IIT Working Paper IIT-10-XXX. First version: March 2010. This version: May 2010.
Submitted to Journal of Regulatory Economics. June 2010.

OPTIMAL INCENTIVES IN THE CONTEXT OF THE SO-CALLED SECURITY OF GENERATION SUPPLY MECHANISMS

P. Rodilla & C. Batlle

Under the market-oriented paradigm, the electric power systems regulation must ensure that the appropriate incentives exist in order to keep the quality of supply at social optimal levels. This is particularly relevant at the generation activity.

When the regulator considers that the market incentives are proven to be insufficient to guarantee some minimum standards, additional security of supply (SoS) mechanism are deemed necessary. Generally speaking, these mechanisms call for the definition of a product to be purchased from generators. This product, which will be referred to as SoS-oriented product, can take many different forms (for instance a capacity credit, operational reserves, a reliability option, a base load energy profile, CO2 emission allowance, etc), and may imply either a physical of financial commitment. By designing and acquiring this product, the regulator seeks to introduce additional signals, which lead the electricity system performance towards the considered as optimal solution that the market is not providing.

In this paper, after briefly introducing the different dimensions of the security of supply problem, and discussing the need to define a metric to evaluate the market performance at each of them, the paper focuses on the analysis of two major issues: first, we study from the theoretical point of view the optimal pricing principles in the context of the security-of-supply-oriented mechanisms and the SoS-oriented products, and second, we show how by introducing any type of additional mechanism, the regulator retakes to a lesser or greater extent its traditional role, since any additional mechanism conditions the agents decisions (scheduling, resource management and investment). A stylized mathematical model is used to support these analyses.

Keywords: Electricity markets · Security of generation supply · Regulatory intervention

JEL Classification L11 · L50 · L51

1 INTRODUCTION

Under the market-oriented paradigm, the electric power systems regulation must ensure that appropriate incentives exist in order to keep the quality of supply at social optimal levels for each activity. This is particularly relevant at the generation activity, where the liberalization process has been more pronounced.

Since the very beginning of the power systems reform process, one of the key questions posed has been whether the market, of its own accord, is able to provide satisfactory security of supply (SoS) at the power generation level, or if some additional regulatory mechanism needs to be introduced, and in the latter case, the subsequent question is which the most suitable approach to tackle the problem is.

* Technological Research Institute, Pontifical University of Comillas, Sta. Cruz de Marcenado 26, 28015 Madrid, Spain.

When the so-called "market reform" started, the expectation was that little by little market agents, especially demand, would be able to learn the market game and therefore no additional mechanism would be needed. But the reality nowadays is that security of electricity supply is more and more turning into a priority in the agendas of electricity regulators, since there is certain evidence that due to a number of factors, there is a market failure[1]. At the time of this writing, most power systems worldwide, as for example the case of UK, France, Ireland, PJM, Colombia, Brazil, Panama, Peru, etc., see (Batlle & Rodilla, 2010), have been recently implemented or are currently in the process of revisiting a mechanism to ensure SoS.

Generally speaking, the introduction of an additional regulatory mechanism calls for the definition of a SoS-oriented product (e.g. a capacity credit, energy forward, renewable energy, etc.) to be provided by generators. By directly purchasing this product on behalf of demand (or compelling demand to acquire it), the regulator seeks to lead the electricity system performance (operation, management, planning and/or investment) towards an optimal solution that the market is not providing.

In this context, and from the point of view of the regulator, this paper focuses on the conceptual analysis of these SoS mechanisms. Its major objective is to contribute to the discussion of how to design optimal incentives in this framework, and how these mechanisms may condition the final market outcomes.

Next, we briefly present the roadmap followed to make this analysis:

- First, in this introduction we set up the general framework in order to properly deal with the problem, identifying the different time dimensions in which the security of supply can be decomposed. This division on the one hand adapts well to the phases in which the exercise of the activity of generation of electricity has been traditionally divided into (operation, planning, investment and long-term expansion policy), and on the other hand, allows to better build the regulatory analysis that follows.
- Second, as a necessary previous step to the analysis of the design of additional regulatory incentives, we discuss the need to properly define a metric to evaluate the market performance. The regulator has to design a methodology to qualitatively and (as far as possible) quantitatively assess if (and to what extent) the market mechanism performance may not be yielding the expected/required outcomes. This is essential not only to determine whether or not an additional mechanism is deemed necessary, but as we show later, also to set the objectives and the incentives of such mechanism.
- Third, once a metric is defined, and the evaluation of the system performance according to it reveals a market insufficiency, the regulator's task is to put in place the means to provide optimal signals to lead the power market performance to (a closer to the considered by the regulator as) the optimal solution. These incentives are built around the definition of a SoS-oriented product to be acquired from generators. We introduce the nature of these products, and then we analyse, from the theoretical point of view, the optimal pricing principles in the framework of security-of-supply-oriented

[1] See for instance (Pérez-Arriaga, 2001), (Stoft, 2002), (Joskow, 2007), (Hogan, 2005), (Rodilla & Batlle, 2010).

mechanisms, in such a way that the additional incentives perceived by the generators actually contribute to improve the system performance according to the regulator's objectives.

- Fourth, we show how the introduction of any type of additional mechanism, implies that the regulator retakes to a lesser or greater extent its traditional central planner role with all its consequences, since by introducing any type of additional mechanism it conditions the market results (scheduling, planning and investments).

The four dimensions of the security of supply problem

From the time dimension perspective, it is helpful for the regulatory discussion to decouple the security of supply at the generation level into its four major components. Breaking down the central problem into its sub-problems facilitates not only its understanding but also the design of a regulatory mechanism (if required). These components (or dimensions) are the following:

- Security, a short-term issue, defined by the North American Electric Reliability Council as the "ability of the electrical system to support unexpected disturbances such as electrical short circuits or unexpected loss of components of the system" (NERC 1997).
- Firmness, a short- to medium-term issue, defined as the ability of the already installed facilities to supply electricity efficiently. This dimension is conditioned by the characteristics of the existing generation portfolio and the medium-term resource-management decisions of the generators (fuel provision, water reservoir management, maintenance scheduling, etc.).
- Adequacy, a long-term issue, defined as the existence of enough available generation capability, both installed and/or expected to be installed, to meet efficiently demand in the long term.
- Strategic Expansion Policy, which concerns the very long-term availability of energy resources and infrastructures. This dimension usually entails the diversification of the fuel provision and the technology mix of generation.

Bearing in mind these four and interrelated dimensions will help us in the following when analyzing the SoS mechanisms.

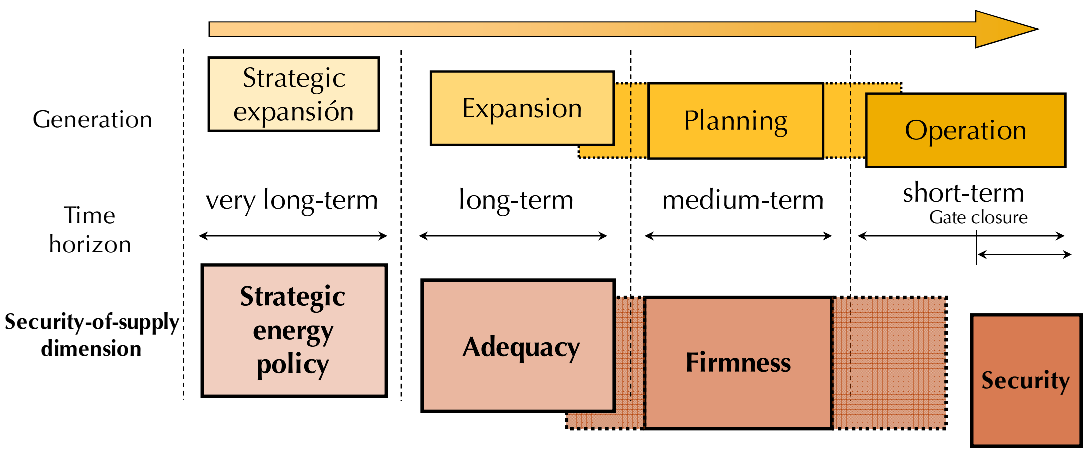

Figure 1. The dimensions of the Security of Supply (SoS) problem

Paper structure

After this introduction, the paper has been structured as follows: next, in section 2 we present the different available alternatives to define the metric to evaluate the system performance at the different security of supply dimensions. Then, in section 3, we review by means of a conceptual model, the principles on how to design incentives to optimally achieve the desired objectives when an additional mechanism is deemed necessary. In section XXX, we reformulate the traditional streamlined conceptual mathematical model to illustrate in a deeper detail the actual functioning of the additional regulatory mechanisms oriented to enhance firmness and adequacy. To do so, we explicitly represent the trading of the reliability product (the name the SoS-oriented product receives in this particular context). We also centre the discussion on the significant impact that this reliability product can have on the future configuration of the generating system. XXXThen, in section XXX, we focus on the most practical difficulties that arise when implementing the mechanisms based on purchasing the reliability product. Paradigmatic regulatory experiences implemented to date are presented so as to illustrate the main points analyzed in the previous discussion. We close the paper gathering briefly the main conclusionsXXX.

