of this assumption in Section 4 below.

Let us consider the following simplified description of the global optimization problem:

$$\begin{aligned} \min_{x,g,u,v,w} \; & \sum_i CI_i x_i + \sum_{t,i} (c_i g_{t,i} + cb_i x_i u_{t,i} + ca_i x_i v_{t,i}) \\ s.\ t.\; & \sum_i g_{t,i} = d_t \\ & u_{t,i} x_i g_i^{min} \le g_{t,i} \le u_{t,i} x_i g_i^{max} \\ & u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} \\ & u_{t,i} \in \{0,1\} \end{aligned} \tag{2}$$

where CI_i and x_i are the parameters that describe the investment decisions, being CI_i the investment cost of plant i ($/MW) and x_i its installed capacity (MW). c_i, cb_i and ca_i are the operational cost parameters of plant i: respectively, variable cost ($/MWh), per-MW no-load cost ($/h/MW) and per-MW start-up cost ($/start/MW). We assume shut-down cost is zero. Note that cb_i and ca_i are expressed in per-unit terms in order to account for their evolution when capacity changes. $g_{t,i}$, $u_{t,i}$, $v_{t,i}$ and $w_{t,i}$ describe the operation of plant i at time t: they are, respectively, its production (MW), commitment (1 if the plant is on-line, 0 if it is not), start-up (1 if the plant is starting-up at time t, 0 if it is not) and close-down (1 if the plant is closing-down at time t, 0 if it is not) decisions. Finally, g_i^{min} and g_i^{max} are the per-MW production limits of plant i (MW/MW), and d_t is system demand at time t. g_i^{min} and g_i^{max} are also expressed in per-unit terms so they can adapt to capacity changes. The third constraint represents a continuity constraint: if the generator is on-line at time t, then either it was on-line at time $t-1$ or it is started-up at time t; conversely, if it is not on-line at time t, then either it was off-line at time $t-1$ or it is shut-down at time t. Only the $u_{t,i}$ variables are explicitly forced to be binary, as the third equation combined with the costly nature of $v_{t,i}$ lead $v_{t,i}$ and $w_{t,i}$ to have only 0 or 1 values without adding any additional constraint to the model. Some other technical characteristics might be added to this model (ramp rates, etc.), which do not change substantially the conclusions of this simplified version.

Re-arranging terms, the previous problem can be expressed as

$$\begin{aligned} \min_{x,g,u,v,w} \; & \sum_i CI_i x_i + \sum_{t,i} (c_i g_{t,i} + cb_i u_{t,i} + ca_i v_{t,i}) \\ s.\ t.\; & \sum_i g_{t,i} = d_t \\ & u_{t,i} g_i^{min} \le g_{t,i} \le u_{t,i} g_i^{max} \\ & u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} \\ & u_{t,i} \in \{0, x_i\} \end{aligned} \tag{3}$$

Moreover, if x_i can be as flexible as required, it is always possible to substitute any generation plant by several smaller units. For instance, consider that a certain marginal plant is producing at 60% of its maximum capacity in one hour and at full capacity in the next hour. It is always possible to replace it by two smaller plants of the same technology, with installed capacities equal to 60% and 40% of the capacity of the first generator, so one of them is producing at its maximum in the first hour and both of them produce at full load in the second one. This implies that problem (3) is equivalent to a linear problem: as we are making x_i arbitrarily flexible, the optimal solution will only contain variables at their integer limits. This is the definition of the integrality property, so the solution of the linear relaxation is the same as the integer solution, (Minoux and Vajda, 1986).

$$\begin{aligned} \min_{x,g,u,v,w} \; & \sum_i CI_i x_i + \sum_{t,i} (c_i g_{t,i} + cb_i u_{t,i} + ca_i v_{t,i}) & \\ s.\ t.\; & \sum_i g_{t,i} = d_t & \lambda_t^{dem} \\ & u_{t,i} g_i^{min} \le g_{t,i} \le u_{t,i} g_i^{max} & \lambda_{t,i}^{min}, \lambda_{t,i}^{max} \\ & u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} & \lambda_{t,i}^{online} \\ & 0 \le u_{t,i} \le x_i & \lambda_{t,i}^{umin}, \lambda_{t,i}^{umax} \end{aligned} \tag{4}$$

In a perfectly-competitive market environment with a complex auction, with the investment assumptions described above, the results of problem (4) can be achieved by: i) clearing the market by solving the optimization model described in (5); ii) adopting λ_t^{dem} as the price for each hour t; and iii) allowing market players to freely decide by themselves on their installed capacities x_i. The Appendix provides a more formal derivation for this result.

$$\begin{aligned}
&\min_{g,u,v,w} \sum_{t,i} (c_i g_{t,i} + cb_i x_i^* u_{t,i} + ca_i x_i^* v_{t,i}) \\
&s.\ t.\ \sum_i g_{t,i} = d_t && \lambda_t^{dem} \\
&\quad u_{t,i} x_i^* g_i^{min} \le g_{t,i} \le u_{t,i} x_i^* g_i^{max} && \lambda_{t,i}^{min},\ \lambda_{t,i}^{max} \\
&\quad u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} && \lambda_{t,i}^{online} \\
&\quad 0 \le u_{t,i} \le 1 \lambda_{t,i}^{umin}, && \lambda_{t,i}^{umax}
\end{aligned} \tag{5}$$

where x_i^* is the actual set of generation plants existing in the system at the moment of clearing the market.

