THE SYSTEM ADEQUACY PROBLEM: LESSONS LEARNED FROM THE AMERICAN CONTINENT

C. Batlle, P. Mastropietro, P. Rodilla, I.J. Pérez-Arriaga

A market-oriented level playing field is ideally supposed to be the most efficient way to expand the generation capacity and to guarantee the adequacy of the system. This belief has been a major driver behind the regulatory reforms of the electricity industry that have taken place during the last three decades.

However, in the American continent, the vast majority of the restructured power systems implemented from their outset some sort of capacity mechanism, thus complementing the short-term market. These mechanisms have undergone in the last decade a profound redesign towards new approaches, mainly based on auctioning long-term contracts of different nature. These same solutions are now in the centre of the regulatory debate in a good number of countries.

In this paper we perform an in-depth regulatory analysis of the design elements of the American new capacity mechanisms that allows us to develop sound guidelines and recommendations to support the design processes currently underway, not exclusively but especially, in the European Union context.

1 INTRODUCTION

The main reason behind liberalizing the power generation activity was to promote the economic efficiency at all levels, in the short-term (at the operation level), but especially in the long-term (at capacity expansion level). This belief was based on fundamental economic theory, which asserts that the short-term market marginal price is all that is needed to remunerate the generators to lead the system expansion towards the optimally adapted generation mix. This market mechanism to rule short-term operation was also seen as the needed level playing field to attract new investors. However, from the very first moment, ever since Chile restructured its power sector with its pioneering reform in 1982, the ability of short-term marginal prices to provide enough incentives for investors on generation to ensure
security of supply was called into question in a good number of countries in which the liberalisation was implemented. This was especially the case in the American continent: most Latin American countries (with the exception of Brazil) and some power systems in the United States (e.g. PJM), introduced in their original designs some sort of capacity mechanism (a capacity market or payment, or sometimes both), aimed to complement short-term marginal prices with a remuneration for the available capacity. On the other hand, the majority of European countries (once again with some exceptions, namely Spain and Ireland) followed the so-called “energy-only market” approach, refusing to implement any explicit capacity mechanism (although, as is discussed in Batlle and Rodilla (2010) the majority of them employed implicit and subtle regulatory safeguards regarding security of supply).

After all these years since the first market mechanism was implemented, there is sufficient evidence that, in the vast majority of cases, the theoretical premises under which the market alone would provide the optimal investment signal are unfortunately absent in practice, (Rodilla and Batlle, 2012). Nowadays, capacity mechanisms are in the minds of national regulators in almost all those countries where they were not included in the original market design. In particular, this issue is today at the core of the regulatory debate in Europe\(^1\). Italy and Portugal have recently redesigned their regulatory schemes in this direction and France - see Finon and Pignon (2008) and the so-called NOME law (JO, 2010)- and the United Kingdom (DECC, 2013) have officially announced the implementation of mechanisms of this nature and are currently in the process of developing the design details, while in Germany the Government has publicly admitted that it is evaluating the different alternatives (Cramton and Ockenfels, 2012).

---

\(^1\) In 2012 the European Commission launched a consultation paper on generation adequacy and capacity mechanisms (EC, 2012), in the framework of which the potential introduction of Capacity Remuneration Mechanisms (CRMs) was discussed. This resulted in 2013 in the publication of a working document on generation adequacy in the internal electricity market, providing guidance on public interventions of the Member States (EC, 2013).
On the other hand, those systems that originally considered a capacity mechanism underwent during the last decade a process of fine-tuning of their schemes. In North America, in 2008 PJM fixed a good number of flaws of its original capacity market during the implementation of the new Reliability Pricing Model, and a new capacity mechanism based on auctions (the so-called Forward Capacity Market) was implemented in New England.

However, it is in South America where most of the activity on this issue has taken place. Here, the fast-paced growth in electricity demand and the difficulties to mobilise financial resources has stressed the importance of properly designing a mechanism that attracts the adequate level of investment. At the beginning of the twentieth century, several South American countries suffered serious shortage situations, which were attributed to the deficiencies of the original market designs and in particular the initial capacity mechanisms, see Batlle et al. (2010). These circumstances resulted in the implementation of a second wave of reforms, which modified significantly the old schemes. All these new mechanisms intend to reduce the investors’ (price and regulatory) risk, through the auctioning of long-term contracts, in order to hedge at least part of their remuneration. Nevertheless, despite this apparent common scheme, these mechanisms present significant differences in terms of several relevant design aspects to be defined by the regulator. The proper identification and analysis of these design aspects, which will be referred to as design elements in the following, are two major objectives of this paper.

The first lesson that has to be extracted from an in-depth analysis of these differences is that the specifics in the design of these elements severely condition the outcome in the implementation of a capacity mechanism. The authors of this paper, after two decades of deeply involvement in the development of these instruments across the American continent, want to share some of the key regulatory lessons that have been learned in the process. This article does not describe the design of the long-term electricity auctions in North and South America country by country (a review of this kind is already available in literature, see Maurer and Barroso, 2011), but it rather identifies the main design elements of these mechanisms and defines guidelines for their determination, a subject which has not been covered in the existing literature to the best of the authors knowledge. This analysis should provide a valuable tool to
those regulators, in Europe and elsewhere, who are presently designing or thinking of introducing a capacity mechanism in their power sectors.

The paper is organised as follows. Section 2 identifies the main design elements of the auctions implemented in the American continent, studies the impact of their determination on the performance of these mechanisms and, based on this, provides guidelines for those systems that are willing to introduce similar schemes. Section 3 concludes the paper by summing up the high level recommendations.

2 DESIGN ELEMENTS: REGULATORY DISCUSSION AND GUIDELINES

The capacity auction mechanisms can be properly analyzed by decomposition and assessment of their several design elements (e.g. the target market, the level of centralisation, the lag period or the contract duration). However, also those design features that are often undervalued and considered as minor or secondary (as for example performance requirements, indexation, warranties, contract type or auction format) can significantly affect the final result of the mechanism.

