IIT Working Paper IIT-07-024I. First version: August 2007. This version: November 2007.

To appear in a special issue on "Capacity mechanisms for long term supply security in liberalised electricity markets", Utilities Policy

Design criteria for implementing a capacity mechanism in deregulated electricity markets

Carlos Batlle, Ignacio J. Pérez-Arriaga

Over twenty years since electricity industry deregulation was first implemented, and nearly ten since many power systems boarded the liberalization train, reliability of supply appears to be the major concern of energy regulators. Drawing from the cumulative experience of systems that have already implemented some manner of security of supply mechanism, the present article reviews the main criteria to be taken into consideration in the design of a regulatory mechanism of this nature.

Keywords: liberalization; electricity markets; regulation; market rules; security of supply

1 INTRODUCTION

Designing a stable regulatory framework for the efficient and reliable delivery of electric power at present and in the long term future is one of the major concerns of electricity market regulation policies. Since the choice of a regulatory framework that is open to competition as far as possible is an accepted principle, the key question now is how to introduce the necessary adjustments in the market designs currently in place. In this context power supply reliability has emerged as the key issue to be addressed, since many markets -in Latin America in particular- are finding it fairly, not to say extremely, difficult to properly adapt their generation capacity to demand growth. The prolonged or systematic service interruptions that arise as a result may lead to political or market model crises, such as in Ontario for instance, see for example Trebilcock (2006).

The context surrounding the reliability of supply question in deregulated electricity systems is described below.

Terminology

The ultimate measure of generation reliability is the quality of supply delivered at the wholesale power level. Although quality of supply only materializes in real time, its attainment involves a number of deregulated activities that must be performed in different time horizons (Pérez-Arriaga, 07). For the sake of greater clarity about the scope of the regulatory measures dealing with this issue, three dimensions of the reliability problem are distinguished here: *security*, *firmness* and *adequacy* (Batlle et al., 2007a):

- Security is understood to be the readiness of existing generating capacity to respond, when needed, to meet the actual load (a short-term issue, i.e. operating reserves prescribed by the System Operator).

Carlos.Batlle@iit.upcomillas.es, Ignacio@iit.upcomillas.es

Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, Sta. Cruz de Marcenado 26, 28015 Madrid, Spain.

- Firmness is defined to be the short-term generation availability resulting from the operational scheduling of installed capacity (a short to mid-term issue, i.e. generator maintenance management, fuel supply contracts, reservoir management, start-up schedules and so on).
- Adequacy means the existence of sufficient available installed capacity, both installed and/or expected to be installed, to meet demand (a long-term issue).

It is widely agreed that the System Operator can deal with security difficulties by resorting to *ad hoc* markets. No such consensus has yet been reached, however, about how to ensure that the existing levels of adequacy and firmness will deliver acceptable service.

The nature of the problem

Theoretical microeconomic analysis of power systems shows that the price resulting from a competitive market suffices to remunerate the total costs of generating units whose investment is well adapted to existing demand and to the presence of all other generation plants.

A number of conditions must be met for such an ideal situation to materialize, however. And this is not usually the case in practice for several reasons, such as the existence of price caps on the income of peaking generation units, the downward pressure on energy prices possibly exerted by mandatory levels of operating reserves, and (primarily country or regulatory) risk aversion. But of all these conditions, the chief flaw is that real demand is not playing its proper market role. This is due, among other reasons, to the existence of a certain implicit assurance that leads consumers to believe that the regulator will never allow their interests to be severely jeopardized by supply shortfalls or inordinately high prices. Consequently, demand does not respond suitably on the long-term market.

Regulator intervention

If the market, left to its own devices, is assumed to be unable to provide sufficient generation availability when needed without regulatory intervention, the solution to the problem necessarily entails the development of additional mechanisms to assure firmness and adequacy of supply. And indeed, some kind of regulatory capacity mechanism has been designed and implemented in a number of power markets around the world.

While no such mechanism has been explicitly implemented in several, particularly European, markets, it may be safely asserted that no system lacks at least an implicit regulatory safeguard: in some, the System Operator enters into long-term reserve agreements, in others the incumbent, now in a market-like context but still under partial (and sufficient) public control, "shares the regulator's concern" about system reliability, and in yet others it is the retailer who it is still publicly controlled in some way (by municipalities in many cases) and therefore the agent that seeks to protect its customers from unexpected annoyance through long-term contracting .

Objectives and roadmap

Over twenty-five years since electrical industry deregulation was first implemented, and nearly ten since many power systems boarded the liberalization train, supply reliability, and more specifically firmness and adequacy, appear to be the major concern of energy regulators.

Several different solutions are already in place in power markets worldwide, and reviews -see for example Finon and Pignon (2006) and De Vries (2007)- have been published on the theory underlying the different approaches and propose new ways to tackle the problem.

Many power markets have recently restructured their capacity mechanisms to a greater or lesser extent-e.g , are currently in the process Spain, Guatemala and Italy (CEER, 2006)- or are considering and debating the suitability of undertaking such a process -such as

Against this backdrop, and based on the experience acquired in this regard in several markets in different countries, the present report reviews the main criteria that regulators should consider when designing a mechanism geared to enhancing electricity system firmness and adequacy in a market context. The discussion hinges on existing experience: the results obtained and obstacles surmounted. The report begins with a brief description of the general objectives of capacity mechanisms, distinguishing between firmness and adequacy.