2 A METRIC TO EVALUATE THE SYSTEM'S PERFORMANCE

Defining the performance metrics in each of the dimensions of the SoS problem

The necessary first step in the process to evaluate the suitability of implementing a SoS mechanism, is to design a methodology to evaluate the market performance. In a market oriented environment[2], how to assess this performance metric is a not-so-clear task.

Although a general overall measure, which we could call Security of Supply Performance Metric ($SOSPM$), can be designed so as to represent a compact metric determining how far regulator's global objectives are from the current market outcomes, usually these measures (or metrics) are better expressed on a dimension by dimension basis. This way, in practice, it is easier to evaluate the system performance at the security level ($SECPM$), at the firmness level ($FIRPM$), at the adequacy level ($ADEPM$), and also at the strategic expansion policy level ($SEPPM$).

The regulator should (and often does) design methodologies and metrics to properly evaluate the system performance. Some examples include:

- if the System Operator, at gate closure, has enough reserves to maintain voltage and frequency within acceptable margins for the system,
- if the generating units in the system are being scheduled and managed (e.g. hydro reservoirs) so as to accomplish with any reliability standard,
- if the system expansion resulting from the market is properly evolving in such a way that long-term security of supply is sufficiently guaranteed,

[2] We assume here a market where the market signals are not artificially distorted by the regulator or the System Operator (see XXX). It would not make sense to introduce a mechanism in order to correct the effect caused by any unnecessary regulatory flawed rule or action.

- if the average emissions rate of the generating system meets the required standards, etc.

The regulator's objective: just securing the supply or maximizing the net social benefit?

In the liberalized context and from the perspective of the regulator, security of supply is undoubtedly one of the major present concerns. However, the meaning and implications of this "security of supply" concept are not so-clear. Traditionally, under this concept it was primarily understood the need to ensure system reliability (with at least a prescribed level of continuity and quality), assigning the optimization of the cost of providing the service a secondary role (and often not even taken into consideration)[3]. In the end, we consider that actually, in its broader and practical sense, ensuring security of generation supply is nothing but maximizing the net social benefit (NSB) at the power system level. Indeed, generally speaking, the four dimensions already presented are the sequential and interlinked levels at which the main problem of optimizing the net social benefit at the generation level can be decomposed.

From this perspective, the more general methodology to determine the performance of an electricity market at any dimension would be to compare its outcomes with those resulting from the ideal (theoretically optimal) system, which represents the most efficient design, management and operation from the net social benefit point of view. Note that comparing both systems (the real and the ideal benchmark) would imply analyzing all parameters that affect the overall system net social benefit, such as costs, prices, total consumption, utility obtained from that consumption, etc. Obviously, the closer the outcome, the smaller is the need to introduce an additional mechanism to reinforce market signals.

But, finally this discussion translates into a specific metric (or a number of them) in each dimension. And always to a greater or lesser extent the definition of these metrics are subject to an intense debate.

In the Security dimension, the $SECPM$ is usually established in terms of the system capability to provide the so-called operational reserves[4]. And defining the technical requirements of these reserves is an often hot topic of discussion, since for instance it means that a minimum ramp speed is required to generating units willing to provide the service, what often impedes some of the units installed to take part in the market for service.

On the other extreme, at the Strategic Expansion Policy dimension it is usual to introduce several metrics, such as CO2 or NOx emissions, energy production from renewable

[3] According to the definition on security of supply of Eurelectric (2006) discussed in (Pérez-Arriaga, 2007), 'security of electricity supply is the ability of the electric power system to provide electricity to end-users with a specified level of continuity and quality in a sustainable manner, relating to the existing standards and contractual agreements at the points of delivery'. Sustainability is an added target included in the definition, implying that a whole bunch of features are considered, being the compatibility with an adequate economic development one of them.

[4] It has been argued (XXXStoftXXX), that cost criteria should also be introduced in the determination of these requirements.

sources, etc. Again, these metrics are not free of certain controversy, for instance, it is often argued which technology should be considered renewable and which would not.

Although there has not been a full consensus about the metric in these two dimensions, (security and strategic expansion policy), they have not been among the main reasons behind the regulatory flaws that in some cases have taken place. This has not been the case, however, in the firmness and adequacy dimensions. This is discussed next.

Firmness and Adequacy performance metric

It has been in these two dimensions where the previously introduced discussion on the regulator objectives (just reliability of net social benefit) has been more acute. Indeed, in the adequacy and the firmness dimensions, the electricity system performance assessment has usually been determined by giving priority to reliability criteria. This way, in many cases the evaluation is just based on the reliability performance of the generating mix. This means that regulators, particularly in the first stages of the reform, have been much more concerned about the continuity of the supply service rather than about the actual cost involved in providing such service. This is the reason why the regulatory mechanisms focused on these dimensions have been often referred to as reliability mechanisms, and the product purchased in this context the "reliability product".

Hence, assuring the continuity of supply with a certain probability has been traditionally prioritized over the actual cost of providing this service up to these quality standards. The main reason behind this is that for regulators the Value Of Loss of Load (VOLL), which in theory it is assumed to represent the marginal demand utility, is still considered to be much higher (at least one order of magnitude larger) than any cost of production.

Among the reliability criteria traditionally used we can mention the Capacity Reserve Margin (the system capacity, often weighted by any sort of estimation of the average availability of the generating units, minus the peak demand consumption), the Energy Reserve Margin (the system energy reservoir minus the maximum expected energy consumption), the loss of load probability (LOLP), the loss of load expectancy (LOLE), the non served energy expectation (NSEE), etc., see (Wood & Wollemberg, 1996). For instance, the reliability criterion in PJM has been based on Loss of Load Expectation (LOLE) not exceeding one occurrence in ten years. The resource requirement to meet the reliability criterion is expressed as the Installed Reserve Margin (IRM) as a percentage of forecast peak load. Another example is Western Australia, where the objective is to limit energy shortfalls to 0.002% of annual energy system consumption.

However, this hypothesis of the marginal demand utility being much higher than the units' marginal production costs is not always an appropriate assumption. This fact is particularly relevant in certain real cases, in which either the generation side or the demand side deviate from this (up to now understood as) traditional underlying assumption. For instance:

- on the generators side, in some markets, and particularly this is the case of Latin American ones (Batlle & Pérez-Arriaga, 2009), there are some highly inefficient units installed presenting extremely high variable costs. These costs sometimes reach values which are explicitly perceived by the regulator as closer (or even above) the value of loss of load[5].

When this is the case, designing a measure and thus a mechanism based just on reliability criteria may in the end attract this type of inefficient units. This is the reason why when it came to implement any sort of additional security of supply mechanism, some systems have introduced the so-called efficiency criteria. We find some examples of this trend in Brazil, Guatemala, Bolivia or Ireland (all take into account either the production costs or the bidding price of each unit in the additional remuneration).

- on the demand side, as its participation in the market rises, an elastic curve expressing their consumption preferences progressively changes the paradigm of the inelastic offer at the VOLL costs. Thus, again, the marginal demand utility (or at least a part of it) may fall within the same range of values stemming from the generating costs.

  Indeed, if all demand side participated in the market, thus expressing their actual preferences, the classic concept of reliability (based on ensuring the supply of a certain inelastic consumption) would no longer make sense.

  Thus, from our point of view, if demand side becomes fully elastic, it would only make sense to evaluate system performance by means of the abovementioned analysis of the optimal benchmark.

3 OPTIMAL PRICING PRINCIPLES FOR SECURITY OF SUPPLY MECHANISMS

The objective of this section is to discuss from a theoretical point of view, the optimal incentives that should be provided by a security of supply mechanism. Using the classic formulation of the optimal pricing problem in electricity markets, we show how this remuneration should ideally be based on each unit's contribution to the regulator's objectives (expressed by means of the $SOSPM$ function). This fact highlights the importance of properly defining these objectives, which as previously described, are usually expressed on a dimension by dimension basis. Let us recall that these objectives as previously described, may include all type of efficiency criteria and not only reliability standards. This way, a reliable unit with extremely high costs may not contribute to the regulator's objective, and thus would receive a low incentive.

3.1 Optimal incentives in security of supply mechanisms

In (Schweppe et al., 1988) the conceptual foundations for the determination of optimal wholesale prices in an electricity market context were presented. Following the same line of thought, Pérez-Arriaga & Meseguer (1997) extended the discussion by not only analyzing the incentives provided by short-term energy prices but also those provided by

[5] In some cases this situation has led the regulator to limit the marginal spot price when these extremely high variable cost units are required. If a unit whose cost exceeds a threshold defined by the regulator is scheduled, it receives the bid price, but the system marginal price is not allowed to rise above such threshold. This was the case of the electricity market of El Salvador prior the reform launched in 2008.

some additional short- and long-term signals: prices for operational reserves, long-term incentives in the form of an extra payment provided to the installed capacity (MW), etc.

