UNIVERSIDAD PONTIFICIA COMILLAS DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

(Instituto de Investigación Tecnológica)

REGULATORY TOOLS TO ENHANCE SECURITY OF SUPPLY AT THE GENERATION LEVEL IN ELECTRICITY MARKETS

Instrumentos regulatorios para mejorar la seguridad del suministro en mercados eléctricos

Tesis para la obtención del grado de Doctor

Director: Dr. D. Carlos Batlle López

Autor: Ing. D. Pablo Rodilla Rodríguez

Madrid 2010

CONSTANCIA REGISTRAL DEL TRIBUNAL DEL ACTO DE LA DEFENSA DE TESIS DOCTORAL

TÍTULO: REGULATORY TOOLS TO ENHANCE SECURITY OF SUPPLY AT THE GENERATION LEVEL IN ELECTRICITY MARKETS

Instrumentos regulatorios para mejorar la seguridad del suministro en mercados eléctricos

AUTOR: Ing. D. Pablo Rodilla Rodríguez

DIRECTOR: Dr. D. Carlos Batlle López

TUTOR-PONENTE:

DEPARTAMENTO: Instituto de Investigación Tecnológica

FACULTAD O ESCUELA: Escuela Técnica Superior de Ingeniería (ICAI)

Miembros del Tribunal Calificador:

PRESIDENTE: Dr. D. Josep Borrell Fontelles **Firma:**

VOCAL: Dr. D. Jean Michel Glachant **Firma:**

VOCAL: Dr. D. Carlos Ocaña Pérez de Tudela **Firma:**

VOCAL: Dr. D. Carlos Álvarez Bel **Firma:**

SECRETARIO: Dr. D. José Ignacio Pérez Arriaga **Firma:**

Fecha de lectura: 27 de septiembre de 2010

Calificación:

AGRADECIMIENTOS

Madrid, a 31 de julio de 2010

REGULATORY TOOLS TO ENHANCE SECURITY OF SUPPLY AT THE GENERATION LEVEL IN ELECTRICITY MARKETS

CONTENTS

PART A

PART A. INTRODUCTION

1. THE NEED FOR REGULATORY TOOLS

The energy industry, and particularly the electricity sector, has been subject to major reforms over the recent decades. Until these processes started, the activity that we now call generation supply was a part of the complete chain of activities of vertically integrated utilities and was therefore performed as a public service or as a regulated monopoly.

Since there are no fully comparable energy systems, these reforms have taken very different forms, but all of them have shared a common approach, consisting in taking steps towards introducing competition at any of the activities in which it was considered feasible. Ideally competition is useful since, if properly implemented, it sends sound economic signals -market prices- both to electricity consumers and suppliers that are supposed to drive their decisions in the direction of efficiency.

These reforms have been traditionally denoted as “liberalization” or “deregulation” processes, terms which might appear to be slightly misleading, since they could easily be understood as just a relaxation of government limitations, leading to a weaker or “lighter-handed” regulation.

From the regulatory perspective, the fact is that in the case of the energy industry, the reform has entailed exactly the opposite: rather than a “deregulatory” process, it has been (and it is still being and still expected to be) an intensely “re-regulatory” one, see (Borenstein & Bushnell, 2000) or (Ruff, 2003). In this sense, the term “restructuring” has sometimes been preferred in some electric power systems (particularly in the United States).

The discussion developed in this work, the need for the regulator’s intervention to complement electricity markets in order to guarantee security of generation supply, is a good illustration of this paradox: the deregulation in electricity systems has accentuated the crucial need for reinforcing regulation.

Broadly speaking, the objective of regulation is *to prevent (produce) inefficient (efficient) outcomes in different places and timescales which might (might not) otherwise occur*. In particular, we show in the present work how, in the (already not-so) new “deregulated” and “liberalized” scheme that governs electricity business, the intervention of the regulator is needed *to guarantee a minimum required level of security of supply in different places and timescales*, since it has been largely demonstrated that *otherwise they will not occur*.

This regulatory intervention, which is carried out in order to drive the market towards the most efficient outcome, is in practice implemented by introducing additional mechanisms, roughly speaking additional sets of rules. These mechanisms are regulatory tools aimed to provide additional (and optimal) signals to market agents.

To perform this role properly, the regulator needs another sort of tools: assessment tools (models of different nature) either to evaluate the market performance and thus detect the need of introducing additional mechanisms, or to assist in the design of these mechanisms (if needed), by evaluating ex-ante the impact of the regulatory solutions that could be considered.

The analyses carried out in this thesis are focused on these regulatory tools (the mechanisms and the models). But prior to going into deeper detail with respect to these regulatory tools, it is first essential to properly set the background of the security of generation supply regulatory problem. We also set in this introduction the basic terminology that will be used in the remainder of this thesis.

2. INTRODUCING THE FOUR DIMENSIONS OF SECURITY OF GENERATION SUPPLY PROBLEM

Modern society depends critically on the availability of electricity. The consequences of a lack of supply are known to affect regions and countries deep into the social, economic and political dimensions. Progress is undoubtedly linked to the availability of sufficient electricity; therefore, it is of no surprise that avoiding emergency situations and ensuring certain quality supply standards represents a major concern for regulators.

The actual physical supply of electricity to end-consumers at a given moment in time is the outcome of a complex and interlinked set of actions -some of which were performed many years before- which have jointly made possible that the right technologies and infrastructures have been developed and installed, provision of fuels have been contracted, hydro reservoirs have been properly managed, power plants have been maintained correctly and at an appropriate time, generators have been started-up and connected to the grid so that they were ready to function when needed, margins of operating reserves were maintained, etc.

Thus, the provision of electricity comprises a multiplicity of actions and measures that have to be performed in different time ranges -from many years to seconds-, by different agents -from investors to regulators or system operators- and involving different types of technologies and equipment. In this thesis, we focus on those measures that have to be carried out at the generation level. Therefore, other levels playing also relevant roles in the security-of-supply problem will be left aside (e.g. an adequate and secure network planning).

Previously, we stated that regulation has to cover different time scales. As with any other problem, decoupling the security of supply problem into its major components (or dimensions) facilitates its understanding, its discussion, and finally the design of proper technical procedures and regulatory measures (if deemed necessary). Different classifications of these dimensions have been presented, one of the most popular being the one provided by the North American Electric Reliability Council (NERC), see for instance (NERC, 1997), where two dimensions were considered, namely security (a short-term issue) and adequacy (a long-term issue). The dimensions presented below are in line with those firstly sketched in (Batlle et al., 2007), where a greater disaggregation level of the time scopes involved is provided.

The four dimensions of security of electricity generation supply

From the "time" perspective one can distinguish four dimensions of security of electricity generation supply:

- Security, a very short-term issue (close to real time), defined by the NERC as the "ability of the electrical system to support unexpected disturbances such as electrical short circuits or unexpected loss of components of the system or suddenly disconnection" (NERC 1997).
- Firmness, a short- to mid-term issue, which can be defined as the ability of the already installed facilities to provide generating resources efficiently (especially when most needed). This dimension is linked to both the generating units' technical characteristics and also their medium-term resource management decisions.
- Adequacy, a long-term issue, which means the existence of enough available generation capability, both installed and/or expected to be installed, to efficiently meet 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.

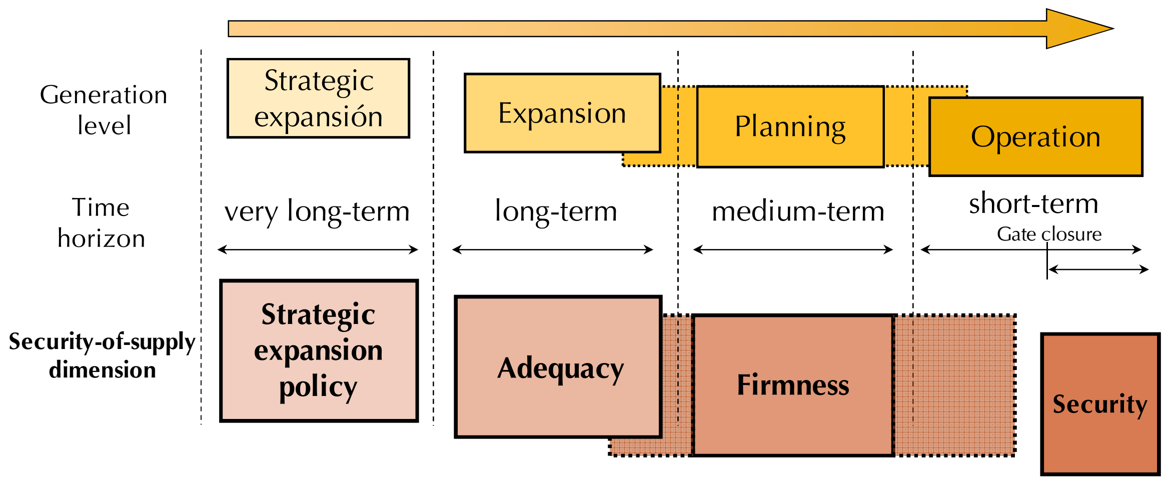


Figure 1. Security of supply dimensions

It is important to note that these four dimensions are to a large extent interrelated and they cannot be completely decoupled from each other. For instance, when defining requirements for the shortest term dimension (security), the longer term dimensions may be affected (this is discussed in Chapter I, section 2 in more detail). In the other direction the same dependence also exists, and thus, when analyzing the new necessary investments that would be required to meet efficiently demand in the following years, it is also essential to take into account whether or not they will be able to meet the future requirements at the shorter term dimensions levels (e.g. requirements at the security level).

The main objective of the reform of the electricity system consists in changing the way decisions at these four stages are made. These four are the sequential and interlinked levels at which the main problem of optimizing the net social benefit at the power system level can be decomposed. Specifying that this optimization refers to the power system level is not superfluous, since as we next discuss, these four dimensions cannot be decoupled from the overall optimization of the global system net social benefit optimization.

In the market-based context, at each point in time, market forces are supposed to make the decisions that were formerly taken on the regulator side: performing the short-term balancing of the system, managing and planning the operation of the existing facilities in the medium term, deciding when and where to start building new investments, choosing among the generation technologies available and transmission alternatives and finally monitoring not only the cheapest options for the present but also those which might be expected to be more efficient in the future.

The fifth dimension

Additionally, from a broader perspective, the optimization of the electric power system can be considered as a sub-problem of the overall policy objective: the maximization of the global net social benefit.

The energy sector (oil, gas, electricity, etc.) plays a key role in the overall development of a country/state, so more often than not, the previously enumerated dimensions of the electricity generation activity are affected by higher level criteria. Thus, there is an additional dimension, which is originated outside the security of supply problem, but which may condition to a large extent the system outcomes at the previous four dimensions:

- Strategic Policy, which concerns the impact and influence on (and from) economic (and social and political) objectives of a single State or a group of States: job creation, economic growth, competitiveness, regional and rural development, etc. This is an inter-sectorial issue affecting the short, medium and long term.

In this sense, for example, the European Commission has long recognized the need to further promote renewable energy given that its exploitation, besides contributing to climate change mitigation through the reduction of greenhouse gas emissions, sustainable development and security of supply, is a key factor in the development of knowledge-based industry which can create jobs, economic growth, competitiveness and regional and rural development.

This last dimension falls outside of the scope of the present work. Here, we focus on the analysis of the four dimensions of the security of electricity generation supply problem, that is, we exclusively analyze the point of view of the sub-problem: the maximization of the net benefit of the electricity system.

3. THE SECURITY OF SUPPLY (SOS) ORIENTED PRODUCT IN THE CONTEXT OF THE REGULATORY MECHANISMS

We have mentioned that regulators sometimes deem necessary the introduction of additional mechanisms to ensure security of supply. These mechanisms represent the core of the different analyses carried out in the thesis.

Continuing with the basic terminology that will be used in the context of the thesis, it is important at this stage to introduce an element playing a key role in the mechanism design: the SoS-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. By directly purchasing this/these product/s on behalf of demand (or compelling demand to acquire it/them), 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.

This security-of-supply-oriented product will focus our attention in a large part of this work, since its definition determines to a large extent the results derived from the implementation of the mechanism. The characteristics of these products depend on the dimension in which the additional mechanism is focused.

The product definition includes several elements, as for instance:

- the underlying asset: e. g. energy, renewable energy, wind energy, capacity (installed or "available", whatever this might mean), capability to provide operating reserves, etc.,
- the financial features: optionalities (forward or option contract, tradable in a secondary market or not, guarantees, penalties, force majeure clauses, etc.), time terms of the commitment (lag period, contract duration), etc.

It is relevant to clarify that by "product" it is often understood just the underlying "physical" asset being traded (i.e. energy production, or availability (or production) in the peaks, or operational reserves, etc.). However, we deem more suitable to understand by product the whole resulting commitment. In other words, the security of supply oriented product is nothing but the regulatory compromise or the contract being signed by the two counterparties, i.e. the demand (or the regulator itself on its behalf) on one side and the generators on the other, with all its inherent characteristics. This way, for instance, in the Colombian Reliability Charge scheme (see Chapter IV), the product, the so-called reliability option which the regulator buys on behalf of demand through an auction, among many other features, consists in a financial call option of energy, with a required physical back-up (guarantee), presenting a lag period of 3 years and a contract duration of up to 20 years. All these characteristics form part of the product definition.

4. Scope and objectives of the thesis

In the context of the aforementioned dimensions of the security of generation supply problem, there is a certain consensus around the idea that the security dimension can be tackled by means of operation reserves markets, where the reserves requirements are prescribed by the System Operator. On the other extreme, it is also commonly agreed that the Strategic Expansion Policy has to be solved through the implementation of additional "out-of-the-electricity-market" mechanisms (e.g. feed-in tariffs or cap and trade mechanisms).

Nowadays, in the liberalized context, the problem is more acute in the other two dimensions, where the debate (particularly on the adequacy dimension) has always been, and still is, quite intense. For many different reasons, ensuring a secure generation supply at these dimensions is still far from being an evident matter as it seems that a system driven by purely market-based mechanisms ensures neither an efficient resource management nor the required long to very long-term expansion.

In the thesis, we have mainly focused on these adequacy and firmness dimensions. Although some general analyses including all the four dimensions are also carried out (see Chapter III), unless explicitly mentioned, it will be considered that the problems being tackled are those stemming from these two intermediate dimensions.

The objectives of the thesis are gathered around three differentiated but interrelated areas of the security of supply regulatory problem: first, the diagnostic of the problem, second, the analysis and design of mechanisms (rules and incentives) to tackle the problem, and third, the development of simulation modeling tools to support both the performance of the system and also to assess the impact of the implementation of these regulatory mechanisms.

Following this division, we next present the roadmap followed in the thesis:

First part (Part B): The diagnostic of the problem

In order to set the basis upon which the subsequent developments of this thesis will be built, the first objective is to analyze how in the "reformed", "deregulated" and/or "liberalized" scheme that governs the electricity business in many systems worldwide, the intervention of the regulator is most of times needed to ensure security of supply.

Under a number of strong simplifying hypotheses, short-term marginal market prices are known to provide optimal incentives for the operation, resource management and investment duties, leading to the maximization of the system's overall efficiency. Nevertheless, as we will review, unfortunately these hypotheses do not hold in real systems.

In this sense, the thesis aims to contribute by developing a renewed discussion of the nature of the so-called market failure in the context of real electricity markets after thirty years of electricity market experience. We show that, although ideally the market itself should be enough to provide with adequate incentives to generation at all dimensions, there are several factors that prevent this result from being achieved.

Once analyzed and described the reasons and scenarios that justify the suitability of intervening in any way to complement the market signals, we analyze in deeper detail the consequences of the so-called market failure at the firmness dimension. We have focused on this dimension, because it represents the dimension which has received less (or even no) attention in the literature. On the basis of a conceptual mathematical model, we seek to demonstrate how in the currently most common scenario (almost in the majority of cases) in which demand does not actively participate in the market, generators' decision-making is naturally affected by their risk aversion in such a way that the overall system efficiency decreases. In this context, we will demonstrate that the regulator intervention on behalf of the overall power system efficiency is fully justified.

Second part (Part C): Security of generation supply mechanisms design: objectives, elements, incentives and outcomes

In the first of the two research fields previously mentioned (the design of regulatory mechanisms to ensure security of electricity supply), the objective of this thesis is to contribute to the discussion of how to properly design the additional incentives provided by these security-of-supply-oriented mechanisms.

With this purpose, first, a stylized mathematical model will be developed to represent and understand how these incentives should be designed to optimally achieve the desired objectives at each one of the dimensions of the problem. The analysis is supported by a qualitative analysis in line with the one developed by Prof. Pérez-Arriaga back in 1994. Analyzing agents' incentives and regulators' objectives will be a major objective of this study. This way, the classic approach will be extended to characterize all four security of supply dimensions. After analyzing the security of supply problem from a broad point of view, we then extend and refine the model to study in detail those mechanisms aimed at solving the inefficiencies detected at the firmness and adequacy dimensions.

In the remainder of the thesis, the focus will be put on these two latter dimensions. The analysis will continue by exhaustively analyzing the international experiences. This analysis is particularly relevant in this context, since when it has come to give solutions to the market inefficiencies at the firmness and adequacy dimensions, actual system regulatory design has often been one step forward than academic theory. The particular characteristics of each power system have led to different requirements, which in the end have translated into many different mechanisms. Thus, it is important to analyze the different resulting outcomes and the major lessons learned.

This study will show that after at least a couple of decades of electric power systems reform, academics and regulators have the means to gather a broader perspective on the whole problem. Although it is clear that there is not one single solution that fits perfectly in the different systems worldwide, there is a common set of criteria that do apply to them all and, in practice, narrows considerably the number of proper parameters and adequate alternatives at the regulator's hand.

The final objective of this PART C is, in the light of the previous two analyses, to set the principles and criteria to take into account when designing a mid- and long-term mechanism aimed at solving inefficiencies detected at the firmness and adequacy dimensions. We will propose a common framework establishing the guidelines to design a well-suited solution.

Third part (Part D): Simulation models to support regulatory design and impact assessment of security of generation supply mechanisms.

The ability to analyze the market performance, set proper objectives and to anticipate and analyze the market response to potential new mechanisms is essential for regulators. In this context, simulation models represent powerful tools to assist the regulator in his duties.

In this research field, we aim to contribute by presenting two different models each one developed to pursue and illustrate different objectives:

- The first model has been developed with the objective to illustrate and deal with the proper definition of metrics to evaluate whether the market is reaching efficient outcomes. We first extend the formulation of the classic Probabilistic Production Cost models (PPC) by proposing a novel algorithm that allows easily and efficiently to introduce demand elasticity and then on this basis we illustrate how traditional reliability measures are not suitable metrics to be used when a non-negligible part of the demand is elastic. This analysis will serve to support some of the discussions developed in Chapter III.
- The second model aims to simulate long-term auction-based security of supply mechanisms in the formats presently in force in some systems (for instance Colombia or New England). As described in PART C, in an attempt to provide electricity generation investors with appropriate economic incentives as well as to maintain quality of supply at socially optimal levels (particularly in the long run), a growing number of electricity market regulators have opted for implementing a security of supply mechanism based on long-term auctions.

  In this context, the ability to anticipate and analyze long-term investment dynamics under this regulatory scheme is a key issue not only for regulators, but also for market agents. A model, inspired on the system dynamics technique, has been developed to serve this purpose. Several new refinements will be introduced so as to allow for better representation in a variety of contexts, among others introducing a multi scenario analysis, the capability to evaluate new hydro projects and a detailed simulation of the generators' bidding curve building process in a context of a long-term auction.

  A real-size simulation based on the Colombian system has been developed to illustrate the capabilities of the model and to show how these models can help to gain insight on the market dynamics in the described context.

5. ORGANIZATION OF THIS DOCUMENT

The document has been organized around the three parts that have been just introduced. After this introduction, PART B aims to analyze the reasons behind the so-called market failure, first in all four dimensions but then mainly focused on the firmness and adequacy ones (Chapter I) and then we particularly focus in some more detail on the firmness dimension (Chapter II).

Then, in PART C we centre our efforts on the analysis of the regulatory mechanism design. First we discuss, from a theoretical point of view, how to design proper regulatory objectives and optimal incentives in this context (Chapter III), then we perform a critical assessment of the international experience (Chapter IV), and finally we summarize the principles and criteria to be taken in to account when designing a security of supply oriented mechanism (Chapter V).

In PART D we develop two simulation models, first in Chapter VI we describe a reliability-oriented model based on the convolution technique, and then in Chapter VII we present a system dynamics based model which is used in the context of long-term auction based mechanisms.

Conclusions and future research have been included in PART E.

6. References

Batlle, C., C. Vázquez, M. Rivier and 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.

Borenstein, S. & J. Bushnell (2000). Electricity restructuring: Deregulation or reregulation. Regulation. The Cato Review of Business and Government 23(2): 46-52.

North American Electric Reliability Council (1997). "NERC Planning Standards", September 1997.

Pérez-Arriaga, I. J. (1994). "*Principios económicos marginalistas en los sistemas de energía eléctrica*", (in Spanish). IIT Working Paper IIT-93-044, January 1994.

Ruff, L. E. (2003). UnReDeregulating Electricity: Hard Times for a True Believer. Seminar on New Directions in Regulation. Kennedy School of Government, Harvard University, Cambridge, MA. May 1, 2003.

PART B

PART B. SECURITY OF ELECTRICITY GENERATION SUPPLY: PROBLEM DIAGNOSIS

Chapter I

Chapter I. General Problem Analysis

* The IIT began working on the security of supply issue back in the beginning of the 90s, led by Prof. Ignacio J. Pérez-Arriaga. Since then, a group of researchers have had the opportunity to be part of a team (Michel Rivier, Carlos Vázquez and Carlos Batlle among others) which has analyzed the matter in more than ten countries, advising regulatory commissions, public institutions, market and system operators and private companies. This work takes as its starting point several pieces of research developed in a series of articles published over the years by the members of this team (some of these references are provided throughout the thesis).

1. INTRODUCTION

Since the very beginning of the restructuring process, back in 1982 in Chile, the ability of an electricity market to provide the system with the required level of security of generation supply has been put into question. The mistrust on the ability of the market, left to its own devices, to provide firstly sufficient (and later also efficient) generation availability when needed, has gradually but inexorably led to the implementation of additional regulatory mechanisms. This matter is undoubtedly gaining importance and it has taken a key role in the energy regulators' agendas.

In this chapter, we develop a renewed discussion of the nature of the so-called market failure in the context of real electricity markets after thirty years of electricity market experience.

After a brief description of the potential inefficiencies that may arise at each one of the dimensions of the problem (security, firmness, adequacy and strategic expansion policy), we focus on the firmness and adequacy dimensions.

In this context, we show that although ideally the market itself should be enough to provide adequate production resource management and investment incentives, there are several factors that prevent this result from being achieved. Ideally, under a number of hypotheses, the market should provide the optimal solution. Here we show how these hypotheses do not hold in practice.

First, as a preliminary step preceding the detailed problem analysis, we delve into larger details on the discussion of the different dimensions of the security of supply problem.

2. THE FOUR DIMENSIONS OF THE SECURITY OF GENERATION SUPPLY PROBLEM

As we have presented in the introduction, for the sake of a better understanding and regulatory design, the security of supply problem can be decoupled from the time dimension perspective into four major components (or dimensions), namely: security (a very short-term issue), firmness (a short to medium-term issue), adequacy (a long-term issue) and strategic expansion policy (a very long-term issue).

Security

The real-time operation of a power system requires a central coordination so as to ensure a continuous match between supply and demand. It is commonly accepted that the System Operator (SO) has to be responsible for such coordination.

As pointed out in (Stoft, 2002), between the real time and the longer term 'there are dividing lines that describe the system operator's diminishing role in forward markets. Where to draw those lines is the central controversy of power-market design'. Each system has traditionally used different criteria to define the point at which the SO takes control of the system so as to ensure security. This point is usually known as the *gate closure*, and it is after this *gate closure* that the security dimension arises.

At gate closure the scheduled generation is transferred to the System Operator to guarantee the quality (maintaining voltage and frequency within acceptable margins), security (short-term uninterruptibility of supply) and financial efficiency (supplying electric power at the lowest possible cost) of supply. In a market environment, the general approach consists in having the SO acquiring (supposedly through a transparent and competitive process) the so-called ancillary services. These products are often classified in three different categories: frequency control (operating reserves: primary, secondary and tertiary); reactive power for voltage regulation; and black-start capabilities (restoration of power).

The ad hoc markets for the quantities of operating reserves -that are prescribed by the SO[1]- are considered to represent a good hybrid (market & regulation) alternative for ensuring security[2]. But although these market mechanisms are considered to be functioning adequately, the intervention of the regulator at this level is manifest:

- First and foremost, it is well-known that operating reserves requirements affect short-term prices and consequently long-term investment signals. Therefore, the determination of the required reserves, usually calculated by System Operator as a percentage of the consumer demand, indirectly interacts with the long-term investment incentives.

---

1 Ideally, an electricity wholesale market would be capable to provide to some extent operating reserves. However, in practice, is unlikely that regulators and the system operators would leave this security issue to the market.

2 It is important to differentiate these operating reserves from the reserves designated for producing electricity during those times when demand is almost greater than the available production capacity, in other words, the reserves that are called upon to supply energy when scarcity in generation capability arises. While operating reserves are a security tool, 'scarcity' reserves are closely-related to firmness and adequacy dimensions. Purchasing reserves with the objective of using them for both purposes is clearly inefficient, since some units capable of helping in the firmness and adequacy dimensions, but not capable of providing operating reserves at reasonable costs, are not be able to compete in the purchasing process.

The importance of this effect has been analyzed in (Stoft, 2003) and (Hogan, 2005), where it is pointed out that changing intentionally this short-term requirements can modify long-term signals and thus, increase the reserve margin. Indeed, in (Stoft, 2003), it is pointed out that no matter if a price cap is in place, the regulator can obtain any reserve margin by just fine-tuning this operation reserve requirements. The reason is that in doing so, the regulator can alter the length of the scarcity periods[3], and thus, the resulting scarcity rents. In this way, by increasing these rents a higher reserve margin would be achieved. The problem with this approach is the fact that it increases the demand for a product which is intended for another purpose, that is, short term system security.

- But the intervention sometimes goes even further, and there are indeed some rules that clearly deviate from the purely market-based principles. For instance, market agents are in most cases compelled to submit bids for all balancing reserves they have available (mandatory) both to increase and decrease energy capacity for the whole scheduling time horizon of the following day.

- Moreover, in many power systems (for instance in most European systems), the imbalance settlement is currently solved through a dual imbalance pricing methodology, where a different price is applied to positive imbalance volumes and negative imbalance volumes (with respect to the resulting system's net imbalance) for each given hour. This dual imbalance pricing itself is supposed to provide incentives for the market agents to try to avoid deviating from their scheduled programs, but this is achieved at the cost of artificially modifying the system's marginal price signal. It is assumed to be a measure intended to improve system security, but it also implies some adverse effects on the development of the market[4] and on overall efficiency, see for instance (Batlle et al., 2007B).

Finally, it is important to highlight that, traditionally, it was implicitly assumed that an electrical system with a high degree of installed and available capacity also presented a high level of available operating reserves, meaning that it had a high degree of flexibility to overcome short term contingencies. Nevertheless, this is not necessarily true, and indeed, the present trend of introducing large amounts of wind energy will require a higher than usual proportion of flexible generation. These requirements should be taken into account in the subsequent dimensions.

---

[3] Note that a scarcity in the operation reserve margin would increase prices in both markets (energy and reserves) without implying a situation of energy supply rationing.

[4] When this dual imbalance pricing is coupled to the fact that imbalances are evaluated on a portfolio basis, a clear entry barrier for new entrants is introduced.

Firmness

Even with abundant installed generation, if, for a variety of reasons (lack of water in the reservoirs or of fuel in the tanks, units out of service for maintenance or because of a forced outage, etc.), a significant part of this capacity is not readily available when needed, then the demand may not be efficiently met.

In the new scheme, 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 was intended to achieve: to leave to the agents tasks that they can perform more efficiently. In the liberalized context, these tasks are driven by the market signals.

From the firmness standpoint, regulators should evaluate whether market signals are capable of ensuring efficient generation resource management, or if it would be appropriate to introduce some additional mechanism to ensure such a result. We present an in depth analysis of this problem in Chapter II, where we use a mathematical model to study how the so-called market failure may result in an inefficient medium term planning.