This is followed by a review of the design alternatives available to regulators implementing capacity mechanisms, addressing separately the two main elements to be defined: the properties and characteristics that determine a reliability product and the pricing mechanisms that will ensure that it acts as a proper incentive for market agents to more earnestly internalize system reliability in their decision-making.

2 GENERAL OBJECTIVES OF CAPACITY MECHANISMS: ENHANCING FIRMNESS AND/OR ADEQUACY

2.1 Firmness: increasing the availability of installed units in critical periods

Many electricity system firmness-related aspects have changed in the last decade. First, unit planning has been decentralized in many markets, leaving System Operators with many fewer resources than they had in the past. More often than not, generating unit management as performed by market agents differs from the scheme that a System Operator would devise. But that in fact is what deregulation intended to achieve: to leave to the agents tasks that they can perform more efficiently. The drawback, however, is that market agents do not have the same overview of the problem as the System Operator did in the past, since they do not necessarily know how their competitors are managing their plants -no generator, for instance, can be expected to reveal to its competitors that it is experiencing difficulties with its gas supply-. Moreover, the new terms included in optimization functions do not always move in the direction of reliability: i.e., while in the past a scarcity episode was heavily penalized in the centralized optimization of the system planning, nowadays, scarcity need not be a hardship for a generator, and in fact might prove to be the contrary, since it could imply larger profits.

Furthermore, new generating technologies have rendered the problem even more complex. The former definition of thermal plant availability, couched primarily in terms of technical outages, is no longer valid. The availability of certain units in the system may change dramatically from one day to the next, depending on the harvest of rice in China -implying an increase in the fuel transportation costs-, the price in power systems in other parts of the world willing to pay more for electricity, or even very complex geopolitical strategies -for instance, the gas supply in the Spanish system can be significantly affected by a sudden change in the bilateral relationship between the Spanish and Algerian governments -. Gas-fired plants can now be modelled as a sort of hydro plants on a reservoir with very fuzzy characteristics. The advent of renewables, in turn, adds to the problem. Wind generation is non-manageable and extremely volatile[1] and biomass is likewise highly seasonal.

From the firmness standpoint, then, regulators must aim to modify the mid-term scheduling of the system's existing generating facilities to reduce the expectation of undelivered energy. On the one hand, this involves providing some manner of incentive for generating plant managers to enhance availability in critical periods, by minimizing the likelihood of outages, adequately planning their fuel supplies and maintenance programs or conducting more cautious reservoir management; and on the other it calls for discriminating against unreliable generators.

In other words, regulators must define a methodology to evaluate and provide consideration for each generating unit's actual contribution to system reliability. This definition is usually based on a measure of the availability of units during critical periods, when the likelihood of scarcity is highest. In more general terms, such a measure might be referred to as "firm supply" which, depending on system requirements and the specific details of the incentive, is termed "firm capacity" (Spain), "firm energy" (Brazil), "adequacy capacity" (Chile) or even "efficient firm offer" (Guatemala), see Batlle et al. (2007b). Defining the measure is much trickier than it might seem at first glance. The regulator must first establish, *ex ante*, an objective rule for determining when a period (a given hourly interval, for instance) is critical, while at the same time assessing the real availability of the generating units should such situations arise. The latter is particularly complicated, since not all units will necessarily be producing at that specific time (see section 4.2 below, where we argue that the spot market price would ultimately appear to be the best indicator of the existence of critical situations).

Once the asset -the contribution to system firmness- is unambiguously defined, the regulator must devise a mechanism for pricing it. The design of any such mechanism should occupy the middle ground between two extremes: a demand-side obligation to acquire this new product from generators, leading to a capacity market price at one end, and an administrative price at the other. The solutions in place in different power markets are reviewed in section 3.

2.2 Adequacy: encouraging new investments in generation

The worldwide development of installed capacity appears to be losing pace. The very high capacity margins existing in most EU countries when deregulation began have

---

[1] For instance, on the eve of the day in 2005 when the reserve margin was tightest in the Spanish system (the System Operator had to shed "interruptible loads"), wind generation produced around 4 GW, but less than 1 GW when load peaked (at around 42 GW).

typically shrunk since competition was introduced (Eurelectric, 2007), (CEER, 2006). In Latin America the problem is more severe: despite the many types of fuel available -hydro, gas, bio-fuel- investments are not forthcoming. Brazil, for instance, had to impose rationing for nine months, cutting load back by 20%. Argentina also experienced serious problems this past winter, and the situation in other countries on the continent is not much more auspicious. In North America, adequacy problems in both Ontario and California occasioned some backtracking in the deregulation process.

Then, in the face of an issue of growing importance, regulators' objective in terms of adequacy is to guarantee an "adequate" long-term reserve margin by strengthening incentives to attract new entrants.

Assuring a suitable reserve margin is not the only reason for implementing this type of mechanism, however. In some cases such as in the Peruvian system, when viewed from the reliability standpoint the reserve margin is much larger than theoretically suitable (around 15%-20%, depending on how it is measured). However, a very significant proportion of these installed units are extremely expensive fuel-fired plants[2]. The resulting high spot prices should in principle be more than sufficient to attract new generating units. This is not the case, unfortunately. Risk aversion, as mentioned above, added to the fact that such prices would disappear as soon as a more efficient generating unit comes on stream, discourages the investment needed to remedy the scarcity episodes to which the system is presently prone.