Here, we use a simplified formulation of the problem presented in this latter paper[6], and we slightly modify the conceptual framework as regards the long-term security of supply mechanism. There, in Prof. Pérez-Arriaga's work, an explicit reliability margin constraint is considered, assuming that the regulator could require a long-term margin larger than the amount strictly required under an economic standpoint. This long-term margin is defined as 'any measure of long term security of supply, i.e. reliability; it is assumed to be a function of the available installed capacity and the demand'. Accordingly, an explicit and externally fixed remuneration of the available installed capacity is proposed.

Here, we redefine the market context and the embedded security of supply objectives as follows:

- We represent an electricity market context in which the regulator has implemented any sort of additional mechanism aimed to guarantee a minimum level of performance, $SOSPM_{min}$ (Security of Supply Performance Metric minimum value). We consider that in order to reach these objectives, the regulator acts on behalf of the demand (since it represents the side of the market not really taking an active role). Thus, from an optimization model perspective, the demand and the regulator are represented as a single agent (the generators' counterparty).
- The embedded system performance objectives are expressed from a broader point of view by means of a Security of Supply Performance Metric that is not just reliability-based objectives, i.e. it is decoupled into the different objectives pursued at each dimension (that is, the metrics at the security, firmness, adequacy and strategy expansion policy dimensions). We first introduce the compact metric to ease the formulation ($SOSPM$) to latter explicitly represent its different components (the $SECPM$, $SEPPM$, $ADEPM$ and $FIRPM$).

As stated, we adopt the traditional problem formulation, so the Security of Supply Performance Metric is represented in a compact and general way as a function of the installed capacity of each type of unit, i, that is, $SOSPM = SOSPM(q^i_{max})$. In return, an additional incentive consisting in an individualized extra payment provided to generators proportional to the installed capacity (MWs) is assumed.

[6] The analysis that follows is based on the same simplifying hypotheses as those assumed in the demonstration which leads to the conclusion that short-term marginal prices provide optimal incentives for the efficient operation and investments towards the maximization of the system's overall efficiency, see (Batlle & Rodilla, 2009). Among others: generators' costs functions are convex, risk neutrality assumption, no economies of scale nor lumpy investments and perfect competitive market with perfect information. We leave aside important characteristics playing a key role in electricity markets, as for instance the inefficient allocation of risk that usually exists in the absence of regulatory intervention, see for instance (Rodilla et al., 2010). But anyway, the so obtained results allow us to give the flavor of the optimal incentives design problem, which is the main objective at this stage of the analysis.

Expressing the system objectives in terms of the just mere installed or available capacity has been traditionally the approach selected when modeling long-term system SoS requirements, see for instance the noteworthy models developed in (De Vries, 2004), (Ehrenmann & Smeers, 2008) or (Hogan, 2009). At the present stage of the analysis, we start from this simple modeling assumption to draw some general conclusions on the principles to design the incentives. Note that this is a rather simplifying hypothesis with the underlying assumption that a certain amount of installed MWs of a certain unit, univocally defines not only all the relevant parameters of that unit (reservoir capacity, for instance), but also the medium-term resource management carried out by that generator, the operational reserves provided to the system, etc.

Based on the conclusions derived from this simplified analysis, we latter propose a refined formulation that better represents the actual interaction in these type of mechanisms, based on the so-called SoS-oriented product upon which the additional mechanism is built. This new formulation will allow us to gain insight on the impact on market outcomes of these mechanisms.

The demand (and the regulator) model

The demand's problem in this context consists in maximizing the utility U_h^d (defined for each hour h) obtained from the total hourly consumption of electricity, Q_h, minus the costs derived from both:

- the energy purchased in the wholesale market (equal to the hourly spot energy price π_h times the total hourly consumption Q_h),
- and the extra payments provided to the generators, which are equal to the prices τ^i (which has been generally modelled as a remuneration to every installed MW on a unit by unit basis, i) times the corresponding installed capacity (q_{max}^i).

This way, the demand's problem can be represented as:

$$\begin{array}{l} \underset{Q_h, q_{max}^i}{Max} \; \sum_h [U_h^d(Q_h) - \pi_h \cdot Q_h] - \sum_i \tau^i \cdot q_{max}^i \\ subject\ to: \\ \quad SOSPM(q_{max}^i) \geq SOSPM_{min} \quad \perp \chi \end{array}$$

The model proposed intentionally lacks of a key constraint which is in many cases the "always-active" one: above any other well-intentioned objective which we have represented by the so-called SOSPM constraint, distressingly (for some, logically for others) there is always a dominant constraint reflecting the regulator's (i.e. the Government's) severe reluctance to allow electricity tariffs (or more precisely electricity service costs, since often the tariffs do not reflect the full cost of the service, as parts of it are deviated to the budget) increasing above a certain limit. We opted for skipping this issue in our analysis, since although it is probably the main feature needed to fully understand what it is behind many of the regulatory flaws inherent to the electricity business[7], it makes no much sense to try to include it: it is rather straightforward that when active, it leads the problem to an infeasibility.

The generators' model

On the other side of the market, the generators' problem consists in maximizing its own profit function, i.e. income minus costs. As previously mentioned, in the described context, there are two sources of income: the one derived from the energy sold and the one derived from the additional payments provided to each generating unit installed capacity (MW).

The hourly energy produced, q_h^i, the installed capacity, q_{max}^i, and the hour availability factor λ_h^i, are represented on a unit by unit basis. Each unit's hourly production cost, $C^i(q_h^i)$, depends on the energy produced. The units' associated investment costs, $IC^i(q_{max}^i)$, are represented as a function of the corresponding installed capacities.

$R_c^i(q_h^i)$ represents schematically the operation constraints relating the production of the unit i along the hours h considered in the time scope being studied. Some examples are constraints on minimum and maximum energy production.

This way, the generators' problem can be schematically written as follows:

$$\begin{aligned}
&\underset{q_h^i, q_{max}^i}{Max} \quad \sum_h [\pi_h \cdot \sum_i q_h^i - \sum_i C_i(q_h^i)] + \sum_i \tau^i \cdot q_{max}^i - \sum_i IC_i(q_{max}^i) \\
&subject\ to: \\
&\quad q_h^i \leq \lambda_h^i \cdot q_{max}^i \quad \perp \psi_h^i \\
&\quad R_c^i(q_h^i) \leq 0 \qquad \perp \zeta_c^i
\end{aligned}$$

The $SOSPM$ ***-based incentive***

The demand's decision variables are both the hourly consumption and the installed capacity of each generating unit. The optimality conditions of this problem are obtained by forming the Lagrangian function and deriving it with respect to the decision variables:

$$\frac{dU_h^d(Q_h)}{dQ_h} = \pi_h, \forall h$$

$$-\frac{\partial SOSPM}{\partial q_{max}^i} \cdot \chi = \tau^i, \ \forall i$$

Some interesting conclusions can be derived from this latter condition. We can rewrite the security-of-supply-based incentive perceived by each generating unit as:

---

[7] Obviously it is not the only one, often the reason why this constraint is active is the market structure *sin of Adam*. A poor market structure (an excessive integration of the electricity activities in all directions, horizontal, vertical, etc.) which the regulator cannot (or is not committed to) mitigate is behind the regulator's mistrust on the market functioning.

$$\tau^i \cdot q^i_{max} = -\chi \cdot \frac{\partial SOSPM}{\partial q^i_{max}} \cdot q^i_{max}$$

The optimal additional incentive that should be perceived by each generating unit depends on the value of the dual variable (χ) associated to the regulator's $SOSPM_{min}$ constraint, multiplied by the product of the installed capacity and the marginal contribution of each particular unit to the Security of Supply Performance measure.

Note that based on the traditional formulation of the problem, the incentive is interpreted:

- $(-\chi \cdot \partial SOSPM \,/\, \partial q^i_{max}) \cdot q^i_{max} \Rightarrow$ Price = $-\chi \cdot \partial SOSPM \,/\, \partial q^i_{max}$; Quantity $= q^i_{max}$

  Where the price incentive is individualized for each generating unit, and should be equal to $-\chi \cdot \partial SOSPM \,/\, \partial \overline{q}_i$. These individualized price incentives remunerate the installed capacity (MW).

However, note that rearranging the terms we can obtain a more adequate (and closer to the practical implementation of the mechanisms in force) interpretation of this incentive:

- $(-\chi) \cdot \partial SOSPM \,/\, \partial q^i_{max} \cdot q^i_{max} \Rightarrow$ Price $= -\chi$; Quantity $= (\partial SOSPM \,/\, \partial q^i_{max}) \cdot q^i_{max}$

  The long-term price used to provide generators with an additional signal is the same for all the different generating units and it is equal to ($-\chi$), but the quantity entitled to perceive the incentive, has to be normalized based on the marginal contribution of each unit to the regulator's objectives.