For regulators, the most inefficient result consists in not having enough generating resources in critical periods. This is the reason why the additional regulatory mechanisms, as analyzed latter in this work, usually consist in providing some kind 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.

In the context of the regulatory mechanisms aimed to tackle the inefficiencies detected at the firmness and adequacy dimensions, it is highly recommended that the regulator defines a methodology to evaluate the production capability of the different generating plants, taking into account current market incentives. This production capability, closely related to the firmness dimension, is often estimated by means of historical data regarding availability, historical production data or sometimes by means of historical availability or production under scarcity conditions. In other cases, it is provided by a mathematical model (which has to be fed again with historical data), etc. This production capability, which will be referred here as *firm supply*, is analyzed in Chapter IV. The assessment of the production capability still represents a challenge from a theoretical perspective, and is considered to be a controversial issue from the generators' point of view, since the determination of these values has often a significant impact on their income.

Adequacy

The regulator's objective in terms of adequacy is to guarantee that appropriate incentives to attract new entrants exist (i.e. incentives to attract new efficient generating units).

As we present latter in PART C, the regulatory instrument basically consists in providing new entrants with an extra source of income and/or the hedging instruments they require to proceed with efficient investments. The product that is received form generators in exchange may present many different forms. The definition of the time terms (lag period[5], the duration of the incentive, etc.) and the volatility associated with this additional income, as we will see, are key factors in this respect.

Additionally, regulators should determine a mechanism that implicitly discriminates "good" from "fair" or even "poor" investments, thus ensuring that correct incentives are provided. The "firm supply" concept for firmness (where any are in place) or a similar measure is often taken as the reference. As has been indicated above, if considered necessary, security criteria could also be taken into account when evaluating the generating units.

There has been much discussion about the convenience of introducing regulatory measures to enable electricity markets to provide an adequacy level (usually just based on reliability criteria) with which the regulator feels comfortable.

Strategic Expansion Policy

Adequacy mechanisms often consist in introducing incentives to enhance the advent of new capacity, in such a way that directly or indirectly, the resulting reserve margin is higher than the one the market would naturally provide if left to its own devices.

This leads to market agents choosing between the different technologies available at each time, as well as the regulator trying to compare the performance of these technologies in the long term.

But the expected firmness of the available technologies may suffer changes in the very long term that are difficult to take into account at the present moment (drastic changes in fuel prices, resources exhaustion, etc.). This is the reason why it may be appropriate to consider the introduction of some very long-term criteria in the long-term mechanisms; these criteria reflect the very long-term view of the regulator.

The introduction of such criteria is usually motivated by the existence of externalities. But the intervention of the regulator at this stage is rather evident. For example, the regulator can decide that it will be profitable for the maximization of the net power system benefit to invest in the development of a new technology given the expectation that after some years it will become an efficient alternative. Wind energy is a good example of this: after years of investing in support mechanisms for wind generation, it seems that the time for cost convergence with traditional alternatives is more than close.

[5] The term 'lag period' is used to refer to the existing time interval between the moment the commitment of deliverability is signed and the moment it has to be delivered. This period allows investors to develop the project.

Security of supply also requires that electricity is supplied in a sustainable manner. Sustainability links the need to provide electricity for present end users whilst caring for the provision for future users, generation after generation. This is not a minor requirement, since the present model of electricity supply –and the entire energy model, for that matter- is not sustainable[6].

The debate about the need for regulatory intervention to ensure security of supply

We have just seen that regulatory intervention is accepted to be necessary at the security and strategic expansion policy dimensions. However, in between these two dimensions the debate on the necessity of intervening to ensure security of supply has always been, and still is, open and quite intense. We focus the remainder of this chapter on this issue, that is, on whether some sort of intervention may be necessary to ensure firmness and adequacy.

3. IS THE MARKET CAPABLE OF ENSURING SECURITY OF GENERATION SUPPLY?

The changes in the regulation of the electric power industry worldwide have modified the traditional security of supply issues and approaches drastically. In the vertically integrated utility, under cost-of-service regulation, security of supply was seen as a major ingredient in the global exercise of centralized utility planning at all levels: generation, transmission and distribution. Under the market-oriented paradigm, the new regulation must make sure that the appropriate economic incentives exist for each one of the activities so that quality of supply is maintained at socially optimal levels.

This thesis only concerns security of supply at the generation level, where the change was more pronounced since, in the new regulation, the generation activity is opened to competition. The theoretical orthodox reason justifying the liberalization process at the generation activity was mainly to promote efficiency at all levels: operation, planning and expansion. However, the market ability to bring efficient results at all these levels in the real world (and especially in the medium to long term) remains as a far-from-being-clear issue.

Since the very beginning of the restructuring process, back in 1982 in Chile, the ability of an electricity market to provide the system with the required level of security-of-supply has been put into question. Some authors, for instance (Pérez-Arriaga, 2001), (Stoft, 2002), (Joskow, 2005) and (Hogan, 2005) contributed to this debate by claiming that, in a number of different contexts, and for a variety of reasons, there is a market failure. This is arguably one of the issues of greatest importance still awaiting a solution under the current regulatory scheme. Although no international consensus has been reached in this regard, with

---

[6] Sustainable development, as defined in (WCED 1987) as development that 'meets the needs of the present without compromising the ability of future generations to meet their own needs'. A sustainable energy model must include some essential features: for instance, lasting and dependable access to primary energy sources and adequate infrastructures to produce and deliver the required amount of energy reliably.

countries opting for one alternative or the other, the more and more accepted existence of this market failure leads to the conclusion that without regulatory intervention, the market, left to its own devices, is unable to provide sufficient generation availability when needed[7].

Next we discuss this issue, demonstrating that the answer to the question is that actual markets need some regulatory intervention. We first briefly review the major results stemming from the marginal theory applied to electricity markets, showing that short-term prices, under ideal hypothesis, are supposed to drive and efficient operation, planning and investments. We then analyze how these hypotheses do not hold in actual markets, and thus, these ideal results are not achieved in practice.

4. IDEALLY THE MARKET SOLVES THE PROBLEM

Under a market-based scheme, driven by demand and supply laws, an equilibrium price[8] and an equilibrium quantity are determined as the result of generators and demand interaction in the market[9].

In perfect competitive short-term markets the equilibrium price (also known as the system's *marginal price*) should ideally represent the demand's marginal utility, and since all plants' supply bids should reflect their actual production costs, except in the case of a scarcity in the generation resources, the equilibrium price should also equal the system's marginal production cost.

Under several strong simplifying hypotheses, these short-term marginal prices are known to provide optimal incentives for the efficient operation and investments that will lead to the maximization of the system's overall efficiency. The most relevant of these hypotheses are:

- The short term spot price reflects the marginal demand utility.

---

[7] Indeed, as shown later in Chapter IV, in almost every electricity market, in one way or another, the regulator has designed some kind of rule to drive or put boundaries to the natural market evolution in an attempt to guarantee supply in the short, medium and long term.

[8] In this work, it will always be assumed that a single price is used to clear the market. In a market where every generator receives its bid price, agents have to estimate the system's marginal bid, in order to bid slightly below this estimated value (as long as their costs are also below this value, this is the profit-maximizing strategy). Thus, the result will be exactly the same as the one provided by the single price clearing market: generators with lower marginal costs will receive a price which is greater than their costs. However, this "pay as bid" scheme may introduce inefficiencies due to the uncertainty involved in this system's marginal bid estimation. This is the reason why a single (marginal) price for the system is the most efficient way to organize the system's operation. A single price market encourages generators to bid their true costs, as it increases the odds that their plants are committed without affecting the price they receive (except for the last accepted bid, which sets the marginal price).

[9] In simple auctions the market equilibrium is directly determined by the intersection of the supply's bid curve and the demand's offer curve.

- Risk is allocated efficiently. That is, there is a well-functioning long-term market[10].
- Generators' costs functions are convex[11].
- There are neither economies of scale nor lumpy investments.
- There is a perfect competitive market with perfect information[12].

It can be demonstrated that within such as context, the expected outcome of a perfect competitive market model (in which market agents make decisions in a decentralized way) equals that which is achieved by the traditional utility model with perfect information (in which the utility centralizes decisions to meet demand, maximizing the system's net social benefit).

Market prices are ideally capable to send the optimal signals: if less generation than the optimal (and efficient) amount is installed, then the market provides higher profits for existing generation. These additional profits act as a signal to attract more generation up to the optimal generation mix. On the contrary, an excessive reserve margin would lead the market to penalize poorest investment decisions.

In Annex A we delve into the classic analytical demonstration of this issue (see section A.i.) and we also use a simplified example to further illustrate how market prices ensure the recovery of both operational and investment costs (see section A.ii.).

5. MARKET IMPERFECTIONS AND FLAWED REGULATORY RULES

Before delving into the discussion of the nature of the so-called market failure in the context of real electricity markets, we briefly present two relevant assumed conditions that will not be discussed here, for they are considered to fall outside of the scope of the present work:

- First, we assume the existence of the ideal conditions to introduce the "reform" at the generation level. We do not analyze here the direct and indirect effects derived from not having the well-known textbook conditions for the introduction of competition at the generation level. These textbook conditions, as discussed in Joskow (2006), include privatization, vertical unbundling, an adequate horizontal structure, etc.

[10] Indeed, most of the theoretical analyses are based on the risk neutrality assumption.

[11] Although this hypothesis does not hold in electricity systems, we will not analyze here the difficulties and changes that non-convexities may introduce. For a comprehensive and detailed description of the problem, see (Vázquez, 2003).

[12] Nonetheless, given the beneficial effects of perfect markets on social welfare, one of the objectives of regulation should be to come as close as possible to creating one. Thus, unless the opposite is specifically mentioned, this is taken as the reference framework in the analyses carried out throughout the document.

In the analyses carried out throughout the document, unless indicated the opposite, it will be assumed that these ideal conditions hold. Nevertheless, it is noteworthy that the non-compliance of these conditions is not always against the generators' interests. For instance, a concentrated horizontal structure may provide agents the possibility to alter prices and thus ease the recovery of investment fixed costs.

- The second condition, usually considered as a relevant imperfection playing a key role in security of supply, is the lack of short-term demand elasticity. Ideally, the most efficient result would be achieved if prices were perceived by the demand-side in real time, and demand could respond to those prices accordingly. However, this is still far from being the case in electricity systems. Nowadays there are still some barriers that avoid this situation from being fully achieved (although less and less as time passes, thanks to the so-called smart meters and demand response programs).

  For the efficiency of the whole scheme, and from a theoretical point of view, it is essential that, in case of scarcity, the price could be determined by the demand offer curve. In case of scarcity, we consider that an administratively price for the non served energy is introduced. As we later discuss, the difference between this administratively determined price and a price cap is weak.

  In the present work, unless indicated the opposite, it will be assumed that there is not demand elasticity in the short term.

Provided that the two previous conditions are in place, we next analyze the impact of the non-accomplishment of the most relevant previously mentioned ideal hypotheses on the efficiency of an electricity market, namely:

- The consequences of the lumpiness problem in generation.
- The consequences of an inefficient risk allocation: the lack of the demand response in the long term.
- The consequences of introducing regulatory flawed rules that distort market signal, and do not allow to appear high prices (ideally the marginal demand utility) when generation is scarce.

5.1. The effect of the existence of the lumpiness investment problem in generation

Investments in generation are lumpy, meaning that certain technologies present a minimum feasible size (installed MWs). This problem has an important implication: short-term prices may not be capable of providing the optimal ideal signal[13] described in the Annex.

[13] This lumpiness investment problem does not impede in rigor that the market reaches the same outcome the regulator would devise, but it does impede the optimal remuneration described in the Annex.

However, this effect is negligible if the size of the system is sufficiently large with respect to this minimum feasible size.

But in small systems the result can be dramatic: high prices in the market cannot provide a correct signal for an investor, since the correct (and optimal) amount of investment (the one that would recover at least both the investment and operation costs) is not feasible. To illustrate this problem with a real example we present below the situation in Peru by March 2009.

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 constituted an incentive for “undesirable” generators (see Chapter IV, section 5.2.6), leading to a dash for extremely expensive (in terms of variable costs) junk peak generation, due to their relatively small capital requirements. This has led to a situation in which, from the standpoint of reliability, the reserve margin is much larger than is theoretically suitable but, at the same time, prices are significantly high. Figure 2 illustrates the current situation.

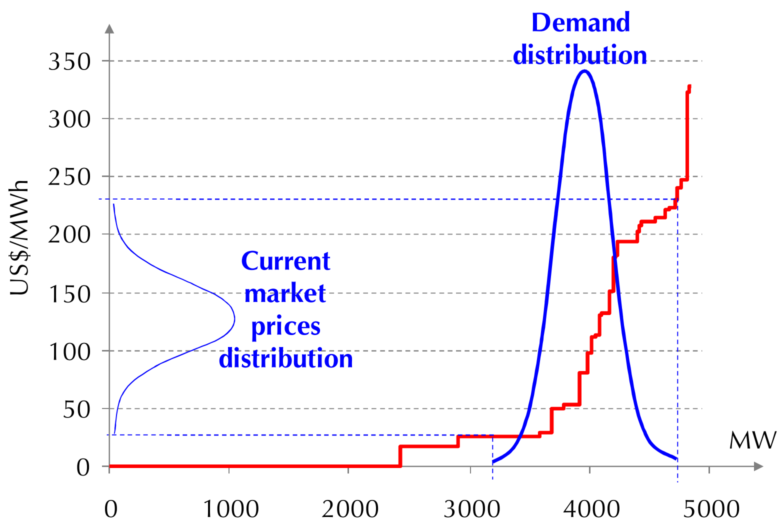


Figure 2. Supply function in March 2009.The distribution of demand and market prices in Peru

A quick look at the previous figure might lead to the conclusion that installing an efficient generating plant would be extremely good business. But unfortunately this is not the case, due to the existence of a (in this case penalizing) lumpiness problem. The fact that such prices will disappear as soon as a more efficient generating unit comes on stream (coupled also with risk aversion), discourages the investment needed to remedy the scarcity episodes to which the system is presently prone. Figure 3 illustrates the consequences of installing a brand new combined-cycle gas turbine of, for instance, 400 MW.

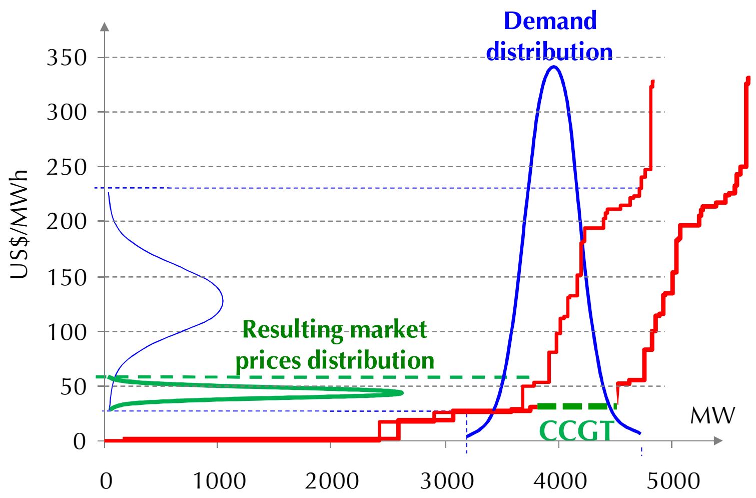


Figure 3. Market price distribution as a result of the installation of new CCGT in Peru

5.2. Agents risk aversion and the consequences of an inefficient allocation of risk

Power generation investment decision-making risk is high and failures[14] are likely. Risk, although to a lesser extent, also

plays a key role in the resource management decision-making process.

The traditional regulatory scheme

In the traditional regulatory scheme, a government-controlled centralized coordinator is responsible for overall electric power system operation decisions, resource management, control and monitoring. This body is likewise entrusted with the formulation of plans for system expansion as regards the installation of both new generating capacity and transmission grid lines and facilities.

Within this context, and from the point of view of the required generation investments, incentives to make decisions efficiently were considered to be weaker, and errors in planning are paid for by (not always all) customers via tariffs[15] or even by the whole society through the general budget. As a consequence, the risks involved are not borne by those who actually invest.

From the resource management perspective, incentives are also weak, since the inefficiencies stemming from over-contracting fuel provisions or from planning an

[14] By 'investment failures' we make reference to those investments that do not maximize the net social benefit as much as other available possibilities would have. The uncertainty involved in the power sector investment decision-making process is the main factor responsible for these suboptimal (when evaluated ex-post) investments.

[15] The former nuclear moratorium in Spain is a good example.

excessively conservative water reservoir management are in the end completely borne by the final consumers.

As we discuss next, in the market context, risk is ideally more efficiently allocated between the different agents through market signals and market mechanisms.

Generation risks in a fully liberalized market

In a business in competition, each generator decides its investments by itself and profitability is in principle not guaranteed ex-ante. Analogously, each agent has to decide the medium term resource management of its generating plants, and again the profitability of its decisions is not guaranteed.

For many different reasons, risk aversion is a particularly relevant characteristic defining generators' behavior in power markets, and as we next comment, it significantly affects a generator's decisions regarding long-term investments and medium-term resource management[16] (e.g. water reservoir management, fuel provision, maintenance scheduling, etc.).

How generators' risk aversion affects long-term investments

New facilities require very large investments, they take time to be installed and operational and there is a lot of uncertainty involved during the typically long economic lifespan (due to, among others, technological, price and regulatory uncertainty). These issues make investment especially risky and also make generators more risk averse than investors in other types of markets. The major consequence is that generators, in their attempt to protect themselves against low price scenarios, tend to install less capacity than if they were risk-neutral.

How generators' risk aversion may affect medium term resource management

In real systems, suppliers have to make important decisions (generally in the medium term) to ensure the capability of existing generation to produce electricity in the future. Thus they have to sign contracts to procure their future fuel requirements[17], they have to decide when it will be more profitable to produce using the limited hydro energy resources available (under the uncertainty of future inflows or the risk of spillage) or they have to decide when to carry out plant maintenance. All these decisions affect the availability of electricity in the future, and thus, the system's future outcomes. But again, in their attempt to protect themselves against risks (low prices, losses derived from water spillages, fuel overcontracting, etc.) generators will be conservative and for instance, they will prefer to

---

[16] The importance and the effect of generators' risk aversion depend on the particular structure and characteristics of the system.

[17] Some contracts may imply rigid constraints as is the case with the "take or pay" or "use it or lose it" modalities.

produce with the limited water resources when prices are moderately high rather than wait for the possible uncertain scarcity in generating resources (implying very high peak prices) in the future.

An example of the risk involved in electricity markets: the case of Brazil.

To illustrate the risks involved in electricity markets, the case of Brazil is presented[18]. Brazil represents an extreme example of a hydro-dominated system, for very large hydrological cycles tend to make generators' income very volatile during a plant's lifespan.

Brazil

In 2005, Brazil had an installed capacity of 91 GW, with hydro generation accounting for 85%, a peak and energy demand near 54 GW and 44 average GW respectively. The hydro system is composed of several large reservoirs, capable of multi-year regulation (up to five years), see (Barroso et al., 2006). The hydrological cycles are usually around six or seven years long, and are characterized by a pattern that includes an extreme wet episode ("El Niño") as well one severe drought ("La Niña"), both difficult to predict with accuracy.

These characteristics lead to the market prices and centrally managed reservoir levels represented in the following figure.

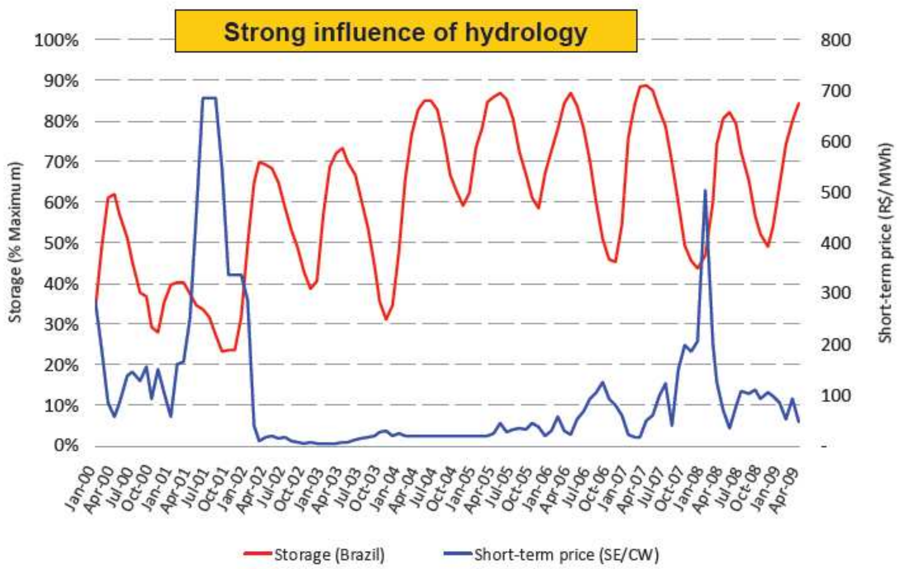


Figure 4. Market prices in Brazil from 2000 to 2009 (Barroso, 2009)

In principle, an energy scarcity such as the one that, due to the exhaustion of hydro reserves, affected the country for nine months during 2001-2002 and resulted in extremely high prices should theoretically be incentive enough for both optimal resource management and investment in suitable generation (not only hydro but also thermal generating units).

[18] We take Brazil as an example, but we could have also resorted to the Colombian or New Zealand cases.

When the risk involved in a fully liberalized context is as large as the one presented here, the medium- and long-term decisions tend to be very conservative. Indeed, with respect to long-term investments, there is no practical way to get any project financing on the basis of an expectation of high profits in perhaps five or seven years' time, if ever.

The other side of the market

This does not mean that demand no longer bears any of the risks involved in generation activity. The generators' risk aversion translates into higher expected profit being required in order to carry out new projects or to make a reliability-oriented medium-term planning with the existing ones. This higher expected profit materializes in a risk premium in long-term markets[19], and where these long-term markets do not exist, as previously mentioned, this risk translates into a lower reserve margin that leads in the end to higher short-term prices.

Demand is also risk averse

Furthermore, risk-averse consumers want to protect themselves against high prices, and would therefore prefer a system with greater installed capacity and greater resource availability than they would prefer if they were risk-neutral.

The ideal market based solution

From a theoretical perspective, there are consistent reasons that support the idea that in a market scheme, both the generation and the demand have enough incentives to hedge their risk and thus allocate the risk efficiently (by signing long-term contracts). In the case of the generator, we have seen how volatile prices may difficult the project finance, or may lead to a sub-optimal resource management (derived from being conservative in the use of the resources).

In the case of the demand there are also clear incentives to enter into long term contracts:

- First, long-term contracting provides demand with the means to hedge against the aforementioned peak prices.
- Second, there are benefits resulting from an efficient management of the generation-side risk. This is most of the times forgotten, and it may represent an incentive even higher than the one of protecting against volatile prices.
  - It is widely accepted that the required expected return on an investment (in any asset but particularly in a generating unit) depends critically on the degree of risk involved (the higher the risk the higher the expected rate of return). Therefore, if demand plays a role in the long-term market and collaborates in the risk management process by signing

[19] Indeed this risk should lead to the development of well-functioning, efficient and liquid long-term markets which make possible determining the price for this risk premium. Unfortunately, for a number of reasons that we will discuss in detail later, long-term markets are illiquid or even non-existent in most power systems.

long-term contracts, it reduces generators' risk exposure, and consequently their required expected rate of return.

- A similar reasoning can be applied with respect to the medium term resource management. Thus, by entering into long-term contracts, a more efficient medium-term resource management can also be achieved, see next chapter for a more detailed analysis on this issue.

Hence, even if the demand were risk neutral, by entering into long-term contracts, a more efficient outcome is achieved. It is a well-known result that, when at least one of the market sides is risk averse, the net social benefit can be maximized by means of well-functioning long-term markets.

Thus, in the market context, both demand and generation have to bear (and suffer the consequences derived from) the risk involved in the generation activity. By liberalizing the generation sector, a means to efficiently allocate the risk between the different agents is found, since both sides have clear incentives to take optimal decisions which should ideally lead to the maximization of the net social benefit.

Therefore, in this context a long-term market should spontaneously arise that would supplement the spot market and solve the risk aversion problem. In this way, agents' risk management would be left to be entirely determined by market forces[20].

This way, in the new liberalized context, both sides of the market also have to bear part of the risk involved in the investment and resource management processes. A serious problem arises when this is forgotten, and this has been the case with most supplier (retail and particularly regulated retail) companies worldwide, which typically commit to purchasing electricity at the price applicable to their wholesale market operations but have yet to enter into long-term commitments (i. e. longer than a year). This modus operandi has been mainly driven by a complete reliance on the fact that "somebody else will ensure the supply"[21].

The need for learning processes of immature electricity markets

Actual electricity markets, even after more than two decades of functioning, cannot yet be considered mature. Even in those rare markets where demand is really exposed to spot prices, long-term contracts (with a duration of more than one year) are not entered into.

---

[20] This does not mean that introducing long-term markets guarantees an efficient outcome. Several experiences (the case of OMIP in the Iberian market is a clear example) have shown that if the regulator decides to put in place (and even provide some funding for) a long-tem market (power exchange), this does not bring demand participation. Efficient long-term markets arise because of the willingness of market participants. Regulators can help by creating a suitable transparent platform for trade, but if the market structure conditions are not adequate, artificially implementing a trading floor makes no difference.

[21] It has been also pointed out that allowing consumers to change retailer without penalization does not provide the right incentive for long-term contracting, see (Neuhoff & de Vries, 2004).

Most consumers are not mature enough to realize the risks involved in electricity markets, and the direct benefits that an efficient allocation of risk would provide. Demand (not only domestic, but also industrial), usually tend to make their decisions using only short to medium term run criteria.

This lack of demand-side response creates a malfunctioning of the long-term market that cannot be solved in the short run, and it causes both a lack of generation investment and also a very conservative (and thus inefficient) medium term resource management, which paves the way for potential future shortages. Note that the need here is not just for consumers to demand less energy from the market when prices are high -this is the typical goal of demand-side management programs- but especially for them to sign efficient hedging contracts to express their need for a higher level of generation reliability (i. e., to express their risk aversion).

The most orthodox solution to this problem would be to do nothing[22] . Consumers, having not signed contracts, would suffer the high prices and the severe consequences which derive from rotating blackouts and, the following year, some of them would realize the need to protect themselves against this situation and would sign some contracts. This process would continue until consumers understood how to operate efficiently in the long-term market.

This reasoning has been defended by various authors in the literature to support the argument that there is no need for any specific security-of-supply regulation. The most common case taken as paradigmatic of this view is the supply shock that hit the Nordic electricity market in 2002-2003 (von der Fehr et al., 2005).

Regulators' risk aversion

Given what we have seen internationally thus far, it is likely that a long learning period, which may include several rationing episodes, would ultimately be considered to be more of a problem caused by the market than a problem caused by consumers that are not acting efficiently.

Electricity is an essential good, without an easy replacement in modern society; shortages of electricity have significant social and political implications which make politicians, regulators and system operators particularly aware of the need for a reliable electricity supply. In most systems, and this was the case for instance in California and Ontario, the market rules will be changed dramatically before consumers have time to complete their learning process. The long-term market will never reach a steady state because it will be completely refurbished before that can happen. In fact, what underlies beneath this problem is the principle that a wise regulator should not assign responsibilities to any individual who

[22] Although some educational programs which provide information about the potential consequences of not contracting may help to reduce the potential impact.

is not prepared to carry them out appropriately. In addition, nowadays there is a common (although arguable) belief that most of the demand is not yet prepared to deal efficiently with the problem of long-term generation reliability.

This discussion can therefore be summarized as follows: politicians' risk aversion is by far larger than that of almost any power consumer. Regulated rates preclude the need for protection against high prices and even consumers initially exposed to spot market prices ignore reliability when making their decisions. There is a certain implicit assurance that leads consumers to believe that the regulator will never allow supply shortfalls or inordinately high prices that would jeopardise their interests.

A consequence of the above is that demand does not respond suitably in the long-term market. Consumers take no interest in a suitable level of adequacy -mainly because there is no real need to respond- and therefore do not include the item in the pricing process. This hinders the installation of generation which is geared to reliability provision.