These elements are not always independent and in any case their design should not be decided autonomously from the others. Therefore they must be seen as pieces of a puzzle, which must be modelled in a way that allows them to fit together so as to form a robust regulatory instrument that is as effective and efficient as possible. In this section, the design elements will be studied one by one, characterising the potential impact that each of the available choices might have on the auction performance and providing guidelines for their determination. Nonetheless, since auctions must be tailored first to the peculiarities of each power system and ultimately to the regulatory and policy objectives pursued, it will not be possible to provide general guidelines valid for every condition.
2.1 Target market

2.1.1 The buying side

When designing a long-term electricity supply auction, or more generally a capacity remuneration mechanism based on auctions, the first element which must be set is the demand to be involved in the transaction, i.e. defining who the buyer is or on whose behalf the regulator is buying. Roughly speaking, the main options are the following:

- Captive demand (regulated customers supplied by distributors or more generally regulated retailers).

- Free demand (free customers, usually large users, who are eligible to procure electricity independently).

- All the system demand.

An additional alternative in this last case is to allow demand to participate in the auction, allowing consumers to opt out depending on the clearing price. When implementing this solution, it is important to allow demand response bids only from those agents for which it is possible to guarantee that the demand reduction is effective (Chao, 2011). Penalties for underperformance, similar to the ones applied to generating units (later described in this paper) should also be applied. This approach has not been considered in South American mechanisms, but it has been implemented in the Forward Capacity Market of ISO New England, where demand response can bid in the auctions (ISO-NE, 2012), and in the PJM Reliability Pricing Model, where demand response is also accepted as capacity services provider (PJM, 2013). In these cases, capacity mechanisms have proven to be capable of involving demand response resources, even if they did not result in a full exploitation of demand response potential.

This design element is significantly affected by the regulatory objectives of the auction. For example, long-term electricity auctions, originally introduced in order to solve problems related to system adequacy and system expansion, are utilised in the South American context
to achieve a secondary objective: hedging the end-user default tariff price. This is the reason why in Brazil, Chile and Peru the auctions cover only the captive demand (though in the Brazilian case also free demand is required to cover 100% of its requirements through long-term contracts). In these countries distribution companies (the regulated retailers) are mandated to take part in the electricity auctions, so that they can also set stable default tariffs for their customers for a large period of time via these mechanisms.

Considering now only the original objective of the electricity auctions, the decision about who has to buy leads us to the following question: who has to pay for system adequacy? Since long-term security of supply benefits all the consumers of the power sector, the only correct answer to this question seems to be the entire demand of the system. Any other arrangement would create an evident situation of free riding, because some users will be taking advantage of the system adequacy without paying the associated costs (it is important to bear in mind that these mechanisms mainly aim at providing investors with a hedge to cover their regulatory and market risk, so the demand who actually buys in these auctions bears the related risk premium).

Since the charge necessary to guarantee the capacity expansion is one of the costs that compose the total cost of the system, if it is paid only by a part of the demand, this can be seen as a cross-subsidy from the customers who must take part in the electricity auction to those who do not have to participate. From this point of view, the choice made in the Colombian

---

2 In the South American context, an argument commonly used to justify the decision to lay only on regulated consumers the obligation to buy in the auctions, is that a contract with a distribution company (in their role of regulated retailer) as counterparty reduces the investors perception of credit and regulatory risk, which would be higher if the counterparty is just “the whole system” (in the form of a regulatory commitment, the capacity payment, instead of a contract) or the free consumers.

3 This approach would give a justification for the regulator to shed the load of free customers in case of a future scarcity event, therefore eliminating the concern of free riding. However, the problem is that it is
and North American auctions, in which the regulator procures some sort of “reliability product” on behalf of the whole spectrum of consumers, seems to be the most adequate.

The most recent Argentinean regulation addresses the same problem in a clearly different way. After the economic crisis suffered at the beginning of the 2000s and the energy scarcity that followed, the Government decided to establish that regulated consumers had full priority of supply and that free users had to cover on their own, through auctions, their expected new capacity requirements. To some extent, it can be stated that there is a free riding issue also with this approach, but in this case in the opposite direction, since free demand bears the cost of attracting new investments and captive consumers benefit from it.

**Level of centralisation**

The level of centralisation of an electricity auction can refer both to the overall auction process (either to organise one centralised auction or to leave this task to each agent subject to the obligation, i.e. free demand or distributors on behalf of their regulated demand) and to the demand forecasting (either to centrally calculate the amount of electricity to be procured or to leave this task to the formerly mentioned agents).

In the American context, different approaches have been applied. Colombia opted for a completely centralised approach, while the Chilean and Peruvian schemes are fully decentralised\(^4\). Brazil has centrally-managed auctions, but the decision about the amount that not always technically feasible to discriminate between categories of consumers when cutting the supply during an emergency event, so a certain level of free riding will always be present.

\(^4\) Actually Peru has implemented two different auctions schemes. The first to be introduced (auctions under the framework of Law 28.832) target only distribution companies and follows a fully decentralised approach. However, a secondary mechanism (Proinversión) was implemented to contract large power projects, mainly aiming at the untapped hydropower potential present in the country, which is completely managed by the State.
needs to be procured is carried out by distribution companies. The ISO New England and PJM capacity mechanisms can be considered as completely centralised auctions.

There are several advantages in a centralised auction approach, see for example Batlle and Rodilla (2010). The most important feature of a centralised auction is that it allows exploiting economies of scale in electricity generation. If all the agents involved in the mechanism organised independent auctions, the amount of electricity to be procured by each one of them would be too small to justify the construction of a new plant, especially if large hydropower projects are considered.