The regulatory instrument basically consists in assuring new entrants an extra payment for a number of years from the time they become operational: the definition of the time terms is a key factor in this respect. It may adopt the form of a mere regulatory settlement with no explicit commitment with respect to duration (e.g., the Spanish capacity payment mechanism, implemented at market start up and currently under review) or of a contract with a certain term, in which the counterparty is the whole system -via a securitized fixed charge included in access tariffs billed to all consumers- or only a distributor or pool of distributors, such as in Brazil.

Such payments/contracts are priced to the same criteria as described for the firmness consideration, and as discussed below, may be set by the regulator (as a sort of "feeding tariff") or driven by the results of market mechanisms such as public auctions.

Regulators must, moreover, define how generating unit quality should be valued to distinguish "good" from "fair" or even "poor" investments. The "firm supply" concept itself enunciated for firmness (where any are in place) or a similar measure is often taken as the reference.

2.3 Key elements of the reliability incentive: the product and its pricing mechanism

The main conclusion to be drawn from the foregoing is that any regulatory design specifically geared to enhancing reliability in a deregulated power system consists of two

---

[2] Paradoxically, in the Peruvian electricity system for instance, the capacity mechanism has led to a *dash* for these inefficient units, whose relatively small capital requirements and short recovery periods lower their exposure to regulatory risk.

main elements: the definition of a reliability product and the determination of how it should be valued. The following section reviews the different ways to price reliability products, ranging from the most to the least liberalized, i.e., from the capacity market to the so-called capacity payment.

This is followed by a discussion of the criteria to be taken into consideration to proceed to define the reliability product.

3 REGULATORY MECHANISMS FOR PRICING THE RELIABILITY PRODUCT

Irrespective of how the reliability product is specifically defined, which will be discussed in section 4 below, the valuation mechanism used to provide a suitable incentive for market agents may be set up in a number of ways. The various options are enumerated and described in decreasing order of liberalization, i.e., from the design with the lightest regulatory content to the alternative with the heaviest. As the following discussion will try to illustrate, the key determinant in the suitability of a given alternative is market structure, namely the degree of retail deregulation, (horizontal and vertical) market concentration and the effectiveness of distribution/generation unbundling (when the distribution company acts as regulated retailer).

Pursuant to this criterion and leaving the decision to essentially take no action whatsoever aside, three alternatives can be distinguished:

3.1 The regulator imposes demand-side purchase of the reliability product

Capacity markets

In the first, to counter the effects of the absence of long-term hedging, the regulator simply obliges market agents to buy/sell a product that (supposedly) ensures security of supply in the mid and/or long term (firmness or adequacy or both). To this end, once the reliability product is defined, the regulator must establish the amount of the demand-side purchase, i.e. the desired reserve margin, and at the same time the amount of the asset that each generator may trade on the resulting market. In other words, the regulator must determine each generating unit's so-called "firm supply".

Examples of this approach are the capacity markets formerly in place in some of the pools in the north-eastern US -PJM, NYPP, the ISO-NE approach, in force until the beginning of 2008, see Stoft (2000)- and in many Latin American markets, such as Chile, Guatemala or Panama.

These capacity markets have been criticized for being insufficiently effective, and more specifically for providing a weak incentive for either reliability-oriented operation or new entrants. The reason that the mechanism has not worked properly lies not in the way the reliability product is traded (on the capacity market) but in the nature of the product itself. The obligations typically associated with these markets have failed in the two aspects of long-term reliability: as far as firmness is concerned, the agreements typically established too few commitments for generators[3] and with respect to adequacy, their duration was not long enough to provide a real incentive for new investments.

The weakest point of this approach, however, even if the product were properly defined, is that it fails if the market structure is not sufficiently competitive and the generation and distribution businesses not sufficiently unbundled. Indeed, where the separation between one of the generators and the regulated retailer – i. e. the distribution company – is imperfect, the latter might not be sufficiently keen on buying at the best price. In such cases, the regulator must be able to monitor purchases and the way they are reflected in tariffs.

Public auctions

Regulators may institute a public auction in lieu of the capacity market to imbue trading with greater transparency. This is not the only aim of this approach, however. Another key reason for organizing centralized long-term auctions is to facilitate the entry of new sizeable generating facilities. In systems where project financing is only feasible if a significant proportion of the investment is hedged via long-term contracts, potential investors in large-scale plants must negotiate with several small retailers and industrial and commercial consumers. This kind of centralized auction seeks to facilitate this task by aggregating the purchases of several regulated retailers and allowing the system to benefit from the resulting economies of scale. Such a scheme is in place in Brazil (Barroso et al., 2006) and in the process of institution in other countries, including Peru and Guatemala.

3.2 The regulator purchases the reliability product on behalf of system demand

The second alternative involves a greater role for the regulator in the capacity market, not only establishing the obligation for retailers to enter into long-term contracts, but also assuming the responsibility of buying a prescribed volume of contracts on behalf of demand. This is one of the two main contributions of the so-called reliability options mechanism, first outlined in Pérez-Arriaga (1999)[4] and subsequently developed in greater detail in Vázquez et al. (2000) and Vázquez et al. (2003)[5].