5.3. The consequences of introducing regulatory flawed rules that distort market signals

It has been long debated in the literature the importance of having an appropriate pricing mechanism to be applied in the event that the market fails to provide enough supply to meet the demand. Indeed, this is considered as one of the cornerstones of the market model.

However, more often than not, we still find many regulators intervening in the short-term marginal signal with the aim of limiting the revenue that generators can extract from the market. These measures are in most cases justified by the absence of adequate demand elasticity, and represent an attempt either to:

- administratively determine the value of non-served energy (as previously mentioned),
- or to limit market power, since in the event of the reserve margin tightening drastically, generators could eventually bid (and be committed) at extremely high prices,
- or to artificially decrease the inframarginal income of generating units. This approach is currently in force in some Latin American markets, in which, for various reasons, the only generating units which have entered the market in recent years are extremely inefficient and therefore expensive fuel plants.

In particular, these regulators' interventions interfering with marginal market price formation have taken many different forms:

Price caps

- Explicit price caps for market prices, for example, 180 €/MWh in the Spanish market or 1000 $/MWh in Alberta, see (AESO, 2009). See section A.iii in Annex A for a simple illustrative analysis of the long-term effect of introducing a price cap.

"Failure price" ("Precio de falla"), in force in certain Latin America power markets. This consists of an administratively defined maximum market price to be paid to the generators committed in hours in which a certain failure has been declared, with the exception of those plants that can certify that their production costs are higher. These plants are paid pay-as-bid.

Offer caps

Offer caps are constraints (explicitly or implicitly) imposed to generators bids. For example:

- in the Spanish electricity market, the law stipulates that generation units are 'obliged to make economic bids', see (CNE, 2005);
- in the Irish market, the "Bidding Code of Practice" stipulates that generators' bids have to based on the "Opportunity Cost", defined as 'the value of the benefit foregone by a generator in employing that cost-item for the purposes of electricity generation, by reference to the most valuable realizable alternative use of that cost-item for purposes other than electricity generation', see (AIP, 2007);
- in California, the Automatic Mitigation Procedure (AMP) implemented in 2002, intended to limit the ability of suppliers of energy in the real-time market to exercise market power. This basically consists in an automatic comparison with previous bids. If an offer price is too high, the AMP reduces it to a price reference that is in accordance with the cost of production in that power plant.

All these previous measures lead to a situation in which in one way or another, the system's marginal prices are only based on generation bids, precluding the participation of the demand in the determination of these prices.

Out-of-the-market interventions

- "Operating reserve shortage" actions. In some cases, when operating reserves fall below a certain level, the SO take actions, such as voltage reductions and non-price rationing of demand (rolling blackouts), to reduce demand administratively see (Joskow, 2007). This type of measures complicates the price formation process in conditions of scarcity, and again affects the proper and expected recovery of generation investments.

The existence of these rules, with their influence on short term market price formation, may affect both the suppliers' medium term resource management and long-term investments. Indeed, these regulatory interventions can hinder the recuperation of the investment costs of those generation units which have already been installed, which, in the longer term, may lead the generation system to expand in ways which are a long way from what is theoretically supposed to be the perfectly adapted situation as described previously.

Long-term markets could alleviate the effect of these flawed short-term regulatory rules

Intervening in short-term prices has severe consequences, but this does not necessarily mean that under this scenario it is impossible for the liberalized market approach to guarantee the recovery of investments. As stated, the fundamental problem is again the lack of demand-side participation. By contracting in the long term, demand could alleviate the effect of these flawed regulatory rules.

6. CONCLUSIONS

The need for regulatory intervention

Up to this point, we have shown that although ideally the market itself should be enough to provide adequate production resource management and investment incentives, there are several factors that prevent this result from being achieved, and some actually existing markets have already experienced problems related with a lack of generation availability (due to lack of production resources that may have been caused by a deficient middle-term resource management and/or by a lack of new investments).

This market failure, sometimes "helped" by some of the aforementioned regulatory interventions regarding short-term price formation, results in the so-called "missing money problem"[23], the "missing signals problem", "missing markets problem", etc. In the end, this has led to the conclusion that in most cases some kind of regulatory intervention is required.

Acknowledgements

The author wants to thank Ignacio J. Pérez-Arriaga for his valuable comments and suggestions, which helped to improve the work developed in this chapter.

7. REFERENCES

AIP, All Island Project (2007). The Bidding Code of Practice. A Response and Decision Paper. AIP-SEM-07-430, 30 July 2007.

Batlle, C., C. Vázquez and J. Barquín (2007B). "A critical analysis of the current balancing mechanism in France". IIT Working Paper, Sep. 2007.

Baughman, M. L., S. N. Siddiqi and J. W. Zarnikau (1997). "Advanced pricing in electrical systems: part I and II". IEEE Transactions on Power Systems, Vol. 12, No. 1, February 1997.

---

[23] By the "missing money problem" we refer to the unrecovered fraction of the investment costs that arises when regulators impose price caps with the objective of limiting prices during scarcity situations. The term was popularized by Shanker (2003).

Bohn, R. E., M. C. Caramanis and F. C. Schweppe (1984). "Optimal pricing of electrical networks over space and time". Rand Journal of Economics, Vol. 15, No. 3, Autumn 1984.

Caramanis, M. C. (1982). "Investment decisions and long-term planning under electricity spot pricing", IEEE Transaction on Power Apparatus and Systems, Vol. PAS-101, No. 12, December 1982.

Caramanis, M. C., R. E. Bohn and F. C. Schweppe (1982). "Optimal spot pricing: practice and theory", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 9, September 1982.

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

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

Joskow, P. L. (2006). "Introduction to Electricity Sector Liberalization: Lessons Learned from Cross-Country Studies," in Sioshansi, F. and W. Pfaffenberger. Electricity Market Reform: An International Perspective, 1-32, 2006.

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.

Neuhoff, K. & L. De Vries (2004). "Insufficient incentives for investment in electricity generations," Utilities Policy, Elsevier, vol. 12(4), pages 253-267, December 2004.

Pérez-Arriaga, I. J. (1994), "Principios económicos marginalistas en los sistemas de energía eléctrica", (In Spanish) Technical Report IIT-93-044, January 1994.

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

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.

Stoft, S. (2003). “The demand for operating reserves: key to price spikes and investment”, IEEE Transactions on Power Systems, Volume: 18, Issue: 2, pp. 470- 477, May 2003.

Vázquez, C. (2003). “Modelos de casación de ofertas en mercados eléctricos” (in Spanish). PhD Thesis, Universidad Pontificia Comillas, September 2003.

Von der Fehr, N. H. M., E. S. Amundsen and L. Bergman (2005). “The Nordic Market: Signs of Stress?” The Energy Journal, Special Issue, 71-98.

Annex A. OPTIMAL SHORT-TERM PRICES UNDER IDEAL HYPOTHESES

A.i. Theoretical results under ideal hypotheses

Here we review the major results stemming from microeconomic marginal theory applied to electricity markets, and we show how short-term prices, under ideal hypotheses, are supposed to drive efficient operation, planning and investments.

The application of microeconomic marginal theory to the electric power systems was first sketched by a MIT research group (Caramanis et al., 1982), (Bohn et al., 1984), (Caramanis, 1982), (Scheweppe et at. 1988) and has been subsequently complemented and refined by some other works, among others (Pérez-Arriaga, 1994), (Pérez-Arriaga & Meseguer, 97), (Baughman et al, 1997) and (Vázquez, 2003).

The classic analysis makes use of a reference model, which consists in an ideal centralized planner having perfect information about costs and agents' preferences, and whose objective is the maximization of the net social benefit. This reference model is compared with the one resulting from a market context where short-term energy prices are the sole signal driving agents' decisions. The main objective is to analyze whether or not both contexts are equivalent, in other words, whether or not short-term market prices are capable of driving efficient operation, planning and investments.

Some ideal hypotheses are considered in this analysis, being the most relevant ones:

- Generators' costs functions are convex.
- Agents' are not risk averse.
- Generators can only get revenue from the sale of their energy in the short term market.
- There are neither economies of scale nor lumpy investments.
- The market is perfectly competitive.

Optimal prices for operation

The optimal centralized operation problem consists in a central planner maximizing the net social benefit. Thus, this problem can be schematically represented as:

$$\begin{aligned}
&\underset{q_{ih}}{Max} \quad \sum_h [U_{dh}(\sum_i q_{ih}) - \sum_i C_i(q_{ih})] \\
&s.t. \\
&\quad q_{ih} \le \overline{q}_{ih} \quad \perp \psi_{ih} \\
&\quad R(q_{ih}) = 0 \quad \perp \zeta_{ih}
\end{aligned} \tag{1}$$

Where:

$C_i(q_{ih})$ represents the variable costs incurred by the unit i when producing the quantity q_{ih} in the hour h

U_{dh} represents the demand utility function in hour h for the total consumption $Q_h = \sum_i q_{ih}$.

$\overline{q}_{ih}$ is the maximum output limit of unit i in hour h.

$R(q_{ih}) = 0$, represents schematically the operational technical constraints of the different generating units.

ψ and ζ are the dual variables of the previous constraints.

By forming the Lagrangian function and then calculating the first order derivative with respect to the decision variables (q_{ih}) we obtain the optimality conditions of the problem:

$$\begin{aligned} &\frac{dU_{dh}(\sum_i q_{ih})}{dq_{ih}} - \frac{dC_i(q_{ih})}{dq_{ih}} + \psi_{ih} + \frac{dR(q_{ih})}{dq_{ih}}\zeta_{ih} = 0 \Rightarrow \\ &\Rightarrow \frac{dU_{dh}(Q_h)}{dQ_h} = \frac{dC_i(q_{ih})}{dq_{ih}} - \psi_{ih} - \frac{dR(q_{ih})}{dq_{ih}}\zeta_{ih}, \forall i,h \end{aligned} \tag{2}$$

Therefore, each generating unit should produce in each hour up to the level in which its marginal costs equals the marginal demand utility, in other words, the cost of producing an additional unit (\$/MWh) should equal the price (\$/MWh) that the demand is willing to pay for the last MWh consumed. Indeed, this relationship will only be true in each hour h for the marginal unit, which is the generating unit i that is producing in that moment and whose technical constraints are not binding (i.e. ψ_{ih} and ζ_{ih} have a zero value).

On the other hand the problems of the generators and demand in a market context can be represented as:

$$\begin{array}{ll} \textit{Demand's problem} & \textit{Generators' problem} \\ \underset{Q_h}{Max} \ \sum_h [U_{dh}(Q_h) - \pi_h \cdot Q_h] & \underset{q_{ih}}{Max} \ \sum_h [\pi_h \cdot \sum_i q_{ih} - \sum_i C_i(q_{ih})] \\ & \quad s.t. \\ & \qquad q_{ih} \le \overline{q}_{ih} \quad \perp \psi_{ih} \\ & \qquad R(q_{ih}) = 0 \quad \perp \zeta_{ih} \end{array} \tag{3}$$

Again, we obtain the first order condition for each one of the corresponding Lagrangian functions with respect to the decision variables (Q and q_{ih} respectively) so as to analyze the optimality conditions of each problem:

$$\begin{array}{ll} Demand's\ optimality\ conditions & Generators'\ optimality\ conditions \\ \dfrac{dU_{dh}(Q_h)}{dQ_h} = \pi_h, \forall h & \pi_h = \dfrac{dC_i(q_{ih})}{dq_{ih}} - \psi_{ih} - \dfrac{dR(q_{ih})}{dq_{ih}}\zeta_{ih}, \forall i,h \end{array} \tag{4}$$

It is straightforward to check how these optimality conditions are equivalent to the conditions obtained in the central planner problem[24]. Therefore, under the ideal hypotheses enumerated above, both contexts should provide the same outcome.

Note that short-term prices should always be determined by the marginal demand utility. These short-term prices are also equal to the marginal costs of the marginal unit. But in the particular case where all existing plants are at their full capacity, the market price will not correspond to any of their marginal costs. This is a very important: when there is not enough generation to meet demand requirements, the price has to be set by the demand (not by any of the marginal costs of the generating plants) so as to ensure an efficient outcome.

Optimal prices for investment

We have seen how short-term prices should drive an efficient operation in a market context. But, in order to conclude that both, the ideal central planner and the market context, lead to the same results, it is essential to prove that short-term market prices send also optimal signals to long-term investments. With this purpose we next extend the previous analysis in order to include the investments in generation.

The new optimal centralized operation and investment problem can be schematically represented as:

$$\begin{array}{l} \underset{\overline{q}_{ih}}{Max} \quad NSB(q_{ih}, \overline{q}_{ih}) - \sum\limits_i IC_i(\overline{q}_{ih}) \\ NSB = \begin{cases} \underset{q_{ih}}{Max} \quad \sum\limits_h [U_{dh}(\sum\limits_i q_{ih}) - \sum\limits_i C_i(q_{ih})] \\ s.t. \\ \quad q_{ih} \le \overline{q}_{ih} \qquad \perp \psi_{ih},\ \forall i,h \\ \quad R(q_{ih}) = 0 \quad \perp \zeta_{ih},\ \forall i,h \end{cases} \end{array} \tag{5}$$

[24] Note that for the sake of simplicity we have assumed that the set of R constraints is the same in both cases (central planner and market). In this respect, a more general representation can be found in (Pérez-Arriaga, 1994).

Where:

NSB is the net social benefit, i.e. the objective function of the centralized scheduling problem.

IC_i represents the investment costs of the generating plant i.

The optimality condition of the investment problem is:

$$\frac{dNSB}{d\overline{q}_{ih}} = \frac{dIC_i(\overline{q}_{ih})}{d\overline{q}_{ih}} \tag{6}$$

Meaning that investments should be carried out up to the point in which the long-term marginal cost equals the short-term marginal increment of the net social benefit.

In the operation problem, if we take into account the relation existing between the objective function and the dual variable ψ_{ih} we have:

$$\frac{dNSB}{d\overline{q}_{ih}} = \psi_{ih} \Rightarrow \psi_{ih} = \frac{dIC_i(\overline{q}_{ih})}{d\overline{q}_{ih}} \tag{7}$$

Thus, if we introduce the previous expression in the first order condition of the operation problem we obtain:

$$\frac{dU_{dh}(Q_h)}{dQ_h} - \frac{dC_i(q_{ih})}{dq_{ih}} + \frac{dR(q_{ih})}{dq_{ih}}\zeta_{ih} = -\frac{dIC_i(\overline{q}_{ih})}{d\overline{q}_{ih}}, \forall i,h \tag{8}$$

On the other hand the generators' and demand problem in a market context can be represented as:

$$\begin{array}{ll}
\begin{array}{c} Demand's \;\; problem \\ \underset{Q_h}{Max} \;\; \sum_h [U_{dh}(Q_h) - \pi_h \cdot Q_h] \end{array} &
\begin{array}{l} \quad Generators' \;\; problem \\ \underset{\overline{q}_{ih}}{Max} \;\; B - \sum_i IC_i(\overline{q}_{ih}) \\
B = \begin{cases} \underset{q_{ih}}{Max} \;\; \sum_h [\pi_h \cdot \sum_i q_{ih} - \sum_i C_i(q_{ih})] \\ s.t. \\ \quad q_{ih} \le \overline{q}_{ih} \qquad \perp \psi_{ih}, \;\forall i,h \\ \quad R(q_{ih}) = 0 \quad\; \perp \zeta_{ih}, \;\; \forall i,h \end{cases} \end{array}
\end{array} \tag{9}$$

Where

B is the generator accumulated benefit (along the period considered) in the short-term market, i.e. the objective function of the generator' s operation dispatch problem in a market context.

The optimality conditions of this problem are:

$$\begin{array}{ll}
\textit{Demand's optimality conditions} & \textit{Generators' optimality conditions} \\
\dfrac{dU_{dh}(Q_h)}{dQ_h} = \pi_h, \forall h & \dfrac{dB}{d\overline{q}_{ih}} = \dfrac{dIC_i(\overline{q}_{ih})}{d\overline{q}_{ih}} \Rightarrow \psi_{ih} = \dfrac{dIC_i(\overline{q}_{ih})}{d\overline{q}_{ih}}, \forall i,h \\
 & \pi_h = \dfrac{dC_i(q_{ih})}{dq_{ih}} - \psi_{ih} - \dfrac{dR(q_{ih})}{dq_{ih}} \zeta_{ih}, \forall i,h
\end{array} \tag{10}$$

Again, it is straightforward to check how these optimality conditions are equivalent to the conditions obtained in the central planner problem. Therefore, under the ideal hypotheses enumerated above, both contexts should provide the same outcome in terms of operation and investments.

A.ii. Inframarginal profits: illustrating how fixed investment costs are recovered in the market context

Next we use a simplified example to further illustrate how market prices ensure the recovery of both operational and investment costs. Two additional ideal assumptions with respect to the former analysis have been introduced for the sake of simplicity in the exposition: no technical constraints are considered in the operation and the marginal demand utility has been considered to be constant.

To show, in a simplified way, how generators can fully recover their investment costs from the income derived from the energy market (although prices are based solely on operating short-term costs and demand's short-term marginal utility), the graphic procedure (also known as the screening curves method) that was used in traditional systems to calculate the optimal generation mix that minimizes overall costs can be used.

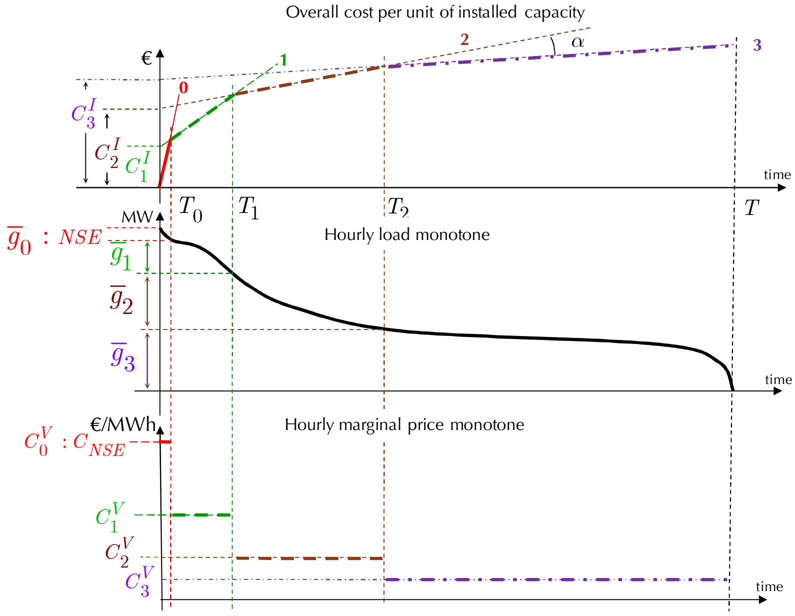


Figure 5. Optimal generation mix

The upper part of Figure 5 (below) represents the evolution (per unit of installed capacity) of different technologies' overall costs as a function of the number of hours of use.

Technology 0 has no investment costs, so there is no cost if it is not used, and it has a high operating cost, so the costs increase rapidly with the hours of use. This “technology” is a means of representing the social cost which derives from the loss of demand surplus when some energy cannot be provided by other existing technologies.

This is a key issue for the success of the overall design, since it is essential to ensure the recovery of the investment costs of the generating units. This is particularly so in the case of the peaking units (traditionally the ones that have the highest variable costs; technology 1 in the case of the example) where, if they are not paid their opportunity cost, which should be related to the cost of the non-served energy (the technology 0 in the case of the example), no investment cost will be recovered at all.

Technologies 1, 2 and 3 do have some fixed investment costs, denoted in the figure as C_1^I, C_2^I and C_3^I respectively, which constitute respectively the total cost when the equipment is not used. From this value, costs grow in proportion to each technology variable’s cost of operation.

The piecewise-linear bold line at the top of Figure 5 shows the most efficient alternative for each value of hours of use. Thus, if a certain megawatt of generation is going to be used for a number of hours greater than T_2, then the best solution is to construct a megawatt of technology 3. If that megawatt is to be used for a period that falls between T_2 and T_1, the most efficient alternative would be to construct a megawatt of technology 2 (and analogously for T_1 and T_0 and technology 1). Finally, if the group is going to be producing fewer hours than T_0, then it is better to provide that consumption with a megawatt of the type 0 “generator”, i. e. it is not worth supplying that energy.

Once the T_2, T_1 and T_0 values are known, by means of the graphical analysis shown in the figure, it is possible to determine, using the system load duration curve, how much power will be consumed for more than T_2 hours, how much will be consumed between T_2 and T_1 hours, and so forth. Thus, the $\overline{g}_1$, $\overline{g}_2$ and $\overline{g}_3$ capacities that must be installed in each of the three production technologies considered can be obtained. This process is illustrated in the second graph in Figure 5, which represents the optimal capacities that ensure overall cost minimization; hence this process also represents the desirable mix under a centralized hypothesis.

From now on, we will assume that this is the generating mix installed in a competitive market and we will assess whether short term market prices allow a full recovery of investments costs.

In the time interval between T_2 to T, technology 3 sets the system's marginal price, which equals its variable cost C_3^V (see the lower graph in Figure 5). That price allows technology 3 generators to recover their costs of operation, but does not provide any compensation for their investment costs. In the interval that ranges from T_1 to T_2 the market price equals the variable cost of the technology 2, C_2^V. Technology 3 obtains, in each of those hours, an operating profit that equals the difference between technology 2's variable cost and its own variable cost. Graphically, this is equal to the difference between the slope of the costs curve, in other words, the tangent of the angle. Thus, group 3 obtains a profit equal to the price spread for the duration of the period, i. e. $\mathrm{tg}\alpha\cdot\left(T_2 - T_1\right)$. In Figure 6, this is equal to the segment a.

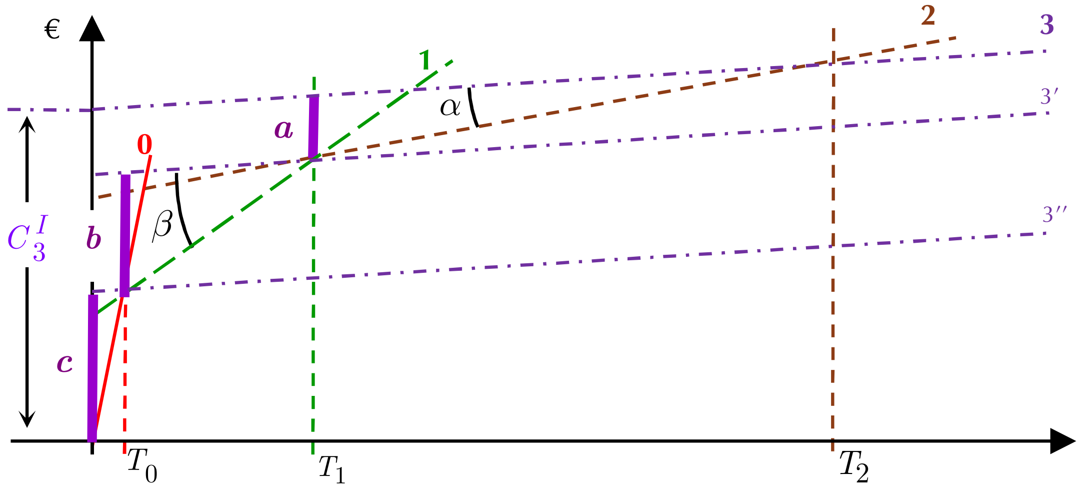


Figure 6. Detail of Figure 5: technology 3's investment cost recovery.

Similarly, in the hours that go from T_1 and T_0, technology 3's income will be equal to $\mathrm{tg}\beta\cdot(T_1 - T_0)$, which is the segment b, and analogously the segment c, for the interval from zero to T_0. As can be seen, the sum of the segments a, b and c (total income) is equal to technology 3's investment cost (C_3^I).

It is important to note the importance of the segment c, which represents the income received when the generation is scarce and, as previously mentioned, the price is set by the demand. If restrictions are imposed on the price during those hours, neither the peak generator nor all the remaining technologies will be able to fully recover investment costs..

The procedure can be repeated analogously for technologies 2 and 1, with equivalent results. This reasoning, which has been presented here with only three (plus one) generators for the sake of simplicity, can be extended without any difficulty to a larger number of energy generation technologies.

Thus the generating mix that minimizes overall costs provides the scenario in which all generation fully recovers both its investment costs[25] and its operation costs. This is known as the generators' break-even position. If less generation than the optimal amount is installed, then the market provides higher profits for existing generation. These additional profits act as a signal to attract more generation up to the optimal generation mix, where the break-even position is restored. On the contrary an excessive reserve margin would lead the market to penalize poorest investment decisions.

A.iii. Analyzing the long-term effect of introducing a price cap

In the short term, the implementation of a price cap affects the income of the generating units in the system. This is illustrated in Figure 7, which shows when the "price cap technology", denoted as "technology 0*", replaces what we called "technology 0" in the previous demonstration (i.e. the technology that served as a means to represent the demand's utility for electricity). The income in the interval from zero to T_0 is represented by segment c^*. Therefore, the immediate consequence is that short term market incomes no longer suffice to fully recover both fixed and variable costs of the desirable (from the social perspective) mix.

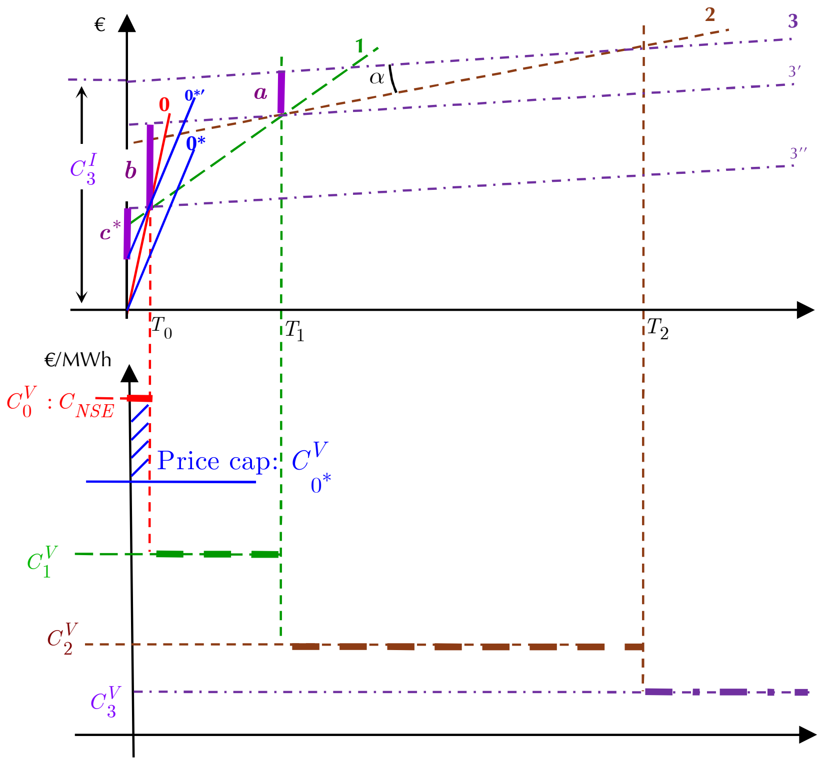


Figure 7. Short-term impact on investment cost recovery of regulatory interventions.

[25] Including depreciation and a rate of return on debt and equity capital

But the impact of this kind of measures goes far beyond the short term, since a regulatory measure of this nature affects the future expansion of the generation system: so as to secure the projects' profitability, the generation system adapts itself to the new regulation by decreasing investment in peaking units, which leads to more frequent scarcity events characterized by these capped prices.

Taking previous Figure 7 as a reference, Figure 8 illustrates this latter issue. The implementation of a price cap changes the resulting prices in the short term as they move from the ones corresponding to the ideal scenario described previously (the black dashed line in the lower graph of the figure) to the ones represented by the continuous red line. But if regulatory intervention continues, peak units withdraw from the system, leading to more frequent scarcity periods in which the price cap is met and to the prices represented in the graph by the dash-pointed blue line.