Prices resulting from problem (5) provide incentives for the producers to build the perfectly-adapted and completely-flexible generation mix that results from problem (4). In practice, real generators will not be able to reach this ideal generation portfolio, but they will have incentives to approach to it as much as they can. Any certain technology that manages to have more flexible sizing options, so it can come closer to the perfectly-adapted capacity mix, will have an advantage in the market. Thus, even if the scalable-generators assumption does not hold, the prices obtained using it are good incentives for capacity expansion, leading the generation mix to be better suited to demand.

Therefore, leveraging on the idea of flexible capacity additions, we have moved from a unit commitment problem with binary variables to a linear problem with a convex cost function. Start-up costs, which are fixed in the model with binary variables and thus are not reflected in prices, evolve in a continuous way in the linear model, which allows for the computation of a marginal cost that reflects the costs related to start-up variables and, in general, with any binary variables.

3.3. Implementation proposal

Therefore, we propose to implement the day-ahead market auction as a three-steps procedure:

1. Solve the optimization problem (6) (the scalable-generators linear problem), which is a particular linearization of the unit commitment problem and was presented as problem (5) above, and obtain market prices. $p_t = \lambda_t^{dem^L}$, where p_t is the market price in hour t and the super-index L refers to this linear problem.

$$\begin{aligned}
&\min_{g,u,v,w} \sum_{t,i} (c_i g_{t,i} + cb_i x_i^* u_{t,i} + ca_i x_i^* v_{t,i}) \\
&s.\ t.\ \sum_i g_{t,i} = d_t && \lambda_t^{dem} \\
&\quad u_{t,i} x_i^* g_i^{min} \le g_{t,i} \le u_{t,i} x_i^* g_i^{max} && \lambda_{t,i}^{min},\ \lambda_{t,i}^{max} \\
&\quad u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} && \lambda_{t,i}^{online} \\
&\quad 0 \le u_{t,i} \le 1 && \lambda_{t,i}^{umin},\ \lambda_{t,i}^{umax}
\end{aligned} \tag{6}$$

2. Solve the optimization problem (7) (the unit commitment problem), which is a standard unit commitment with binary variables, and obtain the quantities that are assigned to each generator in each hour, together with their production costs. $q_{t,i} = g_{t,i}^{UC}$; $C_i = \sum_t (c_i g_{t,i} + cb_i x_i^* u_{t,i} + ca_i x_i^* v_{t,i})^{UC}$, where $q_{t,i}$ is the quantity assigned to generator i in hour t, C_i is the production cost of generator i, and the super-index ^{UC} refers to this unit commitment problem.

Table 1
Demand data.

Time t	t1	t2	t3	t4	t5	t6	t7	t8
Demand d_t (MW)	1000	800	650	200	700	600	500	400
Willingness to pay ($/MWh)	500							

$$\begin{aligned}
&\min_{g,u,v,w} \sum_{t,i} (c_i g_{t,i} + cb_i x_i^* u_{t,i} + ca_i x_i^* v_{t,i}) \\
&s.\ t.\ \sum_i g_{t,i} = d_t && \lambda_t^{dem} \\
&\quad u_{t,i} x_i^* g_i^{min} \le g_{t,i} \le u_{t,i} x_i^* g_i^{max} && \lambda_{t,i}^{min},\ \lambda_{t,i}^{max} \\
&\quad u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} && \lambda_{t,i}^{online} \\
&\quad u_{t,i} \in \{0, 1\}
\end{aligned} \tag{7}$$

3. Compute a side payment for one (or a few) near-marginal generators as the difference between its total operational costs C_i and its total market revenue $\sum_t (p_t \cdot q_{t,i})$. The generators will only receive this side payment when it is a positive value.

Typically a linear version of the unit commitment is obtained as an initial solution for the discrete optimization model, so this process should not imply a significant additional cost in terms of computational effort. Probably, it will be substantially easier to compute than the relaxation proposed in (Gribik et al., 2007).