This mechanism aims to equalize the obligations to be met by regulated retailers, and therefore consumers still under the regulated tariff, and liberalized consumers -often only large consumers and sometimes new retailers-. Otherwise, the latter could easily enjoy a “free ride”. This constitutes a covert cross-subsidy: regulated consumers bear the costs of guaranteeing the entry of new investments -since country and regulatory risk is significant

---

[3] In some cases, they were allowed to export instead of selling into the pool, even as demand was being rationed.

[4] What are now termed “reliability options” were called “price risk-hedging contracts” in Pérez-Arriaga (1999).

[5] In addition to certain special characteristics of the reliability option contract, mainly lag period and long-term duration, the other main contribution was to consider the wholesale market price as the reference to identify system scarcity.

in certain systems, the rates of return may be quite substantial-. Such new investments contribute to the improvement of overall system reliability while at the same time often lowering marginal prices, so consumers who are not compelled to participate in the auctions can avert these costs and benefit for free. In certain markets, this asymmetry has been instrumental in encouraging large customers to by-pass the distributor, i.e., to abandon the regulated tariff. Since regulated customers are the ones that bear the highest costs of the contracts concluded with new facilities, some customers decide to buy directly on the market where prices decline with the entry of new players, thereby averting the fixed costs – including the risk premium – built into such contracts[6].

The regulator's intervention may be limited to purchasing only long-term reserve capacity -either by directly buying peak generators as in Sweden, or by hedging under contracts such as New Zealand's *Dry Year Reserves* mechanism (MED-NZ, 2003)-, or cover full system demand plus a reserve margin -as in the reliability mechanism proposal (Batlle et al., 2007) included in the Spanish White Paper by Pérez-Arriaga et al. (2005) or in the new mechanisms in force in New England, see Bidwell (2005) and Cramton and Stoft (2005)-.

While contracting reserves solves the adequacy problem, it stymies market mechanism development. Firstly, it splits the system in two, with reserve units on one side and the remaining units on the other. The former bear much less risk and therefore constitute a more appealing investment. Moreover, while such market segmentation ensures peaking unit remuneration, it distorts the marginal signal that the other generating units require to recover their investment. In the long term this policy will probably impact investment in base-load plants, and the regulator, through the System Operator, will have to steadily increase its intervention. Covering the full system demand plus a reserve margin solves this problem.

The regulator implements a capacity payment

Under the third and last alternative, generally known as the "capacity payment", the role of regulation is increased to the point that the price of the "reliability product" is determined not by the market, but directly by the regulator.

This more interventionist approach arises in response to the absence of a market structure able to guarantee a sufficiently competitive price for the reliability product. The Spanish "capacity guarantee mechanism" in force since the dawn of the market in 1998 is one of the most prominent examples of this solution (Pérez-Arriaga et al., 2006). A recent

[6] As noted, while this may on occasion constitute covert support for industrial consumers, in some cases regulatory decisions appear to have precisely the opposite effect. The security of supply mechanism currently in force in Argentina is rather peculiar: the Kirchner Administration has decided that industrial consumers should be expected to ensure at least their own system adequacy, and so have been administratively and unilaterally declared "interruptible" when scarcity arises (last winter, for instance, they were systematically shed off). They have likewise been persuaded that they must assume responsibility for their own supply by seeking new investments in generation. This is indisputably an innovative and paradoxical way to support the country's economic growth.

proposal put forward by the Council of Regulators of the Iberian Market[7] (Batlle et al., 2007c), like the one developed for the Spanish White Paper, envisages two capacity payments determined by the regulator, a firmness and an adequacy payment, and an auction mechanism in case the latter proves to be insufficient to attract new entrants.

This approach provides for an incentive in a concentrated market, but obliges the regulator to deduce a suitable price, which is in essence an unsolvable problem. As in the case of any other "feeding tariff" mechanism, there is no assurance that it will be sufficiently appealing to attain the existing capacity to generate the desired reserve margin (firmness) or to encourage new entrants (adequacy), or conversely, that it will not lead to excessive investment (as in the case of Peru discussed above).

Be it said in summary that none of these alternatives can be regarded to be better than the others, although in principle the fewer decisions that have to be made by the regulator as a stand-in for market forces, the better. It may be deduced from the foregoing that the suitability of the final design will depend on the specific features of a given electricity market: demand maturity, unbundling, regulated tariff design, and horizontal and vertical concentration.

4 THE RELIABILITY PRODUCT: DESIGN PRINCIPLES

The aim of any capacity mechanism is to provide an added incentive for market agents to raise their contribution to overall system reliability, and more precisely to system firmness and adequacy. The definition of the reliability product that is subsequently priced to establish the incentive lies at the core of any such solution.

The components of the product definition include: a rule to determine what constitutes a critical period; a criterion to evaluate each generating unit's real contribution during such periods; the details characterizing the reliability product, such as the term of the commitment, i.e. the time the incentive will be in place; and finally a series of additional safeguards, primarily intended to avoid "gaming", free-riding and credit risks. These questions are reviewed below.

4.1 Critical period indicator

The first step is to design a non-arbitrary rule to identify near-rationing conditions when assessing unit availability. After lengthy debate, the spot market price[8] would ultimately appear to be the best indicator of the existence of critical situations, as originally proposed in the reliability options mechanism mentioned above. The other alternatives are clearly less suitable:

- Peak demand: the hourly intervals throughout the year when demand is highest are not necessarily the times when the reserve margin is tightest.