In South America, other advantages of centralised, public and therefore transparent auctions come from the typical structure of the power systems. Only a few countries in the region have unbundled distribution from retailing and the most common scheme is to have local distribution monopolies, which are in charge of purchasing the electricity for the regulated customers connected to their network, i.e. acting as regulated retailers. Furthermore, where this configuration is in place, the main concern comes from the fact that the unbundling between generators and these regulated retailers is at least insufficient (often inexistent). With this structure, distribution companies, acting as regulated retailers, not only have no incentives to design a mechanism that results in the minimisation of prices for the end consumers, but they could even arrange the auction process in order to procure their entire future supply needs from the generation being part of its same holding, at a hardly verifiable price. Within this framework, as highlighted by (Batlle et al., 2010), a centralised auction makes much harder for incumbents to abuse a vertical integration position. Moreover, this design also ensures that all the regulated demand has the same energy price, which fulfils the equity principle of tariff design. In fact, in those systems where the auctions are decentralised, smaller distribution companies are commonly exposed to higher prices.

Concerns regarding vertical integration and entry barriers are in place not only in South America but also, to some extent, in the North American and European contexts. As it has already been observed in the United Kingdom (see for example Ofgem, 2011), vertical
integration between generation and retailing reduces the liquidity in the market and acts as an
entry barrier to new competitors. In fact, incumbents plan their capacity expansion in order to
supply their consolidated portfolio and this reduces the potential market share for new entrants. In this framework, a capacity mechanism designed around decentralised auctions
allows incumbents to exercise their vertical integration. In order to avoid this entry barrier for
new entrants, a transparent and centralised auction should be launched for the coverage of the
entire reliability demand, in which the perfect match between the vertical integrated
generator's supply and retailer's demand is not possible. In the case of the UK, this is
apparently the proposed design of the capacity mechanism to be implemented (DECC, 2013).
However, in France, another market characterised by vertical integration between generation
and retailing, the mechanism to be implemented in the framework of the NOME law (JO,
2010) is focused on capacity obligations for retailers. The capacity certificates needed to fulfil
the obligations are procured through bilateral contracts. Apparently no auction scheme in
foreseen, not even a decentralised one, and exchanges are expected to take place through direct
negotiations. This design stresses the above-mentioned concerns regarding vertical
integration and permits incumbents to create entry barriers for new entrants.

As regards the allocation of the responsibility to estimate the future demand, some authors
(Moreno et al., 2010) have expressed their preference for a decentralised approach, suggesting
that distribution companies should be in charge of forecasting future demand, with a proper
scheme of penalties for over/under predicting. This methodology is based on the assumption
that distributors have a clearer vision of the demand they cover and can better estimate its
growth. While this argument cannot be disproved, first it should not be forgotten that the
ultimate responsibility for security of supply is always with the regulator (or the system
operator often acting on its behalf). Moreover, it is important to remember that one of the key
reasons behind the implementation of capacity mechanisms is the lack of ability or will of
demand (represented by the companies uncharged of retail) to properly hedge beyond the
short to medium term. If the design of efficient penalties for generating units for non-
compliance is in any case an intricate issue (and always controversial), setting the right
incentives and credible penalties for retailers to properly estimate their demand in the very long run is an even more troublesome task. So, this decision should be made by the regulator itself.

2.1.2 The selling side

The regulator, or the agent calling the auction, has to decide who is allowed to present bids. Two major decisions have to be made: first, whether existing and new investments should be both allowed to participate; and second, if any kind of technology discrimination should be explicitly applied.

The role of existing plants

The regulatory decision to implement an auction mechanism to bring in new generation could be perceived as a market intervention that affects existing plants negatively, as this addition of generation depresses prices in the spot market. From the perspective of the rigorous economic theory, it can be argued that, in order to guarantee in the long run the fully efficient signal to drive generation expansion, existing generation plants, at the time to call the tender for new capacity, should also be allowed to compete in equal terms, so they could receive the same price (for example, as it is the case in most markets, in real estate market nobody questions that old and new houses have to compete at the same level).

However, the above mentioned reasoning is often refuted with two objections. Firstly, political interests often lead to the simple conclusion that, since their fixed cost are sunk, there is no need to allow existing units to participate in the auction or to receive the resulting price. This decision certainly minimises the cost of supply for consumers in the short term, but, as stated, it is far from being clear that this decision will not end increasing the cost in the long run, as investors might get to the conclusion that they will not be able to capture the long-term marginal price. Secondly, the fact that in most cases the vast majority of the existing generating units were installed under the former traditional cost-of-service scheme, which guaranteed the recovery of the capital costs, and that often they are still in some way publicly-
owned, provides some justification to the regulators’ decision to discriminate between existing and new generation, especially in the South American context. However, even if this turns to be the regulatory decision, it would be important that also existing plants, which have fully recovered investment costs would be given an economic signal encouraging them to manage their units in order to enhance their availability in scarcity conditions. This is the underlying motivation behind the proposal developed in Batlle et al. (2008) to define two different reliability-oriented payments (an adequacy-oriented one only for new entrants and a firmness-oriented one aimed at every generating unit in the system).

Therefore, the most common distinction for this design element is between mechanisms where different auctions are organised for new and existing plants and schemes that mix the two categories in the same auction. Brazil has implemented from the beginning separate auctions, while in US-ISOs, Chile, Colombia and Peru existing and new plants participate in the same tenders -even if in Colombia and in ISO-NE they are then discriminated through the contract duration, which is one year for the existing plants and up to 20 years (Colombia) or 5 years (ISO-NE) for new projects-. Another approach is to allow existing generation units to take part in the auction only as price-takers (i.e. allowing them to bid only at zero or in a very tight range around it), while leaving the task of setting the auction price to the new power projects (who act as price-makers). This is the current design in ISO-NE Forward Capacity Market, while in Colombia a system of different price caps to be applied in special auction conditions is used to further discriminate between existing and new facilities.