It is relevant to note that not every linear problem that derives from a unit commitment is suitable for computing the prices we are describing. Throughout Section 3 we have developed an approach to define this pricing model, which is quite restrictive, and which is based on the idea of flexible additions of capacity. This means that when the optimization results in a commitment variable that is, for instance, at 60%, we are assuming that there is a generator that is fully committed with an installed capacity of 60% the capacity of the original generator. And this new generator needs to have the same technical characteristics as the previous larger one, all of them scaled down accordingly. A generator committed at 60% requires to have 60% maximum output, but also 60% minimum stable load, 60% ramp rate, 60% no-load cost, etc. It is the entire generator, with all of its features, what we are scaling up or down. Only if this notion of scalable generators is properly modelled in the pricing algorithm, the resulting prices will display the properties we are describing as optimal long-term investment signals.

Obviously, the formulation in problem (6) does fulfill this requirement. If additional constraints were added to the model, such as ramp rates, or minimum on-line time, or others, one should be careful to ensure that the new constraints are modelled consistently with scaling up or down the size of the generator. For instance, one should make sure that the modelling of ramp-rates works well if the installed capacity is half the initial one.

Besides, under the process we are describing, it may happen that some generators that are not dispatched in the 'quantities' model (the unit commitment problem) are dispatched in the 'prices' model (the scalable-generators linear problem). Typically, a plant with low average production cost but large in size, may be replaced in the 'quantities' model by some smaller, although more expensive, generators. In our view, this should not be a cause of concern: the smaller but expensive

Table 2
Generation technologies data.

	Variable cost cv_i ($/MWh)	No-load cost cb_i ($/h/MW)	Start-up cost ca_i ($/start/MW)	Investment cost CI_i ($/MW)	Max generation (MW/MW)	Min generation (MW/MW)
$i1$	60	0	70	809	1	0.4
$i2$	20	0	60	1050	1	0.4

Table 3
Dispatch and costs for the global solution.

	Installed capacity (MW)	t1	t2	t3	t4	t5	t6	t7	t8
$g_{t,i1}$ (MW)	300	300	300	150	0	200	100	0	0
$g_{t,i2}$ (MW)	500	500	500	500	200	500	500	500	400
$v_{t,i1}$ (MW)		300	0	0	0	200	0	0	0
$v_{t,i2}$ (MW)		500	0	0	0	0	0	0	0
Operation/total cost $i1$ ($)		98,000/340,700							
Operation/total cost $i2$ ($)		102,000/640,000							

plants will receive a side payment to cover for their fixed costs, the large plant will have an incentive to be more flexible, while the infra-marginal generators will receive a price that is based on the aspirational expansion plan, where the size of the cheaper unit is better adapted to demand. See further discussion in Section 4 below.

3.4. Illustrative example

An example might be useful to gain insight into the effects of prices on the generation mix. Let us assume we have the following demand (Table 1) and the following two generation technologies: $i1$ and $i2$ (Table 2). We will be using the names defined in problem (2).

For these data, the global optimum; i.e., the results that minimize total operation-plus-investment costs are presented in Table 3.

Note that there is a rationing of 200 MW in the first hour. This is the optimal decision taking into account the willingness to pay defined in our example. That is, we are defining that is optimal to ration in one hour. The decision of building more capacity from technology $i1$ instead of having 200 MW of rationing is related with the adequacy topic that we are not covering in this paper. The decision between technologies $i1$ and $i2$ depends critically on the definition of the prices. This latter decision is the focus of this work. In this regard, our simple example also shows that the adequacy problem and the price computation problem are different.

Using the results from the cost-minimization problem, let us analyze the impact of different pricing alternatives on the investment incentives. We will start by testing the "price equal to the variable cost of the marginal unit" assumption.

3.4.1. Price equal to the variable cost of the marginal unit

For the cases where the generator is at its maximum capacity, so the marginal cost is undefined, we use the convention of choosing the price of the next unit. When we set prices to the variable cost of the marginal unit, we obtain the market results in Table 4.

Table 4
Economic results for the "price equal to the variable cost of the marginal unit" pricing rule.

	t1	t2	t3	t4	t5	t6	t7	t8
Price ($/MWh)	500	500	60	0	60	60	20	20
Side payment $i1$ ($/MW)	0	0	0	0	70	0	0	0
Side payment $i2$ ($/MW)	0	0	0	20	0	0	0	0
Total revenue $i1$ ($)	348,000							
Total revenue $i2$ ($)	612,000							

By comparing Tables 3,4, we observe that market results imply that generator $i2$ is under-remunerated in this scheme (it is receiving 612 k$ while its total operation+investment costs are 640 k$). From an investment point of view, these market results mean that investments in technology $i2$ are less profitable tan required. Consequently, market players would tend to build less capacity of this kind, which in turn implies that the generation mix would adapt in response to the economic signals contained in the prices.

The next step thus is to analyze the generation mix induced by the price signals defined by market prices. The new generation mix that arises with those prices is shown in Table 5.

Table 5
Dispatch and costs of the generation mix that is adapted to the "price equal to the variable cost of the marginal unit" pricing rule.