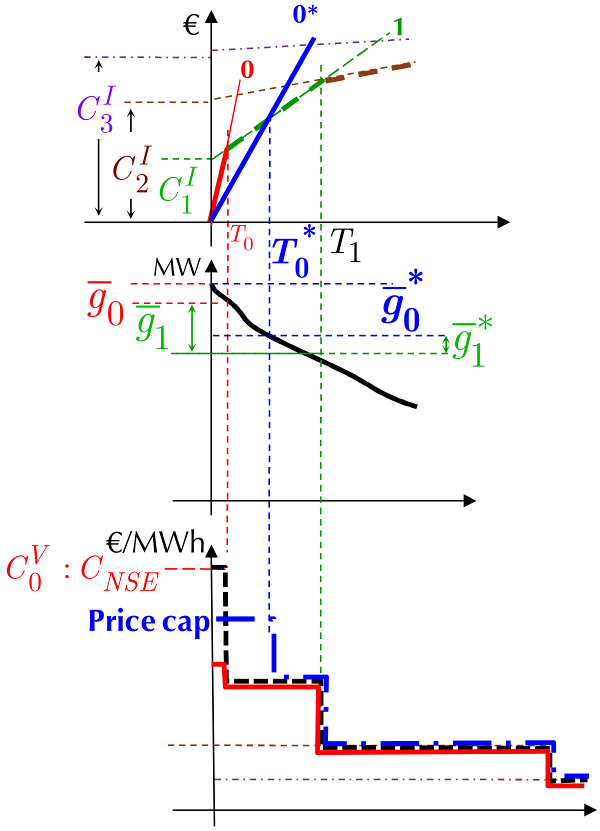


Figure 8. Long-term impact of regulatory interventions on generation structure, reliability and prices

Chapter II

Chapter II. ANALYSIS OF THE MARKET FAILURE AT THE FIRMNESS DIMENSION

* The work developed in this chapter would have not been possible without the key contribution of Álvaro Baíllo and Santiago Cerisola, who are to all intents and purposes co-authors of it.

1. INTRODUCTION

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 at the power generation level or if some additional regulatory mechanism needs to be introduced, and in the latter case, which is the most suitable approach to tackle the problem.

Long has been debated on how to ensure system security, system adequacy, and (more recently) how to ensure sustainable resources in the very long term (what has been termed strategic expansion policy). However, little attention has been devoted in the literature to the firmness dimension[26]. We have defined previously firmness as the ability of the already installed facilities to supply electricity efficiently. This capability of providing electricity efficiently depends on the units' technical characteristics and on the medium-term resource management (e.g. hydro reserves scheduling). Since the decision on the characteristics of the generating units is a longer term concern, from the firmness standpoint, efficiency can only be addressed by ensuring that market signals provide the right incentives to manage resources optimally.

This dimension is particularly relevant in those systems where the medium-term resource management plays an important role (e.g., in hydro dominated systems). Furthermore, taking care of the firmness dimension is a matter of the utmost importance from the regulator's perspective, when trying to ensure quality of supply within a short to medium time scope (e.g. two years, which practically leaves no room for bringing new investments to the system). In this chapter we contribute to the analysis of the market failure at this particular dimension from a theoretical point of view.

In this context, the first question is whether the introduction of competition at the generation level guarantees a satisfactory medium term resource management of the generating units. If this is not the case, some additional regulatory mechanism may be needed.

[26] Indeed, traditionally no distinction has been made between firmness and adequacy, implicitly considering that both can be solved through the same regulatory mechanism. 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. For instance, the SoS mechanism implemented in Spain includes two differentiated services: the availability service and the investment service, see (Batlle et al., 2008b).

The goal of this chapter is to focus on the analysis of how generator's risk aversion coupled with the generators' inability to efficiently hedge medium to long-term positions[27] may compromise to a large extent the system firmness. In other words, we study how this situation may alter the medium term resource management reducing the net social welfare. In order to illustrate all this, we focus on the medium-term management of hydro reserves[28].

From the point of view of generating companies a number of analyses on how risk aversion may affect medium term planning can be found in the literature, see for instance (Unger, 2002) or (Fleten et al, 2002). Here we focus on the social consequences, analyzing how generators' risk aversion coupled with incomplete markets may affect the social benefit.

Here we contribute to the theoretical analysis of the firmness dimension. In particular:

1) We identify the main factors that have an impact on the ability of already installed units to provide generating resources when needed.

2) We analyze in detail the impact of generators' risk aversion. We study how in the absence of any kind of instrument providing generators with a hedge (i.e. a well-functioning long-term market), generators risk aversion affects critically the efficient (from a social perspective) management of generation resources.

3) Then we show how the existence of well-functioning markets for risk (modeled in this chapter by means of forward markets for the sake of simplicity), helps the system to provide a greater degree of firmness (i.e. a more efficient resource management).

4) Finally, we discuss whether, under the light of this analysis, if a well-functioning market for risk does not arise of its own accord, some intervention of the regulator may be recommended in order to guarantee an adequate level of firmness.

The analysis follows the steps of a classical regulatory analysis, see (Pérez-Arriaga and Meseguer, 1997), for the particular case of firmness:

1) Define the concept subject of analysis and identify the relevant factors (section 2).

2) Formulate and solve the benchmark optimization problem (section 3). In this case, the benchmark solution is the ideal central planner problem.

3) Formulate and solve the problem of the individual agents, paying particular attention to the ingredients of interest, in this case how risk aversion affects the medium term resource

[27] This happens because of market incompleteness. If markets are incomplete, they will be in general Pareto inefficient, see for instance Magill and Quinzii (2002).

[28] To ease the model formulation, we build the analysis upon hydro reservoir management, but it is important to highlight that it is directly applicable to any sort of limited energy plant (LEP), for instance CCGTs, which could eventually be subject to gas transmission constraints.

management in the absence of well-functioning long-term markets. Compare with the benchmark and consider the introduction of new market signals (section 4).

4) Formulate and solve the problem of the individual agents introducing the new market signals, i.e. the forward market (section 5). Compare with the benchmark and draw conclusions (sections 6 and 7).

2. SYSTEM FIRMNESS

We have defined firmness as the ability of the already installed facilities to supply electricity efficiently[29].

In the context of an electricity market, this capability of providing electricity efficiently depends principally on two factors:

- The characteristics of the generation units: the amount of load-following units, the percentage of the so-called intermittent generation resources, etc.
- The medium-term management carried out by the generation companies, i.e., the management of fuel stocks, of hydro reserves and of scheduled maintenances. In the context of an electricity market, these medium-term decisions are mainly driven by market signals. These planning decisions condition to a large extent the availability of resources when most needed.

Since the decision of the characteristics of the generating units is a longer term concern, from the firmness standpoint, efficiency can only be addressed by ensuring that market signals provide the right incentives to manage resources optimally.

The deregulation of the power industry is based on the assumption that the decisions of the individual agents lead to an outcome similar to the one that a central authority with perfect information would devise. Thus, in order to evaluate the performance of an electricity market design, the benchmark is usually the optimization problem that the central authority would solve (the solution that maximizes the net social benefit). If the market deviates significantly from the benchmark, then the regulator may consider adequate to introduce an additional mechanism to complement market signals.

In the context of the medium-term management of generation resources, agents may adopt decisions that are far from the efficient solution (from overall system perspective). For example, as we show next, in the absence of well-functioning long-term markets, generation risk averse companies may shift hydro production from one time period to a different one in order to hedge their market risk exposure.

In the next sections we provide an example of how to carry out a regulatory analysis of the system firmness. We show how, due to risk aversion and in the absence of long-term

[29] That is, in such a way that the net social benefit is maximized.

markets, generation companies may change the management of hydro reserves and deviate from the benchmark solution. We also show that the existence of a forward market may mitigate this undesired result. We finally discuss that the mere existence of a forward market is not a sufficient condition, since in some systems worldwide these markets are not functioning properly. If this is the case, then additional regulatory intervention may be required.

3. THE BENCHMARK PROBLEM

In this section we formulate and solve the optimization problem that we use as a benchmark for our regulatory analysis.

3.1. General modeling assumptions

It is not our goal to analyze all the possible issues involved in the firmness dimension[30]. Therefore, our model of the power system only includes the essential ingredients for the purposes of our study, which is the analysis of how risk aversion may affect the medium term planning. We intentionally avoid unnecessary details that may obscure the regulatory analysis[31].

We consider the following setting:

- The demand side is constituted by a large number of consumers with no bargaining power. At each time period t they consume an amount of electricity q_t which results in a certain degree of satisfaction or utility $U_t^D(q_t)$. We assume demand utility functions to be strictly increasing ($dU_t^D \,/\, dq_t > 0$) and concave ($d^2U_t^D \,/\, dq_t{}^2 \leq 0$)
- The generation side is constituted by a large number of generation companies with no ability to affect the spot price of electricity (no market power). At each time period t they produce an amount of electricity q_t, which is the sum of the output of their thermal units (T) and the output of their hydro units (H). Thus the total amount of electricity produced is at each time period is given by the following expression:

$$q_t = q_t^T + q_t^H \tag{1}$$

- We assume thermal generation costs $C_t(q_t^T)$ to be strictly increasing and convex.

[30] For instance regulatory flawed rules distorting short term prices may prevent the efficient result from being achieved. Here we do not focus in all the possible flaws preventing the market to reach the most efficient outcome.

[31] A much more detailed benchmark problem should be developed to assist regulatory decisions. Here, for the sake of simplicity and clarity, we focus on the classic formulation where no risk aversion on the regulator side (assumed to act on behalf of the demand) is considered.

3.2. Problem formulation

We formulate the benchmark problem as the maximization of the expected social welfare, which is given by the difference between the expected utility of demand and the expected cost of generation:

$$\max_{q_t^T, q_t^H} \quad E\left[\sum_{t \in \mathcal{T}} U_t^D(q_t) - C_t(q_t^T)\right] \tag{2}$$

subject to hydro reserve balance equations.

A usual approach to solve optimization problems under uncertainty is to adopt a discrete representation of the probability distribution, see for instance (Birge & Louveaux, 1997). Thus we assume a representation based on a multistage scenario tree such that:

- The set of scenarios is denoted by $\mathcal{S}$. Each scenario s has a probability p_s. The set of nodes belonging to scenario s is $\mathcal{I}_s$.
- The set of time periods is denoted by $\mathcal{T}$. At each time period t we have a set of nodes $\mathcal{I}_t$. $\mathcal{I}_{ts} = \mathcal{I}_t \cap \mathcal{I}_s$ is the node of period t that belongs to scenario s.
- The probability of reaching node $i \in \mathcal{I}_t$ is $p_{t,i}$.
- Each node i has a set of descendant nodes, $\mathcal{D}_i$. The transition probability between i and $j \in \mathcal{D}_i$ is $p_{i,j}$.

Under this framework we reformulate the benchmark problem as follows:

$$\max_{q_{t,i}^T, q_{t,i}^H} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}_t} p_{t,i} \left(U_{t,i}^D(q_{t,i}^T, q_{t,i}^H) - C_{t,i}(q_{t,i}^T) \right), \tag{3}$$

$$\text{s.t.:} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}_{ts}} q_{t,i}^H = Q_s^H, s \in \mathcal{S} \tag{4}$$

We have included an explicit formulation of the hydro reserves balance for each scenario s, where the total hydro production is given by Q_s^H.

3.3. Optimality Conditions

In order to obtain the first-order necessary conditions we formulate the Lagrangian function, $\mathcal{L}(q_{t,i}^T, q_{t,i}^H)$, and compute its first partial derivatives with respect to the decision variables.

$$\mathcal{L}^G = \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}_t} p_{t,i} \left(U_{t,i}^D(q_{t,i}^T, q_{t,i}^H) - C_{t,i}(q_{t,i}^T) \right) + \sum_{s \in \mathcal{S}} p_s \cdot \lambda_s \left(Q_s^H - \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}_{ts}} q_{t,i}^H \right) \tag{5}$$

where λ_s can be interpreted as the marginal value of water in scenario s (notice that we have intentionally extracted the probability from this Lagrange multiplier).

The optimality conditions stemming from the partial derivatives with respect to the output of thermal units, lead to a usual result: at each node i the marginal thermal costs are equal to the marginal utility of demand:

$$\frac{\partial \mathcal{L}}{\partial q_{t,i}^T} = 0 \rightarrow \frac{dU_{t,i}^D}{dq_{t,i}} = \frac{dC_{t,i}}{dq_{t,i}^T}, t \in \mathcal{T}, i \in \mathcal{I}_t. \tag{6}$$

Assume that hydro reservoirs do not reach their limits in any time scenario. Then the hydro production in each of the nodes of the last stage is given by the total inflows of its scenario minus the hydro production in the rest of the nodes of that scenario. This leads to a link between the marginal utility in each node i and the expected marginal utility in its descendant nodes:

$$\frac{\partial \mathcal{L}}{\partial q_{t,i}^H} = 0 \rightarrow \frac{dU_{t,i}^D}{dq_{t,i}} = \sum_{j \in \mathcal{D}_i} p_{i,j} \frac{dU_{t+1,j}^D}{dq_{t+1,j}}, t \in \hat{\mathcal{T}}, i \in \mathcal{I}_t, \tag{7}$$

where $\hat{\mathcal{T}}$ is the set of decision stages excluding the last stage. Notice that this condition can also be expressed in terms of marginal costs:

$$\frac{dC_{t,i}}{dq_{t,i}^T} = \sum_{j \in \mathcal{D}_i} p_{i,j} \frac{dC_{t+1,j}^D}{dq_{t+1,j}^T}, t \in \hat{\mathcal{T}}, i \in \mathcal{I}_t, \tag{8}$$

This means that hydro reserves are managed in order to balance thermal marginal costs between periods. In each node, i, the marginal cost is equal to the expected (observed from i) marginal cost in the descendant nodes.

3.4. Numerical example

For illustrative purposes we define a numerical example:

- Two time stages, $t \in \{1,2\}$.
- Two scenarios with equal probabilities and total hydro productions $Q_1^H = 4$ GWh and $Q_2^H = 5$ GWh.

- Thermal generation costs function are the same in every node of the tree and equal to $C(q^T) = (q^T)^2 + 0.01(q^T)^3$.

Demand utility varies depending on the node of the tree: $U^D_{t=1,i=1}(q) = 10q - 0.5q^2$, $U^D_{t=2,i=1}(q) = 10q - 0.5q^2$ and $U^D_{t=2,i=2}(q) = 12q - 1.1q^2$.

The results obtained for the benchmark problem in this numerical example are shown in table 1.

Table 1: Results for the benchmark problem

t	i	q^H	q^T	$dU^D / dq = dC / dq^T$
1	1	2.90	2,31	4,78
2	1	1.10	2,88	6,01
2	2	2.10	1,73	3,55

4. MARKET EQUILIBRIUM WITH RISK AVERSION IN THE GENERATION SIDE

4.1. The demand side

As mentioned, the demand side has no influence on the spot market price ($d\pi_{t,i} / dq_{t,i} = 0$, where π stands for the spot price). We also assume that consumers are risk neutral. As a result, the joint decisions of all consumers can be modeled through the following optimization problem:

$$\max_{q_{t,i}} \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{I}_t} p_{t,i} \left(U^D_{t,i}(q_{t,i}) - \pi_{t,i} q_{t,i} \right). \tag{9}$$

Let $\mathcal{L}^D(q_{t,i})$ be the Lagrangian function for the demand side. The optimality conditions are given by:

$$\frac{d\mathcal{L}^D}{dq_{t,i}} = \frac{dU^D_{t,i}}{dq_{t,i}} - \pi_{t,i} = 0, t \in \mathcal{T}, i \in \mathcal{I}_t, \tag{10}$$

which is the well-known result that the demand side consumes electricity until the marginal utility obtained is equal to its price.

4.2. The generation side

We assume that generators cannot affect the price of electricity, but that they are risk averse. For each generator we model risk aversion through a utility function U^G of its profit. This utility function gives more weight to scenarios with lower profits and less weight to scenarios with higher profits.

4.3. Modeling generators' risk aversion by means of their utility functions

We assume that the plants management decisions made by generators are not aimed to maximize their bare market margin functions (defined as the difference between the market income and the production costs) but rather their own utility functions. These utility functions implicitly consider their risk aversion.

Although the traditional approach of expressing the generators' risk aversion consists in imposing constraints over the probability distribution function of their market profit margin[32], we have rather preferred to use the utility function since it embodies a compact (and differentiable) representation of the generators' preferences[33], thus easing the conceptual model formulation.

For each generator G, this utility function, denoted as U^G, is defined over the market profit margin in scenario s, (P_s^G). As illustrated in Figure 9, a risk averse utility function is assumed to be strictly monotone ($dU_g / dP > 0$) and concave ($d^2U_g / dP^2 < 0$).

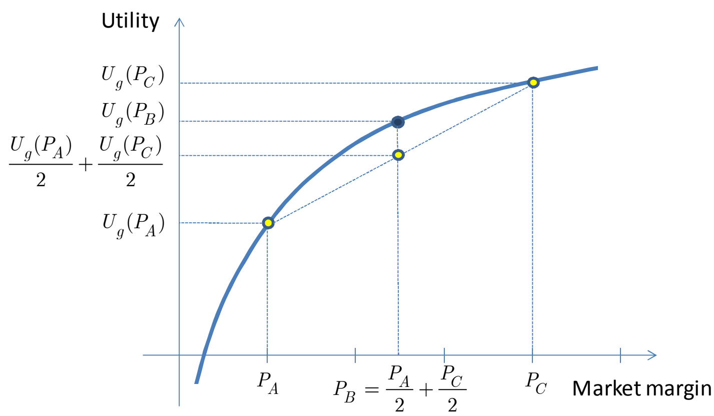


Figure 9. Generator utility function

As it can be observed in Figure 9, a concave utility function penalizes uncertainty. For example, let us suppose two different scenarios, A and C, each of them presenting the same probability of occurrence and characterized by the resulting profits P_A and P_C

---

[32] Such as constraints imposed on the resulting variance, the Value at Risk (VaR) or the Conditional Value at Risk (CVaR), see for instance (Cabero et al, 2010).

[33] Similar approaches can be found in the literature, see for instance (Fan et al, 2009), where the generator's utility function is used to make a conceptual analysis of the generation investment problem (and particularly under regulatory uncertainty regarding CO2 policies).

respectively. The expected utility under such an uncertain situation would be $U^G(P_A)/2 + U^G(P_C)/2$.

This results in a lower utility than the one that would have been obtained if the average profit, $P_B = P_A/2 + P_C/2$, had been received with probability 1. This can clearly be observed in Figure 9.

By means of these utility functions, the problem of each generation company can be formulated as follows:

$$\max_{q_{t,i}^T, q_{t,i}^H} \sum_{s \in \mathcal{S}} p_s U^G\left(P_s^G(q_{t,i}^T, q_{t,i}^H)\right), \tag{11}$$

$$\text{s.t.:} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}_{ts}} q_{t,i}^H = Q_s^H, s \in \mathcal{S}, \tag{12}$$

where $\mathcal{I}_{ts} = \mathcal{I}_t \cap \mathcal{I}_s$ is the node of period t that belongs to scenario s. $P_s^G(q_{t,i}^T, q_{t,i}^H)$ is the profit of the generation company in scenario s, which is given by:

$$P_s^G(q_{t,i}^T, q_{t,i}^H) = \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}_{ts}} \pi_{t,i}(q_{t,i}^T + q_{t,i}^H) - C_{t,i}(q_{t,i}^T). \tag{13}$$

Notice that in this case we have adopted a scenario-wise representation because of the previously described way we have defined the utility function for the generation side.

The Lagrangian function for the generation problem is given by:

$$\mathcal{L}^G = \sum_{s \in \mathcal{S}} p_s \left(U^G(P_s^G) + \lambda_s \left(Q_s^H - \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}_{ts}} q_{t,i}^H \right) \right), \tag{14}$$

where again, λ_s can be interpreted as the value of water in scenario s, and we have intentionally extracted the probability from this Lagrange multiplier.

The first optimality condition is:

$$\frac{\partial \mathcal{L}^G}{\partial q_{t,i}^T} = \left(\pi_{t,i} - \frac{dC_{t,i}}{dq_{t,i}^T} \right) \left(\sum_{s | i \in \mathcal{I}_s} p_s \frac{dU^G\left(P_s^G\right)}{dP^G} \right) = 0, \tag{15}$$

with $t \in \mathcal{T}, i \in \mathcal{I}_t$. Since U^G has been defined as an increasing function, this condition is only fulfilled if $\pi_{t,i} = \dfrac{dC_{t,i}}{dq_{t,i}^T}$ $\forall t \in \mathcal{T}, i \in \mathcal{I}_t$. In other words, the generator increases the output of its thermal units until its marginal cost is equal to the price of electricity.

The second optimality condition is:

$$\frac{\partial \mathcal{L}^G}{\partial q_{t,i}^H} = \pi_{t,i}\left(\sum_{s|i\in\mathcal{I}_s} p_s \frac{dU^G(P_s^G)}{dP^G}\right) - \sum_{s|i\in\mathcal{I}_s} p_s\lambda_s = 0, \tag{16}$$

with $t \in \mathcal{T}, i \in \mathcal{I}_t$. We interpret this optimality condition both under the assumption of risk neutrality and under the assumption of risk aversion:

4.3.1. The generation company is risk neutral:

If the generation company is risk neutral, ($U^G(P_s^G) = P_s^G$) then, at any of the last nodes of the tree we have $\pi_{\overline{t},i} = \lambda_s$, $i \in \mathcal{I}_{\overline{t}}, s \mid i \in \mathcal{I}_s$, which means that the value of water for scenario s is equal to the price of electricity at the final node of that scenario.

If we move backwards in the tree, at each node we have:

$$\pi_{t,i} = \sum_{j\in\mathcal{D}_i} p_{i,j} \cdot \pi_{t+1,j}, t \in \hat{\mathcal{T}}, i \in \mathcal{I}_t, \tag{17}$$

It means that water resources are managed in such a way that the resulting spot price at each node is equal to the expected future spot price seen from that node. Taking into account the relationship between the prices and the marginal demand utility, this is a similar condition to (7). Then if generators are risk neutral, the medium term planning decisions are exaclty the same as those observed in the benchmark problem.

4.3.2. The generation company is risk averse:

Again we first particularize in the last nodes of the tree and then we go backwards (substituting the value of the Lagrange multiplier). Condition (16) takes then the following form:

$$\pi_{t,i} = \sum_{j\in\mathcal{D}_i} \tilde{p}_{i,j} \cdot \pi_{t+1,j}, \ t \in \hat{\mathcal{T}}, i \in \mathcal{I}_t, \tag{18}$$

$$\tilde{p}_{i,j} = \sum_{j\in\mathcal{D}_i} \frac{\sum_{s|j\in\mathcal{D}_i} p_s \cdot \frac{dU_s^G}{dP_s^G}}{\sum_{s|i\in\mathcal{I}_s} p_s \cdot \frac{dU_s^G}{dP_s^G}} \tag{19}$$

Where $\tilde{p}_{i,j}$ takes the form of a risk modified probability. Using these new probabilities (see the expression above) an analogous interpretation can be drawn: the water resources are used so as to make the price in each node i, equal to the expected price, using the risk modified probabilities, in the descendant nodes j.

Since we defined the utility function as concave, the lower the income of a certain scenario the higher the value of the corresponding derivative dU_s^G / dP_s^G, and consequently the higher the value of the associated risk modified probability. This way, the prices in any of descendant nodes leading to low profits scenarios are overweighed.

This condition obviously implies a different resource management that the one analyzed in the benchmark problem. Thus, when generation companies with hydro capacity are risk-averse they may use their hydro reserves in order to hedge their risk exposure. This leads to a market outcome different from the benchmark solution, which can be interpreted as an inefficient situation. We therefore need to explore additional mechanisms or a regulatory intervention in order to take the market back to the benchmark solution.

4.4. Computing market equilibrium

The market equilibrium can then be computed by simultaneously solving the optimality conditions of the demand side (10) and the generation side (15) and (16).

In order to gain intuition, we assume that the generation side is constituted by companies with similar portfolios and similar utility functions. Under this assumption we can treat them as a single generation company in order to compute the market equilibrium.

Equation (10) together with equation (15) are equivalent to equation (6) and have the usual interpretation for a perfectly competitive market: At equilibrium, the marginal cost of electricity is equal to the marginal utility of demand and this determines the price that consumers should pay and generators should receive.

As pointed out previously, if we assume that generation companies are risk neutral, the optimality condition (17) is equivalent to (7). In other words, if generation companies are risk neutral the market equilibrium yields the same outcome as the benchmark problem.

In contrast, if generation companies are risk averse, a different market equilibrium is reached. As we have discussed, we expect a shift of hydro production with respect to the benchmark solution, in order to hedge against low-profit scenarios and thus maximize their expected value of their utility function.

4.5. Numerical example

In order to illustrate the impact of risk aversion on the management of hydro reserves by generators we solve a numerical example.

We consider the same setting as in the previous example. In order to obtain comparable results, we assume that generation companies have similar portfolios and similar utility functions and treat them as a single price-taking generation company.

If we assume that generation companies are risk-neutral we obtain an equilibrium identical to the result obtained for the benchmark problem. The price at each node of the scenario tree is given by the marginal cost at that node. The profits obtained by the generation side in the two scenarios are $P_1^G = 34.89$ and $P_2^G = 30.07$.

We then introduce a utility function for the generation side given by $U_G = -145.92 + 9.489P^G - 0.12288(P^G)^2$[34]. The results obtained are shown in table 2. As expected, the generation companies shift hydro production with respect to the benchmark solution. In this case, it implies moving resources from period 2 to period 1 in order to improve the utility obtained in scenario 2. The profits obtained are $P_1^G = 34.34$ and $P_2^G = 30.42$.

Table 2: Results for the market equilibrium with risk-averse generation

t	i	q^H	q^T	$\pi = dU^D / dq = dC / dq^T$
1	1	3.14	2,24	4,63
2	1	0,86	2,96	6,18
2	2	1,86	1,86	3,82

As can be seen in the next figure, risk aversion changes the market equilibrium by reducing the generation side profits in the scenarios where they were higher and increasing them in the scenarios where they were lower. The average generation profit is thus lower, but the average utility perceived by generation companies is higher.

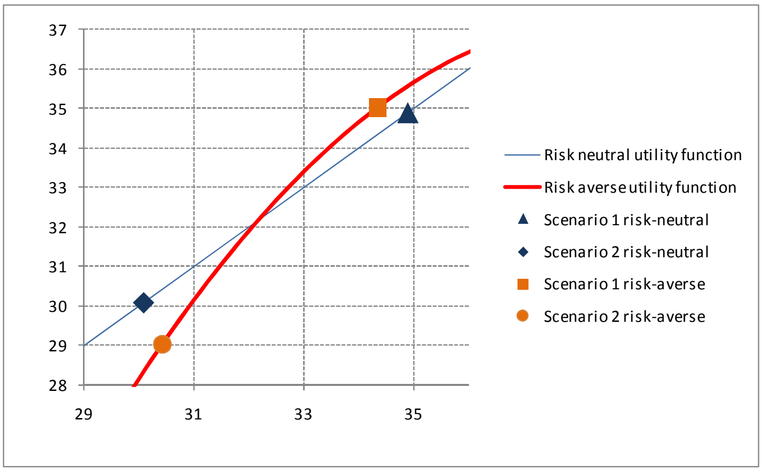


Figure 10. The influence of risk aversion on generation profits and expected utility

Since this new dispatch deviates from the benchmark solution, this translates into a reduction of the social benefit.

[34] Note that in the interval of interest, this function can be used as a utility function, since it meets the conditions defined previously.

4.6. Analysis of the results

We have shown that generator's risk aversion may introduce social inefficiencies in the medium-term management of hydro reserves (the firmness of the system) due to market incompleteness (the impossibility to hedge risks).

In the particular case analyzed, risk averse generation companies tend use more hydro reserves in the first period in order to improve their profit in low income scenarios. This implies a change in the firmness of electricity supply and a reduction on the social efficiency (with respect to the benchmark).

The regulator should be aware of this effect and evaluate the possibility of intervention in order to accommodate it.

5. MARKET EQUILIBRIUM WITH RISK AVERSION IN THE GENERATION SIDE AND A FORWARD MARKET

5.1. The joint forward-spot equilibrium

In this section we analyze the impact that the existence of a well-functioning forward market has on the medium-term management of hydro reserves under risk aversion from the generation side.

Computing the joint forward-spot equilibrium of an electricity market requires making assumptions with respect to the behavior of market participants. For example, (Allaz & Villa, 1993) formulates and solves the joint forward-spot equilibrium for the case of two competing firms as a two-level game. This leads to a structure known as equilibrium problem with equilibrium constraints (EPEC). In contrast, (Cabero et al, 2010) argues that the problem should be addressed as a repetitive game in which forward and spot decisions are at the same level. This implies assuming that, not only current decisions in the forward market affect future decisions in the spot market, but also current decisions in the spot market affect future decisions in the forward market.