**Technology discrimination**

Another issue related to this design element is the introduction of technology-specific auctions, which have been launched by the regulators of several Latin American countries, either to promote renewable energy projects -e.g. the case of the RER auctions in Peru or the Brazilian ones to add biomass or wind generation, see for example Cunha et al. (2012) or Mastropietro et al. (2013) -, or to foster the exploitation of the large untapped hydropower resources of which the continent is endowed (the Proinversión auctions in Peru or the project-
specific hydropower auctions in Brazil). From a theoretical point of view, in liberalised power systems the capacity expansion, and more in particular the technology choice, are left to the decisions of market agents, who are expected to bear the related risk of a bad assessment of the system future needs. This approach is at the base of the original decision to liberalise electric power systems, and it is thought to result in the entrance to the system of the most economic-efficient mix of technologies.

This kind of intervention is usually accepted if it aims at very long-term strategic objectives of the country, which would not otherwise be achieved because they are not considered by market agents when taking their decisions. In this case, these long-term objectives should be specified at the moment of launching the technology-specific auction and the subsequent regulation should be consistent with the goals identified.

However, it is important to bear in mind that when there is a broad range of potential technologies, the auction design details, especially in terms of contract provisions, often determine the final outcome. Although it is discussed in the next item in more detail, just to illustrate this statement with a very simple example, defining a lag period (i.e. the maximum time available for construction) of three years makes almost impossible for large hydro plants to compete against conventional thermal generation.

### 2.2 Lag period (or lead time)

The lag period is the time that separates the contract signature from the date when the contract enters into force. Unless they are involved in previous contractual commitments, existing plants need no lag period, because from a technical point of view they can start producing electricity immediately (however, for administrative reasons, contracts with existing plants usually consider a lag period between a few months and one year). On the other hand, for new generation projects the lag period represents the maximum time available for construction.
Obviously this parameter heavily conditions the competitiveness of the different plants and technologies in the auction. A critical look at the length of the lag periods in the South American experiences clearly illustrates this fact. In Brazil, three different kinds of auctions are implemented, with the main difference between them being precisely the lag period. The so-called A1 auction (one-year lag period) is meant for existing generating units, while the A3 auction (three-year) clearly aims at adding thermal generation and A5 is intended for large hydro projects. In Peru, the first design defined a three-year lag period, clearly in line with the governmental desire to exploit the Camisea gas pipeline constructed in the late nineties through the installation of gas plants. Few years later, the Government considered the need of new large hydro investments, and put into question the auctions design. This concern led to an out-of-the-electricity-regulation call for tenders aimed specifically to attract this generation technology⁵.

In the reliability charge mechanism in force in Colombia, in an attempt to attract both thermal and hydro plants, two related but separated auctions are called, a first one (OEF⁶ auction) with a lag period of four and a half years, more suitable for and implicitly targeting thermal plants, and a second one with a lag period of seven years, focusing on new hydropower projects⁷. Thus, the only apparent way to design an auction in which more than one technology could equally compete would be to allow for different lag periods.

---

⁵ In 2002 and 2003, Decree no. 027/2002/PCM and Decree no. 095/2003/EF created and renamed the agency finally called Proinversión. This institution is in charge of fostering private investment in strategic infrastructure development, hydropower plants and other large power projects within this scope.

⁶ OEF stands for Obligación de Energía Firme (Firm Energy Obligation in English).

⁷ The so-called GPPS auction is organised in Colombia after the OEF auction and it is designed for those plants with construction times exceeding the OEF lag period. The reserve price in the GPPS auction is the price cleared in the OEF auction.
Therefore, if no strategic objective results in a preference towards certain technologies, the only approach to solve this issue would be to allow all kind of plants to bid in the auction with the assurance that enough time will be conceded for the installation. However, this immediately leads to another relevant problem, since it is very difficult to define proper criteria to transparently compare the different bids (e.g. it would not be clear how to compare a plant bidding a short lag period and a price 10% lower than another bidding a longer one).

Therefore the regulator has a trade off in defining whether a single or multiple lag periods are to be considered. In the United States, the most common approach has been that of considering one single lag period. In particular, the value considered has been three years (PJM and ISO-NE). However, it is worth noting that these mechanisms target principally thermal plants (mainly CCGTs, which have short construction times) and demand response, which do not need large lag periods.

### 2.3 Contract duration

The duration of the contract offered in the tender is one of the most relevant design elements of a long-term electricity auction. As mentioned throughout the paper, the main objective of electricity auctions is to hedge the generators' risk against the high volatility of spot market prices and more importantly, against regulatory risk, in order to facilitate their access to financing and to improve the overall attractiveness of the investment. Keeping this in mind, the contract duration should be large enough to provide new generation projects with the stability they require to carry out the investments.

The long-term signal in charge of attracting new investments is enclosed in this parameter and its determination has a dramatic impact on the results of the auction. As for the lag period, the selection of the contract duration may not be unique and it is possible to differentiate between plants (new or existing) and technologies (e.g. hydro or thermal) also in the same action.
The Brazilian experience clearly illustrates how tenders can be “guided” through their design elements. The A1 contracts, since they are aimed only at hedging market prices according to the El Niño Southern oscillation, have duration of 5 to 8 years. Conversely, the length of the contracts offered in the A5 auctions ranges from 15 to 30 years, confirming our previous statement that these auctions are clearly designed to bring in large hydro plants. Similar conclusions can be for example put forward when reviewing the Peruvian auctions (the contracts in the Proinversión auctions have durations larger than 10 years).