	Installed capacity (MW)	t1	t2	t3	t4	t5	t6	t7	t8
$g_{t,i1}$ (MW)	500	500	500	350	200	400	300	200	200
$g_{t,i2}$ (MW)	300	300	300	300	0	300	300	300	200
$v_{t,i1}$ (MW)		500	0	0	0	0	0	0	0
$v_{t,i2}$ (MW)		300	0	0	0	300	0	0	0
Operation/total cost $i1$ ($)		194,060/598560							
Operation/total cost $i2$ ($)		76,000/391,000							

As before, we calculate the market results corresponding to the technology mix in Table 6.

Table 6
Market results for the "price equal to the variable cost of the marginal unit" pricing rule and the technology mix adapted to it.

	t1	t2	t3	t4	t5	t6	t7	t8
Price ($/MWh)	500	500	60	0	60	60	60	20
Side payment $i1$ ($/MW)	0	0	0	60	0	0	0	0
Side payment $i2$ ($/MW)	0	0	0	0	60	0	0	0
Total revenue $i1$ ($)	609,000							
Total revenue $i2$ ($)	394,000							

Note that the prices corresponding to the new technology mix are the same but side payments change. Generator $i2$ has decreased its installed capacity as a response to prices. This allows her to capture higher prices and compensate for the previous under-remuneration. Both generators receive their full costs.[2] We observe that total costs in Table 6 are higher than total costs in the global solution (Table 3). This means that the generation mix that has been induced by the "price equal to the variable cost of the marginal unit" pricing rule is not optimal, and is more expensive for consumers than the ideal one. The modifications in the technology mix induced by these prices have

[2] If our example had a larger number of hours, total revenues in Table 5 would be equal to total costs in Table 6; small errors remain due to the simple nature of the example.

Table 7
Market results for the pricing rule proposed in this paper.

	t1	t2	t3	t4	t5	t6	t7	t8
Price ($/MWh)	500	500	60	−115	130	60	60	20
Side payment $i1$ ($)	7000							
Side payment $i2$ ($)	0							
Total revenue $i1$ ($)	348,000							
Total revenue $i2$ ($)	640,000							

increased total costs; this pricing rule is failing to minimize costs and should not be adopted. These differences in the investments decisions come from the fact that the global optimization is detecting that an additional megawatt from technology $i2$ reduces the expenses in variable costs from the marginal unit, but also part of the start-up cost of other generators. With the "price equal to the variable cost of the marginal unit" pricing rule, only the savings in variable costs are considered for the investment decisions, as the rest of the costs are included in the side-payments and does not have an influence in the remuneration of the base-load generators.

3.4.2. *Price equal to the price proposed in this paper*

In order to understand the basic characteristics of the price proposal developed in this paper, let us consider again the technology mix obtained from the global optimization described in Table 3. Using that technology mix, we calculate the market results corresponding to a price definition as the one proposed in Section 3.3, which are shown in Table 7.

Comparing to the costs in Table 3, both generators recover their investment costs, so neither of them regret their investment decisions.[3] This set of prices induces generators to build the ideal capacity mix and is therefore optimal from the investment point of view.

Comparing to the prices in Table 6, one of the most relevant changes is the inclusion of the start-up cost of generator $i1$ in the price of hour $t5$. Since the plant is starting-up in that hour, and commitment variable are continuous, the costs of the start-up is included into the marginal cost. This allows for an additional remuneration for plant $i2$, as the start-up cost moves from the side payment to the marginal price. And this is a price signal for the baseload generator, pointing out that an additional investment in technology $i2$ would also save the start-up costs of $i1$. Besides, the price in $t4$ is modified to reflect the willingness of generator $i1$ to avoid shutting down during that hour. In effect, if there were an additional megawatt of demand in hour $t4$, generator $i1$ could produce one megawatt of its minimum stable load, which means committing 1/0.4 MW, and thus saving 175 $/MW in avoided start-ups during the following hour. Its total savings would be lower, since it would have to spend 60 $/MWh in variable costs, so she finally is willing to pay 175-60=115 $/MWh for producing in hour $t4$. This is the marginal cost that sets the price in this hour (Table 7).

4. Discussion

We have moved from a unit commitment problem with binary variables, where prices could not be computed in a straight-forward way, to a linear problem with a convex cost function, where prices are easy to compute; and we have been able to do so by leveraging on the idea of flexible capacity additions. We are assuming that in the long run it is possible to add capacity of any size. Start-up costs, which are quasi-fixed if the plant is already built, evolve in a continuous way as installed capacity increases, and this feature allows us to compute a marginal cost that reflects the costs related to start-up variables and, in general, with binary variables.

[3] Again, if our example had a larger number of hours, total revenues would be exactly equal to total costs.

Intuitively, having start-up costs (and other similar costs related to binary variables) included into the price is required in order to have correct expansion decisions in base-load generation. Otherwise, the net income perceived in the market by base-load generators -which would not include the start-up costs of the marginal generator- would be lower that the savings in operational costs that the existence of those base-load plants brings to the system throughout the year -which does include to some extent start-up costs-, and therefore base-load generation would be underinvested.