  Indeed, the security of supply mechanism (either price- or quantity-based) leads to a single price for the security-of-supply-oriented product, and the amount each unit delivers of this product (if properly defined) represents the normalized quantity just mentioned. This clearly suggests that installed capacity is not the variable expressing the degree of fulfillment with the regulator objectives, but the amount of the reliability product delivered instead. This is a strong result, once depending on the system characteristics (capacity and/or energy constrained) the contribution of each generation unit for the system reliability might not be its nameplate capacity

From these optimality conditions, it is important to remark the evident but important existing relationship between the regulator's objectives and the optimal incentives provided to each unit.

We now analyze in deeper detail the former expression, by decoupling the SoS metric into the metrics expressed in each of the four dimensions. This allow us to illustrate on the one hand how the incentive can be decoupled into dimension-based incentives and also to highlight the importance of not forgetting the fact that these dimensions are interrelated (an incentive in one direction can have a significant impact on others).

Optimal incentives explicitly representing the regulatory objective at each dimension

We now present a conceptual framework to decouple the general security of supply performance aims into the most usual particular objectives imposed at each of the dimensions involved. This disaggregation allows us to study how the different requirements affect the optimal remuneration a unit should receive.

The demand's problem can be now represented as[8]:

$$\begin{array}{l} \underset{Q_h, q^i_{max}}{Max} \sum_m \sum_h [U^d_{m,h}(Q_h) - \pi_{m,h} \cdot Q_{m,h}] - \sum_i \tau_i \cdot q^i_{max} \\ subject\ to: \\ SoS(q^i_{max}) \left\{ \begin{array}{ll} SECPM(q^i_{max}) \geq SECPM_{min} & \perp \chi_{sec} \\ FIRPM(q^i_{max}) \geq FIRPM_{min} & \perp \chi_{fir} \\ ADEPM(q^i_{max}) \geq ADEPM_{min} & \perp \chi_{ade} \\ SEPPM(q^i_{max}) \geq SEPPM_{min} & \perp \chi_{sep} \end{array} \right. \end{array}$$

Notice that we have reformulated the demand's SoS constraint on the basis of the performance metric constraints in each of the four dimensions. The optimality conditions now become:

$$\frac{dU^d_h(Q_h)}{dQ_h} = \pi_h, \forall h$$

$$-\frac{SECPM}{\partial q^i_{max}} \cdot \chi_{sec} - \frac{FIRPM}{\partial q^i_{max}} \cdot \chi_{fir} - \frac{ADEPM}{\partial q^i_{max}} \cdot \chi_{ade} - \frac{SEPPM}{\partial q^i_{max}} \cdot \chi_{sep} = \tau^i,\ \forall i$$

Hence, the additional price incentive provided to each generator can be decomposed in several terms. Each term can be interpreted as the particular incentive received by the generators with respect to each of the regulator's constraints.

Thus, the conclusions derived previously can then be generalized as follows: a different price signal can be associated to each objective, and these price signals take the value of the dual variables associated to the corresponding constraints. For each particular objective constraint, the corresponding price signal perceived by all the generating units is the same, but the quantity entitled to perceive this incentive has to be normalized based on its marginal contribution to the fulfillment of the objective being considered. Observe also that there is some overlapping between the contributions in different dimensions provided by a generation unit: a unit such as a coal-fired plant might contribute more to adequacy and firmness and less to a strategic policy view or decarbonizing the system. A renewable generator contributes, on the other hand, to the latter dimension but less to the security dimension because of its non-dispatchable characteristic.

[8] The performance of the system at any of the dimensions is still expressed as function of the installed capacities. Thus, the simplifying hypothesis, of installed capacity (MWs) of a certain unit being capable to explain the contribution of that unit at all levels still applies.

$$\tau^i \cdot q^i_{max} = \left(-\frac{SECPM}{\partial q^i_{max}} \cdot \chi_{sec} - \frac{FIRPM}{\partial q^i_{max}} \cdot \chi_{fir} - \frac{ADEPM}{\partial q^i_{max}} \cdot \chi_{ade} - \frac{SEPPM}{\partial q^i_{max}} \cdot \chi_{sep} \right) \cdot q^i_{max} =$$
$$= (-\chi_{sec}) \frac{SECPM}{\partial q^i_{max}} q^i_{max} + (-\chi_{fir}) \frac{FIRPM}{\partial q^i_{max}} q^i_{max} + (-\chi_{ade}) \frac{ADEPM}{\partial q^i_{max}} q^i_{max} + (-\chi_{sep}) \frac{SEPPM}{\partial q^i_{max}} q^i_{max}$$

Before continuing with the analysis, and developing a more detailed formulation of the problem based on the major results derived in this section, it is relevant to briefly highlight some relevant issues related to the design of these metrics and incentives.

In most cases, regulators design different mechanisms to tackle the different dimensions of the problem. The incentives introduced in a particular dimension, sometimes are designed disregarding the potential impact that has on the other dimensions objectives. Since the four dimensions are interrelated, it is essential to coordinate objectives. This way, if for instance a greater amount of renewable energy is desired for Strategic Expansion Policy criteria, it is important to take into account the impact it will have on system security, firmness and adequacy. If these objectives are not carefully coordinated, undesired and inefficient outcomes will arise.

Furthermore, if the regulator decides to change the objective in one dimension, for instance considering a more ambitious renewable objective, $SEPPM^*$, it does not only change the corresponding payment needed to fulfill the renewed objective, χ^*_{sep}, but also affects in particular the payments related to the other dimensions, $\chi^*_{sec}, \chi^*_{fir}, \chi^*_{ade}$, and more generally, the market outcomes of all the installed units in the system, since also the hourly energy produced by all the generating units, q^i_h, and the spot energy prices π_h change, so the income does not correspond with the one expected when for instance the decision of installing a new unit was made[9].

The conclusion is therefore that although we have thoroughly stated that decoupling the overall SoS into different dimensions eases the regulatory discussion and design, it is at the same time of utmost importance not to forget that they are all interrelated parts of a common objective function.

[9] This is for instance a hot topic of debate in the Spanish electricity market at the time of writing this analysis. The regulator has repeatedly changed the "Plan for the promotion of renewable energy", increasing significantly the target for RES share in the primary energy supply quota. Currently, in a complicated scenario due to the crisis that implied an abrupt drop of demand, generators owning a significant proportion of conventional thermal generating units in their portfolios claim that the intervention of the regulator in this dimension, $SEPPM^*$, puts the recovery of their investments at a clear risk, and claim for a compensation, for instance through an increase of the adequacy payment (χ^*_{ade}).

4 ANALYSING THE MECHANISMS FOCUSED ON ENHANCING THE SYSTEM FIRMNESS AND ADEQUACY: THE ROLE OF THE RELIABILITY PRODUCT

The previous conceptual results have shown how a SoS mechanism should be designed and remunerated. We now introduce some more detail in the modeling hypotheses to better represent actual SoS mechanisms implemented worldwide and how they may affect generation decisions. Two major refinements have been introduced in this respect:

- The explicit representation of the product being purchased in the mechanism. As we describe next, we also reformulate the model on the basis of the so-called SoS-oriented product
- The introduction of all generation decision parameters. As it was previously mentioned, expressing the system's requirements as a function of just the installed capacity turns to be a rather simple way to model the problem. We have reformulated the stylized model schematically representing all parameters and decisions involved on the generators side.

Before delving into the discussion, we briefly introduce the underlying product being purchased within the framework of a SoS mechanism.

The definition of a security-of-supply-oriented product

Generally speaking, all additional mechanisms require the regulator to define one or more security-of-supply-oriented products to be provided by generators. The characteristics of these products depend on the dimension in which the additional mechanism is focused.

The product definition includes several elements, such us the underlying asset (energy, renewable energy, capacity, capability to provide operating reserves, etc.), the financial characteristics (forward, option, etc.), the time terms, the guarantees (usually physical guarantees), etc.

Some examples in the security dimension are operational reserves (for instance reserves to restore frequency or black start reserves) purchased either in the short or in the long term. In the strategic expansion policy dimension one typical product is renewable production (remunerated through feed-in tariffs or renewable obligations, for instance), although there are also many others as energy efficiency related products (such as white certificates) or CO2 emission rights.

In the other two dimensions (firmness and adequacy) it is difficult to find in practice differentiated products to tackle individually the potential deficient system performance at each dimension. Indeed, traditionally no distinction has been made between firmness and adequacy, implicitly considering that both can be solved through a single mechanism and the same product(s). Although this is partially true, there are some proposals aimed to discuss the possibility of introducing separated incentives to secure each of the previous dimensions[10].

In the context of the mechanisms focused on solving deficiencies at the firmness and adequacy dimensions, the associated product is usually known as the "reliability product", see for instance (Batlle & Pérez-Arriaga, 2008), for it is usually considered to be oriented to increasing reliability. Among the possible different designs of this reliability product we find: capacity credits (PJM or Western Australia), energy forward contracts (Brazil, Peru or Chile[11]), strategic reserves (Nord Pool), reliability options (Colombia, Brazil or New England), etc.