As regards new plants, the contract duration should reflect the capital intensiveness of the technology. Obviously, thermal plants need shorter terms for mitigating their risk than hydropower plants. In any case, at the moment of determining the contract duration, it must be taken into consideration how the project financing works. In fact, due to the effect of the discount rate, the impact of future income on the decision-making becomes less and less important, depending on how far this income is from the present. Therefore, contract durations of thirty years are seldom justified with the high discount rates used in the generation sector in South America. On the other hand, in North America, common contract durations are significantly lower. In the ISO-NE Forward Capacity Market, existing plants are entitled only one-year contract, while new plants can choose contract durations from one to five years. This is probably due, once again, to the different technological targets (thermal plants, in this case), but also to the lower country risk.

Moving for a while from the regional scope of this article, the most recent official regulation concerning the capacity mechanism to be implemented in Italy (AEEG, 2011) considers a scheme based on reliability option contracts, to be procured in a centralised auction with no discrimination between new and existing plants. These contracts, according to the current design proposal, have a lag period of at least 4 years, thus clearly targeting new generation projects. Nevertheless, the contract durations foreseen are equal to 3 years, which could even be reduced to 1 year in some cases. While these durations could be acceptable for existing generation, they are clearly not sufficient to hedge investors’ risk and, therefore, they are not suitable for attracting new plants. Such design could hamper the effectiveness of the
mechanism in guaranteeing the system adequacy. Actually, in the current framework of overcapacity that the Italian system is experiencing, the objective of a capacity mechanism with such short contract durations, more than to foster new investments, seems to be to remunerate the fixed costs of the existing generation mix, which are not being recovered in the energy market due to prices lower than expected.

2.4 Defining the requirements associated to the “reliability product”

Traditionally, the spectrum of capacity products that the regulator can purchase on behalf of demand (aka “reliability product”) has received several denominations, such as capacity credits/obligations, firm energy contracts, reliability contracts/options, strategic reserves, etc. Today the use of these names is still vague and can be misleading, since it is not clear which actual characteristics and commitments are associated to these names.

In general, the product can be more properly defined and classified based on the following questions:

• Is the product energy-based or capacity-based?

• Does the contract imply a physical commitment, financial or both?

• When are agents selected in the auction required to fulfil their commitment?

• How are they penalised in case of underperformance?

• Is there any regulatory limitation on the quantity of product that each agent of the power system can sell in the mechanism?

Answering these questions allows in this section to identify different kinds of reliability product.

Reliability in capacity- and energy-constrained systems

In order to properly define the product that is required from the reliability providers in exchange for the remuneration they receive, it must be understood which kind of scarcity
conditions can be expected in the system. The main classification that can be applied divides systems into capacity-constrained and energy-constrained.

- In capacity-constrained systems, scarcity problems arise because there is not enough installed capacity available (MW) to satisfy demand at a given moment (e.g., due to the forced outage of thermal plants and/or minimum wind output); aggregating all the hours, the system could certainly have enough energy available to satisfy demand on that day (more than enough thermal capacity in the valley), but it lacks installed capacity to satisfy peak demand. This type of potential scarcity conditions is usually found in the European and North American systems, where this led operators and market participants to be primarily concerned about modelling the very short term in great detail.

- In energy-constrained systems the situation is quite the opposite: rationing is applied due to a lack of available energy; the system could certainly satisfy peak demand, but would not be able to supply the demand during the remaining hours of the day/week. A large proportion of Latin American systems have traditionally fallen in this category, due to the large share of hydropower in their generation mixes. The availability of reservoirs with a large storage capacity has historically reduced the necessity of considering the short-term operation in much detail. However, in these systems it has been of critical importance to suitably represent uncertainty and to determine the optimal management strategies for hydro resources and their medium-, long- and very long-term interaction with thermal facilities.

This distinction must obviously be reflected in the design of the reliability product. On the one side, a capacity-constrained system has to remunerate the ability of the agents to provide instantaneous power to cover the peak demand. On the other side, an energy-constrained system should reward the capacity of the agents to manage their resources in order to guarantee their availability in those periods when energy scarcity takes place, as it could happen during a dry year. This consideration is confirmed by the diversity of capacity mechanisms introduced in different systems. In the United Kingdom, a system with a very large share of conventional thermal plants, the capacity mechanism currently under design is
completely focused on the capacity that the agents can inject in the network. On the other hand, the energy-constrained Brazilian system ensures the security of supply through long-term electricity auctions, in which generators offer full-energy contracts, with yearly settlements. An intermediate solution between these two extremes can be found in the Colombian energy-constrained system. The capacity mechanism implemented in this country (Firm Energy Obligations scheme) requires the agents to deliver the energy committed in the auction during those days when the spot price exceeds the strike price at least once.

**When are agents selected in the auction required to fulfil their commitment? The role of the critical period definition**

Some designs of capacity mechanisms consider the definition of a critical period (also called scarcity conditions, or near-rationing conditions, depending on the nomenclature used), during which each agent with a reliability commitment must deliver the product sold in the auction. As regards the countries analysed in this paper, this does not always apply. In Brazil, Chile and Peru, where standard future supply contracts are auctioned, the generators selected by the tender mechanism have to deliver electricity according to the contracts they sign. No critical period is defined and if the system has some method to identify near-rationing conditions, this does not affect the contract provisions. On the other hand, the Colombian firm energy obligation (OEF) mechanism is based on the definition of scarcity conditions. Following Vázquez et al. (2002), the spot market price is used as critical period indicator and the scarcity conditions are defined as the period of time during which the spot market price exceeds a predetermined strike price8. Since the Colombian mechanism is based on option contracts (the seller, in exchange for a fee, commits to provide the buyer with electricity not at the actual spot price, but rather at the strike price), the strike price has two functions: on the one side, it

---

8 Actually, the Colombian mechanism, differently from what suggested in Vázquez et al. (2002), considers a daily obligation. Also if the spot price exceeds the strike only for one hour, scarcity conditions are applied to the entire day.
identifies the critical period, on the other side it acts as a soft price cap (only for the generation awarded with a reliability commitment in the auction). In Colombia the strike price calculation is linked to a basket of fuel price indexes -updated through the Platts US Gulf Coast Residual Fuel No. 6 1.0% sulfur fuel oil, see CREG (2006)-.