Among the various potential alternative pricing schemes that somehow include start-up costs in their prices, we are proposing a particular one, constructed upon the scalable generators assumption. We are not implying that continuous capacity additions are actually feasible, but we are considering that the scenario with all of the capacities perfectly adapted to demand -i.e., plants of any size are built if required to achieve least-cost supply- is our reference scenario. It is not a realistic expansion plan; it is an ideal one intended to be used as an investment signal. Our pricing proposal is thus devised to provide long-term signals that encourage investors, when deciding their capacity additions, to approach as much as possible to this reference generation capacity mix. This is an extra feature, when compared with most pricing options in the literature, which do not care about investment, and is in line with the assumptions used in the standard case with no binary variables, see Caramanis (1982) for instance. In effect, investment incentives in the standard model are also based on a perfectly-adapted capacity mix, to which investors try to approach, which is a non-realistic aspirational scenario; among other idealizations, it ignores that many generators are already in operation.

Comparing our findings with the proposals in the literature, a first conclusion is that the alternatives in the line of O'Neill et al. (2005) are not good investment signals, since their prices do not include any cost related to binary variables. They will tend to lead to underinvestment in base-load generators.

Formally, our proposal is similar to other linear models, such as the dispatchable model of Gribik et al. (2007). However, one of the implications of the reasoning we have used to derive our pricing scheme, as a long-term signal, is that it provides us with a very precise way of relaxing the integrality conditions of the binary variables in the unit commitment. The scalable generators assumption, which emanates directly from the long-term analysis, implies that all of the technical characteristics of any certain generator have to be scaled up or down when a partial commitment decision is determined. Other linear problems that are not defined under this condition may fail to reflect the impact of linear commitment variables in the overall behavior of the plants and thus may result in prices that are substantially different from our proposal. These kind of issues might explain some of the results described by Gribik et al. (2007) about their dispatchable model.

In fact, the prices we are proposing are closer to those of the convex hull in Gribik et al. (2007). Hogan (2014) points out that the dispatchable model yields prices that are workable approximations to the convex hull prices. Our intuition is that both prices would be identical if our scalable-generators linear problem were used instead of the dispatchable model that (Gribik et al., 2007) are considering. Though we have failed to find an example of a dispatch situation where the two solutions differ, we do not have a formal proof yet. In any case, our results tend to reinforce the convex hull approach, both by providing a stronger motivation -the idea of providing optimal long term signals is in our view more solid that just minimizing uplifts- and by defining a simpler method for price computation.

Finally, our discussion is connected with the critique in Scarf (1994) to the analysis of optimality based on marginal costs when there are discrete variables (indivisibilities in his naming). He points out that the result of the problem with binary variables is different from the result of the linear problem, so dispatch cannot be determined just by using the linear model. This is straight-forward in our case: some more expensive units may come on line if they fit better with demand size. Our two-steps process is intended to provide a solution for this situation; whenever (Scarf, 1994) critique holds, the price-setting generator will be different from the most expensive unit in operation. Using the unit commitment problem for dispatching ensures that the best solution is always adopted. On top of that, we add a pricing scheme, where the extra-costs derived from the size feature are not paid to all of the infra-marginal generators, but only to those plants that were called to produce due to their adequate size. In the Scarf (1994) example, where only the high tech plant should be constructed when demand is very high, our pricing proposal would result in a market price that is determined by the cost of the high tech plant, and if some smokestack plant were required, the additional income would be provided as a side payment.

5. Conclusions and policy implications

This paper discusses the question of which pricing mechanism should be used for day-ahead electricity auctions in the presence of binary variables. In general, the pricing methods in the literature are concerned with satisfying short-term competitive equilibrium conditions and ensuring that, once prices are known, all of the generators are willing to produce (or not) according to the volumes determined by the optimal dispatch. Our analysis shows that it is not possible to reach to a conclusion using only this short-term competitive equilibrium condition: if a pure marginal price is used, then there is no price that satisfies the equilibrium condition; but if an additional side payment is included, then any price could be a competitive equilibrium.

We propose to add a second criterion to help with the discussion: prices have to be also signals for generation expansion. Therefore, we propose to design prices with the objective of providing investors with the incentives to build new capacity in such a way that the generation equipment is as close as possible to the ideal perfectly-adapted capacity mix.

Using this condition, we arrive to a single solution to the pricing problem, where prices are computed as the shadow prices of the balance equations in a linear version of the unit commitment problem.

Importantly, not every linearization of the unit commitment provides the prices we are describing. In order to attain the optimal long-term signals, modelling of the linear model has to be done according to the notion of scalable generators, ensuring that all of the technical characteristics of the generators are scaled up or down when a partial commitment decision is determined, as if the model were using a generator that is completely equal to the original one but just smaller.