[7] The Council of Regulators of the Iberian Market gathers the Spanish and Portuguese stock markets and energy national commissions.

[8] By "spot market" we mean the mechanism reference market, the day-ahead one, as discussed in Vázquez et al. (2003).

- Unavailability: conditions need not necessarily be critical during the period when the amount of unavailable installed capacity is largest.
- Dry season: the season when water reserves are lowest need not necessarily concur with the most critical periods.

An example of this approach would be the scheme presently in force in Guatemala (Batlle et al., 2007c), whereby system units' contributions are calculated as their average production in the four peak hours of working days in what it is known as the dry season, from December to May. One example of the problems that this approach poses is to be found in the bitter debate that arose in that market, with some units arguing that the criterion should be five rather than four hours per working day, because of the significant difference that would make in their remuneration.

Linking a capacity incentive to this kind of criteria often leads to undesirable results. The Colombian experience in the past is a case in point. The critical period used as a reference for reliability remuneration was the dry season. Since the payment was significantly higher than the spot price, hydro units were given an incentive to maximize their hydro reserves in that season, which led to inefficient reservoir management (obviously this implied a minimization of the probability of scarcity, but at a completely ludicrous cost.

One alternative might be to use some other measure of reserve margin defined by the System Operator or to allow the Operator to announce that the system is nearing a critical situation a few days in advance; Italy's supposedly transitory mechanism would be an example of this approach (CEER, 2007). Such measures are undesirably arbitrary, however.

Disregarding potentially oligopolistic behaviour, and taking into account the actual peculiarities of power markets discussed at the start of this paper, the best -and indisputably most market-oriented- indicator of an impending scarcity episode is an abnormally high market price. Obviously, asserting that this is the best solution is not tantamount to saying that it is perfect. Some of the problems that arise around its implementation in certain power markets are described in the following section. Moreover, to avoid free-riding issues, demand is not allowed to participate fully in the market game (Vázquez et al., 2003). Such considerations aside, however, there is no apparent market rationale for a very high price in the absence of a generation shortage, nor any reason why an available generating unit would refrain from producing under those conditions, as seen in the following section.

Consequently, the regulator must determine a regulatory price to draw a boundary between the "normal operation" and the "near rationing" segments of the market. As this is an *ex post* measurement, units are obliged to identify, in advance, the hourly intervals when the system will be strained. While this entails an obvious risk for generators, they are able to manage it appropriately, and the scheme precludes the distortion that may be introduced by ex ante forecasts, as illustrated by the Colombian experience just mentioned.

4.2 Measuring generator unit compliance in critical periods

After designing the rule to clearly identify critical periods, the regulator must devise a way to assess each unit's contribution under such circumstances. At first glance, the procedure would appear to be as simple as checking unit availability under critical conditions. Unfortunately, this is not a straightforward task.

Traditionally, availability has been subject to rather light-handed supervision, in which generators' good faith was mostly taken for granted. No mechanism is usually at hand to verify the accuracy of generators' availability statements. When part of their remuneration depends on these data, however, supervision must be reinforced to prevent gaming. Indeed, unavailability can be rather readily masked by simply bidding at a high price on the spot market, which significantly reduces the chances of dispatch. Moreover, the higher a unit's costs, the easier it is to conceal unavailability.

In markets where capacity payments are or have been in place, some attempts have been made to counter blatant "gaming". In Spain, generating units had to produce power for at least 480 hours a year to be entitled to the capacity payment. In some Latin American countries (Brazil and Guatemala, for instance), the regulator conducts random inspections to ascertain the actual availability of generating units. Such measures, especially the one implemented in Spain, create greater problems than they were intended to solve, for they interfere with market functioning, obliging a number of expensive plants to generate when they are not needed while by no means guaranteeing that they will be available in critical periods.

Obviously, the only way to check whether a unit is available to "see" it producing. This is the approach originally adopted in the reliability options design, which provides that when the spot market price is over a certain threshold, the generating units receiving the reliability option premium are committed to produce. Whenever the spot price exceeds the strike price, generators having sold their options would have to reimburse consumers for at least the difference between the two (additional penalties might be also considered), regardless of whether or not the unit is producing[9]. This rule works because the proposal is to set the price above the variable cost of the most expensive generator expected to be in operation.

Unfortunately, however, real life sometimes renders regulatory design much more complex than described above. For instance: what happens when certain generating units' variable operating costs are higher than the scarcity threshold set by the regulator? Take for example the very poorly adapted power markets in Latin America mentioned above, characterized by large numbers of ultra-expensive generating units. A reliability option would be the best alternative to provide additional remuneration and at the same time a firmness incentive for new and efficient generation, but it would also benefit the inefficient units already installed or any others of the same type that investors might be

[9] To prevent "gaming" (a generator that sells power on the daily market and buys it back in subsequent markets to mask unavailability), the commitment covers all prices from the daily market to the real time, i.e. all the power sold in shorter term markets such as intradaily or ancillary services markets, see Vázquez et al. (2003) for a detailed discussion.

considering. If efficient generating units are ultimately installed, many of the other plants will in all likelihood never be dispatched; the risk for the system would therefore be that they would earn the reliability option premium in exchange for almost nothing. Indeed, the scarcity price might often be below their operating costs[10], and even in this case these generators might be willing to sell their reliability option claiming that they would bear the opportunity cost of occasionally having to operate below their operating costs. This situation is by no means acceptable. In most such electricity markets, the schedule is calculated by the Market Operator based on (supposedly) audited costs, so there is no actual way to commit a unit if the system marginal cost (price) is lower than the unit's costs (the units can not self-commit in any case). To date, then, there appears to be no "least worse" solution to this dilemma than setting an artificial boundary between the generating units the regulator wants and those it does not want to reward. This is obviously is a complex and controversial task.