Some general recommendations on the selection of the critical period indicator were already provided by Batlle and Pérez-Arriaga (2008), who identified the short-term market price as the best “thermometer” of scarcity conditions in a market environment. This consideration should turn to be more valid in the future, in a scenario of increased elasticity of the demand. In fact, as long as the amount of completely inelastic demand in the market (i.e. the demand that bids at the price cap) decreases, it will become more and more difficult to define the demand which “must” be served and, consequently, the identification of near-rationing conditions based only on the comparison of peak demand and generation available. This critical period indicator obviously assumes the presence of a liquid reference short-term market in the system. The selection of the reference market also affects the “kind” of scarcity conditions that are covered, and that the regulator wants to be covered, by the capacity mechanism. On the one hand, day-ahead markets are only capable of capturing emergency situations related to the combination of high loads (as peak winter demand) and reduced supply (due to fuel constraints or a dry year that limits hydro production), i.e. pure adequacy issues. On the other hand, intraday and balancing markets are also subject to price fluctuations due to more or less sudden events (as the outage of a nuclear plant or, in those systems with high renewable penetration, the fall of intermittent generation due to bad forecasting), which provoke temporary generation scarcity even if the load is far from the peak and which have a time horizon larger than the one covered by ancillary services, i.e. firmness and flexibility issues. However, the selection has also to take into account the signal that the capacity mechanism is providing to the generation mix. While all units are more or less technically capable of producing if notified one day ahead, certain technologies (base-load technologies, as coal power plants) would not be able to take part to the balancing market, because they cannot respond in such a short term, due to ramp constraints. Therefore, a capacity mechanism using the balancing as the reference market
provides agents with a signal that discourage the installation of new base-load units and this may not be the objective.

In case the short-term market price is selected as critical period indicator and a strike price has to be determined in order to set the frontier of the scarcity conditions, two aspects must be taken into consideration (Vázquez et al., 2002):

- The strike price must be predetermined by the regulator based on a formula available to all interested agents and must be unique for all the agents taking part in the auction. Allowing agents to bid both the strike price and the option fee associated to it, would add significant complexities to the auction mechanism. In fact, since the bids would not be simple price-quantity pairs, their comparison would be possible only through a model simulating the market behaviour and this reduces the transparency and can add somehow arbitrariness to the process.

- The strike price must be determined in a way that it does not interfere with the normal functioning of the short-term market. This can be achieved fixing it much above the short-term production costs of the most expensive generator expected to produce.

In any case, the selection of the critical period indicator should be carried out considering the long to very long term involved in the contracts signed under the framework of capacity mechanisms. A critical period indicator that may appear appropriate in the present could become unsuitable in the future because of changes in the generation mix or in fuel prices. Eluding for a while the regional scope of this paper, the capacity mechanism proposed in the framework of the NOME Law in France, see Finon (2011) for instance, apparently identifies the scarcity conditions using the temperature as the critical period indicator, because it is when the temperature is higher that scarcity conditions usually occur (due to the increased demand for air conditioning as well as to the reduced cooling capacity in the condensers of nuclear plants). However, this violates what was mentioned above, because the correlation between scarcity conditions and temperature may change in the future (e.g. due to a large penetration of solar PV generation) and there is a risk of having long-term contracts that are
not providing the expected reliability anymore. Also the capacity market outlined in DECC (2013) for the United Kingdom proposes a critical period indicator not aligned with the recommendations presented here. In fact, in this design, the scarcity conditions are replaced by the concept of “system stress”, defined as “any settlement periods in which either voltage control or controlled load shedding are experienced at any point on the system for 15 minutes or longer” and communicated to the agents with a reliability commitment at least four hours in advance through a “capacity market warning”. This way, the scarcity conditions are not linked to the reserve margin anymore and they could also occur when the demand is for example 70% of peak demand. In this case, each reliability provider “will only be required to be generating electricity or reducing demand up to 70% of their raw capacity obligation”. This approach has two clear drawbacks. Firstly, the determination of system stress is somehow arbitrary and completely unpredictable, and this creates a challenge for agents at the moment of evaluating the impact that taking part in the capacity market can have on their revenues, therefore in the definition of their bids in the auction. Secondly, it hampers demand side response (DSR) participation in the capacity market, since DSR is only available to provide a certain demand reduction during peak demand conditions.

**Penalties for non compliance**

The non-fulfilment of the contract commitments must be penalised by the regulator. The penalty should be high enough to dissuade the selected bidders to manage their generation plants for them not to accomplish their obligation. On the other hand, the penalty should not be excessive in case of prolonged technical unavailability. This could be achieved through an hourly penalisation associated to a correction factor that diminishes with the duration of the unavailability or through the use of cumulative penalty caps. In the latter case, “soft caps” should be used, which allow the agents who have reached the penalty cap to reduce their
Another distinction to be made is between implicit and explicit penalties for improper performance. Without entering into details of the different possible capacity mechanisms, when an agent does not fulfil its commitment, it can be penalised in two ways: first, it could be required, in case of “non-delivery” of the reliability product, to procure an equivalent amount of the reliability product in the electricity market, in order to honour its contract, even if not through its own assets (implicit penalty). On the top of this, when designing the capacity mechanism, the regulator can introduce an explicit penalty, in the form of an extra fee for non-compliance calculated with a predetermined formula (explicit penalty). Vázquez et al. (2002), in presenting the reliability option principles, claimed for the necessity of an explicit penalty in order to discourage those agents that are not backed by reliable generation capacity from participating in the auction.