The pricing scheme proposed provides incentives to build new capacity that are sensible. It is, in our view, a reasonable approach to the problem of designing prices that are optimal long-term signals. The solution proposed is substantially different from the prices in O'Neill et al. (2005) and is close to the convex hull pricing of Gribik et al. (2007). Compared to the later, our results provide a stronger motivation for the pricing scheme (the idea of providing optimal long term signals) and a simpler method for price computation.

5.1. Implications for power market design

Short-run marginal cost is a central coordination mechanism. It defines the dispatch and remuneration of power plants, so it defines also long-run decisions. One of the problems where the study of short-run marginal costs is relevant is the design of day-ahead power auctions. In that context, there is a fundamental difference among market designs: who decides the short-run marginal cost. The question can be cast in terms of auction design: depending on the clearing methodology, the coordination responsibility, or equivalently the responsibility of defining the short-run marginal cost, rests with the auctioneer or with the auction participants. Most US systems rely on a short-run marginal cost defined by the auctioneer. Most of EU systems rely on power producers to define their own short-run marginal cost. Most Latin American countries rely on the system operator. From that viewpoint, this paper have been largely devoted to calculating short-run marginal costs in purely complex auctions, i.e. auctions where marginal costs are defined by the auctioneer. Nonetheless, the logic could be applied to sequential auctions with complex rules, as the EU ones. Besides, most EU markets have some kind of complex rules. The block bids adopted in NordPool and EEX or the minimum-income conditions of the Spanish pool are examples.

From a general point of view, the choice between complex and sequential auctions implies trade-offs. The immediate advantage of these mechanisms is that they capture the inter-relation of the different hours and eliminates the need for internalization. On the negative side, their complexity makes their results difficult to explain and this may raise some credibility problems. Besides, eliminating the need for internalization of power producers means that the associated risk is borne by the auctioneer, which is typically equivalent to being borne by consumers. The auctioneers have more information than individual power plants about the system costs. However, the auctioneer must decide before realization of demand evolution over the next hours. The trade-offs associated with each solution are not clear and need to be investigated carefully. In any case, the results obtained in this paper need to be taken into account in either case. In particular, including complex rules in sequential auctions, as we have shown, need to consider that prices will be redefined accordingly.

Moreover, the impact of short-run marginal costs on long-run decisions (investment) can be a direct or an indirect one. The indirect link between short- and long-term markets is established through the expectation on short-term market results, i.e., potential investors calculate their expected cash flows using short-term marginal costs. But there might be also a direct link, as the one present in markets based on long-term auctions. This is of significant relevance for markets including capacity mechanisms. An extreme case of this is the Brazilian power market, where the market is built around a long-term Power Purchase Agreement that is allocated through long-term auctions. The system operation (including generation assets) is the responsibility of a central operator. In that context, there is no short-run price to define. However, the need for a short-run marginal cost signal remains. In fact, in the Brazilian system, one of the components of the index that decides the auction winners (the ICB, Cost-Benefit Index in Portuguese) is precisely the CMO (Operation Marginal Cost in Portuguese). That operation marginal cost (actually, several scenarios of it) is calculated by the system operator and received by auction participants in order to prepare their bids for Power Purchase Agreements. Therefore, even in long-run-only markets the need for defining short-run marginal costs remains.[4] Therefore, the main objective of this paper is to contribute to the discussion of the economics of electricity short-run marginal costs. The importance of this definition can be justified from two viewpoints: it may be needed to calculate the auction price (if we consider the system marginal cost) or to place bids in an auction (if we consider a producer's portfolio marginal cost).

Acknowledgements

The ideas developed in this paper took shape within a research project sponsored by the CSEN (the former Electricity Regulatory Commission of Spain) in 1997–98. It has been partly funded by CAPES/BRASIL through the Science without Borders research program. Over the years, the authors have benefitted from many fruitful

[4] Currently, there is intense debate on the design of flexibility signals for gas-fired power plants in the Brazilian long-term auctions, as they are expected to play a greater role in the future of the industry. Such debate can be seen as the design of forward-looking flexibility signals. We will tackle this problem in the paper.

discussions, especially those with Michel Rivier, Ignacio Perez-Arriaga, and Alberto de Frutos. We also thank Dominique Finon, three anonymous reviewers, and the participants of the 26th USAEE/IAEE North American Conference in 2006, of the 23rd European Conference on Operational Research (EURO) in 2009, and of the 2015 CEEM/Dauphine conference, where preliminary versions of this work were presented. We are, naturally, responsible for all remaining errors.

Appendix A

In this appendix we show formally the relationship between the dual variable λ_t^{dem} and the market price. Specifically, we show that a market with a price equal to λ_t^{dem} has the same long-term incentives as a global cost-minimization problem under the same hypotheses.