A similar problem arose when a new capacity mechanism was being designed for the Iberian power market. The initial scheme envisaged a capacity payment in exchange for a reliability option. However, on the Spanish market at least, electricity generation units are 'obliged to make economic bids' (CNE, 2005), which the regulator interprets to mean that they must bid their marginal operating cost for each scheduled interval. Coupled to this requirement, the reliability options mechanism would end in paradox. The strike price (the scarcity price) may not be set higher than the most expensive generator's marginal variable cost, for according to the regulator's interpretation of the legislation, no such price can ever be reached. Given that the regulator is not considering any change in the bidding rule, the solution to this problem proposed by the Council of Energy Regulators (Batlle et al., 2007c) is: on the one hand to maintain the market price as the indicator, while setting as the so-called "scarcity price", the price of one of the most expensive units (fuel-fired plants, not necessarily inefficient such as described above for some Latin American markets); and on the other to strike out the generators' obligation to return the difference between the market price and the established threshold if they are not producing electricity. The generator's penalty consists in having to reimburse part of the capacity payment and accepting recalculation of its unit's firm capacity, on which its future remuneration will be based. This is little more than a dynamic methodology to set the artificial frontier described in the preceding paragraph.

4.3 The discreet return to centralized planning via the definition of reliability product details

In most cases, at least explicitly, regulatory intervention is intended to attract efficient -meaning reliable and obviously the cheaper the better-– generation, with no detriment to any manner of generation technology. The aim is to enhance availability and investment while at the same time presumably attempting to minimize regulator-mediated alteration of the market-driven technology mix[11]. There is ample proof, however, that regulators'

---

[10] The variable operating cost of such units often ranges from $150 to $300 per megawatt hour.

[11] Although peaking plants were the technology that was expected to suffer most from market failure and therefore the one that would be the most reluctant to join the system, the original

decisions on product details inescapably lead to a specific mix and hence constitute a significant market intervention.

The most relevant of these details is probably the timing of the commitment associated with the capacity mechanism, whether it be a reliability (option) contract or a capacity payment. This is briefly illustrated below. That discussion is followed by a short comment on a similar effect that may be exerted by the reliability option strike price (or scarcity reference price in the case of capacity payments).

Timing in reliability contracts or payments

In accordance with the definition given in the original reliability options design, duration of the commitment is divided into two terms:

- The lag period, defined as the time between when the commitment is effective and when it is operational, i.e. the time allowed for investors awarded capacity contracts or payments to build their plants.
- The obligation period (or maturity in the case of contracts), which starts when the lag period ends and during which the plant commits to produce power when the market price exceeds the reliability mechanism reference price.

The definition of the obligation period and even the lag period constitutes an implicit selection of technologies by the implementation scheme adopted.

In some ways, the reliability mechanism subject to the greatest uncertainty is the traditional capacity payment. On the one hand, the lag period is zero, *de facto*, and on the other, since it constitutes a mere regulatory commitment, the regulator may be expected to revoke it with little or no advance warning. That situation has spawned extremely undesirable results in some cases, particularly in Latin America. Such regulatory uncertainty reduced the incentive for capital intensive base-load investments and prompted the entry of far too many “ultra expensive” peaking units, characterized by minimum investment costs.

The new capacity payment mechanism proposed for the Iberian Market (Batlle et al., 2007c) envisages two different payments, both linked to the commitment of producing when the spot market price is above a predetermined “scarcity price”:

- a firmness payment for all the units in the system, subject to a commitment on the part of the regulator to announce any change one or two years in advance, so generators adjust unit management accordingly (reservoir or gas contracts, for instance);
- an adequacy payment for new facilities only, assured for five to seven years -to be determined-, also subject to the commitment to announce any change at least two years in advance; this would constitute a kind of lag period, since generating units

capacity payments were supposed to compensate all technologies, since all were risk averse or subject to revenue loss due, for instance, to explicit or implicit price caps.

installed less than two years after the announcement of the change are entitled to remuneration for the five- to seven- year period provided[12].

Defining the duration in the above terms may impact the outcome by prompting some regulators to adopt the opposite approach, i.e., to try to determine the duration of reliability contracts that would attract the kind of technology they prefer for their system. The Brazilian experience provides an example of this. In addition to other more explicit tools considered to steer the mechanism toward specific results, such as introducing "handicaps" for certain technologies at auctions, to capitalize on the country's huge hydroelectric potential, the Brazilian regulator auctions so-called "energy call options" (Bezerra et al., 2006) with three- and five-year lag periods and a maturity of from fifteen to twenty-five years[13].

Another subtle way of steering the capacity mechanism toward certain results is based on the reference (strike, scarcity) price definition. The closer this level is to a certain technology's variable costs, the smaller is its price risk compared to others. Setting a strike price of around $50 per megawatt hour improves the position of base-load investments, for instance.