The quantity of the reliability product delivered by each unit is referred in this work as *firm supply*, a concept that is always present in one way or another in most of the firmness- and adequacy-oriented mechanisms worldwide[12]. This concept appears under different and very varied names, for instance, it is termed "capacity credits" in PJM or Western Australia, "firm capacity" in Spain or Peru, "firm energy" in Brazil or Colombia, "adequacy capacity" in Chile, "efficient firm offer" in Guatemala, "guaranteed capacity" in Bolivia, "long-term firm capacity" in Panama, etc.

Next, for the sake of simplicity we center the discussion on those mechanism focused on enhancing the system adequacy and firmness. This way, the problem is reformulated introducing the trading of the reliability product.

The objective of the analysis that follows is to show how, by introducing these additional incentives, the regulator actively retakes to a greater or lesser extent part of its central planner role, significantly conditioning not just the amount of plain installed capacity, but also the nature (technologies) and characteristics (capacity, reservoirs, etc.) of the units that will enter the system (adequacy), as well as the way system generation resources are managed (firmness) and scheduled.

4.1 A closer representation of the adequacy and firmness-oriented mechanisms: the impact of the reliability product on the overall market outcomes.

We have seen in the previous section how the optimal (additional) incentives provided on a unit by unit basis should be based on the particular contribution of each unit to the fulfilment of the overall regulator's objectives. We present next an equivalent formulation of the problem, which allows us to introduce a more realistic representation of how these additional regulatory mechanisms work in practice.

---

[10] For instance, the SoS mechanism implemented in Spain includes two differentiated services: the availability service and the investment service, see (Batlle et al., 2008b).

[11] In some systems, particularly in Latin American ones, it is common that two products are defined to tackle the adequacy and firmness problem. For instance, in Chile or Peru, on the top of the energy forward purchasing, additional payments are provided based on a capacity-based product (capacity payments).

[12] Although in some cases this firm supply makes reference to the quantity of the product actually delivered, and in some other cases it makes reference to the ex-ante expected capability to deliver the product.

Demand side problem reformulation

The formulation of the problem is similar to the one presented previously, except for the following changes:

- Now the extra payments provided to the generators are equal to the unitary price of the reliability product, τ_{rp}, times the total quantity of the reliability product purchased, Q_{rp} [13].
- The regulator's performance objectives are now expressed as a function of the reliability product.
- To allow better representing the decision-making process on the generator side, besides the short-term time index, h, we now introduce a medium-term time index, m.

This way, the demand (and the regulator) model on the basis of the introduction of a security of supply mechanism consisting in the acquisition of a reliability product is:

$$\begin{aligned}
&\underset{Q_h, Q_{rp}}{Max} \sum_m \sum_h [U_{m,h}^d(Q_h) - \pi_{m,h} \cdot Q_{m,h}] - \tau_{rp} \cdot Q_{rp} \\
&subject\ to: \\
&\quad ADEPM(Q_{rp}) \geq ADEPM_{min} \quad \perp \chi_{ade} \\
&\quad FIRPM(Q_{rp}) \geq FIRPM_{min} \quad \perp \chi_{fir}
\end{aligned}$$

Therefore, it is clearly represented that the regulator is not just worried about the amount of installed MWs. Note that this variable does not explicitly appear in its own decision problem, since as stated it is not the only factor determining the quantity of reliability product of generating units.

Thus, in this context, this side of the market has now to decide the hourly consumption and the amount of the reliability product to be purchased. The optimality conditions stemming from this reformulation are:

$$\begin{aligned}
&\frac{dU_h^d(Q_h)}{dQ_h} = \pi_h, \forall h \quad , \\
&-\frac{\partial ADEPM}{\partial Q_{rp}} \cdot \chi_{ade} - \frac{\partial FIRPM}{\partial Q_{rp}} \cdot \chi_{fir} = \tau_{rp}
\end{aligned}$$

Thus, the price signal provided by the mechanism, τ_{rp}, should be equal to the dual variable associated to the constraint imposed over the system performance (expressed through the adequacy and firmness performance function), multiplied by the marginal contributions of that reliability product to those objectives.

Note that now a homogeneous price is defined, since an additional unit of a certain reliability product contributes in the same way, no matter the type of unit that provides it.

[13] Different reliability products may be traded in the mechanism, but we next particularize for the case of a single one.

Observe that instead of paying to each generator a variable amount for each MW of its available capacity, a homogeneous price is defined for each unit of the generators' *firm supply*. The same amount of the firm supply represents the same contribution to the SoS objectives, no matter the type of unit that provides it.

The necessity to properly define the reliability product

The problem is that in real life defining properly the reliability product is a complicated task. An obvious necessary condition is that it should be assured that any marginal increment of the reliability product provided, results in an improvement of the system performance ($\partial SOSPM / \partial Q_{rp} > 0$). But, at the same time, the regulator should ideally get to the optimal definition of the reliability product from the overall efficiency perspective, i.e. the one among the many that fulfill the necessary condition which leads to the least-cost option. For instance, defining the reliability product (i.e. firm supply) as installed capacity to reduce the probability of scarcity meets the necessary condition in any case, but it does not guarantees that the new entries are the most efficient, , see for instance the Peruvian case example mentioned later in section 5.2.

However, although perfection is not possible, optimizing the definition of the performance metrics and the reliability product is essential to guarantee an efficient outcome. In section 5 we review some real regulatory design examples that illustrate this issue.

The generators capability to deliver the reliability product

As repeatedly stated, the quantity of the reliability product that each unit can deliver depends, in its more general form, on the physical design parameters, the random factors affecting each unit production capability, the medium-term resource management and also on the short-term operation (hourly) decisions.

We designate the unit's investment decisions by Id^i. This variable represents all decisions regarding the technical characteristics of the units, as for instance the maximum generation capacity, q^i_{max}, the reservoir (hydro or gas) capacity, rc^i, etc. Strictly speaking, these parameters are decided at the time the investment is carried out (e.g. it is possible to install a turbine with the ability to burn different fuels).

The medium-term management decisions have been represented by Md^i_m, which involves, among others, the provision of fuel (e.g. signing contracts in advance). Some of this medium-term decisions have an associated cost $MC^i(Md^i_m)$, for instance paying for an extra risk premium derived from entering into any sort of firm fuel supply contract, see description of the Colombian case next in section 5.1.

We also consider the random variables affecting each unit production capability. This random factors are represented by ε^i, which includes for instance hydro inflows, HI^i, hourly failure factor, $\lambda^i_{m,h}$, etc.

Thus, the reliability product is expressed as a function of the investment and planning decisions and also the production scheduling, $q^i_{rp}(Id^i, Md^i_m, q^i_{m,h})$. For example, the capability of a hydro plant (the unit), to deliver a certain amount of energy in the dry period (the reliability product), depends on the maximum output and the reservoir capacity (physical parameters), on the resource management carried out up to that moment (in this case, the past production decisions) and on the hydro inflows (random factors).

Generator side problem reformulation

Taking into account the previous details, we modify the generators problem to introduce the additional payment provided by the mechanism as an explicit remuneration in exchange for the delivery of the reliability product.

$R_c^i(Id^i, Md_m^i, q_{m,h}^i, \varepsilon^i)$ still represents schematically the operation constraints relating the production of the unit, i, along the index, m, and the corresponding hours, h, considered in the time scope being studied. We have now introduced the effect of the investment and medium-term planning decisions as well as the random factors. The maximum capacity constraint has now been embedded under this schematic formulation (the inequation $q_{ih} \le q_{max}^i$ of the previous formulation is a particular case, where q_{max}^i is now considered one of the investment design parameters).

Again, just one reliability product has been represented, since it represents the most real case.

Thus, the generator's problem can be rewritten as follows:

$$\begin{aligned}
\underset{Id_i, Md_i, q_{ih}, q_{rp_i}}{Max} \quad & \sum_m \sum_h [\pi_{m,h} \cdot \sum_i q_{m,h}^i - \sum_i C^i(q_{m,h}^i)] - \\
& - \sum_m \sum_i MC^i(Md_m^i) - \sum_i IC^i(Id^i) \\
& + \tau_{rp} \cdot \sum_i q_{rp}^i(Id^i, Md_m^i, q_{m,h}^i) \\
\textit{subject to}: & \\
& R_c^i(q_{m,h}^i, Id^i, Md_m^i, \varepsilon^i) \le 0 \quad \perp \zeta_c^i
\end{aligned}$$

The generator's maximizes its own profit. But as pointed out previously, the source of income derived from the additional payments provided by the mechanism is now expressed as a direct remuneration of the reliability product delivered.

Thus, the decision variables are now the hourly energy produced $q_{m,h}^i$, the design parameters of the generating units, Id^i, the medium term resource management decisions, Md_m^i, and the quantity of the reliability product to be delivered by each unit in the context of the regulatory mechanism, q_{rp}^i.