To that end, we begin by considering the global optimization problem defined in problem (4), which is the critical element of our derivation of a mechanism for price computation. The optimality condition for x_i in problem (4) is $CI_i = \sum_t \lambda_{t,i}^{umax}$, assuming that generally $\lambda_{t,i}^{umin} = 0$. In other words, the model will keep investing in technology i until the savings in operational costs caused by the installation of one additional megawatt of technology i throughout the model horizon ($\sum_t \lambda_{t,i}^{umax}$) are equal to the investment cost of that additional unit of capacity (CI_i).

We can transform problem (4), using Benders decomposition, into the following two problems:

$$\min_x \quad \sum_i CI_i x_i + \theta(x_i) \tag{A1}$$

$$\begin{aligned}
\theta(x_i^*) \;=\; &\min_{g,u,v,w} \sum_{t,i} (c_i g_{t,i} + cb_i u_{t,i} + ca_i v_{t,i}) \\
&s.\,t.\ \sum_i g_{t,i} = d_t && \lambda_t^{dem} \\
&u_{t,i} g_i^{min} \le g_{t,i} \le u_{t,i} g_i^{max} && \lambda_{t,i}^{min},\ \lambda_{t,i}^{max} \\
&u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} && \lambda_{t,i}^{online} \\
&0 \le u_{t,i} \le x_i^* && \lambda_{t,i}^{umin},\ \lambda_{t,i}^{umax}
\end{aligned} \tag{A2}$$

where θ is the total operational cost of the system. For every set of generation plants (x_i^*), problem (A2) computes a value for the minimum operation cost that can be attained with that capacity mix. x_i^* is fixed in problem (A2). Then, the master problem (A1) decides the optimal values of x_i, considering investment costs and the impact of capacity decisions on the operational costs. The optimality condition for x_i in problem (A1) is $CI_i = \frac{\partial\theta}{\partial x_i} = \sum_t \lambda_{t,i}^{umax}$, the same as in problem (4). Therefore, the combination of problems (A1) and (A2) leads to the same investment solution as problem (4). This is also true for operational decisions.

Furthermore, we can operate in problem (A2) by relaxing the first constraint and dualizing it. Then, arranging some terms, problems (A1) and (A2) can be expressed as follows:

$$\max_x \quad \sum_i \theta(x_i) - CI_i x_i \tag{A3}$$

$$\begin{aligned}
\theta(x_i^*) = &\max_{g,u,v,w} \sum_{t,i} (\pi_t g_{t,i}) - \sum_{t,i} (c_i g_{t,i} + cb_i u_{t,i} + ca_i v_{t,i}) \\
&u_{t,i} g_i^{min} \le g_{t,i} \le u_{t,i} g_i^{max} && \lambda_{t,i}^{min},\ \lambda_{t,i}^{max} \\
&u_{t,i} = u_{t-1,i} + v_{t,i} - w_{t,i} && \lambda_{t,i}^{online} \\
&0 \le u_{t,i} \le x_i^* && \lambda_{t,i}^{umin},\ \lambda_{t,i}^{umax}
\end{aligned} \tag{A4}$$

Here θ is interpreted as the total operational profits of the generators. π_t is the market price at time t. These two problems (A3) and (A4) replicate in format the decision structure in a competitive market: generators determine their operational variables (production, start-up, etc.) with the aim of maximizing the net income they receive from the market (income minus cost), which is represented by problem (A4); on the other hand, firms decide their capacity by maximizing their expected revenue (market net revenue minus investment costs), which is represented by problem (A3). Both problems (A3) and (A4) could be decomposed into several smaller ones, one for each market player i.

Problem (A4) is only equivalent to problem (A2) if the price in problem (A4) π_t is equal to the shadow price of the balance equation in problem (A2) λ_t^{dem}. In effect, since problem (A4) is constructed by relaxing the balance equation in problem (A2), this condition is required for the relaxation to provide the same results as the non-relaxed problem (Minoux and Vajda, 1986). Therefore, if $\pi_t = \lambda_t^{dem}$ holds, then all of the shadow prices in (A4) are equal to those in (A2) and both problems are equivalent. As a consequence, if this condition held, the optimality for x_i in problem (A4) would be $CI_i = \frac{\partial\theta}{\partial x_i} = \sum_t \lambda_{t,i}^{umax}$, which would be the same as in problem (A1), which was in turn equivalent to that in (4). In other words, for the market to provide incentives for the generators to invest in such a way that the capacity mix approaches as much as possible to the reference mix defined in Section 3.2, prices have to be computed as the shadow prices of the balance constraints in problem (A2), λ_t^{dem}.

References

Andrianesis, Panagiotis, Liberopoulos, George, Kozanidis, George, Papalexopoulos, Alex D., 2013. Recovery mechanisms in day-ahead electricity markets with non-convexities-Part I: design and evaluation methodology. IEEE Trans. Power Syst. 28 (2), 960–968.