The main conclusion that can be drawn is that between the energy-only market, paradigm and initial objective of the liberalization process, and the former centralized planning based on power purchase agreements, there is a large grayscale of security of supply or reliability mechanisms. The latter in turn are the tools to which some markets, mainly in Latin America, are resorting to retard or even reverse deregulation, in an attempt to remedy the sometimes desperate lack of capacity to which they are presently subject.

4.4 Additional design details

In addition to the three issues described in the foregoing, defining the reliability product calls for a whole series of other relevant decisions, none of which is straightforward. Depending on the particular characteristics of a given power system, the regulator must deal with a host of what might be referred to as safeguard conditions, geared primarily to

---

[12] In the beginning of September, days before the final version of this paper was completed, the Ministry of Industry, Tourism and Trade of Spain has developed a primary draft of the Royal Decree (the final version will be in force by October 1st), containing the final details of a new mechanism that has very little to do with the one recommended by the Council of Energy Regulators three months before. It basically consists of allowing the System Operator to bilaterally contract reserves one year in advance plus a "plain vanilla" capacity payment for system units -apparently linked to any availability commitment whatsoever- during their first ten years of operation the ones already installed, i.e. all the Combined Cycle Gas Turbines, installed under the liberalized scheme, and the ones to come.

[13] Unfortunately, despite all the effort deployed in auction and long-term contract design to encourage investment, the adequacy problem appears to be far from being solved. According to the results of the latest auctions for new entrants, of the many projects that qualifying to participate, only a very small number have been awarded an "energy call option", since most of the bids were higher than the safeguard prices determined by the regulator.

preventing gaming, free-riding and credit risks. These conditions range from determining how to deal with imports to defining a reference market -see Vázquez et al. (2003) for a detailed discussion of both-, while establishing rules for dealing with credit risk, designing auctions where appropriate or otherwise calculating the capacity payment, laying down penalties for non-compliance, deciding whether certain types of consumer can take part in the mechanism by offering to shed net load and so on. Each of these issues merits full development in a separate paper. The sections below summarize a few ideas on how to estimate a unit's "firm supply" (or "firm capacity" as it might be termed, see above), see Batlle et al. (2007b) for a more comprehensive discussion.

Firm supply

Regardless of the mechanism in place for pricing the reliability product, the regulator needs a methodology to evaluate the contribution to system reliability that each generating unit is reasonably able to provide to be able to calculate the maximum amount that each generating unit is allowed to offer (in a market-based mechanism) or be remunerated for (in the event of an administrative capacity payment).

This need is obvious under the capacity payments scheme, inasmuch as it is the parameter that defines a unit's remuneration rights. In other more market-oriented schemes, however, it would not be necessary, in principle. Indeed, one of the main objectives and advantages to such schemes is that they leave it up to market agents themselves to decide their own ability to contribute to the system reliability. It is nonetheless most recommendable, for these mechanisms are highly vulnerable to free riding.

Traditionally, this has been a concern only in power systems with a significant proportion of hydro plants, but this situation is changing due to the rise in wind production in some countries and to the growing proportion of gas-fired plants, which can be increasingly viewed as a new sort of limited energy plants, comparable in this regard to hydro stations in light of international and sometimes national gas market spasms. Spain is a paradigm in both respects.

Another key difficulty lies in the fact that this value depends crucially on mechanism design itself and most particularly on the magnitude of the economic incentives deriving from it. In principle, theoretically at least, almost any unit would plan its maintenance and manage its fuel stocks to maximize its net capacity at any given time, if the price is right. Conversely, the higher the penalties for non-compliance and credit risk hedges, the smaller is the need to be strict on defining the maximum value each unit can trade or ask to be remunerated.

Evaluating "firm supply" therefore entails making assumptions about unit management and planning criteria. Two approaches may be adopted in performing this task (Batlle et al., 2007b):

- *Ex ante*: In Latin America, the regulator typically runs a long-term stochastic model and attempts to estimate what generating unit dispatch will look like in critical periods.

This approach is obviously questionable, but relies on the fact that planning decisions are still centralized in these markets, so the same model that is used to run the system is used to perform the calculations (Granville et al., 2003). Inevitably, the regulator's assumptions when performing the calculations (scenarios considered, hydro inflows, plant failures and their probabilities) are permanently subject to heated controversy.

- *Ex post*: Consists of updating the parameter value on the grounds of previous years' results. Where remuneration is adequate and the algorithm to update the value is sufficiently sensitive, this may be an incentive for generators to improve their performance.

  The Spanish capacity payment provides an example of this approach. Since to date ″firm capacity″ has been calculated monthly as a sort of mean between installed capacity and the energy produced over the last five years, generators were afforded virtually no incentive to change their plans. This is one of the shortcomings that the current restructuring aims to remedy, by updating this value on the grounds of units' production in critical periods only (Batlle et al., 2007c).

5 CONCLUSIONS

Capacity mechanisms, which started out in life as a regulatory addition, can currently be regarded to be the vehicle that will drive the second phase of the deregulation process, in which regulators apply the lessons learnt in the past to change or steady the chosen course. This paper reviews the major issues around such mechanisms, drawing from cumulative worldwide experience.

6 ACKNOWLEDGMENT

The authors sincerely thank the helpful and brilliant discussions received from Virginie Pignon and Steven Stoft.