The optimality conditions obtained from this problem are the following:

$$\pi_{m,h} - \frac{dC_i(q^i_{m,h})}{dq^i_{m,h}} + \tau_{rp} \cdot \frac{\partial q^i_{rp}}{\partial q^i_{m,h}} + \sum_c \left[\frac{\partial R^i_c}{\partial q^i_{m,h}} \cdot \zeta_{c,i} \right] = 0, \ \forall i, m, h$$

$$-\frac{dMC(Md^i_m)}{dMd^i_m} + \tau_{rp} \cdot \frac{\partial q^i_{rp}}{\partial Md^i_m} + \sum_c \left[\frac{\partial R^i_c}{\partial Md^i_m} \cdot \zeta_{c,i} \right] = 0 \ , \ \forall i$$

$$-\frac{dIC(Id^i)}{dId^i} + \tau_{rp} \cdot \frac{\partial q^i_{rp}}{\partial Id^i} + \sum_c \left[\frac{\partial R^i_c}{\partial Id^i} \cdot \zeta_{c,i} \right] = 0 \ , \ \forall i$$

- Interpretation of the first optimality condition: Note that the production schedule is optimized taking into account the short-term profit, $\pi_{m,h} - dC_i(q^i_{m,h}) \,/\, dq^i_{m,h}$, the effect of the technical constraints on the marginal profit, $\sum_c \left[\zeta_{c,i} \cdot \partial R^i_c \,/\, \partial q^i_{m,h} \right]$, and also how the production may affect the firm supply and thus, the marginal income perceived by the mechanism, $\tau_{rp} \cdot \partial q^i_{rp} \,/\, \partial q^i_{m,h}$. Thus, the production schedule can be affected by the regulatory mechanism. Special attention should be devoted in order not to distort the efficient short-term decisions (see the brief description on the Colombian case in section 5.2).
- Interpretation of the second and third optimality condition: These two optimality conditions are completely analogous. They respectively show how under these modeling hypotheses, the drivers to make planning and investment decision are the following:
  - the marginal costs of taking the corresponding long-term or medium-term decision;
  - the marginal income that could be perceived due to relaxing the limits of the binding constraints (R^i_c) as a consequence of the new investment or planning;
  - the marginal additional income that could be perceived in the additional regulatory mechanism.

In summary, investment and management decisions should be carried out by generators in a way to comply with the firm supply requirement. Under a regulatory standpoint, the previous development has shown these decisions are affected by that short-term price signals, which cannot be distorted and by the marginal income perceived from the additional SoS mechanism. The latter issue highlights that when implementing a SoS mechanisms, the regulator conditions the market functioning.

5 IMPACT OF THE RELIABILITY PRODUCT DESIGN ON THE MARKET FUNCTIONING: LESSONS LEARNED FROM THE REGULATORY EXPERIENCE

In the previous section, a conceptual mathematical model of an electric power system has been used to develop the expressions and a deeper understanding of the so-called reliability product. The discussion was mainly divided in two points:

- On the basis of one of the optimality conditions of the demand's problem, we showed that the optimal additional remuneration provided by an additional security of supply mechanism should be based on each unit's contribution to the regulator's objectives. As also stated, when it comes to implement such a mechanism, assessing this contribution is an extremely complex issue.

- On the basis of the optimality conditions of the generators' problem, we illustrated how when implementing this kind of mechanisms, the regulator conditions the market functioning. We also highlighted how a poor or flawed definition of the reliability product can lead to undesired outcomes.

In the next two points in this section we examine these two issues in the light of the lessons learned from the most real case examples implemented to date.

5.1 Reliability product definition: assessing the actual contribution of the units

As detailed in (Batlle & Rodilla, 2010), the different security of supply mechanisms can be classified based on whether the regulator's main objective has been to ensure a certain quantity of the "reliability product" (via any sort of the traditionally so-called capacity market or auction) or to administratively set a price for the product itself (i.e. a capacity payment). Irrespective of the alternative selected, the regulator has in any case two main duties:

- First, as thoroughly discussed before, a methodology to assess the actual contribution of a generating unit to the regulator's objective pursued with the implementation of the additional mechanism has to be designed.

  Once the regulator is able to evaluate the expected contribution of the generating units in (and to enter) the system (i.e. the firm supply), it is possible to assess how far the market outcomes are from the desired ones.

  Then if the conclusion is that an additional mechanism is needed, the methodology has to allow calculating the firm supply to be awarded to each unit in the system (in case a capacity payment is implemented) or the maximum firm supply each unit is allowed to offer/trade (in case a quantity-based mechanism is implemented, e.g. capacity credits in a capacity market).

- Second, once the mechanism is in force, to supervise and assess the compliance of each generating unit with respect to the delivery of the reliability product (i.e. the firm supply).

Thus, from the regulators' point of view there are two relevant measures to be determined associated to the firm supply: the maximum amount of the reliability product each unit is capable to provide and the reliability product actually delivered.

5.1.1 Regulatory estimation of the maximum firm supply

The existence of quantity risks

In some mechanisms, the generators may acquire a commitment (subject to certain penalties) to deliver a certain quantity of the reliability product in advance. When this is the case, the generator may face a quantity risk regarding the associated production. A characteristic example of these mechanisms implying a commitment in advance are the quantity mechanisms based on long-term auctions described in (Batlle & Rodilla, 2010), in which generating units ask for a premium in exchange of the reliability product, as it is for instance the case in Colombia, New England or Brazil, where a sort of option contract (the so-called reliability options) is acquired.

This quantity risk involved mainly depends on the unit's characteristics. The ability of the generator to fulfil the commitment of providing the reliability product can be subject on the one hand to more or less uncertainty (for instance, due to failure rates or limited fuel availability, as it can be the case of a hydro plant subject to rainfalls or a gas plant exposed to network constraints), and on the other to the level of dispatchability of the plant and thus to the generator's management policy (for instance it may depend on the medium-term resource management carried out by the generator). This is the reason why, depending on the type of generating unit, there may be a larger or lower risk of non-compliance on the generator's side.

In principle, any market-based solution would ideally allow generators to assess their own capability to provide the reliability product (the firm supply committed in advance). Indeed, the market should ideally leave to market agents the decisions which they are supposed to make more efficiently (since they are clearly the ones in the best position to do so).

The risk perceived by the regulator would not be a problem as long as the penalties for non-compliance and the credit risk hedges would be optimally determined. The higher the penalties and the higher the hedges the easier it would be in principle for the regulator to rely on agents assessments with respect to their own firm supply capability. However, more often than not it is considered that the best guarantee that can be asked to a generating company is to require them to have enough physical back-up to cover its commitments, i.e. to have enough firm supply to provide the reliability product (which is always linked to electricity production).

This is the reason why the regulator prefers to set an upper limit on the firm supply each unit is "reasonably" capable to provide, in order to limit the quantity the generator can commit, and consequently reduce the risk of non-compliance. This applies either if the mechanism is priced-based or quantity-based. Sometimes (see below the case of Colombia), the regulator also determines a minimum limit to be offered in the mechanism.

The calculation of this firm supply limit represents a huge challenge and also an extremely controversial issue, for it puts boundaries to a significant source of income for generators. One of the key difficulties lies in the side of the necessary assumptions about the behavior of the generators (medium-term resource management, maintenances, etc.). Furthermore, this generators' behavior depends crucially on the magnitude of the economic incentives derived from the mechanism. In principle, theoretically at least, almost any unit would plan its maintenance and manage its fuel stocks to maximize the system performance at any given time, if the price is sufficiently right. In practice, there are two major alternatives to face the problem of determining the firm supply limit ex-ante:

- to use a simulation model to forecast the potential and expected firm supply,
- or to determine it by just inspecting the past (historical) performance of the unit (or the past performance of similar units).

Estimation of the firm supply based on a simulation model

This approach consists in running a long-term (preferable stochastic) model in order to estimate how generating unit dispatch will look like in the period being evaluated. Based on this information, the regulator determines the expected contribution of each unit to the system performance, or more precisely, their contribution with a certain probability. In this case, the regulator (acting on behalf of the demand) risk aversion determines to a large extent this limit.

The main advantage of this approach lies on the side of both its capability to consider scenarios that have not been observed in the past and its capability to adapt to changing conditions easily (for example estimating the future contribution of generating units when other new installations enter the system and change the system dispatch). Another clear advantage is that this approach allows introducing the effect of the incentive provided by the regulatory mechanism.

The main problem is that it requires making strong assumptions on the units' future behavior, since it is necessary to model certain critical aspects (as for instance the risk aversion in the generating business, which for instance impacts on the way limited energy plants are managed). This problem, which also represents the most controversial characteristic of this approach, is not a so relevant issue in those markets in which the central planner is in charge of scheduling the production of the different units (i.e. most Latin American ones, except for instance Colombia). In fact when this is the case, using the model the market operator employs to determine the system medium-term planning and short-term dispatch to evaluate the future firm capacity seems to be a quite reasonable (and even the more suitable) approach. Anyway, inevitably, the regulator's assumptions when performing the calculations (scenarios considered, hydro inflows, plant failures and their probabilities) are permanently subject to heated controversy.