In South American mechanisms the only penalty usually considered is the implicit one. The Colombian firm energy obligations for example do not consider any explicit penalty. According to several experts, this represents a current vulnerability of the Colombian capacity mechanism. This is for example the view of the Colombian Wholesale Electricity Market Monitoring Committee (CSMEM, 2010) after the scarcity conditions occurred in 2009/2010. In this report it is stated that hydro generators prefer to assume a future non compliance of their reliability option contract than to face a certain economic loss in the present. This behaviour is clearly related to a lack of an explicit penalty, which would increase the

---

9 It must be underlined that a soft penalty cap can be applied only in those schemes where the agents can actually over-perform with respect to their commitment. This could happen because of a production that exceeds the maximum quantity the agent was allowed to trade in the capacity mechanism or due to the specifications of the product to be provided during scarcity conditions.
magnitude of the effect of future non-compliances, consequently increasing its “weight” in present decisions.

Conversely, in the North American context it is possible to find explicit penalties in case of underperformance during scarcity conditions. In the Forward Capacity Market of ISO-NE, every time a shortage event occurs (a shortage event is any period of thirty or more contiguous minutes of system-wide deficiency in operating reserves), each agent holding a capacity supply obligation has a shortage event availability score calculated. Based on the latter, a shortage event availability penalty is defined as:

\[(\text{Annualised Forward Capacity Payment}) \times (\text{Shortage Event Penalty Factor}) \times (100\% - \text{Shortage Event Availability Score}),\]

where the shortage event penalty factor is 5% for events lasting 5 hours or less and increased by 1% for each hour above 5 hours. If the shortage event availability score is 100% (i.e. capacity obligation fully fulfilled), the agent is not penalised, otherwise the explicit penalty is applied. These penalties are subject to a number of daily, monthly and annual caps, for them not to exceed the annualised forward capacity payment.

**Constraints on tradable quantities**

Theoretically, if properly designed explicit penalties for under-performance are set, aimed at discriminating in favour of more reliable generation technologies, and if high enough collaterals are required, providing the system with a financial guarantee in case the estimations of the agents reveal to be wrong, there should be no need of setting such constraints. Nevertheless, most regulators and system operators of those countries that have implemented a capacity mechanism, have introduced some methodology for setting an upper and a lower limit on the quantities tradable by each agent. These decisions are based on the following reasoning:

- Without an upper limit, it is possible that an overestimation by part of the agents on the quantity of reliable product they are capable of producing in scarcity conditions compromises the financial stability of the entire power sector, with potential repercussions
on security of supply. In order to avoid this, very large and costly warranties should be introduced, but this could dramatically reduce the participation in the capacity mechanism.

- Without a lower limit, some agents could be tempted to behave strategically, withholding part of their capacity from the auction, with the objective of increasing the clearing price. Therefore, setting a lower amount that the agents are “required” to trade, helps manage market power issues.

These methodologies resulted in the creation of the concepts of firm energy or firm capacity, depending on the specific design of the reliable product auctioned, and the establishment of some sort of prequalification phase, during which these parameters are calculated for all the agents willing to participate in the auction. In South American auctions, most of the regulators have introduced such methodologies, being Chile the main exception\(^\text{10}\). A detailed description of these calculation methods exceeds the scope of this article, but as a way of example, we can explain how it is done in the Colombian case for South America and in ISO-NE for North America.

In Colombia, generators willing to take part in the OEF auction must be backed by firm energy certificates (also called ENFICC), whose calculation methodology is outlined in CREG (2006). The ENFICC of hydraulic plants is calculated using a computational model (HIDENFICC), which determines the maximum production that can be obtained monthly from a hydro plant during dry periods. The lower ENFICC limit that a generator or investor can trade is termed ENFICC Base and corresponds to the minimum energy obtained by the maximisation model. The upper ENFICC limit corresponds to the energy that a generator can produce with a probability of 95%, called ENFICC 95%. If the generator or investor is willing

\(^{10}\) Theoretically, the contracts signed in Chilean auctions are not required to be covered by any firm certificate and the distribution companies have to assess the bidders’ credibility on their own. However, generators have to specify to the regulator, on a yearly basis, which plants will be used to cover the contracted demand.
to trade an ENFICC higher than the ENFICC Base in the auction, without exceeding the ENFICC 95%, the generator should back this difference with a financial warranty. On the other hand the ENFICC of a thermal plant is calculated based on the generation capacity of the plant, the fuel availability, the number of hours per year and an index that incorporates the historical restrictions imposed on the plant, which limits its maximum energy generation. As regards renewable energy, CREG (2011) has recently outlined the methodology for calculating the ENFICC Base and the ENFICC 95% for wind farms.

In the ISO-NE Forward Capacity Market a different approach, based on historical availability, is followed. Existing plants are recognised summer and winter qualified capacities to be calculated as the median of the most recent five summer and winter claimed capability ratings. For hydro plants with daily cycling, the qualified capacities are based on the seasonal average, calculated using the 50th percentile flow rate. As regards new plants, project sponsors must submit a show-of-interest to the System Operator, which contains the requested summer and winter qualified capacities. Based on the financial reliability of the project, the construction schedule, and the requested qualified capacities, the show-of-interest can be accepted or rejected. The qualified capacities are then adjusted during the following auctions based on real data from the plant.

Finally, it must be underlined that the firm energy and the firm capacity are the basis on which plants and projects are remunerated in the capacity mechanism market. Therefore a revision procedure that punishes underperformance by administratively reducing the firm energy or capacity of a plant can represent a very effective penalty scheme.