In this thesis we use the same approach as in Cabero et al. (2010) and formulate a single-level equilibrium problem in which the agents take their decisions for the forward and the spot market simultaneously. The link between the spot price of different periods is reinforced by hydro reserves. Thus, it is reasonable to assume that the decisions for the spot market and for the forward market are taken simultaneously.

Our goal is to obtain the optimality conditions that define the joint forward-spot equilibrium. To that end we introduce the forward market into the equations that we formulated in section 4. A full representation of the dynamic hedging process is required to represent the forward market. Introducing this full representation obscures to a large extent the formulation we have presented so far. Thus, for the sake of clarity in the exposition, we have rather preferred to resort to a simple two-period scenario tree representation to analyze this case. This tree presents the structure shown in Figure 11. However, these results can easily be generalized to include any number of periods.

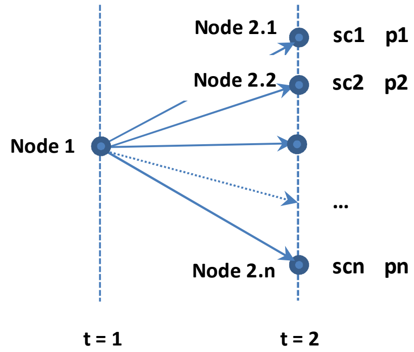


Figure 11. The two period model

We consider that the demand side buys from the generation side a certain amount of forward electricity contracts, q^F, in the first period, with delivery in $t=2$ at a price π^F.

5.2. The demand side

The optimization problem that represents the behavior of the demand side is formulated in this simple case as follows:

$$\max_{q_{t,i},q^F} \quad U_{1,1}^D(q_{1,1}) - \pi_{1,1}\cdot q_{1,1} - \pi^F\cdot q^F + \sum_{i\in\mathcal{I}_{t=2}} p_{2,i}\cdot\left(U_{2,i}^D(q_{2,i}) - \pi_{2,i}(q_{2,i}-q^F)\right) \tag{20}$$

Let $\mathcal{L}^D(q_{t,i},q^F)$ be the Lagrangian function for the demand side. The first optimality condition is equation (10). The second optimality condition is:

$$\frac{d\mathcal{L}^D}{dq_t^F} = -\pi^F + \sum_s p_s \pi_{2,s} = 0, t\in\mathcal{T}, \tag{21}$$

which means that the demand side will interchange electricity in the forward market up to the point at which the forward price is equal to the expected spot price value.

5.3. The generation side

The optimization problem that models the generation side is again given by equations (15) and (16) except for the inclusion of a new decision variable, q^F. The profit of each generation company in each scenario s also includes the effect of the forward market. Again, in this particular two period case the expression takes the following form:

$$P_s^G(q_{t,i}^T,q_{t,i}^H) = \pi_{1,1}\cdot(q_{1,1}^T+q_{1,1}^H) - C_{1,1}(q_{1,1}^T) + \pi^F q^F + \sum_{i\in\mathcal{I}_{t=2}} \pi_{2,i}\cdot(q_{2,i}^T+q_{2,i}^H-q^F) - C_{2,i}(q_{2,i}^T)$$

(22)

A new optimality condition arises when we derive the Lagrangian function with respect to the new variable q^F:

$$\frac{\partial \mathcal{L}^G}{\partial q^F} = \sum_{s \in \mathcal{S}} p_s \frac{dU^G(P_s^G)}{dP^G} \left(\pi^F - \pi_{2,i}\right) = 0, \forall t \in \mathcal{T}, \tag{23}$$

which leads to the following relationship between forward and spot prices:

$$\pi^F \sum_{s \in \mathcal{S}} p_s \frac{dU^G(P_s^G)}{dP^G} = \sum_{s \in \mathcal{S}} p_s \frac{dU^G(P_s^G)}{dP^G} \pi_{2,i}, \forall t \in \mathcal{T}, \tag{24}$$

We interpret this optimality condition both under the assumption of risk neutrality and under the assumption of risk aversion:

5.3.1. Generators are risk neutral:

If the generation side is risk neutral then equations (24) and (21) are the same. In this situation both the demand side and the generation side do not distinguish between the forward market and the spot market and multiple equivalent equilibria exist (since q^F is not determined).

5.3.2. Generators are risk averse:

Note that equation (18) can be written as follows:

$$\frac{\partial \mathcal{L}^G}{\partial q^F} = \sum_{s \in \mathcal{S}} p_s \frac{dU^G(P_s^G)}{dP^G} \left(\pi_{1,1} - \pi_{2,i}\right) = 0, \forall t \in \mathcal{T}, \tag{25}$$

In can be checked how equations (25) and (23) can only be fulfilled simultaneously if the forward price is equal to the price of the node at the node where that forward is signed, in this case $\pi^F = \pi_{1,1}$.

If we also introduce the optimality condition of the demand (21), we finally obtain the following relationship:

$$\pi_{1,1} = \sum_s p_s \pi_{2,s} \tag{26}$$

Thus, in this case, note that $\tilde{p}_{i,j}$ is exactly de same as $p_{i,j}$.

Expression (26) is nothing but the optimality condition defining the hydro resource management of the benchmark problem. As previously pointed out, this result can be extended to show that in under the hypothesis presented here, a well-functioning long-term market brings back the optimal resource management solution.

It is important to bear in mind that our analysis of the optimality conditions does not take into account the impact of constraints. We have intentionally formulated the optimization problems without including constraints in order to obtain meaningful and reasonably simple optimality conditions. If the solution were limited by one or more of the constraints, then the optimality conditions would take a different form.

5.4. Numerical example

We consider the same numerical example as in 3.4 and 4.5 and introduce the optimality conditions given by equations (21) and (23). The results obtained are shown in table 3

Table 3: Results for the market equilibrium with risk-averse generation and a forward market

t	i	q^H	q^T	q^F	$\pi = dU^D / dq = dC / dq^T$	π^F
1	1	2.90	2,31	-	4,78	-
2	1	1.10	2,88	1.96	6,01	4,79
2	2	2.10	1,73	1.96	3,55	4,79

The generation side is selling in advance energy in period 1 for delivery in period 2 at a forward price $\pi_2^F = 4.79$. This produces a gain in the low-profit scenario 1 (where the spot price is 3.55) and a loss in the high-profit scenario 2 (where the spot price is 6.01).

Eventually the profit obtained by the generation side in both scenarios is the same, $P_1^G = P_2^G = 32.48$. This is a solution with a higher average utility for the generation side.

As previously pointed out, this result is exactly the one we obtained when solving the benchmark problem. This means that the firmness of the power system in both cases is also the same if a well-functioning long-term market is in place.

5.5. Analysis of the results

In section 4 we saw that if the generation side is risk averse and there is no forward market, the market equilibrium can deviate from what we have defined as the benchmark solution. The reason is that generation companies use their hydro resources to hedge against low-profit scenarios.

In this section we have shown that the existence of a well functioning forward market brings again the optimal social solution provided by the benchmark problem. The generation side would hedge against by purchasing or selling forwards electricity and then selling or buying electricity at the spot price.

In an isolated system (with no additional arbitrageurs), we have seen that a risk-neutral demand has incentives to participate in these markets up to the point at which no arbitrage is possible. Depending on the case this may imply taking long or short positions in the forward markets.

In the particular case of the numerical example that we have solved, generation companies make forward purchases of electricity (this represents probably the less intuitive scenario).

6. THE INTERVENTION OF THE REGULATOR TO ENSURE AN EFFICIENT MEDIUM-TERM PLANNING

Up to this point, it has been studied how social efficiency increases when market agents (generation and demand) enter into long term contracts. The risk neutral demand side clearly benefits from helping generators' to hedge their risk.

The problem is that up to now, due to a number of different possible reasons demand does not yet actively play this important role, thus inducing regulators to intervene by administratively introducing any kind of reinforcement of the long-term signal perceived by generators. This can be achieved either by compelling demand to enter into long-term contracts or by directly acting on its behalf.

This intervention has commonly been aimed and justified as a way to enhance system adequacy, but a rigorous analysis of the mechanisms put in place shows that in many cases, the objective is also to affect the way generators operate and plan their generating units in the medium term planning, particularly the one corresponding to the so-called LEPs (limited energy plants, hydro in most cases).

7. CONCLUSIONS

With the advent of electricity markets, new regulation is required to supervise that the market is capable to ensure an adequate security of supply level. This is particularly relevant at the generation activity, where the liberalization process has been more intense.

Among the four dimensions of the security of supply problem, in this chapter we have focused on the firmness dimension. Particularly, we have isolated the medium term resource management problem (e.g. how to manage water resources).

In this context, the first question is whether the introduction of competition at the generation level guarantees a satisfactory medium term resource management of the generating units. If this were not the case, some additional regulatory mechanism may be needed.

We have shown that when the generation side is risk averse and there is not a well-functioning long-term market, the market equilibrium can deviate from what we have defined as the efficient benchmark solution. The reason is that generation companies use their hydro resources to hedge against low-profit scenarios.

A well functioning forward market would replicate the efficient benchmark solution. This would imply the generation side hedging against risk by purchasing or selling forwards electricity and then selling or buying it at the spot price. The demand would have clear incentives to be the counterpart of such contracts, even if it were risk neutral (the case analyzed).

The problem is that up to now, due to a number of different possible reasons demand does not yet actively play this important role. This lack of participation supports the regulators intervention to avoid inefficient market outcomes.

8. REFERENCES

Allaz, B. & J. L. Vila (1993). Cournot competition, forward markets and efficiency. Journal of Economic Theory, 59:1-16, 1993.

Batlle, C. & I. J. Pérez-Arriaga (2008). Design criteria for implementing a capacity mechanism in deregulated electricity markets. Utilities Policy, 16(3):184-193, 2008.

Birge, J. R. & F. Louveaux (1997). Introduction to stochastic programming. Springer Verlag, 1997.

Cabero, J., M. Ventosa, S. Cerisola & A. Baíllo (2010). Modeling risk management in oligopolistic electricity markets: A Benders decomposition approach. IEEE Transactions on Power Systems, 25(1):263-271, 2010.

Fan, L., B. F. Hobbs & C. S. Norman (2009). “Risk Aversion and CO2 Regulatory Uncertainty in Power Generation. Investment: Policy and Modeling Implications”. North American Electric Reliability Council. Planning standards. 1997.

Fleten, S. E., S. W., Wallace and W. T. Ziemba (2002) "Hedging Electricity Portfolios via Stochastic Programming," in Decision Making Under Uncertainty: Energy and Power, vol. 128, C. G. a. A. Ruszczynski, Ed., Springer-Verlag, New York, 2002, pp. 71-93.

I. J. Pérez-Arriaga & C. Meseguer (1997). Wholesale marginal prices in competitive generation markets. IEEE Transactions on Power Systems, 12(2):710-717, 1997.

Magill, M & M, Quinzii. (2002). Theory of incomplete markets. MIT Press.

North American Electric Reliability Council (1997) “NERC Planning Standards”, September 1997.

Schweppe, F. C., M. C. Caramanis, R. D. Tabors & R. E. Bohn (1988). Spot pricing of electricity. Kluwer Academic Publishers, 1988.

Unger, G. (2002), "Hedging Strategy and Electricity Contract Engineering," Ph.D. Thesis, Swiss Federal Institute of Technology, ETHZ, Zurich, 2002.

PART C

PART C. SECURITY OF SUPPLY MECHANISMS DESIGN: OBJECTIVES, ELEMENTS, INCENTIVES AND OUTCOMES

Chapter III

Chapter III. Regulatory Objectives and Optimal Incentives in the Context of Security of Supply Mechanisms

1. INTRODUCTION

We have shown that although ideally the market itself should be enough to provide adequate incentives to keep security of supply at efficient levels, there are several factors that prevent this result from being achieved, and that some existing markets have already experienced problems. This so-called market failure, sometimes “helped” by some regulatory interventions regarding short-term price formation, results in the so-called “missing money problem”, the “missing signals problem”, “missing markets problem”, etc. In the end, this has led to the conclusion that in most cases some kind of regulatory intervention is required.

Indeed, Ofgem (2010) and CEER (2009), two consultation processes underway at this writing, are indicative of the importance attached to this concern. Both initiatives are designed to receive feedback from all stakeholders on how to ensure security of supply in every time horizon[35]. Another clear example is that most power systems worldwide, as for example the case of UK, France, Ireland, PJM, Colombia, Brazil, Panama, Peru, etc, have recently implemented or are currently in the process of revisiting a mechanism to ensure security of supply (SoS in the following). In Chapter IV and Chapter V we analyze the different mechanisms that have been implemented worldwide, as well as the most relevant design elements. In this chapter we introduce the fundamentals on the objectives that these mechanisms should seek to achieve, as well as the principles to be fulfilled by the corresponding incentives so as to optimally reach its goals.

Generally speaking, the introduction of an additional regulatory mechanism calls for the definition of an SoS-oriented product (e.g. a capacity credit, energy forward, renewable energy, etc.) to be provided by generators. The characteristics of these products depend on the dimension on which the additional mechanism is focused.

By directly purchasing this/these product/s on behalf of demand (or compelling demand to acquire it/them), 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 chapter 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. A secondary objective is to show how these mechanisms condition the final market outcomes. In other words, we show that by introducing any additional mechanism, the regulator is recovering to some extent its former central planner role.

---

[35] In the case of Ofgem the consultation process was motivated by a previous analysis (Ofgem, 2009), which highlighted the possibility of a future shortage on supply in the UK in the near future (around 2015).

The analysis is structured around three main themes. Section 2 contains a discussion of the metric needed to evaluate market performance as a preliminary step prior to designing additional regulatory incentives. Regulators must design a methodology to qualitatively and (as far as possible) quantitatively assess whether (and to what extent) market mechanisms are failing to yield the expected/required outcome. Section 3 shows that once a metric is defined and its application reveals a market shortcoming, the regulator's task is to introduce proper incentives to guide the power market towards optimal performance. The theoretical analysis of the optimal incentives required to improve system performance in keeping with the regulator's objectives evinces the need to properly define the type of SoS-oriented product that must be acquired. In section 4 the existence of these additional mechanisms is shown to entail a return, to a lesser or greater extent, to regulator-governed central planning with all that it involves. As we discuss, the specific design details of these additional mechanisms will clearly condition market results (scheduling, planning and investment). In this section, some more detail in the modeling hypotheses to better represent actual SoS mechanisms implemented worldwide and how they may affect generation decisions are introduced.The main conclusions are set out in section 5.

2. A METRIC TO EVALUATE THE SYSTEM PERFORMANCE

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

The necessary preliminary step in evaluating the suitability of introducing an SoS mechanism is to design a methodology to evaluate market performance. The procedure for measuring such performance is far from obvious, however, a general overall measure for evaluating the level of security of supply should be defined to determine the deviation between current market outcome and the regulator's global objectives.

For the present intents and purposes, this overall measure will be termed the security of supply performance metric ($SOSPM$). Defining this $SOSPM$ involves a number of problems, ranging from the very short to the very long term. For this reason, any such overall metric is better expressed in terms of the dimensions described in the introduction, making it easier, in practice, to evaluate system performance at the security ($SECPM$), firmness ($FIRPM$), adequacy ($ADEPM$), and strategic expansion policy ($SEPPM$) levels.

The regulator should (and often does) design methodologies and metrics to properly evaluate system performance in connection with the aforementioned dimensions. At the security level, for instance, the System Operator checks the amount of available operational reserves to ensure that system voltage and frequency remain within acceptable margins. In some electric power systems the firmness dimension metric is the level of (certain) reservoirs. Adequacy is measured in terms of a long-term capacity margin (a function of installed capacity of whatsoever nature and expected demand). Finally, the CO2 emissions attributable to the generation system in a year's time is one of the metrics considered when dealing with strategic expansion policy.

2.2. The regulator's performance metric: classic reliability-based standards versus optimisation of net social benefit

In a liberalised context, security of supply is undoubtedly one of the regulator's major concerns. The meaning and implications of the "security of supply" concept are not completely clear, however.

Traditionally, this concept has been understood to mean the need to ensure system reliability (defined in terms of a certain minimum level of continuity and quality), while optimisation of the cost of providing the service has been regarded to be a secondary issue[36]. Broadly and practically speaking, however, when regulators seek to ensure security of generation supply they are actually looking for the maximisation of the net social benefit (NSB) at the power system level. Indeed, as a rule, the four dimensions mentioned above are the sequential and inter-related levels into which the main problem of optimising the net social benefit afforded by power generation can be broken down.

From this perspective, the general method for determining electricity market performance in any dimension would be to compare its outcome to the results found for the ideal or theoretically optimal system, i.e., the one with the most efficient design, management and operation from the standpoint of net social benefit. Note that comparing the two systems (the real situation and the ideal benchmark) would entail analyzing all the parameters that affect the overall net social benefit deriving from the system, such as costs, prices, and total consumption and associated utility. Obviously, the smaller the difference between the two, the smaller is the need to introduce an additional mechanism to reinforce market signals.

In practice, the metrics used to date have always been subject to intense debate:

The $SECPM$ is usually established in terms of the system's ability to provide so-called operational reserves. The definition of the technical requirements for these reserves is often heatedly discussed, for it involves establishing a minimum ramp speed for the generating units willing to provide the service, among other factors. Such requirements often prevent some of the units installed from taking part in the service market.

For the strategic expansion policy dimension, in turn, several metrics may be used, such as CO2 or NOx emissions or energy production from renewable sources. These metrics also

[36] Eurelectric (2006), cited in Pérez-Arriaga (2007), defines security of electricity supply as "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". The inclusion of sustainability in this definition is an indication of the breadth of features considered.

subject to some degree of controversy, in particular respecting which technologies should and which should not be granted renewable status.

A metric to evaluate performance in terms of firmness and adequacy

Discussion on how to define the regulator's metrics (whether or not to use reliability criteria only) has been most intense in these two intermediate dimensions. In the adequacy and the firmness dimensions, electricity system performance has usually been based primarily or even solely on the reliability of the generating mix. Particularly in the earliest stages of reform, then, regulators tended to show much more concern about supply continuity than about the actual cost of providing such a service. Regulatory mechanisms focusing on these dimensions have therefore been often referred to as reliability mechanisms.

The main reason for prioritizing the continuity of supply with a certain probability over the actual cost of meeting these quality standards is that the value of loss of load (VOLL), which is assumed to represent marginal demand utility, has traditionally been regarded by regulators to be much higher (at least one order of magnitude larger) than any production cost.

The reliability criteria traditionally used include the capacity reserve margin (system capacity, often weighted by an estimate of the average availability of the generating units, minus peak demand), the energy reserve margin (system energy reservoir minus the maximum expected load), the loss of load probability (LOLP), the loss of load expectancy (LOLE) and the non-served energy expectation (NSEE), see Wood and Wollemberg (1996). The reliability criterion used in PJM, for instance, is a loss of load expectation (LOLE) of not over one day in ten years. The resource requirement to meet this reliability criterion, the installed reserve margin (IRM), is expressed as a percentage of estimated peak load. In Western Australia, to cite another example, the objective is to limit energy shortfalls to 0.002 % of annual energy system consumption.

Assuming that marginal demand utility is much higher than units' marginal production costs is not always accurate, however. This fact is particularly relevant in certain real cases where either generation side or demand side behaviour deviates from this traditional assumption.

On the generation side, some markets, particularly in Latin America (Batlle et al., 2010), are characterised by the existence of highly inefficient units installed with extremely high variable costs. These costs sometimes reach values explicitly perceived by the regulator to be close (or even over) the value of loss of load[37]. In such circumstances, designing a

[37] This situation has led some regulators to limit the marginal spot price when such extremely high variable cost units are required. When a unit whose cost exceeds the threshold defined by the regulator is scheduled, it is paid its bid price, but the system marginal price may not exceed the threshold. This provision was in place in the Salvadorian electricity market prior the reform launched in 2008.

measure and thus a mechanism based solely on reliability criteria may in the end attract these inefficient units. This is why some systems have introduced so-called efficiency criteria in conjunction with a security of supply mechanism. Examples of this trend can be found in Brazil, Guatemala, Bolivia and Ireland (as presented in detail in Chapter IV, all take unit production costs or bidding price into account in the additional remuneration).

As demand side participation in the market rises, an elastic curve expressing consumption preferences progressively changes the VOLL paradigm. Here also, marginal demand utility (or at least a part of it) may fall within the range of values resulting from generation costs. Indeed, if the entire demand side participated in the market, expressing its actual preferences, reliability as traditionally conceived (based on ensuring the supply for a certain inelastic load) would make no sense. In other words, if the demand side becomes fully elastic, the above-mentioned analysis of the optimal benchmark would be the only sensible grounds for evaluating system performance. This fact is illustrated in Chapter VI, making use of a traditional Probabilistic Production Cost model, in which the classic approach has been extended to introduce the possibility of considering an elastic demand.

3. OPTIMAL PRICING PRINCIPLES FOR SECURITY OF SUPPLY MECHANISMS

This section discusses the optimal incentives that should be provided by a security of supply mechanism. Using the traditional formulation of the optimal pricing problem in electricity markets, we show that this remuneration should ideally be based on each unit's contribution to the objectives set by the regulator (expressed here in terms of $SOSPM$).

That finding reinforces the importance of properly defining these objectives, which are usually expressed on a dimension-by-dimension basis. As noted above, any manner of efficiency criteria and not only reliability standards may serve as grounds for such definitions. Reliable units with extremely high costs, for instance, may not contribute to attaining a given objective, and would therefore qualify for only a small incentive.

3.1. Optimal incentives in security of supply mechanisms

Schweppe et al. (1988) introduced conceptual grounds for determining optimal wholesale prices in electricity markets. Pérez-Arriaga and Meseguer (1997) extended that discussion by analyzing both short-term energy price incentives and the incentives associated with supplementary short-term and long-term signals: namely, the prices for operational reserves and extra payments for installed capacity (MW), respectively.

The formulation of the problem presented in the latter paper[38] has been simplified for the present study, while the conceptual framework for the long-term security of supply

[38] The analysis that follows is based on the same simplifying assumptions as the discussion proving that short-term marginal prices provide optimal incentives for both efficient operation and investments that maximise overall system efficiency. These include: generator cost function convexity, risk neutrality, absence of

mechanism has been modified slightly. Pérez-Arriaga and Meseguer (1997) assumed an explicit constraint in the form of a reliability margin, in the understanding that the regulator might require a larger long-term margin than strictly required, economically speaking. This long-term margin is defined to be "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, these authors proposed an explicit remuneration for available installed capacity.

In the following analysis, the market context and the embedded security of supply objectives are redefined. An electricity market is represented in which the regulator has implemented some manner of additional mechanism to guarantee a certain minimum level of performance, $SOSPM_{min}$ (minimum security of supply performance metric). To reach these objectives, the regulator is assumed to act on behalf of demand (as discussed in Chapter I, the side of the market with no real active role). In the optimisation model, then, demand and regulator are represented as a single agent (the generators' counterparty).

This way, the embedded system performance objectives are expressed from a broader point of view, using a security of supply performance metric ($SOSPM$) with more than just reliability-based objectives. In the Annex, this metric is broken down into the objectives pursued in each dimension: security, firmness, adequacy and strategic expansion policy.

Since the problem is formulated in traditional terms, the security of supply performance metric can be can be generally and succinctly represented as a function of the installed capacity for each type of unit, i, that is, $SOSPM = SOSPM(q^i_{max})$. Moreover, an additional incentive consisting of an individualised (each unit receives a different price) payment to generators proportional to the installed capacity (MWs) is assumed.

Expressing system objectives solely in terms of installed or available capacity is the traditional approach for modelling system long-term SoS requirements, see for instance the models developed by De Vries (2004), Ehrenmann and Smeers (2008) and Hogan (2009). At this stage of the analysis, this simple modelling assumption is the basis for drawing a series of general conclusions about the principles on which to design the incentives. Note that this is a rather simplified hypothesis in which a certain portion of the of a unit's capacity defines not only all the parameters relevant to that unit (reservoir capacity, for instance), but also factors such as the medium-term resource management conducted by the generator or the operational reserves added to the system.

economies of scale and lumpy investments and existence of a perfectly competitive, perfectly informed marketplace. In other words, characteristics that play a key role in electricity markets, such as the inefficient allocation of risk that usually exists in the absence of regulatory intervention are excluded, see for instance Rodilla et al. (2010). Nonetheless, the findings provide clear insight into the optimal incentive design problem, which is the main objective at this stage of the analysis.

The conclusions derived from this simplified analysis are subsequently used as a basis for a refined formulation that better represents the actual interrelationships in this type of mechanisms.

3.1.1. The demand side (and regulator) model

The demand problem in this context consists of maximising the utility U_h^d (defined for each hour h) obtained from the total hourly consumption of electricity, Q_h, less the derived costs. These include the energy purchased on the wholesale market (equal to the hourly spot price for power, π_h, multiplied by total hourly consumption Q_h), and the extra payments to generators, i.e., the price (normally modelled as the remuneration per installed MW on a unit-by-unit basis, i) multiplied by the respective installed capacity (q_{max}^i).

Note that the amount of installed capacity is also a demand side (and regulator) decision variable. The number of MWs in which each type of group decides to invest may be influenced by demand through the additional incentive.

The demand problem can be represented as follows:

$$\begin{aligned}
&\underset{Q_h, q_{max}^i}{Max} \quad \sum_h [U_h^d(Q_h) - \pi_h \cdot Q_h] - \sum_i \tau^i \cdot q_{max}^i \\
&subject\ to: \\
&\quad SOSPM(q_{max}^i) \geq SOSPM_{min} \quad \perp \chi
\end{aligned} \tag{1}$$

The model proposed intentionally lacks a key constraint which is often “always active”. Above and beyond any other objective contained in the so-called $SOSPM$ constraint, distressingly (for some) and logically (for others), one predominant constraint is always present: the regulator’s (i.e., the Government’s) extreme reluctance to allow electricity tariffs (or more precisely, electricity service costs, since tariffs often fail to cover the full cost of the service, with some portion being subsidised by the national budget) to increase above a certain level. This issue has been disregarded in the present analysis. While it is very likely the main feature behind many of the regulatory flaws inherent in the electricity business[39], any attempt at addressing it here would make little sense, for it would obviously lead to the conclusion that the problem, stated in these terms, may not have a feasible solution.

[39] It is obviously not the only such feature. The reason for implementing this constraint often lies in the *original sin* of market structure. A poor market structure (excessive horizontal and vertical integration of electricity industry activities) that the regulator cannot (or is not committed to) correct underlies its mistrust of market operation.

3.1.2. The *SOSPM*-based incentive

The demand-side decision variables are hourly consumption and each generating unit's installed capacity. The optimality conditions for this problem are obtained by calculating the derivative of the duly formulated Lagrangian function with respect to the decision variables:

$$\frac{dU_h^d(Q_h)}{dQ_h} = \pi_h, \forall h \tag{2}$$

$$-\frac{\partial SOSPM}{\partial q_{max}^i} \cdot \chi = \tau^i, \ \forall i \tag{3}$$

This last condition (3) leads to several interesting conclusions. The security of supply incentive for each generating unit can be rewritten as follows:

$$\tau^i \cdot q_{max}^i = -\chi \cdot \frac{\partial SOSPM}{\partial q_{max}^i} \cdot q_{max}^i \tag{4}$$

The optimal additional incentive that should be paid to each generating unit depends on the value of the dual variable (χ) associated with the demand side (or regulator) $SOSPM_{min}$ constraint, multiplied by the product of installed capacity and the marginal contribution of each unit to the security of supply objective. The value of this dual variable obviously depends on the generator side of the market, disregarded at this stage. If the market, left to its own devices, meets this condition, this dual variable adopts a value of zero and no additional payment is necessary.