7 REFERENCES

Barroso, L. A., B .Bezerra, J. Rosenblatt, A. Guimarães, M. V. Pereira, 2006. Auctions of Contracts and Energy Call Options to Ensure Supply Adequacy in the Second Stage of the Brazilian Power Sector Reform. IEEE PES General Meeting 2006, Montreal, Canada.

Batlle, C., C. Vázquez, M. Rivier, I. J. Pérez-Arriaga, 2007. Enhancing power supply adequacy in Spain: migrating from capacity payments to reliability options. Energy Policy, 35 (9), pp. 4545-4554, Sep. 2007.

Batlle, C., M. Vázquez, P. Rodilla, 2007b. Principles and criteria to estimate the contribution of the groups to the power system firmness: the not-so-well-known "firm capacity". IIT Working Paper IIT-07-022I, available at www.iit.upcomillas.es/batlle.

Batlle, C., C. Solé, M. Rivier, 2007c. A new security of supply mechanism for the Iberian Market. IIT Working Paper IIT-07-023I, available at www.iit.upcomillas.es/batlle. To apper in The Electricity Journal.

Bidwell, M., 2005. Reliability options. A market-oriented approach to long-term adequacy. The Electricity Journal, Vol. 18, Issue 5, June 2005.

Cámac, D., V. Ormeño, L. Espinoza, 2006. Assuring the efficient development of electricity generation in Peru. Power Engineering Society General Meeting, 2006. IEEE, 18-22 June 2006 Page(s):5 pp.

CEER, Council of European Energy Regulators, 2006. Survey of Capacity Support Mechanisms in the Energy Community. Ref: C06-SEE-15-06, 6 June 2006. Available at www.ceer.eu.

CNE, Comisión Nacional de Energía, 2005. Spanish Electric Power Act. Unofficial English translation. Comisión Nacional de la Energía, Volume 7, 3rd Edition, 2005.

Cramtom, P. and S. Stoft, 2005. A Capacity Market that Makes Sense. The Electricity Journal Vol. 18, Issue 7, Aug./Sept. 2005.

Cramtom, P. and S. Stoft, 2007. Colombia Firm Energy Market. Proceedings of the Hawaii International Conference on System Sciences, January 2007. Available at www.cramton.umd.edu.

De Vries, L J., 2007. Generation adequacy: Helping the market do its job. Utilities Policy 15, 2007) 20-35

Eurelectric, 2006) Security of Electricity Supply. Roles, responsibilities and experiences within the EU. Working Group on Security of Electricity Supply, Union of the Electricity Industry (EURELECTRIC), Ref: 2006-180-0001, January 2006.

Finon, D. and V. Pignon, 2006. Electricity and Long-Term Capacity Adequacy. The Quest for Regulatory Mechanism Compatible with Electricity Market. Working Paper LARSEN 2006-2bis, available at www.gis-larsen.org.

MED-NZ, Ministry of Economic Development of New Zealand, 2003. Discussion Paper: Reserve Generation. Available at www.med.govt.nz/.

S. Granville, G. C. Oliveira, L. M. Thomé, N. Campodónico, L. A. Barroso, M. L. Latorre e M. V. F. Pereira. Stochastic Optimization of Transmission Constrained and Large Scale Hydrothermal Systems in a Competitive Framework. Proceedings of the IEEE General Meeting 2003,Volume 2, Toronto, Canada.

Pérez-Arriaga, I. J., 1999. Reliability and generation adequacy. In Reliability in the new market structure (part 1), by Read, W. S., W. K. Newman, I. J. Pérez-Arriaga, H. Rudnick, M. R. Gent, A. J. Roman. IEEE Power Engineering Review 19 (12): 4-5.

Pérez-Arriaga, C. Batlle, C. Vázquez and M. Rivier, 2005. White paper for the reform of the regulatory scheme of the power generation in Spain. (In Spanish) for the Ministry of Industry, Tourism and Trade of Spain.

Pérez-Arriaga, I. J., C. Batlle, C. Vázquez, M. Rivier, 2006. Diagnosis of the White Paper for the reform of the regulatory scheme of the power generation in Spain. IIT Working Paper IIT-07-003, available at www.iit.upcomillas.es.

Pérez-Arriaga, I. J., 2007. Security of electricity supply in a short, medium and long-term perspective. To appear in European Review of Energy Markets, Special Issue on Security of Energy Supply.

Stoft, S., 2000. PJM's capacity market in a price-spike world. University of California Energy Institute. POWER Working Paper PWP-077, May 2000. Available at http://www.ucei.berkeley.edu/ucei/.

Trebilcock, M. and R. Hrab, 2006. Electricity Restructuring in Canada, in Sioshansi and Pfanffeyer (eds.), Electricity Market Reform: An International Perspective (Elsevier).

Vázquez, C., M. Rivier and I. J. Pérez-Arriaga, 2002. A market approach to long-term security of supply. IEEE Transactions on Power Systems 17 (2): 349-357.

Vázquez, C., C. Batlle, M. Rivier and I. J. Pérez-Arriaga, 2003. Security of supply in the Dutch electricity market: the role of reliability options, for The Office for Energy Regulation (DTe) of The Netherlands. Presented at the workshop CEPR Competition & Coordination in the Electricity Industry, Toulouse, January 2004. Available at www.iit.upcomillas.es/batlle.