The use of a model seems to be the only reasonable approach to estimate the firm supply of a new technology entering the system, since no past performance data regarding this type of unit are available yet.

The markets using this approach have traditionally been the hydro dominated ones. Among some of the most relevant experiences we find:

- Brazil: the so-called "firm energy", which represents the maximum production a generating unit can commit in the auction mechanism[14], is calculated following the procedure described next:
  - By means of a hydrothermal stochastic model (using an iterative process) the maximum demand that allows to meet the reliability criteria (5% of Non Served Energy Expectancy, for instance) it is first determined.
  - Using the previous demand consumption, the results of the simulation provide information on the system marginal prices and also the system units' production. Then, the income perceived by each one of these latter is calculated.

[14] See (Batlle & Rodilla, 2010) for a detail description of the security of supply mechanisms in each one of the markets cited hereafter.

- The firm energy is proportional to the average spot income perceived in the simulation of the stochastic model.

Thus, this methodology takes into account the opportunity cost for the system of the energy provided by each unit. Under this scheme, a thermal plant that is always available to produce, but whose costs are extremely high would receive a close to zero firm energy.

- Colombia (CREG, 2006): the firm supply is known as "firm energy for the reliability charge" (Energía Firme para el Cargo por Confiabilidad in Spanish).

  For the hydro units, this value is estimated by means of a computational model (HIDENFICC) that seeks to optimize the minimum energy to be produced every month when the inflows are extremely scarce (a dry year). There is a minimum value of the reliability product to be offered in the auction, which is known as "base firm energy". This minimum quantity is also calculated with the model.

- Panama (ERSP, 2006): the firm supply is known as "long-term firm capacity" (potencia *firme de largo plazo* in Spanish).

  In the case of the hydro plants, the value is calculated using a hydroelectric model (PLAN-H), where the relevant data are obtained from historical series (average inflows from 1963 are considered). It first determines the average and also the maximum output every month. The long-term firm capacity of a hydro unit is calculated as the hourly output capacity which can be guaranteed with a probability of 95% during all the hours corresponding to the period of maximum demand requirement (8 every day).

- Chile: There are two different reliability products: one energy-based, the firm energy (*energía firme* in Spanish) and another capacity-based, firm capacity (*potencia firme*). The firm capacity and energy of hydroelectric plants is computed using a model, w*hich is fed w*ith the two driest historical series (Barroso et al, 2007).

In some cases the model is substituted by just the reasonable expectation of the Regulator's or the System Operator:

- Western Australia: in the Market Rules (IMO, 2009), only general criteria to be applied by the Independent Market Operator are provided to determine the firm energy (known as certified reserve capacity). There is not any formula or explicit model described in the procedure: 'The Certified Reserve Capacity for a Facility for a Reserve Capacity Cycle is not to exceed the IMO's reasonable expectation as to the amount of capacity likely to be available from that Facility, after netting off capacity required to serve Intermittent Loads, embedded loads and parasitic loads, at daily peak demand times in (...)"

Past performance: historical series analysis

This approach consists in updating the firm supply value on the grounds of previous years' results. Its main advantage is that there is no need to estimate the behavior of the different units, since it is directly captured through the past performance data. Past data captures all: the resource management, the effect of the incentive provided by the regulatory mechanism of interest, etc.

The clear disadvantage is the existing lag with which different and new system conditions are captured. This lag makes this approach a not well suited one when relevant changes are happening in the system. For instance, in the former UK capacity payments, the LOLP calculation (a relevant parameter determining the additional remuneration) was calculated using historical pre-1990 data for pre-1990 plants. Since these data did not take into account the strong incentives provided by the mechanism, this led to overestimating the probability of losing load (Roques et al, 2005).

It is also important to note that this approach gives an additional indirect incentive to provide firmness, since a poor performance may condition the future firm supply limits imposed by the regulator in the future, and thus future income derived from this mechanism.

This has been the approach preferred in most systems to assess the maximum firm supply capability of thermal units, but in some cases it has been also used to evaluate the performance of hydro units. Sometimes the past performance is evaluated in terms of reliability, as in the following examples:

- Colombia (CREG, 2006): The firm energy of a thermal plant is calculated based on its installed capacity and its past availability taking into account just (detected) forced outages (caused by maintenances, failures, fuel unavailiability, etc.). No factors which are not under the control of the agent are considered to calculate this availability factor, except in the case of gas-fired plants (generators are supposed to enter into firm contracts with the gas operator company to hedge against potential unavailabilities due to restrictions in the gas transmission network).
- Peru: The firm supply is called "firm capacity" (*potencia firme*). Again, the thermal units' firm supply is calculated by multiplying the effective capacity by the monthly availability factor (calculated based on past performance). The hydro units' firm capacity is calculated as an average production.
- Italy: the firm supply is referred to as "available production capacity" (*capacità produttiva disponibile* in Italian). This value is calculated by the TSO taking into account principally the maximum capacity and unavailabilities (Terna, 2008).
- PJM, NYISO and New England: although there are slight differences, the firm supply is known in all the systems as UCAP, and it is calculated based on the EFORd factor[15]. These EFORd values are calculated for each season, weighting with different factor the different hours depending on the system requirements.

But there are some other cases, where the production costs play an important role in the determination of the firm supply:

- Ecuador (CONELEC, 2004): In the regime in force previous to the reform started in 2008, the firm supply was called "available capacity" (potencia remunerable puesta a disposición). For hydro plants, it is calculated as the average past production in the "dry season" (in between November and February), considering the historical data from the

[15] EFORd is a measure of the probability that a generating unit will not be available due to forced outages or forced deratings when there is a demand on the unit to generate.

previous decade. In case of thermal plants, the assigned firm supply was the effective capacity discounting unavailabilities (which were declared by the generating companies and approved by the market operator). Thermal units were considered in an increasing cost merit order up to the point at which the accumulated capacity reached the thermal requirement at the peak expected consumption in the dry season.

- Bolivia: The firm supply is called “guaranteed capacity” (*potencia garantizada*). This firm supply is determined by using a simulation model and observing the units required to cover the peak demand consumption in a dry year (considering an economic dispatch).

Firm supply settlement

Obviously, in order to measure the units’ compliance the straightforward alternative is to directly observe it. This way, the regulator has to check directly the past behavior of the different units in order to settle the corresponding remuneration.

Nevertheless, depending on the definition of the reliability product, sometimes the regulator has to rely on the units declared compliance instead of checking it by him/herself. For instance, if the reliability product is related to any sort of “availability” rate[16], it would be uneconomical for the regulator to check in situ every unit’s hourly availability; this is the reason why in this particular case it is difficult to find another way that does not entail fully relying on the generators’ declaration. In many of the price-based mechanisms, this has been the only necessary measure with respect to the firm supply (for instance, in the former Argentinean capacity payment, the former Spanish capacity payment or the current Italian or Irish capacity payments).

In particular, the Irish capacity payment combines all the previous approaches, since the incentive is divided in a variable, a fixed and an ex-post component[17]. The incentive also depends on the forecasted LOLP and also on the actual ex-post corrected LOLP.

5.2 The reliability product as a partial planning tool for the regulator

Once the regulator has decided to intervene and undertake the task of “helping” the market to reach what it considers to be an efficient outcome, the next key question is how to introduce the necessary adjustments in the market design in place so as to achieve the objective pursued in the long term. This is particularly complicated and controversial, because in the end, all long-term planning may, directly or indirectly, fall again into the hands of a central planner, and we should not forget that avoiding the potential inefficiencies stemming from the central planner scheme was one of the principal motors behind the liberalization wave that started a few decades ago.

[16] As it has been the case in some price-based mechanisms as the capacity payments in Spain, Korea or Italy.

[17] The weights used to ponder the variable (determined a month ahead), fixed (determined a year ahead) and ex-post (determined at the end of the month) terms are to be defined every year. The mechanism started with 30%, 40% and 30% respectively.

In this sense, the security of supply mechanism design, and particularly the reliability product design turns to be a key element to channelize the role the regulator wants to play in the generation system planning and expansion. Sometimes (indeed, in most cases), the reliability product design is explicit or implicitly performed to favor certain results:

- Examples of explicit ways to do it are for instance in the firmness dimension an specific way of managing the plants (e.g. contracting strategic hydro capacity reserves to be dispatched at the regulator's own discretion) or, in the adequacy dimension, the installation of a particular technology (in the long-term auctions held in Peru, when clearing the auction, bids submitted by new hydro plants are multiplied by 0.85).

- Since by definition there is no "aseptic" reliability product design, there is also a handful of ways to get to the same results through implicit details. For instance, in some cases, the evaluation of the units' firm supply is made by simulating the expected operation of hydro plants in a very dry year and at the same time the availability of thermal plants considered is based only on average values of failure rates (and for instance no gas network constraints are considered), clearly the regulator is setting larger incentives for the installation of new thermal units.