Note that, based on the traditional formulation of the problem, the price incentive, as illustrated in the equation below, is individualised for each generating unit.

$$(-\chi \cdot \partial SOSPM \,/\, \partial q_{max}^i) \cdot q_{max}^i \Rightarrow \text{ Price} = -\chi \cdot \partial SOSPM \,/\, \partial q_{max}^i; \ \text{ Quantity} = q_{max}^i \tag{5}$$

These individualised price incentives remunerate installed capacity (MW). A rearrangement of the terms yields a better interpretation of this incentive, however, closer to its actual practical implementation.

$$(-\chi) \cdot \partial SOSPM \,/\, \partial q_{max}^i \cdot q_{max}^i \Rightarrow \ Price = -\chi; \ \ Quantity = (\partial SOSPM \,/\, \partial q_{max}^i) \cdot q_{max}^i \tag{6}$$

While the pricing signal, which is the same for all generating units, is equal to ($-\chi$), the amount qualifying for the incentive must be normalised based on each unit's marginal contribution to the regulator's objectives.

In principle, the security of supply mechanism leads to a single purchase price for a security-of-supply-oriented product, and the amount delivered by each unit (if properly defined) represents the normalised quantity mentioned above. This clearly suggests that installed capacity is too simple to reflect the degree of fulfilment of the regulator's objectives. Rather, the basis should be the actual amount of product delivered. This is a highly significant conclusion since, depending on system characteristics (capacity and/or energy constrained), a generation unit's contribution to the SoS objectives might not be its nameplate capacity.

As stated, in Annex B we analyze in deeper detail the former expression, by decoupling the SoS metric into the metrics expressed in each of the four dimensions. This allows 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).

4. ANALYZING THE MECHANISMS FOCUSED ON ENHANCING THE SYSTEM FIRMNESS AND ADEQUACY: THE ROLE OF THE SOS ORIENTED PRODUCT

In the previous section we have shown how an SoS mechanism should be designed and remunerated to provide optimal incentives. 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. We have reformulated the model on the basis of the so-called SoS-oriented product
- The introduction of all generation decision parameters, including investment parameters (maximum capacity, reservoir capacity, ramp-up capability, etc.) and medium to short term decision parameters (e.g. reservoir planning), and the relationship between these parameters and the capability to provide the reliability product.

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

4.1. 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 on which the additional mechanism is focused. Product definition includes several elements, such as the underlying asset (energy, renewable energy, capacity or capability to provide operating reserves), financial characteristics (forward, option or other), the time periods involved in the contract and the guarantees (usually physical guarantees), to name a few.

It is relevant to clarify that by product it is often understood just the underlying "physical" asset being traded (i.e. energy production, or availability in the peaks, etc.). However, we deem more suitable to understand by product the whole resulting commitment. In other words, the SoS-oriented product is nothing but the regulatory compromise or the contract being signed by the two counterparts, i.e. the regulator and the generators, with all its inherent characteristics. This way, in the Colombian Reliability Charge scheme the reliability product, the so-called reliability option, among many other features, consists in a financial option of energy, with a required physical back-up (guarantee), presenting a lag period of 3 years and a contract duration of up to 20 years. All these characteristics form part of the product definition.

Operational reserves constitute an example of a product geared to the security dimension (reserves to restore frequency or black start reserves, for instance) that may be purchased either in the short or the long term. Renewable production (remunerated through feed-in tariffs or renewable obligations, for instance) is a typical product in the strategic expansion policy dimension, although there are many others, such as energy efficiency-related products (white certificates) or CO2 emission rights.

In the other two dimensions (firmness and adequacy), differentiated products designed to individually tackle potentially deficient system performance are difficult to find in practice. In fact, no distinction has traditionally been made between firmness and adequacy, inasmuch as a single mechanism and the same product(s) have implicitly been thought to suffice for both. Although this is partially true, some authors have proposed designs that introduce separate incentives to reach the objectives sought in each of these dimensions[40].

In the context of the mechanisms focusing on overcoming firmness and adequacy dimension deficiencies, the associated product is usually known as the "reliability product" (see Batlle and Pérez-Arriaga, 2008), for it is usually regarded to be oriented to increasing reliability. Capacity credits (PJM or Western Australia), forward contracts (Brazil, Peru or Chile[41]), strategic reserves (Nord Pool) and reliability options (Colombia, Brazil or New England) are a few examples of designs for this product.

The quantity of the product each unit is capable to provide 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[42]. This concept appears under different and very

[40] The SoS mechanism in place in Spain, for instance, includes two services: the availability service and the investment service (see Batlle et al., 2008b).

[41] In some systems, particularly in Latin America, two products are normally defined to tackle the adequacy and firmness problem. In Chile and Peru, in addition to forward energy purchases (through long-term auctions), payments are made for a capacity-based product (capacity payments).

[42] 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.

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 mechanisms focused on enhancing the system adequacy and firmness. This way, the problem is reformulated introducing the trading of the SoS-oriented product.

The analysis that follows shows that by introducing these additional incentives, regulators actively recover their former central planning role to a certain extent. This circumstance significantly conditions not only the amount of installed capacity, but also the nature (technologies) and characteristics (such as capacity or reservoirs) of the units entering the system (adequacy), as well as the way system generation resources are managed (firmness) and scheduled. It goes without saying that the optimal incentive for each unit is directly related to the regulator’s objectives.

4.2. A closer representation of the adequacy and firmness-oriented mechanisms: the impact of the SoS-oriented product on the overall market outcome.

The foregoing conceptual findings identify the need for designing a remunerated SoS mechanism. The present section describes the details introduced in the modelling hypotheses to better represent actual SoS mechanisms implemented worldwide and their possible effect on generation-related decisions.

Two major refinements were included in the conceptual model. First, to explicitly represent the product purchased under the mechanism, the model was reformulated for what is referred to here as the SoS-oriented product. As noted earlier, expressing system requirements in terms of installed capacity only oversimplifies modelling. And secondly, the stylised model was reformulated to schematically represent all generator side parameters and decisions.

The new formulation, described below, allows for a more realistic representation of how additional regulatory mechanisms work in practice.

4.2.1. Demand side problem reformulation

This problem is addressed in terms similar to the ones described above. The extra payments for generators are equal to the unit price of the SoS-oriented product, τ_{sosp}, times the total amount of product purchased, Q_{sosp} [43]. Accordingly, the regulator’s performance objectives

[43] Several products may be bought and sold under the mechanism, although the present discussion focuses on one only.

are expressed as a function of the SoS product. At the same time, to better represent generator side decision-making, a medium-term index (m) is introduced to supplement the short-term index included in the basic model.

Consequently, where the security of supply mechanism consists of the acquisition of an SoS-oriented product, the demand side (and regulator) model can be expressed as:

$$\begin{aligned}
&\underset{Q_{m,h},Q_{sosp}}{Max} \quad \sum_m \sum_h [U^d_{m,h}(Q_{m,h}) - \pi_{m,h} \cdot Q_{m,h}] - \tau_{sosp} \cdot Q_{sosp} \\
&subject\ to: \\
&\quad SOSPM(Q_{sosp}) \geq SOSPM_{min} \quad \perp \chi
\end{aligned} \tag{7}$$

Under this new formulation, the number of installed MWs is clearly not the regulator's sole concern. Note that this variable does not explicitly appear in its own decision problem, because it is not the only factor determining the amount of SoS-oriented product that the generating units can supply.

The next demand side parameters to be determined are hourly consumption and the amount of the SoS-oriented product to be purchased. The optimality conditions derived from this reformulation are:

$$\frac{dU^d_h(Q_{m,h})}{dQ_{m,h}} = \pi_h, \forall m,h \tag{8}$$

$$-\frac{\partial SOSPM}{\partial Q_{sosp}} \cdot \chi = \tau_{sosp} \tag{9}$$

Hence, the price signal emitted by the mechanism, τ_{sosp}, should be equal to the product of the dual variable associated with the constraint imposed on system performance times the marginal contributions to those objectives made by the product purchased.

Note that this defines a standard price, since the contribution of one additional unit of a given SoS product (if properly defined) is the same regardless of the type of unit involved.

Note also that instead of each generator being paid a variable amount for each MW of available capacity, a standard price is defined for each unit of SoS-oriented product delivered.

The need to properly define the SoS-oriented product

In the real world, however, properly defining the SoS-oriented product is complex. One obvious condition is that any marginal increase in the product must entail an improvement in system performance ($\partial SOSPM / \partial Q_{sosp} > 0$). At the same time, the regulator should ideally define the SoS-oriented product in terms of overall efficiency, i.e., determine the lowest cost option of the many that meet the necessary condition. For instance, while installed capacity, as an SoS-oriented product, can solve the firmness and adequacy dimension problems (lowering the likelihood of scarcity), it does not guarantee that new entrants will be the most efficient units, see for instance the Peruvian case example (Chapter IV).

Although perfection is not possible, optimising the definition of performance metrics and the product is essential to achieving the regulator's objectives.

Generator ability to deliver the SoS-oriented product

As repeatedly stated, the quantity of the product that can be delivered by each unit depends on its design parameters, the random factors affecting its output, medium-term resource management and short-term operation (hourly) decisions.

The investment decision vector for a given unit is represented as Id^i. This variable represents all decisions regarding technical specifications, such as maximum generation capacity, q^i_{max}, or (hydro or gas) storage capacity, rc^i. Strictly speaking, these parameters are defined when the investment is made (installing a turbine able to burn different fuels, for example).

Medium-term management decisions are represented as Md^i_m, which involve, among others, the provision of fuel (e.g., concluding forward contracts). Some of these medium-term decisions have an associated cost, $MC^i(Md^i_m)$, such as extra risk premiums to hedge firm fuel supply contracts.

The random variables affecting each unit's production capability are also taken into consideration. These random factors, represented as ε^i, include hydro inflows, HI^i, and the hourly failure factor, $\lambda^i_{m,h}$ for instance.

The product is therefore expressed as a function of investment and planning decisions as well as production scheduling, $q^i_{sosp}(Id^i, Md^i_m, q^i_{m,h})$. For example, the ability of a hydro plant (the unit) to deliver a certain amount of energy in the dry period (the product) depends on maximum output and reservoir capacity (physical parameters), resource management up to the time the decision is adopted (past production decisions) and hydro inflows (random factors).

4.2.2. Generator side problem

In light of the above details, the generation side problem is modelled to introduce the additional payment provided by the mechanism as explicit remuneration for delivery of the SoS-oriented product.

Moreover, the problem facing generators is to maximise their own profit function, i.e. income less costs. As noted earlier, the two sources of income in the context described are the sale of energy and additional payments (sale of the SoS-oriented product).

Each unit's hourly production cost, $C^i(q^i_{m,h})$, depends on the energy produced (hourly energy produced, $q^i_{m,h}$).

The operating constraints on unit production are schematically represented as $R^i_c(Id^i, Md^i_m, q^i_{m,h}, \varepsilon^i)$. This expression includes the effect of investment and medium-term planning decisions as well as the random factors. The maximum capacity constraint is embedded in this schematic formulation (the inequality $q_{ih} \le q^i_{max}$ is a specific case, where q^i_{max} is regarded to be an investment design parameter).

The generation problem can consequently be rewritten as follows:

$$\begin{array}{ll} \underset{Id_i, Md^i_m, q^i_{m,h}}{Max} & \sum_m \sum_h [\pi_{m,h} \cdot \sum_i q^i_{m,h} - \sum_i C^i(q^i_{m,h})] - \\ & -\sum_m \sum_i MC^i(Md^i_m) - \sum_i IC^i(Id^i) \\ & +\tau_{rp} \cdot \sum_i q^i_{rp}(Id^i, Md^i_m, q^i_{m,h}) \\ subject\ to: & \\ \quad R^i_c(q^i_{m,h}, Id^i, Md^i_m, \varepsilon^i) \le 0 \quad \perp \zeta^i_c & \end{array} \tag{10}$$

Generators maximise their own profit. As noted previously, however, the income deriving from the additional payments provided by the mechanism is expressed as direct remuneration for the SoS-oriented product delivered.

In this formulation, then, the decision variables are hourly energy produced $q^i_{m,h}$, generating unit design parameters, Id^i, medium term resource management decisions, Md^i_m, and the quantity of the SoS-oriented product to be delivered by each unit under the regulatory mechanism, q^i_{sosp}.

The optimality conditions obtained are shown below:

$$
\begin{aligned}
&\pi_{m,h} - \frac{dC_i(q^i_{m,h})}{dq^i_{m,h}} + \tau_{sosp} \cdot \frac{\partial q^i_{sosp}}{\partial q^i_{m,h}} + \sum_c \left[\frac{\partial R^i_c}{\partial q^i_{m,h}} \cdot \zeta_{c,i} \right] = 0, \ \forall i, m, h \\
&- \frac{dMC(Md^i_m)}{dMd^i_m} + \tau_{sosp} \cdot \frac{\partial q^i_{sosp}}{\partial Md^i_m} + \sum_c \left[\frac{\partial R^i_c}{\partial Md^i_m} \cdot \zeta_{c,i} \right] = 0 \ , \ \forall i \\
&- \frac{dIC(Id^i)}{dId^i} + \tau_{sosp} \cdot \frac{\partial q^i_{sosp}}{\partial Id^i} + \sum_c \left[\frac{\partial R^i_c}{\partial Id^i} \cdot \zeta_{c,i} \right] = 0 \ , \ \forall i
\end{aligned}
\tag{11}
$$

Note that under the first optimality condition, the production schedule is optimised on the grounds of short-term profit, $\pi_{m,h} - dC_i(q^i_{m,h}) / dq^i_{m,h}$, the effect of the technical constraints on marginal profit, $\sum_c \left[\zeta_{c,i} \cdot \partial R^i_c / \partial q^i_{m,h} \right]$, and the possible impact of output on firm supply and therefore the marginal income earned from the mechanism, $\tau_{sosp} \cdot \partial q^i_{sosp} / \partial q^i_{m,h}$. In other words, the production schedule may be affected by the regulatory mechanism.

The second and third optimality conditions are wholly analogous to the first. They show that under these modelling assumptions the driving forces behind planning and investment decision-making are the marginal costs of the respective long- or medium-term decision, the marginal income that may be earned if the limits of the constraints (R^i_c) are expanded as a consequence of new investments or planning and the marginal additional income that may be earned under the additional regulatory mechanism.

In short, generators take the potential incentive provided by the additional mechanism into consideration when adopting investment and management decisions. This is tantamount to saying that by implementing SoS mechanisms, regulators indirectly condition market outcome, at least partially recovering their former role as central planners.

4.3. The SoS-oriented product as a partial planning tool for the regulator

Security of supply mechanism design, and particularly SoS-oriented product design, constitute a key element in the regulator's role in generation system scheduling, planning and expansion. The SoS product may be (an indeed often is) explicitly or implicitly designed to favour certain results. An example of an explicit approach would be, in the firmness dimension, specific plant management strategies (contracting hydro capacity reserves for dispatch at the regulator's discretion) or in the adequacy dimension, the installation of a given technology (in Peru, bids submitted by new hydro plants for long-term auctions are multiplied by 0.85 in the clearing process).

Unfortunately, however, regulatory experience shows that more often than not, security of supply mechanisms yield results that were neither expected nor desired by the regulator. When designing an SoS product, regulators must carefully analyse generators' foreseeable response to ascertain whether that response will yield an efficient result or otherwise. The consequences of failing to evaluate product definition beforehand are highly inefficient situations, affecting either unit operation and planning or system expansion.

5. CONCLUSIONS

Under the market-oriented paradigm, regulation must ensure that the incentives in place guarantee efficient sources of short-, medium- and long-term security of supply. The key question is how to introduce the necessary adjustments in market design to achieve the objective pursued. This is a particularly sensitive issue because in the end, all short, medium and long-term planning may directly or indirectly revert to a central planner. And surmounting the potential inefficiencies of central planning schemes was one of the chief reasons for liberalising electric power generation in the first place.

The present chapter has reviewed the fundamental criteria underlying these regulatory approaches and the respective price incentives that should be designed, calculated and managed to emit optimal signals. More specifically, it describes the alternative methods for defining the metric with which to evaluate generation system performance. A stylised conceptual mathematical model have been used to provide a deeper understanding of SoS-oriented products.

The conclusions drawn are, firstly, that the optimal additional remuneration provided by any security of supply mechanism must be based on each unit's contribution to the regulator's objectives. Secondly, assessing this contribution is an extremely complex issue. And thirdly, when implementing this kind of mechanism, the regulator conditions market operation, planning and investment.

For all these reasons more attention should be paid to the precise definition of system performance objectives and the evaluation of each generating unit's contribution to the fulfilment of these objectives. The poor results obtained with many schemes the world over may usually be attributed to the lack of such precision.

Acknowledgements

The authors wish to thank Professors Ignacio J. Pérez-Arriaga and Michel Rivier, as well as Luiz Barroso for their valuable comments on an earlier version of this chapter.

6. REFERENCES

AIM, All Island Project (2007). The Value of Lost Load, the Market Price Cap and the Market Price Floor. A Response and Decision Paper. AIP-SEM-07-484. 18 September 2007.

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.

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.

Concept Economics (2008). Investigation of the value of unserved energy. Report for the Electricity Commission of New Zealand. December 2008.

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

Cramton, P. & J. Lien (2000). "Value of Lost Load". University of Maryland Working Paper.

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.

Ofgem (2009), "Project Discovery - Energy Market Scenarios", Reference number 122/09, Office of the Gas and Electricity Markets (Ofgem), October, 2009.

Ofgem (2010), "Project Discovery Options for delivering secure and sustainable energy supplies. Consultation document", Reference Number 16/10, Office of the Gas and Electricity Markets (Ofgem), February, 2010.

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.

Roques, F., D. M. Newbery & W. j. Nuttall (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.

Stoft, S. (2003). "The demand for operating reserves: key to price spikes and investment", IEEE Transactions on Power Systems, Volume: 18, Issue: 2, pp. 470- 477, May 2003.

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.

Annex B. OPTIMAL INCENTIVES EXPLICITLY REPRESENTING THE REGULATORY OBJECTIVE AT EACH DIMENSION

The optimal incentives problem can be analysed by breaking the SoS metric down into the metrics corresponding to each of the four dimensions. In other words, the general incentive can be unbundled into dimension-based incentives without losing sight of the importance of the interrelations among these dimensions (an incentive geared to one dimension may significantly impact the other three).

From the perspective of the four dimensions of SoS, the demand side problem can be represented as follows[44]:

$$\begin{array}{l} \underset{Q_h, q^i_{max}}{Max} \sum_m \sum_h [U^d_{m,h}(Q_h) - \pi_{m,h} \cdot Q_{m,h}] - \sum_i \tau_i \cdot q^i_{max} \\ subject\ to: \\ SOS(q^i_{max}) \left\{ \begin{array}{l} SECPM(q^i_{max}) \geq SECPM_{min} \quad \perp \chi_{sec} \\ FIRPM(q^i_{max}) \geq FIRPM_{min} \quad \perp \chi_{fir} \\ ADEPM(q^i_{max}) \geq ADEPM_{min} \quad \perp \chi_{ade} \\ SEPPM(q^i_{max}) \geq SEPPM_{min} \quad \perp \chi_{sep} \end{array} \right. \end{array} \tag{12}$$

Reformulating the demand side SoS constraint on the grounds of the performance metric constraints in each of the four dimensions yields the following optimality conditions:

$$\frac{dU^d_h(Q_h)}{dQ_h} = \pi_h, \forall h \tag{13}$$

$$-\frac{SECPM}{\partial q^i_{max}} \cdot \chi_{sec} - \frac{FIRPM}{\partial q^i_{max}} \cdot \chi_{fir} - \frac{ADEPM}{\partial q^i_{max}} \cdot \chi_{ade} - \frac{SEPPM}{\partial q^i_{max}} \cdot \chi_{sep} = \tau^i, \ \forall i \tag{14}$$

The additional price incentive for each generator can therefore be broken down into several terms. Each term can be interpreted to be the specific incentive corresponding to one of the regulator's constraints.

Stated more generally, the above conclusions mean that each objective can be associated with a different price signal, which adopts the value of the dual variables assigned to the corresponding constraints. While the price signal for a given objective is the same for all generating units, the value of this incentive must be normalised on the basis of each one's

[44] Since system performance in any dimension is still expressed as a function of installed capacity, the simplifying assumption to the effect that a unit's installed capacity (MWs) can be used as a measure of its contribution at all levels is consistently valid.

marginal contribution to fulfilment of the objective in question. Moreover, the contributions of a given generating unit to different dimensions may be interrelated. A coal-fired plant may contribute more to adequacy and firmness and less to strategic policy or to system decarbonisation, whereas a renewable generator may contribute to this fourth dimension but less to security due to its non-dispatchability.

$$\begin{aligned}\tau^i \cdot q^i_{max} &= (-\frac{SECPM}{\partial q^i_{max}} \cdot \chi_{sec} - \frac{FIRPM}{\partial q^i_{max}} \cdot \chi_{fir} - \frac{ADEPM}{\partial q^i_{max}} \cdot \chi_{ade} - \frac{SEPPM}{\partial q^i_{max}} \cdot \chi_{sep}) \cdot q^i_{max} = \\ &= (-\chi_{sec})\frac{SECPM}{\partial q^i_{max}} q^i_{max} + (-\chi_{fir})\frac{FIRPM}{\partial q^i_{max}} q^i_{max} + (-\chi_{ade})\frac{ADEPM}{\partial q^i_{max}} q^i_{max} + (-\chi_{sep})\frac{SEPPM}{\partial q^i_{max}} q^i_{max}\end{aligned} \quad (15)$$

In most cases, regulators design different mechanisms to tackle the various dimensions of the problem. The incentives introduced to correct one dimension are sometimes designed with no regard for their potential impact on the objectives sought in other dimensions. Since the four dimensions are interrelated, objectives must necessarily be coordinated. For instance, if strategic expansion policy calls for more renewable energy, the impact of the respective incentives on system security, firmness and adequacy must be borne in mind. If these objectives are not carefully coordinated, the outcome will be undesirable and inefficient.

Furthermore, if the regulator decides to change the objective in one dimension, by adopting a more ambitious renewable objective ($SEPPM^*$), for instance, this should not only change the payment associated with the new objective, χ^*_{sep}, but also affect the payments corresponding to the other dimensions ($\chi^*_{sec}, \chi^*_{fir}, \chi^*_{ade}$) and more generally, the market outcome of all the units installed in the system. This is because the energy produced hourly by all the generating units, q^i_h, and spot energy prices, π_h, would also change in the wake of the regulator's action, reducing actual income to levels lower than the income expected when, for instance, the decision to install a new unit was made[45].

Consequently, while overall SoS can be broken down into different dimensions to facilitate regulatory discussion and design, the interrelationships among these dimensions as parts of a common objective function must not be overlooked.

[45] This issue is being heatedly debated in the Spanish electricity market at this writing. The regulator has raised the RES target share of primary energy supply on a number of occasions. In a scenario in which the economic crisis that has led to an abrupt decline in demand, generators having a significant proportion of conventional steam units in their portfolios claim that regulator intervention in this dimension, $SEPPM^*$, clearly jeopardises their ability to recover their investments, and consequently demand compensation in the form of higher adequacy payments (χ^*_{ade}), for instance.

Chapter IV

Chapter IV. A Critical Assessment of the Different Approaches Aimed to Secure Electricity Generation Supply

1. INTRODUCTION

In Chapter I and Chapter II we have seen how, for many different reasons, the market if left to its own devices does not seem to be capable of ensuring an efficient supply (especially in the long term). Then, in Chapter III, we carried out a theoretical analysis from the regulator's perspective to highlight the importance of properly defining security of supply (SoS) objectives and providing optimal incentives (if deemed necessary). This analysis was first focused on the whole security of supply problem and latter particularized to the firmness and adequacy dimensions. The objective of this chapter is to continue with this latter analysis, by providing a categorization and a critical review of the different approaches that have been used in practice in the different contexts to deal with the problem of ensuring the desired results at the firmness and adequacy dimensions, since as pointed out in the introduction, the debate on the necessity for regulatory intervention is more acute at these dimensions. This analysis, coupled with the previous study, will help us to set the conceptual background that is later used in Chapter V, where the different design features the regulator should consider when introducing a long-term security-of-supply-oriented mechanism are presented.

As we introduced previously (see PART A, section 2), the security of supply concept can be decoupled into four major components (or dimensions), namely: security (a very short-term issue), firmness (a short to medium-term issue), adequacy (a long-term issue) and strategy expansion policy (a very long-term issue). It was also highlighted that there is a certain consensus around the idea that, on one extreme, the security dimension can be tackled by means of an operation reserves market, where the requirements are prescribed by the System Operator, and that, on the other, the Strategic Expansion Policy has to be solved through the implementation of additional "out-of-the-electricity-market" mechanisms (e.g. feed-in tariffs or cap and trade mechanisms). But in between these two dimensions, the debate on the necessity of regulatory intervention at the firmness and adequacy (particularly this latter) has always been, and still is, quite intense.

2. SECURITY OF SUPPLY MECHANISMS TO SOLVE FIRMNESS AND ADEQUACY REQUIREMENTS

The way to design regulation to ensure security of power generation supply has evolved in the last decades, and although it is not possible to establish a universal solution to deal with the problem, international experience has helped to narrow the range of possible measures a regulator should consider. Roughly speaking, the regulator can adopt two opposed strategies:

- Do nothing; in the belief that the market will provide the efficient medium- to long-term outcome. The regulator's lack of intervention would be supported by the expectation that demand will (or will learn in the end to) manage the long-term risk involved in electricity markets (for example, by hedging and guaranteeing their future needs). This is often known as the "energy-only market" approach.

- Do something on behalf of the demand; in the opposite belief. In this case, the regulator designs a security of supply mechanism which entails the definition of a certain reliability-oriented product (the "reliability product" or the "SoS-oriented product", as we will call it hereafter) aimed to ensure system security of supply (i.e. mainly to avoid scarcities). This reliability product is provided by the generators, who receive in exchange the extra income or the hedging instruments they require to both proceed with efficient investments (adequacy) and make resources available when most needed (firmness). The other counterparty is either directly the demand, compelled to purchase the product by the regulator, or the regulator itself (i.e. the system, the tariff) acting on behalf of the demand.

  Thus, several key elements of the mechanism have to be carefully designed (this is analyzed in deeper detail later in Chapter V):

- The counterparties, the regulator has to decide whether to act on behalf of all the demand or just a proportion of it.

- The way to set the reliability product price, i.e. whether the regulator administratively sets the price or just the quantity and allows for a market-based mechanism to reveal the price.

- The reliability product characteristics: for instance, time terms, how to identify near rationing conditions or how to assess each unit's contribution to overall system reliability. There are also other relevant characteristics, such as product optionalities (forward or option contracts), penalties, financial guarantees, force majeure clauses, etc. This reliability product can take many different forms, as for example a capacity credit, a long-term energy contract, etc.

Price mechanisms and quantity mechanisms

The different mechanisms can be classified based on whether the regulator's main objective has been to ensure a certain quantity of the "reliability product" or to administratively set a price for the product itself[46].

- Price mechanisms: an administratively determined payment, often known as the "capacity payment", additional to the income derived from the energy (spot) market, is provided in exchange of the reliability product. In this scheme, the reliability product is in practice the so-called "firm capacity".

---

[46] Generally speaking, all mechanisms require the definition of a price-quantity offer curve for the reliability product purchasing process. For the sake of simplicity, we will classify here the different experiences around the two extreme approaches (quantity-based and price-based).

- Quantity mechanisms: the regulator imposes on (or buys itself on behalf of) the demand the purchase of a specific quantity of the reliability product. In this context, this product takes a variety of formats, e.g. an energy long-term forward, a short or long-term capacity credit, etc. Depending on the system, the product may be traded bilaterally, within an auction (centralized or not) or by means of additional and organized short-term markets.

A good number of analyses from different points of view can be found in the literature, see for instance (Wolak, 2004), (Roques et al, 2005), (Cramton & Stoft, 2005) (Joskow, 2007) or (Finon & Pignon, 2008). Here we take the taxonomy we just outlined as a guide to review and categorize the whole scope of approaches. We exhaustively detail and critically evaluate the broad range of international experiences throughout the years up to the present moment, emphasizing the lessons we have learned so far in each particular context.

The role of demand

It can be considered the possibility of allowing consumers to participate in security of supply oriented mechanism (in some of them is more straightforward than in others) by offering a kind-of-symmetric product to the one required to generating units. Although this paper is focused on the generation side, it is important to note that there is a growing trend on integrating demand in the security of electricity supply mechanisms. Indeed many of the problems regarding security of supply will be more suitably tackled thanks to the efficient demand participation.

First we consider necessary to call into question the so-called "energy-only market" approach, which in principle consists in relinquishing any way of directly or indirectly intervening. In other words, this "energy-only market" approach consists in leaving the market exclusively to its own devices. We show that in practice it is very difficult to find electricity markets in which the regulator does not resort to any explicit (or implicit) safety measure to ensure system reliability.