2.5 Indexation and warranties

There are several other design elements that, though often considered as secondary, can have a significant impact on the final results of the auction. In this paragraph three of them are reviewed.
Indexation formulas

In the South American experiences, in line with what has been the traditional way to remunerate utilities and independent power producers in the previous regime, all the economic figures set by the contract are usually subject to indexation formulas that determine their future evolution. Commonly, these formulas, besides being linked to the retail price index (in the US) or the exchange rate of the dollar with the local currency (in South America), use as a reference the international price of fuels, trying to foresee how this parameter will affect the operation costs of power plants. However, economic theory recommends assigning risks to the agents who can better manage it. Indexing the energy price to fuel prices implies allocating this risk among electricity consumers, who have no ability whatsoever to properly manage it. Thus, avoiding this fuel indexation in the contracts would be a better practice, since in principle generators should be able to hedge this risk more efficiently by signing contracts on the international term markets for commodities. It is also true that unfortunately these markets present a low liquidity in the long- to very long-term (i.e. larger than five years), therefore the optimal solution would be to design an incremental “indexation weight”, i.e. the per cent dependence of the contract prices with respect to international fuel prices.

Another choice concerning indexation formulas is whether to use a unique formula to index all the contracts or to allow agents to include the required indexation within their bid. The latter approach, used for example in Chile\(^{11}\), creates a challenge at the moment of comparing different bids, because their competitiveness in the long term varies broadly according to this

\(^{11}\) As mentioned in Maurer and Barroso (2011), in Chile the indexation formulas are determined and published by the regulator in the form of a multivariable linear function of fuel and inflation indexes in which each multiplying factor is ultimately adjusted by each bidder, thus creating different indexation requirements. However, indexation formulas are not taken into account by the auctioneer during the allocation process, thus affecting the overall economic efficiency of the process, since the set of winners could dramatically change if indexation formulas were incorporated into the clearing mechanism.
parameter. A general recommendation is to define a unique indexation formula for all the generators involved in the auction, in order to increase transparency and to keep the auction format as simple as possible.

**Financial warranties**

As regards warranties, as in other markets, the bidders selected through the auction should provide a monetary endorsement to cover at least part of the potential penalties for non-fulfilment. In order not to introduce large warranties that could limit the participation to the auction, this endorsement could be achieved by retaining part of the first contract payments, until a targeted figure is reached. However, for both penalties and warranties, it must be highlighted that their determination is strictly related to the selection of the reliability product and in particular to the constraints on the tradable quantities, as mentioned in the dedicated paragraph. In fact, as specified by Batlle and Pérez-Arriaga (2008), the higher the penalties for non-compliance and the warranties required, the smaller is the need to be strict on defining the maximum value each unit can trade or ask to be remunerated.

3 **SUMMARY & HIGH-LEVEL RECOMMENDATIONS**

Long-term auctions have been selected by a good number of regulators in the American continent as the most effective instrument to guarantee the adequacy of the system. They reduce the risk associated to the long-term volatility of spot market prices and to potential regulatory interventions by guaranteeing and fixing ex-ante part of the future income of generators, through an in-principle efficient (if competitive) and transparent mechanism. This economic signal facilitates project financing and fosters the installation of new generation capacity. However these instruments must be designed carefully. In this paper, we have identified, reviewed and discussed the key design elements of long-term system-adequacy auctions, extracting guidelines from the South and North American experiences. The main lessons learned are briefly summarised in what follows.

On the buying side, the best option seems to be to involve all the system demand in the auction, since all the system demand benefits from the increased system reliability, thus
avoiding free riding issues. The demand which is not willing to be covered by the reliability mechanism can then offer demand response (or even energy efficiency) bids in the auction, but this option should be only given to those users which can actually be disconnected during scarcity events and after a careful determination of the customer baseline.

As regards the centralisation level, the most efficient solution appears to be a centralised auction that covers the whole system demand. This approach allows to exploit economies of scale in generation and to mitigate the impact of vertical integration of electric companies, which is still an issue in several power sectors, giving a chance to new entrants. Furthermore, a centralised auction enhances transparency and guarantees an equal “system-adequacy” price among all the consumers.

On the selling side, existing and new generators can compete in the same auction, as long as specific measures are taken in order to differentiate between the two categories, such as defining existing generators as price-takers in the auction or setting different price caps. As regards technology-specific tenders and the determination of all these parameters, as for example the lag period for new power plants, which can implicitly include or exclude certain technologies, it must be understood that a certain degree of discrimination will be always present. A mechanism which accurately defines the reliability product and the associated parameters will always favour a certain technology, but the alternative, i.e. letting bidders specify the parameters of the product (lag period, contract duration, strike price, etc.), leads to the undesirable result of having to “compare apples and oranges” in order to clear the auction.

Regarding the critical period indicator, the best choice is the short-term price of the reference market in the system, which is the most suitable “thermometer” of scarcity conditions in a market environment. This critical period indicator obviously assumes the presence of a liquid power exchange in the system, but this is nowadays considered as an essential feature of any efficient wholesale market. Therefore, in those systems where such a reference market is missing, the implementation of a capacity mechanism of this sort could turn to be an opportunity to foster the development of a liquid short-term market (day-ahead or balancing).
Finally, as regards penalties, it is essential to define a robust explicit scheme, which adds an extra fee to the implicit penalty of having to purchase in the market the electricity necessary to fulfil the commitment if the agent is not able to produce it with its own assets. This discourages those bidders which are not backed by reliable generation capacity and encourages those agents selected in the auction to manage their units in order to improve their availability in scarcity conditions, thus enhancing the firmness of the generation mix.

4 REFERENCES


CREG, Comisión de Regulación de Energía y Gas, 2006. Resolución No. 071 de 2006 por la cual se adopta la metodología para la remuneración del Cargo por Confiabilidad en el Mercado Mayorista de Energía. Anexo 1, apartado 1.4. Released on 3 October 2006.


Ofgem, Office of Gas and Electricity Markets, 2011. The retail market review - findings and initial proposals. Consultation paper released in March 2011.