But unfortunately, the regulatory experience shows that more than often the outcomes resulting from the implementation of a security of supply mechanisms do not correspond with the ones (explicit or implicitly) expected and desired by the regulator when designing it. When defining the reliability product, the regulator has to be careful with the foreseeable response of generators, so as to analyze whether this response leads or not to an efficient result. Sometimes the consequences of the product definition are not evaluated beforehand, and highly inefficient situations are the result, either affecting the system units operation and planning or the expansion.

- For instance, two examples of this matter on the adequacy dimension are:

  - If the regulator decides to buy just plain installed capacity, it will probably get the capacity which presents the lowest investment costs, but maybe with low availability rates. Take for instance the Peruvian case: when the market started, a capacity payment (additional fixed annual remuneration to reward installed capacity) was implemented. The value of this payment was determined by taking as a reference the investment cost of a new investment in an efficient peaking plant (an open-cycle gas turbine). This payment led to a dash for extremely expensive junk peak generation, due to their relatively small capital requirements. Thus, from the standpoint of reliability, the reserve margin is much larger than is theoretically suitable but at the same time, prices are significantly high.

  - Another clear example of how these mechanisms can condition the design of new investments is the case of Guatemala. In this market, since the capacity payment is related to average production of the generating units in the four peak hours of each of the working days in the dry season (from December to May), new investments in small hydro plants are designed so as to have a reservoir to allow daily regulation whose energy capacity (MWh) is the capacity of the turbine (MW) times four (h).

- Conversely, on the firmness dimension, other two paradigmatic cases are:

- If the regulator decides to pay for the water reservoir level in the "dry season", it will fill reservoirs to their full capacity in that season. Sometimes the consequences of the product definition are not evaluated beforehand, and highly inefficient situations are the result. This was the case in the Colombian market back in the late nineties.
- If the value of the reliability product is tightly linked to the actual scheduling of the plants, the market unit commitment can be severe and negatively affected: in Argentina, back in the nineties, the capacity payment was only provided to those generating units producing in each hourly interval. The problem that rapidly appeared was that the size of this payment was large enough to condition the optimum strategies of generators, which internalized the payment in the bid sent to the market, bidding below marginal costs in order to receive the extra payment.

6 CONCLUSIONS

Under the market-oriented paradigm, the new regulation must make sure that the appropriate incentives exist so as to ensure an efficient long-term security of supply level. Once the regulator has decided to undertake the task of "helping" the market to reach what it considers to be an efficient outcome, the next key question is how to introduce the necessary adjustments in the market design in place so as to achieve the objective pursued in the long term. This is particularly complicated and controversial, because in the end, all long-term planning may, directly or indirectly, fall again into the hands of a central planner, and we should not forget that avoiding the potential inefficiencies stemming from the central planner scheme was one of the principal motors behind the liberalization wave that started a few decades ago.

We have reviewed the fundamental criteria that support the way how these regulatory assessments and the corresponding price incentives should be designed, calculated and managed in order to provide optimal signals. In particular, we have presented the different available alternatives to define the metric to evaluate the generation system performance. Also, a conceptual mathematical model of an electric power system has been used to develop the expressions and a deeper understanding of the so-called reliability product.

We have shown first that the optimal additional remuneration provided by an additional security of supply mechanism should be based on each unit's contribution to the regulator's objectives, and also that assessing this contribution is an extremely complex issue. Then, we illustrated how when implementing this kind of mechanisms, the regulator conditions the market functioning.

The exhaustive and critical review of the international experience illustrates that the design of a long-term mechanism to acquire a certain reliability product presents challenges that if not properly solved may result in the end in undesired market outcomes.

We claim that more attention has to be devoted first on the ways to properly define the system performance objectives and afterwards evaluate the actual contribution of each generating unit to the fulfillment of these objectives. Many experiences worldwide have failed to define either the former or the latter, and this is usually the reason behind the poor results obtained.

7 REFERENCES

Barroso, L. A., H. Rudnick, R. Moreno, B. Bezerra (2007), “Ensuring Resource Adequacy with Auctions of Options and Forward Contracts”, IEEE PES General Meeting 2007, Tampa, USA.

Batlle, C., I. J. Pérez-Arriaga (2008). “Design criteria for implementing a capacity mechanism in deregulated electricity markets”. Special issue on “Capacity Mechanisms in Imperfect Electricity Markets”. Utilities Policy, volume 16, issue 3, September 2008, Pages 184-193 (doi:10.1016/j.jup.2007.10.004).

Batlle, C., C. Solé, M. Rivier (2008b). “A new security of supply mechanism for the Iberian Market”. The Electricity Journal, vol. 21, iss. 2, March 2008, pp. 63-73, doi:10.1016/j.tej.2008.02.003.

Batlle, C. & P. Rodilla (2009). “Policy and regulatory design on security of electricity generation supply in a market-oriented environment. Problem fundamentals and analysis of regulatory mechanisms”. IIT Working Paper IIT-09-057A.

Batlle, C. & P. Rodilla (2010). A critical assessment of the different approaches aimed to secure electricity generation supply. IIT Working Paper IIT-10-017A. March 2010. Submitted to Energy Policy.

CEER (2009), “Generation Adequacy Treatment in Electricity. A CEER Call for Evidence”, Ref: C09-ESS-05-03, Council of European Energy Regulators (CEER). December 2009.

CONELEC, Consejo Nacional de Electricidad (2004), “Regulación Nº. CONELEC 003/04. Cálculo de la Potencia Remunerable Puesta a Disposición”. October, 2004.

CREG (2006), “Resolución CREG-071 del 3 octubre de 2006” (In Spanish), Comisión de Regulación de Energía y Gas, 2006.

De Vries, L.J. (2004). Securing the public interest in electricity generation markets, The myths of the invisible hand and the copper plate. Ph.D. dissertation, Delft University of Technology, Faculty of Technology, Policy and Management.

Ehrenmann, A. & Y. Smeers (2008). Energy Only, Capacity Market and Security of Supply. A stochastic Equilibrium Analysis. ECORE discussion paper. Available at www.ecore.be/DPs/dp_1203678053.pdf. January 2008.

ERSP, Ente Regulador de Los Servicios Públicos (2006), “Reglas Comerciales del Mercado Mayorista de Electricidad, Reglas Comerciales”. Panamá, 2006.

Eurelectric (2006). “Security of electricity supply: Roles, responsibilities and experiences within the EU”. January 2006.

European Commission (2009). Impact Assessment Guidelines. SEC(2009)92, 15.01.2009.

Hogan, W. (2005). “On an ‘Energy-Only’ Electricity Market Design for Resource Adequacy,” paper prepared for the California ISO, September, 2005.

Hogan, W. (2009). A Model for a Zonal Operating Reserve Demand Curve. MIT presentation. October 15, 2009. Available at www.hks.harvard.edu/fs/whogan.

IMO (2009), “Wholesale Electricity Market Amending Rules”, Electricity Industry Act, Independent Market Operator (IMO), February, 2009.

Joskow, P. L. (2007). Competitive Electricity Markets and Investment in New Generating Capacity. In: The New Energy Paradigm, Dieter Helm, ed., Oxford University Press, 2007.

Pérez-Arriaga, I. J. & C. Meseguer (1997). “Wholesale marginal prices in competitive generation markets”. IEEE Transactions on Power Systems, vol. 12, no. 2, pp. 710-717, May 1997.

Pérez-Arriaga, I. J. (2001). “Long-term reliability of generation in competitive wholesale markets: A critical review of issues and alternative options”. IIT Working Paper IIT-00-098IT, June 2001.

Pérez-Arriaga, I. J. (2007). Security of electricity supply in Europe in a short, medium and long-term perspective. European Review of Energy Markets, vol. 2, iss. 2, December 2007.

Rodilla, P & C. Batlle (2010). “Security of electricity supply at the generation level: problem analysis”, IIT Working Paper IIT-10-017A, April 2010.

Rodilla, P, C. Batlle, A. Baíllo, S. Cerisola (2010). “Regulatory intervention to ensure an efficient medium-term planning in electricity markets”, IIT Working Paper IIT-10-008A, February 2010.

Roques, F., Newbery, D.M., Nuttall, W.J. (2005). Investment Incentives and Electricity Market Design: the British Experience. Review of Network Economics. Vol.4, Issue 2. June 2005.

Schweppe, F. C, M. C. Caramanis, R. D. Tabors and R. E. Bohn (1988). “Spot pricing of electricity”. ISBN 0-89838-260-2, Kluwer Academic Publishers, Boston, MA, 1988.

Stoft, S. (2002) “Power System Economics: Designing Markets for Electricity”, IEEE Press & Wiley-Interscience, ISBN 0-471-15040-1, 2002.

Terna (2008), “Capacità Produttiva Disponibile: Nota Informativa” (In Italian), available at www.terna.it., 2008.

Wood, A. J. & B. F. Wollenberg (1996). Power generation, operation and control. John Wiley & Sons ed., New York, 1996.