Then we discuss the complementary approaches which entail the implementation of an explicit regulatory mechanism, beginning with the price-based mechanisms and ending with the quantity-based ones. We examine the key elements and put forward the criteria the regulator should take into account when designing this sort of regulatory mechanisms. The different experiences are also presented in a rather chronological way, so as to allow the reader to understand why the first approaches failed, and therefore the reason behind the design features of the new ones.

3. DO NOTHING: THE SO-CALLED ENERGY-ONLY MARKETS

The first alternative is doing nothing. By doing nothing we mean a regulator's long-term commitment to refrain from intervening in securing the supply. As previously mentioned, the regulator's lack of intervention would be supported by the expectation that demand will (or will learn in the end to) manage the risk involved in electricity markets (for example, by signing long-term contracts). The regulator's position will have to remain unchanged even though things may not have turned out as initially expected.

As reviewed in Chapter I and also in the Annex A, theoretical microeconomic analysis of power systems shows that, under a number of strong ideal conditions, the short-term price resulting from a competitive market provides efficient outcomes both in the short and long run, see (Caramanis et al., 1982), (Bohn et al., 1984), (Caramanis, 1982), (Scheweppe et at. 1988), (Pérez-Arriaga, 1994) or (Vázquez, 2003). Inframarginal energy revenues (the so-called scarcity rents being particularly important[47]) provide the necessary income for the recovery of both operational and investment costs.

Using this argument (amongst others), some experts (less and less as time passes) suggest that only purely market-based approaches would provide an efficient outcome regarding long-term security of supply.

This approach, focused on not interfering with the market, has often been termed the "energy-only market" approach, see for instance (Hogan, 2005). However, among regulators and academics, it is not always clear what is and what is not considered to be market intervention, and as a consequence, what it is meant by the term "energy-only market" usually depends on the system, the context or the point of view of the author.

When using the term "energy-only market", some authors simply make reference to the absence of some kind of "capacity-based mechanism" (like the well-known capacity markets or capacity payments), while countenancing the possibility of many other types of regulator's actions/interventions regarding long-term security of supply. Some examples of these actions include, for instance:

- The long-term contracting of energy and/or reserves (not only operational but also "load reserves" to be used in scarcity situations and described later in more detail) by the regulator or the System Operator.
- Giving the System Operator full control of the operation in those cases in which a scarcity period is bound to happen.
    - In other cases, when operating reserves fall below a certain level, the SO takes actions, such as voltage reductions and non-price rationing of demand (rolling blackouts), to reduce demand administratively while avoiding prices to reflect the scarcity situation, see (Joskow, 2007). Another similar example is the Maximum Generation Service contracted by the SO in UK (NGET, 2010). These types of "out-of-the-market" measures complicate the price formation process in conditions of scarcity, and affect the proper and expected recovery of generation investments.

[47] That is, the income perceived when the generation resources are not sufficient to supply the demand, and so, the price is set by the demand above the variable cost of any of the generators.

- Allowing the regulator to call an auction to encourage new investments as a backstop mechanism to ensure security of supply, etc.

From our point of view, these actions have to be included among the mechanisms to ensure long-term security of supply, since it is obvious that any of them is a clear indication that the regulator does not fully rely on the market to naturally do all that is required, and any of them clearly distorts the purely market based signals. In order to avoid confusion, we rather prefer to use the term "energy-only market" to make reference to the strict "do-nothing" alternative.

Let us note that using the term "energy-only market" with the meaning of "leaving the market to its own devices" may be quite misleading in some situations. In order to clarify what is and what is not an energy-only market, let us consider two paradoxical situations: if it is the market the one that opts for introducing a capacity market, with no regulatory intervention, then the so-obtained scheme would be an energy-only market. Conversely, if the regulator compels the demand to acquire their energy via forward contracting, then, such an approach could not be considered as an energy-only market.

But, from this perspective, it is very difficult in practice to find a market in which the regulator is really able and committed to just "waiting and seeing", having renounced the possibility of resorting to any explicit (or implicit) form of intervention, especially when the system is already suffering (or it is expected to suffer) a period with a tight or even scarce reserve margin. This form of intervention is sometimes subtle, especially in those cases where the regulator allows a third party to intervene or just suggest that this third party should do so. Such a third party might be the SO or even an incumbent utility.

While, in several (particularly European) markets, no security-of-supply mechanism has been explicitly implemented, it may be safely asserted that no system lacks at least an implicit regulatory safeguard regarding security of supply. In some systems the incumbent (now in a market-like context but still under partial, and sufficient, public control) "shares the regulator's concern" about system reliability (France[48], Italy or Portugal are some examples)[49].

---

[48] Indeed, on April 2010, in the preamble of the (NOME, 2010), to justify the proposal of creating a capacity market, it is said: 'It is about ensuring that all suppliers assume all their industrial and energy responsibilities on behalf of their customers and do not rely on an implicit guarantee of delivery of the incumbent'.

[49] In this sense, one of the arguments presented by the Spanish authorities to stop the German E.ON's takeover bid for Endesa was the nation's need to guarantee its own security of supply (paradoxically the process ended with the successful takeover bid by the Italian company Enel). In some other systems it is the retailer who it is still publicly controlled in some way (by municipalities in many cases) and is therefore the agent that seeks to protect its customers from unexpected annoyance through long-term contracting.

Indeed, in the European case there are "latent" security of supply mechanisms thanks to Directive 2005/89/EC[50], which states that 'The guarantee of a high level of security of electricity supply is a key objective for the successful operation of the internal market and that Directive gives the Member States the possibility of imposing public service obligations on electricity undertakings, inter alia, in relation to security of supply', and also that 'Measures which may be used to ensure that appropriate levels of generation reserve capacity are maintained should be market-based and non-discriminatory and could include measures such as contractual guarantees and arrangements, capacity options or capacity obligations. These measures could also be supplemented by other nondiscriminatory instruments such as capacity payments'.

In some cases, another (not always confessable) reason why certain regulators do not implement an explicit security-of-supply mechanism is the existence of horizontal concentration. A concentrated market allows generators to ensure the recovery of a "reasonable" rate of return.

Thus, strictly speaking, it is not clear that purely energy-only (competitive) markets do exist. That said, it has to be acknowledged that certain systems are greater "market believers" with respect to the market capability to ensure long-term security of supply. Among the most representative systems that are usually included in the literature in this "energy-only" approach we find ERCOT (Texas), NEM (East Australia), Alberta, Ontario, UK and the Nord Pool.

However, in our view, strictly speaking it cannot be considered that any of them has fully relied on the "left-to-its-own-devices" ideal market mechanism approach. Indeed, all of them present some kind of implicit or explicit security of supply mechanism:

- In the case of ERCOT, the emergency program known as EECP (Emergency Electric Curtailment Plan) allows the system operator to use reserves and out-of-the-merit units through "out-of-the-market" protocols. The objective is to avoid load shedding, which is carried out as the last step (the fourth) of the emergency protocol. The resulting short-term prices during these emergency interventions have been criticized for not reflecting the opportunity cost of providing the service. The System Operator may also enter into Reliability Must Run contracts with uneconomical units for many different reasons.
- In the case of Ontario, the Ontario Power Authority can enter into long-term contracts in order to secure an adequate reserve margin.
- In the UK, under the BETTA, the TSO is responsible for the long-term purchasing of the operating reserves. It is well-known that operating reserves requirements affect both short-term prices and consequently long-term investment signals. Thus, artificially modifying

[50] Directive 2005/89/EC of the European Parliament and of the Council of 18 January 2006 concerning measures to safeguard security of electricity supply and infrastructure investment. Official Journal of the European Union, 4.2.2006.

these requirements, above the actual needs to face exclusively very short-term security issues[51] can therefore alter medium- to long-term market outcomes. This has been the case in the UK operating reserve purchasing process. For instance, Roques et al. (2005) state that under the Supplemental Standing Reserve Tender (SSRT) called on October 2003 to increase the reserve capacity, there was evidence that the role of this supplementary tender (requiring a much larger quantity than it usually deems necessary to hold system frequency, so as to bring back some mothballed units) caused an immediate increase in forward market prices (i.e. in longer-term signals).

- In Nord Pool, the SO takes an active role resorting to a long-term "strategic reserves" contracting that it is later discussed in the quantity-based mechanisms.
- Moreover, ERCOT, NEM, Alberta, Ontario and the Nord Pool present a considerable degree of public ownership (either in the generation- and/or the retailer-side). In the NEM in particular, around 63 % of generation capacity is government-owned or controlled (AER, 2007).

4. PRICE MECHANISMS: CAPACITY PAYMENTS

Roughly speaking, capacity payments are a price-based incentive that seeks to achieve both an efficient resource management (firmness) and investment (adequacy and sometimes also strategy expansion policy).

The mechanism entails mainly two problems: first, to properly define the reliability product, second, to fix the price (right enough to avoid falling too short or too long).

In the price-based mechanisms context, the product is usually the firm capacity. Each unit firm capacity is aimed to represent the unit's contribution to the overall system's security of supply. In practice, depending on the system, we find many different alternative methodologies to define the firm capacity. In most of the cases it is mainly based on the (expected) availability of each generating unit when most needed, but sometimes other parameters are used in its calculation as for instance the units' variable costs (e.g. the smaller the variable costs, the larger the firm capacity assigned, as for example it is the case in Guatemala, Ireland or Brazil).

The firm capacity can be estimated ex-ante (according to historical data, with or without ex-post corrections) or directly settled ex-post (based on actual performance in terms of its contribution to the system reliability).

Next we delve into the analysis of the different capacity payment experiences.

[51] Note that a scarcity in generation supply is not a very short-term issue.

4.1. Chile

Administratively determined capacity payments were first introduced in Chile back in 1982. The first capacity payment design was aimed to provide an extra payment to ensure the full recovery of generators' investments and production costs. This payment was provided to each unit based on its firm capacity. The firm capacity was calculated using probabilistic models, and it represented each unit contribution to overall system reliability.

However, the lack of a market for ancillary services led to a subsequent redesign, where security criteria were also taken into account in the firm capacity calculation. Thus, the mechanism introduced in 1997 attempted to introduce incentives to those units capable of providing operating reserves in the following way: the firm capacity consisted of an adequacy component, which represented around 70 to 80 % of the firm capacity value, and a security component, which represented around 20 to 30 % of the firm capacity value.

After realizing that a market signal for security was necessary, a later redesign (2004) introduced markets for these services. The capacity payment was once again only tied to the adequacy concept (Huber et al., 2006).

4.2. UK (1990-2001)

The market model in the UK during the period 1990-2001 was based on a compulsory day-ahead pool, where system marginal prices were calculated on a half-an-hour basis. Additionally, capacity payments were paid to all generating plants declared available in each half hour, and the value was equal to the Loss of Load Probability (LOLP)[52] of the period considered, multiplied by the difference between the Value of Lost Load (VOLL, i.e. the regulator estimate of the cost of non-served energy) and the plants' bid price (if not dispatched) or the system marginal price (if dispatched)

This mechanism was criticized for many different reasons; for instance, Newbery (1995) pointed out that some companies artificially increased the LOLP, and thus capacity payments, by declaring unavailable certain units. Green (2004) highlighted that most of these abnormal payments were rather the result of a deficient definition of the method used to determine the new units' availability factors. In (Roques et al, 2005) a thorough analysis of the major shortcomings of the mechanism is carried out.

These capacity payments disappeared with the introduction of the decentralized NETA model in 2001 (BETTA, since 2005).

[52]The LOLP value represents the probability of rationing. Another relevant measure that can be calculated straightforwardly from the LOLP value is the expected amount of hours of rationing in a given period of time. Many systems define their reliability standards using this latter measure; for example, US power systems usually establish a maximum accumulated rationing period of one day in ten years.

4.3. Argentina

When the market started in Argentina back in 1995, two different capacity payments were implemented: one for dispatched capacity and another focused on remunerating those plants which did not produce on a regular basis but whose availability was essential for system reliability during dry years. A plant could not receive both payments during the same dispatch period.

The formula determining the first capacity payment (denoted as PPAD[53]) was similar to the one already presented in the former UK mechanism, that is, the higher the expected value of loss of load the higher the payment received by the generators. The difference laid in the fact that it was only provided to those generators producing in each hourly interval. The problem that rapidly appeared was that the size of this payment was large enough to severely affect the system dispatch. As remuneration was linked to production, the optimum strategy for generators was to internalize the payment in the bid sent to the market, what induced generators to bid below marginal costs in order to receive the extra payment.

Due to these inefficiencies, it was decided to introduce certain modifications in the design, guided to eliminate dependence on the actual dispatch and on the expected value of the non served energy. The new scheme remunerates not just the plants producing but also the ones available during hours of higher demand (90 hours every week) (CAMMESA, 2005), these hours being predetermined *ex ante* by the regulator. In order to evaluate each plant's availability, historical data regarding production and availability are taken into account together with unavailabilities due to network constraints.

4.4. Spain

The original "capacity guarantee mechanism" implemented in Spain when the market started in 1998 provided an extra remuneration based on average availability rate in the case of thermal plants (plus a minimum annually production requirement[54]) and based on the average historical production in the case of hydro units. This scheme was largely criticized, see for instance (Pérez-Arriaga et al, 2006), for being too simplified on the firmness side, lacking effective incentives for generators to be (or penalties for not being) available when needed and on the adequacy side for being extremely unstable: the total volume of

---

[53] PPAD (*Potencia Puesta A Disposición* in Spanish) makes reference to the capacity being committed in the market by the generation unit.

[54] Plants had to produce at least 480 equivalent hours every year to be entitled to receive the capacity payment. The measure was designed in an attempt to prove a minimum reliability of the plant. But the inefficiencies that resulted from such a rule were obvious, since it led high-cost peaking units to uneconomically forcing being committed to receive the payment.

payments for this item declined from an initial €7.8/MWh of system demand at market start-up in 1998 to €4.8/MWh in 2006[55].

After a two-years period of discussions, see (Batlle et al., 2007) and (Batlle et al., 2008), the Ministry of Industry came up with a redesign consisting of two differentiated services (MITyC, 2007):

- The availability service, aimed at allowing the system operator to enter into bilateral contracts, lasting no longer than one year, with peaking units (as, for instance, fuel-fired plants and the hydro limited energy plants).
- The investment service, for units larger than 50 MW and during their first 10 years of operation[56] an annually capacity payment (expressed in euros) per installed megawatt.

This investment incentive depends on the value of a so-called ''reserve margin index'' ("*índice de cobertura*" in Spanish, or IC) that has to be calculated by the system operator. While the value of this index is below 1.1, the payment is set at 28,000 Euros per installed megawatt and year. If the value of the index is above 1.1, this payment is reduced.

4.5. Italy

In Italy, an administratively determined fixed capacity payment is in place. This mechanism was initially conceived as transitory, but after almost ten years it is still in force. The payment is focused on providing additional remuneration to all the power plants sited in Italy whose production can be considered to be manageable (i.e. wind or run-of-the-river generating units are some of the technologies excluded). The quantity paid depends on the availability during the "high-critical" and "mid-critical" days, which are identified by the Transmission System Operator.

The payment consists of two different components (Benini, 2006):

- A capacity remuneration component which is calculated by the TSO on the basis of the units' estimated power capacity available (Terna, 2008).
- An additional amount that only applies in the event that the unit's revenues obtained from the energy sold on "critical" days are lower than those that would have been obtained on the basis of the administrated tariffs.

---

[55] The "hidden" reason behind these changes was that originally the purpose of the capacity payments was also to reward stranded costs to the existing generating units. As market prices began to rise above the expectations, the regulator (i.e. the Government) began to reduce the capacity payments.

[56] This 10-years condition is aimed to reward only CCGTs, which have entered the system after the market started in 1998. This design is clearly "contaminated" by the windfall profits discussion that puts into question the income of mainly nuclear and hydro plants (installed under the former regulated context) are receiving in the new market scheme.

The mechanism has been criticized for not ensuring the recovery of the investment fixed cost (Benini, 2006), but it seems that for the time being, the main objective of this transitory measure is to avoid mothballing (keeping existing units in operation), and not to foster new investments.

4.6. Ireland

Under the SEM (Single Electricity Market) established in 2007, in order to complement the market income derived from the wholesale price of the centralized pool, generators receive capacity payments. To determine the annual capacity payments, the regulator calculates every year, among others, three parameters:

- The system's capacity requirements to comply with security standards.
- The annual carrying cost of the best new entrant (generally, the most efficient peaking unit).
- The value of the loss of load (VOLL).

The product of the first two parameters determines the total amount assigned to cover the capacity payments, and this amount is distributed among all generating units[57] in accordance with complex criteria that seek to reflect their contribution to overall security of supply.

These payments depend on the declared availability in each of the hours, and each of the hours is weighted depending on the ex-ante expected LoLP and the ex-post calculated LoLP. The payments also depend on the price bid: a unit whose price is above the Value of Lost of Load (VoLL) is not held to contribute to the security of supply. For further details on the (long) formula that determines the payments' distribution see (SEM, 2009).

Due to the continuous yearly updating, this mechanism does not provide a stable source of income to generators, and thus, it may not be helping to significantly reduce their risk exposure problem, since there is no way of knowing how the payment will evolve in the future years. It would be much more appropriate to stabilise the payment for a minimum number of years for those units entering the system.

4.7. South Korea

In South Korea, the electricity market price is composed of the system marginal price (SMP), where offers are based on audited costs, and a capacity payment (CP). A price cap is imposed on the energy price of base load generating units such as coal and nuclear energy

[57]It is important to note that although this method has been considered as a price mechanism, the price is flexible and indeed is a function of the quantity. In this mechanism the regulator is implicitly defining a (hyperbolic) dependence between quantity and price.

(KPX, 2009). This way, in practice, the market is segmented, since for most of the time base-load and peak-load units are receiving different energy prices.

The value of the capacity payment depends on the hour considered (peak, medium, off-peak), and it is paid to all generating units that have declared to be available, whether or not they are finally dispatched. The objective of these capacity payments is to ensure the recovery of fixed capital costs. The exact value of the payment is determined annually by the Generation Cost Evaluation Committee taking into account the long-term marginal fixed cost of the generators, which depends on the fuel type. In 2006, the base load Capacity Payment (20.49 won/kWh[58]) was derived from the capital and a fixed operation and maintenance cost of a 500 MW coal unit, whereas the peak load Capacity Payment (7.17 won/kWh) reflected the capital and a fixed operation and maintenance cost of a standard gas turbine (Park et al., 2007).

For all these characteristics, although some market based principles have been introduced, this hybrid scheme is still quite close to the classic cost of service regulation.

4.8. Others

Besides the ones mentioned, capacity payments were implemented in some other electricity systems, mainly Latin American ones, which reformed their regulatory scheme to introduce a market-based design: in Colombia they were replaced by the Reliability Charge mechanism described later, and they are still in force in others (Peru, Chile, Dominican Republic, etc.). In some cases, these capacity payments coexist with other long-term security of supply mechanism (as mandatory long-term energy contracting).

5. QUANTITY MECHANISMS

The security of supply mechanisms we include in this category of "quantity mechanisms" differ from the previous ones in the fact that the regulator relies on a market-based mechanism to set the price for the reliability product. This approach in principle solves the main problem of the price mechanisms just described: instead of setting administratively a price and then expect (hope) for the right amount to come into the system, the regulator declares the quantity expected and lets the market mechanism reveal the right price.

This relies on the belief that only market-based solutions lead to efficiency. But unfortunately, real life is in most cases away from ideal (the structure of the markets is often far from the fully competitive one, there are significant entry barriers, and/or regulatory design is more than often flawed). This lack of "ideality" of markets is the reason why short-term markets do not lead to the "secure" solution and require an additional regulatory

[58] As for year 2006, 1 euro ~ 1200 won.

support. This is the same reason why it is far from obvious that these market-based "quantity mechanisms" necessarily solve the matter.

We next analyze the different quantity-based experiences, beginning by reviewing the initial mechanisms, the so-called "capacity markets", and highlighting their flaws. Then we delve into the discussion of the current re-designs, aimed to solve these problems.

5.1. Capacity markets

The term "capacity market" was originally used to denote the pioneering markets introduced by some regulators so as to trade a (reliability) product "artificially" created by him (MW of installed capacity) under certain particular conditions (short-term markets, either bilateral or centralized, and short contract commitments). Again, as it was the case with the "energy-only market" term, the "capacity market" designation can be quite misleading, since it is not always clear which are the underlying characteristics of such a mechanism: just a mechanism which introduces the obligation to buy "capacity" or also a mechanism that, on the top the previous requirement, has introduced short term markets and short-term commitments, etc.

In the first capacity markets, demand had the obligation to contract the power capacity (expressed in MW) required to supply its future consumption. This solution was born in the context of a fully thermal (thus capacity-constrained) power system. The problem that rapidly arose when trying to implement this approach in other non-fully-thermal power system was obvious.

In fully-thermal (non-fuel constrained) systems, it can be assumed that the availability of a thermal plant is uncorrelated from the availability of the rest of the plants in the system and also (sufficiently) uncorrelated from the peak demand. But this is not the case at all of hydrothermal systems, which are mostly energy constrained. The direct consequence is that defining "capacity", as the ability to produce energy when needed, is not so obvious.

Under this "capacity markets" category, we will not only include just the obligation to buy MW's of installed capacity but we also extend the category to those cases in which it is imposed the obligation to buy the capability of producing a certain "MWh in certain hours, season, etc." (i.e. in those periods in which the regulator considers the risk of scarcity is higher).

5.1.1. The former ICAP in the Eastern USA (PJM, NYISO and ISO-NE)

ICAP markets have been the most debated case in terms of capacity markets. Nowadays, they are still an inevitable reference, mainly due to the poor results obtained.

These mechanisms consisted in having every Load Serving Entity (LSE) to back up its expected peak-load capacity requirements (plus a reserve margin) with capacity credits. At first, all generating units received credit for all their installed capacity (that is what ICAP stands for: Installed CAPacity). Hence, each LSE had to purchase a certain amount of the product (the credits), that was supposed to serve to guarantee that there would be enough installed capacity to satisfy their expected peak demand (plus the mentioned margin) at peak hours.

However, capacity is not always available, and soon the different ISOs, beginning with PJM, became aware of the firmness problem and developed the concept of UCAP (Unforced Capacity). By using this new UCAP concept, each ISO was able to discount, depending on each unit's actual historical availability, the capacity for which it was given credit (that is, the quantity of the product that the generator was entitled to sell). Therefore, the former ICAP rating was modified in all three ICAP markets by a new measure that took into account each generating unit forced outage.

Nevertheless, this UCAP was calculated as an average of the available capacity over long periods of time (typically a season or even a whole year) irrespective of whether the unavailability did or did not occur during a scarcity period. Thus, the incentive to be available during tight reserve periods was exactly the same as for any other given hour of the year.

The different ISOs, in their attempt to encourage generators to make their installed capacity available, looked for additional rules. As a consequence an additional condition was introduced for those generators willing to participate in the ICAP mechanism: a must-offer requirement in the day-ahead market. Unfortunately, this did not solve the problem, the reason is that it is difficult to find a means that does not entail fully relying on self reporting by generators (the must-offer requirement is not effective, since unavailabilities can be hidden behind high-priced bids)[59].

The controversial performance of the mechanism

In (PJM, 2006) the PJM Market Monitor conducted an analysis in an attempt to assess whether the fixed costs of the different units were covered by the prices received by generators from the PJM markets plus the ICAP payments, and concluded that investments costs were not being recovered.

In addition to this lack of investment cost recovery, there was another relevant problem linked to the design of these capacity markets: the extreme volatility of prices. Capacity market prices tended to alternate between very low prices, during the large periods where the system's reserve margin was large, and extreme high prices when not enough capacity resources were available (Chandley, 2005). In the next figure this variability can be clearly observed within the PJM ICAP market context.

[59] Although some sort of monitoring is possible, the only alternative implies making *in situ* random tests on a unit by unit basis, as it is the common methodology in Latin American designs, as for instance in Guatemala.

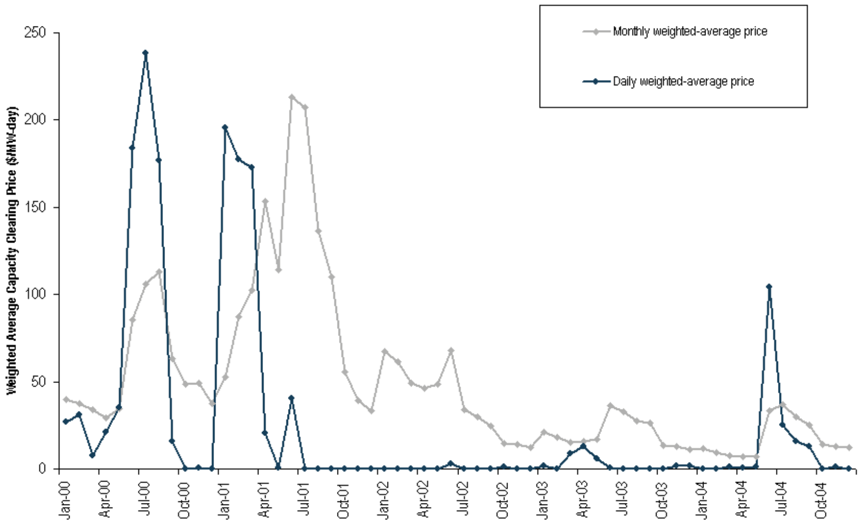


Figure 12 Average capacity prices in PJM - Source: PJM

Capacity demand inelasticity was pointed out as the main reason behind this price volatility, and soon some proposals were made with the objective of determining a downward-sloping demand curve that could better represent demand interests. Nowadays, each of these capacity markets has defined the so-called Variable Resource Requirement demand curve (a price-quantity elastic curve).

But demand inelasticity was neither the only reason behind these market results nor probably the most important. These results were the inevitable consequence of other design flaws that we review next.

Why prices bounced from near-zero to extreme high values?

Once the fixed investment cost of a generating unit has been incurred, it becomes a stranded or sunk cost. According to microeconomic theory these costs should not be considered when facing future decisions. This is the reason why, in (fully competitive) spot markets, it is not rational for generators to internalize their investment costs in their bids.

If the capacity auction is called a short period in advance of the delivery date, only the units already installed can participate. Since these units cannot internalize their investment costs in their bids, the generators have to bid the cost associated with the provision of the reliability product being purchased by the regulator. But, what is the additional cost of keeping an already installed capacity (ICAP) of a unit operational? Obviously near-zero in most cases. This was the reason why prices tended to fall dramatically during long periods.

Conversely, when the capacity reserve margin becomes tight, there is a reliability product scarcity, and prices do reflect this. An additional issue is that under these scenarios, the market becomes prone to market power exercise, especially in those cases where the demand curve is inelastic and does not respond to prices.

This was the case in the early ICAP markets, and therefore, the price that served to complement investors' remuneration in the energy markets presented either near zero or very high peak values.

Thus, instead of providing the stable price signal sought by investors, in the end, another (even more) volatile short-term market was created. The feared energy scarcity periods were replaced by the installed capacity scarcity periods[60].

In the end generators received an extra income that complemented the income received via their sales of energy, but this remuneration was neither certain nor capable of guaranteeing in advance the recovery of investment fixed costs.

Was this volatility unavoidable?

This volatility might have been reduced if the auctions' time terms had been increased so as to allow potential new entrants to participate in the auctions. As pointed out in (Vazquez, 2002) in a general context, and also in (Chandley, 2005) within the PJM ICAP framework, the solution entails allowing for a longer time interval between the moment the commitment of deliverability is signed and the moment it has to be delivered (the so-called "lag period", so as to allow the new entrants to build the generating project). It is also recommended to make some other additional changes: for instance, allowing for longer contract durations, since generation investments usually require long contract durations to ease their project finance.

Lack of locational signals for capacity

In all these Eastern US markets, the effect of network congestions is important, and consequently, reflecting the different value of energy at the different consumption locations has been considered to be a necessary market design characteristic. In this context, if enough liquidity and market competition can be ensured, locational signals should be provided by these capacity markets mechanisms so as to acknowledge the different values of capacity at the different locations.

As we review later, the subsequent designs in all the three systems try to correct all the major design flaws.