Bjørndal, Mette, Jörnsten, Kurt, 2008. Equilibrium prices supported by dual price functions in markets with non-convexities. Eur. J. Oper. Res. 190 (3), 768–789.

Bohn, Roger E., Caramanis, Michael C., Schweppe, Fred C., 1984. Optimal pricing in electrical networks over space and time. Rand J. Econ., 360–376.

Bouffard, Francois, Galiana, Francisco D., 2005. Generalized uplifts in pool-based electricity markets. In: Analysis, Control and Optimization of Complex Dynamic Systems, Springer, pp. 193–214.

Caramanis, Michael, 1982. Investment decisions and long-term planning under electricity spot pricing. IEEE Trans. Power Appar. Syst. 12, 4640–4648.

Ferreira, Luis A.F.M., 1993. On the Duality Gap for Thermal Unit Commitment Problems. In: Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS '93.

Finon, Dominique, Pignon, Virginie, 2008. electricity and long term capacity adequacy:

the quest for regulatory mechanism compatible with electricity market. UtilPolicy 16 (3), 143–158.
Galiana, Francisco D., Motto, Alexis L., Bouffard, François, 2003. Reconciling social welfare, agent profits, and consumer payments in electricity pools. IEEE Trans. Power Syst. 18 (2), 452–459.
Gomory, Ralph E., Baumol, William J., 1960. Integer programming and pricing. Econometrica, 521–550.
Gribik, Paul R., Hogan, William W., Pope, Susan L., 2007. Market-Clearing Electricity Prices and Energy Uplift. Working Paper, Harvard Electricity Policy Group.
Herrero, Ignacio, Rodilla, Pablo, Batlle, Carlos, 2015. Electricity market-clearing prices and investment incentives: the role of pricing rules. Energy Econ. 47, 42–51.
Benjamin F. Hobbs, Michael H. Rothkopf, Richard P. O'Neill, Hung-po Chao, editors. The Next Generation of Electric Power Unit Commitment Models, Kluwer Academic Publishers, Dordrecht, 2001.
Hogan, William W., 2014. Electricity market design and efficient pricing: applications for New England and beyond. Electr. J. 27 (7), 23–49.
Hogan, William W., Ring, Brendan J., 2003. On Minimum-Uplift Pricing for Electricity Markets. Working Paper, Harvard Electricity Policy Group.
Joskow, Paul L., Schmalensee, Richard, 1983. Markets for Power: an Analysis of Electrical Utility Deregulation. MIT Press, Cambridge, MA.
Kahn, Alfred E., 1970. The Economics of Regulation, Vol. 1: Economic Principles. Wiley, New York.
Liberopoulos, George, Andrianesis, Panagiotis, 2016. Critical review of pricing schemes in markets with non-convex costs. Oper. Res. 64 (1).
Minoux, Michel, Vajda, Steven, 1986. Mathematical Programming: Theory and Algorithms. Wiley, New York.
O'Neill, Richard P., Sotkiewicz, Paul M., Hobbs, Benjamin F., Rothkopf, Michael H., Stewart, William R., Jr., 2005. Efficient market-clearing prices in markets with nonconvexities. Eur. J. Oper. Res. 164 (1), 269–285.
Scarf, Herbert E., 1994. The allocation of resources in the presence of indivisibilities. J. Econ. Perspect. 8 (4), 111–128.
Schiro, Dane A., Zheng, Tongxin, Zhao, Feng, Litvinov, Eugene, 2015. Convex Hull Pricing in Electricity Markets: Formulation, Analysis, and Implementation Challenges. Working Paper. ISO New England.
de Sisternes, Fernando J., Webster, Mort D., Perez-Arriaga, Ignacio J., 2015. The impact of Bidding rules on electricity markets with intermittent renewables. IEEE Trans. Power Syst. 30 (3), 1603–1613.
Toczylowski, Eugeniusz, Zoltowska, Izabela, 2009. A new pricing scheme for a multi-period pool-based electricity auction. Eur. J. Oper. Res. 197 (3), 1051–1062.
Van Vyve, Mathieu, 2011. Linear Prices for Non-Convex Electricity Markets: Models and Algorithms. Working Paper. CORE (Center for Operations Research and Econometrics), Université Catholique de Louvain.
Vazquez, Carlos, 2003. Modelos de Casación de Ofertas en Mercados Eléctricos. (Auction Models for Electricity Markets, in Spanish) (Ph.D. thesis). Universidad Pontificia Comillas, Madrid.
Vazquez, Carlos, Rivier, Michel, Pérez-Arriaga, Ignacio J., 2002. A market approach to long-term security of supply. IEEE Trans. Power Syst. 17 (2), 349–357.
Wilson, Robert B., 1998. Activity rules for the California PX electricity auction. In: Proceedings of the 3rd Annual Research Conference on Electricity Industry Restructuring, University of California, Berkeley.