5.1.2. Guatemala

In the electricity market design in Guatemala the regulator also imposes on demand the obligation to hedge their future consumption. These so-called "monomial contracts" consist

[60] Although the consequences of a scarcity period in this new market also had an undesired economic impact, since a certain reserve margin with respect to expected peak consumption was defined by the regulator, it did not imply energy rationing.

of two related (but not necessarily correlated) products: their future energy consumption (linked to a profile) plus their "firm offer" needs (a sort of capacity credits).

There is a "market for capacity imbalances" (*mercado de desvíos de potencia* in Spanish) which sketchily works as follows: demand has to acquire ex ante their (expected) future firm offer consumption in the "peak blocks". Then, the counterparty responsible of any firm offer imbalance in a peak block (either the demand if its consumption is above the one contracted or the generating unit if its production is below the one sold) has to pay the capacity payment price to the system. Thus, this capacity payment price represents both a penalty and the implicit price cap of the capacity market. The market operator collects these penalties and allocates them among the units that covered the imbalances.

In the initial design, the firm offer the regulator assigned to each generating unit was in principle based on the theoretical availability of the unit in the dry periods. For hydro units, this value was based on the production that the plant was able to theoretically assure in 95% of the cases (according to historical series) in the "peak blocks" (the four peak hours of each of the working days), from December to May (corresponding to the dry season). On the other hand, for thermal units, the "firm offer" depended on the historical average failure rate.

Similarly to what happened in the Peruvian case, see (Batlle et al., 2009), highly inefficient "junk" generation plants took advantage of such a definition of the reliability product. Plants with very low investment costs, although presenting extremely high production costs (and even high failure rates) are the most competitive units in this context. Indeed, they find it profitable to enter the system with the major (and even sole) objective of receiving the payment provided by the long-term mechanism. This fact made firm offer prices to drop, the desired incentive for new (efficient) entrants vanished and only this junk generation entered the system.

Years later, the regulator came up with a new design aimed to solve the matter. The "firm offer" was replaced by the "efficient firm offer", OFE (*Oferta Firme Eficiente* in Spanish): the regulator simulated the future dispatch of the system in the dry season assuming hydro production corresponded to the 5% percentile, assigning OFE only to the capacity needed to supply the expected future demand plus a 10% reserve margin. The idea was to reward only the so-called efficient plants (i.e. the cheapest). Surprisingly, the proposal did not take into account that no capacity market was possible, since demand was required to contract their expected demand plus a 10% margin, and only the units which had been assigned OFE could sell, so no competition could be found (offer equaled demand). The hasty patch the regulator came up with consisted in assigning the rest of the installed units (the "inefficient" ones) half of the value of the "firm offer" previously assigned under the original design.

Additionally, the fact that the OFE was assigned according to the expected production of the generating units in the four peak hours of each of the working days in the dry season, implied that new investments in small hydro plants were designed so as to have a reservoir to allow daily regulation whose energy capacity (MWh) was the capacity of the turbine

(MW) multiplied by the aforementioned period (just four hours). This is a very clear example about how a poor design of any of the issues defining the mechanism can dramatically lead to inefficient results.

5.1.3. Western Australia: The Reserve Capacity mechanism

The Reserve Capacity procurement process began in Western Australia in late 2004. All the demand is required to buy capacity credits to cover their share of the system future capacity requirements. Both the system requirements and the capacity credits assigned to the generating facilities (and also to some demand side management resources) are determined by the Independent Market Operator (IMO) on a yearly basis.

Capacity credits can be traded either bilaterally or in a centralized auction, which is only held (and conducted by the IMO) when bilateral contracts have not covered the total requirements.

New facilities entering the market can apply for special conditions, being the most relevant the capability of locking the capacity credit price in longer-term periods (up to 10 years) (IMO, 2006). Units applying for this long-term special arrangement are required to certify their capacity on a yearly basis.

5.1.4. France

As previously introduced, on March 2010, the "New Organization of the Electricity Market Bill" (NOME, 2010), proposes the implementation of a capacity market. Thus, it imposes the obligation on retailers to acquire "guarantee certificates" (*certification de garantie*). The SO will allocate this certificates among generating units (and also "elastic demands") according to the "total technically available capacity", and the CRE (*Commission de Régulation de l'Energie*) will calculate annually the penalties to be paid by market agents in case on non-fulfillment 'to provide agents with incentives to invest on new capacities of demand response or generation'.

It is a preliminary proposal, and thus lacking the final development details, but according to the writing of the Bill, apparently it shares the main characteristics of the just discussed ICAP design (no lag period, no specified long-term contract duration, etc.). Hopefully the final design will not fall into the same flaws.

5.2. Long-term auctions for lagged reliability products

Roughly speaking this renewed approach consists in (often centralized) auctions for longer-term contracts, with the additional feature of postponing the moment to start delivery (the so-called lag period, a number of years) so as to allow the winners of the auction to build the plant.

5.2.1. Colombia: the Reliability Charge

The Colombian power system experience pioneered the major wave of changes regarding regulatory design of security-of-supply mechanism, and directly or indirectly influenced heavily some other redesigns that are reviewed afterwards here.

The Colombian electricity system presents a large dependence on hydro-generation, and it is significantly sensitive to the cyclical climate period known as El Niño-Southern Oscillation, which implies suffering one severely dry year once out of five or eight years.

The first scheme, in force during the period 1996-2006, was an administratively determined capacity payment known as "Capacity Charge". Although there were no scarcities during those years, the reserve margin was tightening severely and concerns about the malfunctioning of the scheme were growing.

In fact, the effectiveness of the scheme was called into question almost from the very beginning. The capacity payments assigned to the generation plants in the Colombian electricity market not only a proper incentive to efficiently manage availability in scarcity periods[61] but also did not constitute a stable and trustworthy long-term signal to potential investors[62]. A consultation process on the different flaws of the mechanism was launched in 1999, and as a result some alternatives were proposed. The approach that finally was chosen consisted in replacing the capacity payment by a quantity mechanism, but correcting the already observed flaws of the ones already implemented (mainly PJM, discussed previously). The original proposal was put forward back in 1999 in response to a requirement of ACOLGEN (the generators association) later on described by the consulting team that developed it in (Vázquez et al., 2002). The two major features of this proposal were the introduction of the so-called "reliability option" as the new reliability product and its acquisition through a centralized auction.

[61] Indeed, it affected negatively the efficient planning of the system. Since the firm capacity of the hydro plants depended critically on the water reservoir level in the "dry season", generators managed their reserves in such an uneconomical way that reservoirs were at their full capacity in that season.

[62] A capacity payment is nothing but a regulatory compromise. In the Latin American context, in which regulatory risk perception for investors is large, the regulator realized that it was a better solution to provide long-term contracts with the distribution companies as counterparties, so as to mitigate this credit risk aversion.

The reliability option[63]

The reliability option is a call option contract with the particularity that the strike price is calculated so as to serve as a threshold for determining scarcity situations[64]. In other words, every time the spot price goes above the defined "scarcity price", all the sellers (generating units) have to sell the committed energy at the strike price instead of selling it at the spot market price. The main objective followed with this design was to get to a better way to identify when the security of the system is in danger, since the best and indisputably most market-oriented indicator of an impending scarcity episode is an abnormally high market price.

The other key design parameters of the reliability option were the time terms: both a lag period (denoted as planning period in the Colombian regulation) and a contract duration long enough. The first, to give them time to build the project, and the latter to reduce risk exposure and thus easing the project finance.

These time terms determine critically the type of generators that will enter the system. For instance, a lag period of three or four years can ease the entry of thermal generation but it is completely irrelevant for large hydro plants, whose construction period exceeds this term. Analogously, a very long contract duration (e.g. fifteen years) matches better the project financing needs of a large hydro than a thermal low-capital-intensive peaking unit. In this sense, the finally implemented design includes special rules to cope with this reality (Cramton & Stoft, 2007): the regulator defines different contract duration for the different generation technologies (e.g. shorter for thermal plants than for large hydro ones). Also, as detailed in (CREG, 2006a) and (CREG, 2006b), for large hydro projects the regulator allows the investor to lock-in the auction price from the 4-year ahead auction up to seven years ahead.

The auction

The other main feature introduced in the original proposal was to centralize the acquisition of the reliability option by means of an auction. The objectives are to increase competition and to benefit from economies of scale (gathering together the different sometimes small and numerous regulated retailers, so as to make possible for large investments to participate) and to enhance transparency (to avoid vertical integrated companies taking advantage of obscure agreements).

---

[63] In the finally implemented mechanism, commissioned to Cramton & Stoft (2007), the reliability product was named Firm Energy Obligation.

[64] In practice, this results in a strike price that is set at a level slightly higher than the most expensive unit's marginal costs.

However, contrary to what the original proposal suggested, which claimed that the regulator should acquire the reliability options on behalf of the whole demand, the regulator just acquires Firm Energy Obligations on behalf of the "domestic demand". This approach, the usual one in Latin American power markets, such as some of the ones reviewed next, Brazil, Peru, etc. makes the mechanism highly vulnerable to free riding[65].

The final auction design consists in a descending-clock auction, including among other details, a downward-slopping curve to specify how the purchased quantity of the reliability product depends upon price. Also, a relevant characteristic of the process is that different rules apply to new and existing plants (e.g. existing plants are price takers in the auction).

5.2.2. Brazil

Electrical energy in Brazil comes mainly (80-85%) from hydro-generation plants with multi-annual reservoirs. The first market-based design, in force from 1996 to 2004 consisted in a centralized system marginal cost calculation to remunerate generating units, with added sort-of security of supply mechanism: regulated retailers were compelled to contract in the long-term 85% of their expected future energy needs and also a very peculiar kind-of capacity payment was implemented, since a floor price existed to overcome the fact that the market price is zero almost 80% of the time. Although the concern of keeping enough reserves for the draught period existed, it was not until the 2001 and 2002 rationing episodes, imposed on all types of consumers in a geographical region representing 80% of consumption, that ensuring long-term security of supply truly became a principal objective. The situation led to several thorough analyses, as a result of which experts concluded that there were some imperfections regarding expansion and efficient contracting. This led to a proposal that to some extent was inspired in the abovementioned solution proposed years before for the Colombian system and that resulted in the mechanism currently in place (Barroso et al., 2006).

The main differences from the Colombian case are:

- Different auctions are called for existing units and new entrants. In the first ones, the lag period and the contract duration are significantly shorter (1-year lag instead of 5, up to 15 years instead of up to 30).
- There are two different reliability products: a financial forward energy contract for hydro units and an "energy call option" (which in very general terms presents the characteristics

[65] That is, those being represented by the regulator in the mechanism are not always the only ones enjoying a higher level of reliability. For instance, on many occasions, particularly in Latin American power markets, large consumers are exempted from long-term contracting or defraying the capacity payment. The problem is that often, in a scarcity event, there is no technical way to discriminate between consumers and therefore these large customers receive an equal supply. This could be easily fixed by simply better defining the product, for instance by charging large customers an extremely high cost for their consumption in this situation.

of the reliability option previously described in the context of the Colombian case) for thermal plants.

- The regulator has a backstop mechanism that allows the government to carry out specific energy auctions driven by energy policy decisions. In 2008, for instance, a special auction for this mechanism was held for 1200 MW of co-generated power produced with sugar cane biomass), see (Batlle et al., 2010).

5.2.3. ISO New England

In ISO-NE, the so-called Forward Capacity Market (FCM) replaced the previous ICAP mechanism, see (Cramton & Stoft, 2005) and (ISO New England, 2006). This new framework shares the major characteristics of the mechanism described in the Colombian context, without entering into some of Colombian complications led to cope with different generation technologies, but including locational signals, i.e. different zones are defined in which the capacity requirements and clearing prices are calculated.

We do not delve into the description of the auction-based mechanism, since as stated is very similar to the Colombian mechanism. However, as it is the case in other designs, it is remarkable the degree of integration of demand as a potential provider of the reliability product.

5.2.4. PJM's Reliability Pricing Model (RPM) and the new NYISO's ICAP

The poor performance of the capacity markets originally implemented in PJM and NYISO led to significant redesigns aimed to correct most of the shortcomings that have been analyzed previously. Sketchily, the new design (PJM, 2008) consists in an auction for a reliability product, which in this case differs from the described Colombian mechanism. The product design is an evolution of the former UCAP: mainly the way availability is measured is much more detailed. The time terms consider both a longer lag period and longer contract durations; in Figure 13, the timing of the organized auctions in the PJM Reliability Pricing Mechanism is shown. Furthermore, as in the NE mechanism, locational signals are provided to capacity.

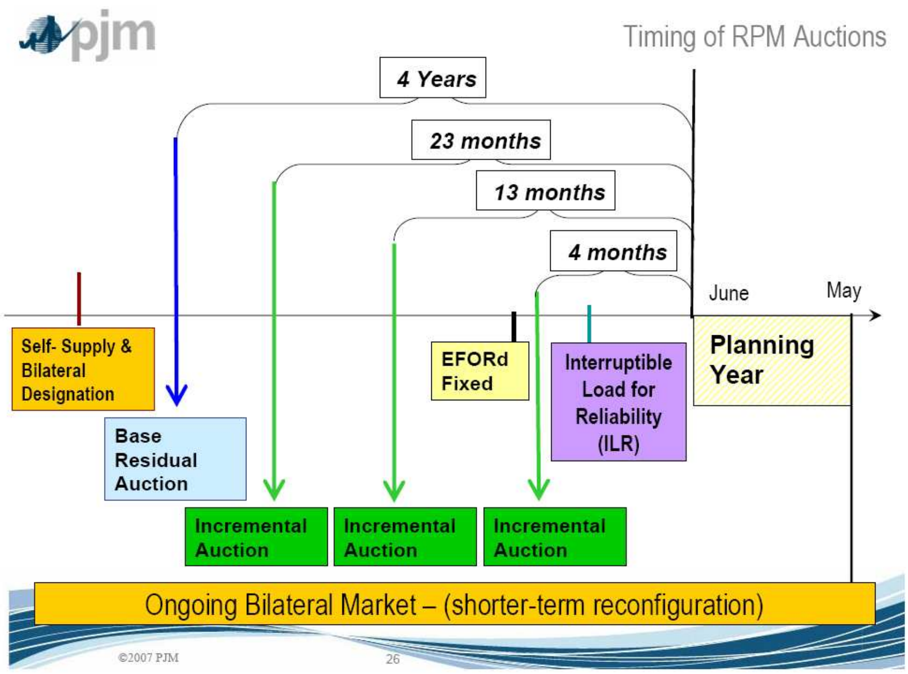


Figure 13 Timing of the RPM Auctions. Source: (PJM, 2007)

5.2.5. Chile

In the Chilean system, distributors (in their role of retailers of the regulated demand) must commit all their consumption at all times and at least three years in advance. The relevant contracts have also to be assigned through public auctions (Ministerio de Economía, 2008). Each distributor must design its own contract characteristics as well as manage its own auction, although other distributors may join the process so as to take advantage of economies of scale. Additionally, the capacity payment is fixed over the duration of the contract.

5.2.6. Peru

The current design in force in Peru, see (Batlle et al., 2009), is rather similar to the Chilean one. In the first auctions under the new scheme established by the 28832 Law in 2006, called in April 2010, the supply contracts were for the period 2014-2021. The design also includes an administratively determined capacity payment that complements generation retribution.

The auctions have to be celebrated at least three years before delivery. In principle, it is up to distributors to decide when and how to hold the auctions. If the delivery time frame is longer than three years, distributors receive an extra payment. The problem is that this extra payment has been defined as a percentage of the costs associated with the contracts committed in the auction, which may give some perverse incentives to vertically integrated companies. Additionally, in order to promote hydro investments, when clearing the auction, bids from this type of installations are multiplied by 0,85, and if winning they are rewarded their full bid price.

5.2.7. Panama

In Panama, distributors have to sign contracts in advance via public auctions for both their expected energy supply and their capacity requirements (peak consumption), taking into account a safety margin determined by the regulator, see (2008, CND) and (Urrutia, 2008). They can opt for auctioning both products (the energy and the capacity ones) in a joint auction or separately.

These auctions have to be called by distributors themselves, at least two years in advance, although three or four years are recommended so as to allow new entrants to participate and thus increase competition in the acquisition process. After suffering some scarcity episodes, the mechanism is about to be corrected by implementing public auctions to acquire long-term supply contracts allowing also for a sufficient lag period.

5.2.8. Argentina

The Argentinean Energy Plus Program (EP, 2006), imposes large industrial consumers (above 300 kW) the obligation to cover their energy requirements exceeding 2005 historical consumption. The counterparts of this Energy Plus Contracts are new generating projects, which are offered a stable remuneration in exchange (this remuneration is calculated so as to cover operational and investment costs). These energy contracts are verified every three months by the Energy Ministry. Penalizations are defined in case of non-compliance. This service came into force in 2006 and the major characteristics are defined in (SdE, 2006).

5.3. Strategic reserves as the reliability product

In this last category of the quantity-based mechanisms, we have included those schemes focused on purchasing the so-called "strategic reserves". The "strategic reserves" have traditionally represented a particularly controversial type of reliability-based product, since it consists in tearing apart a certain amount of generation capacity which does not take part in the energy market, unless the regulators or the SO considers that it is necessary (according to more or less objective criteria).

Different works, as the remarkable analysis carried out in (Finon el al, 2008), have shown that if both the criteria used to trigger the production of the reserves and the purchasing process are well-designed, the mechanism can provide suitable results.

5.3.1. Finland, Norway and Sweden

In Finland, Norway and Sweden, the SO is in charge of purchasing "load reserves". It is important to clarify that these "load reserves" are completely separated from the standard operational reserves, i.e. primary, secondary and tertiary (balancing) reserves to restore frequency. Load reserves (also known as strategic reserves) are defined as reserves designated for times when demand is close to exceeding the available production capacity; in other words, they are called upon to supply energy when a generation scarcity scenario appears.

The objective of this approach is to prevent some old units from mothballing, but in some cases, it is the responsibility of the SO to also define the rules for offering the electricity of these reserves on the market. Obviously, this can result in a significant distortion of the price signals (since the SO may become into an irregular market agent). However, if the price at which the load reserves produce is set at a value that reflects non-served energy, there would be no price distortion (as it is the case in the experience of New Zealand).

Norway

The Norwegian approach adds a market for peak-load reserves, where the TSO buys call options for regulating power (increase in generation or reduction of consumption) in order to always have a minimum of 2000 MW at its disposal.

In the beginning, these options had a duration of 12 months. This changed in 2004, when the time period changed to weekly contracts. Interested players may bid in this market for up to 8 weeks in advance.

The market clears at the marginal price.

Finland

The Finnish long-term security of supply mechanism is based on requiring the TSO to both purchase and define the rules for operating a separate peak load reserve.

The mechanism is regulated by the Power Reserve Law, which establishes the temporary nature of the measure (it is expected to end at February 2011).

At this point, Fingrid (the TSO) is responsible for procuring this type of production capability, which is called upon under near generation-scarce scenarios. Those power plants accepted[66] in a competitive process receive compensation for being available when needed.

The objective of the mechanisms is to prevent some old units from mothballing, mainly due to profitability problems. Fingrid is entitled to reject some offers in the belief that the units will have revenues that suffice to recover its fixed investment costs.

The law also states that the produced electricity from the Peak Reserve Capacity units must be bid in the market at a cost-based price. As previously noted, it is the responsibility of the

[66] The Power Reserve Law states a set of requirements to be met by the units if they wish participate in the mechanisms (e.g. a maximum time for startups or certain minimum ramp-up values are defined).

TSO to define the rules for offering the electricity on the market, but it must bear in mind that, those rules should be designed so as not to distort market prices unless absolutely necessary.

Sweden

The Swedish approach is focused on having a separate peak-load reserve. Again, the TSO is responsible for procuring a maximum of 2000 MW for, this reserve by purchasing capacity (production or reductions). The units selling this capability receive a compensation payment in exchange for their services.

When these reserves are needed they have to be offered through the Swedish regulating market.

5.3.2. New Zealand

New Zealand is a hydro-dominated system (65% of the energy in an average year). Thus, as in most hydro-dominated systems, the concern has been to ensure enough production resources during dry years[67.] The mechanism designed to overcome energy constraints during those years consists in contracting strategic reserves, which may include either new or old equipment (MED, 2003).

The contractors are selected by means of centralized public auctions and their responsibility is only to supply energy and capacity during scarcity periods (dry years). The design of the strategic reserve mechanism includes the price at which the reserve capacity has to be offered on the wholesale market. This price is set at a high value, which ideally serves as a threshold for detecting scarcity situations[68], since if the price of reserve capacity were too low it would compete with ordinary generation, which would deter new investment by generators

Prices above that of the reserve generation are expected to be very rare, although not impossible. Very high peak spot prices could still occur if extreme circumstances exhaust the available reserve capacity.

It is important to note that although this trigger price plays the same role than the strike price in the reliability option (the financial option that has been described within the NE, Brazilian or Colombian mechanisms), the strategic reserves and the reliability options are two very different products. The main difference is that in the case of the strategic reserves, the energy can only be sold if the spot market price reaches (or goes above) the unit trigger price

---

[67] New Zealand's Electricity Commission objective is to ensure that supply remains secure even in a 1-in-60 dry year event, that is, in a hydro drought of a severity that can be expected to occur every 60 years

[68] For instance, the Whirinaki reserve energy trigger price was set at 38.7c/kWh ($387/MWh) in its December 2008 update.

established by the Commission while in the reliability options context, the production can also be sold during the rest of the time. Thus, the strategic reserve represents generation that is kept outside of the market under normal circumstances.

Additionally, it is also recognized that a 100% reliable system is not feasible economically, and thus, under unusual circumstances, there may be a risk of shortages (e.g. a year which is worse than a 1-in-60 dry year). In such an event, the Electricity Commission would activate a conservation campaign at the appropriate time. For further details see (Concept Consulting Group, 2004).

6. REFERENCES

AER (Australian Energy Regulator) (2007). "State of the Energy Market", July, 2007

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

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., 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., Barroso, L. A., Pérez-Arriaga, I, J. (2010). "The changing role of the State in the expansion of electricity supply in Latin America". Energy Policy (2010), doi: 10.1016/j.enpol.2010.07.037. Available at www.upcomillas.es/batlle.

Batlle, C., L. A. Barroso, M. Uribe (2009). "Long-term energy auctions design criteria and flaws in Latin America. The case of Peru". IIT Working Paper, Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, Madrid.

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.

Benini, M., F. Cremonesi, M. Gallanti, A. Gelmini , R. Martini (2006). "Capacity payment schemes in the Italian Electricity Market", CIGRE General Session 2006.

CAMMESA (2005), "Argentine Power Sector Capacity payments, markets and new Generation", Compañía Administradora del Mercado Eléctrico Mayorista, APEX Conference Orlando, USA, November 2005

Caramanis, M. C. (1982), "Investment decisions and long-term planning under electricity spot pricing", IEEE Transaction on Power Apparatus and Systems, Vol. PAS-101, No. 12, December 1982.

Caramanis, M. C.Chandley, J. (2005). "ICAP Reform Proposals in New England and PJM," LECG, Report to the California ISO, September 2005.

CND (2008). "Reglas Comerciales" (In Spanish), Centro Nacional de Despacho (CND), 2008. Available at www.cnd.com.pa.

Cramton, P. & S. Stoft (2005). "A capacity market that makes sense". The Electricity Journal, 18(7), 43–54.

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

CREG (2006a). "Reliability Charge: regulatory scheme to guarantee the reliability in the supply of electric energy in Colombia, a long term vision". Comisión de Regulación de Energía y Gas, 2006. Available at http://www.creg.gov.co/cxc/download/documentos/ReliabilityCharge.pdf

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

EP (2006), http://www.energiaplus.com.ar/, Energy Plus Service Web page.

Finon, D., V. Pignon (2008). "Capacity mechanisms in imperfect electricity markets". Editorial of the special issue on Capacity Mechanisms in Imperfect Electricity Markets, Utilities Policy, Volume 16, Issue 3, September 2008, Pages 141-142.

Finon, D., G. Meunier, V. Pignon (2008)., "The social efficiency of long-term capacity reserve mechanisms", Utilities Policy, Volume 16, Issue 3, Capacity Mechanisms in Imperfect Electricity Markets, September 2008, Pages 202-214.

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

Huber, E. R., V. R. Espinoza, R. Palma-Behnke (2006). "Hydrothermal Coordination and Capacity Payment Schemes in Chile: Current Discussion and Future Challenges". Working Paper. Mimeo.

ISO New England, (2006) "Market Rule 1 Standard Market Design", Section III, ISO New England, Inc. FERC Electric Tariff No. 3, October 2006.

IMO (2006), "Wholesale Electricity Market Design Summary" ,Independent Market Operator, September, 2006.

KPX, Korean Power Exchange (2009). Web page: http://www.kpx.or.kr/english/.

MITyC (2007) "Orden ITC/2794/2007 de 27 Septiembre, por la que se Revisan las Tarifas Eléctricas a Partir del 1 de Octubre de 2007" (Order ITC/2794/ 2007, of Sept. 27, Revising the Electricity Tariff from Oct. 1, 2007), Ministerio de Industria Turismo y Comercio (Ministry of Industry, Tourism and Trade), 2007, available at www.boe.es/boe/dias/ 2007/09/29/pdfs/A39690-39698.pdf (in Spanish).

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.

MED, Ministry of Economic Development (2003). Electricity Supply Security: Questions and Answers. 20 May 2003. Available at www.med.govt.nz/templates/Page____17418.aspx.

Ministerio de Economía, Fomento y Reconstrucción (2008). Nº 39.048, Diario Oficial de la República de Chile. Abril 2008.

NGET, National Grid Electricity Transmission (2010). The Grid Code. Issue 4 Revision 2. 22nd March 2010. Available at . www.nationalgrid.com

NOME (2010). "Projet de loi de nouvelle organisation du marché électrique". (In French, "New Organization of the Electricity Market Bill"). March 2010. Available at www.energie2007.fr/actualites/fiche/2538.

North American Electric Reliability Council (1997) "NERC Planning Standards", September 1997.

Ofgem (2009), "Project Discovery - Energy Market Scenarios", Reference number 122/09, Office of the Gas and Electricity Markets (Ofgem), October, 2009.

Ofgem (2010), "Project Discovery Options for delivering secure and sustainable energy supplies. Consultation document", Reference Number 16/10, Office of the Gas and Electricity Markets (Ofgem), February, 2010.

Park, J., N. Ahn, Y. Yoon, K. Koh, D. Bunn (2007). "Investment incentives in the Korean electricity market", Energy Policy 35 5819–5828, June 2007.

Pérez-Arriaga, I. J. (1994)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, 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., C. Batlle, C. Vázquez, M. Rivier, P. Rodilla (2005). "Libro Blanco sobre la reforma del marco regulatorio de la generación eléctrica en España". ("White Paper for the reform of the regulatory scheme of the power generation in Spain" for the Ministry of Industry, Tourism and Commerce of Spain), ISBN 978-84-4785-6, July 2005.

Pérez-Arriaga, I. J., C. Batlle, 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/batlle.

PJM (2007), "NJ BPU Energy Master Plan", PJM Presentation, February, 2007.

PJM (2008), "PJM Manual 18: PJM Capacity Market", PJM Forward Market Operations, 2008

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.

SdE (2006), "Resolución 1281/2006" (In Spanish), Boletín Oficial del 05-sep-2006, Número: 30983, Página: 5, Secretaría de Energía (State Secretary for Energy), 2006, available at http://www.infoleg.gov.ar/infolegInternet/anexos/115000-119999/119455/norma.htm

SEM (2008), "Designated SEM Trading and Settlement Code v4.5", Single Electricity Market, 2008, available at http://www.allislandmarket.com/MarketRules

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", available at http://www.terna.it/LinkClick.aspx?fileticket=Qmyh8X33TCQ%3d&tabid=2578 (In Italian).

Urrutia, V. C. (2008),"Cobertura de la demanda de energía eléctrica en Panamá: Perspectivas y Mecanismos para asegurar el abasto suficiente a los usuarios" (In Spanish), XII Reunión de la Asociación Iberoamericana de Entidades Reguladoras de Energía

San Luis Potosí, México, Abril 2008.

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, (2003) "Modelos de casación de ofertas en mercados eléctricos", PhD Thesis, Universidad Pontificia Comilla, September 2003.

Wolak, F. (2004). "What's Wrong with Capacity Markets," Mimeo, Stanford University.

Chapter V

Chapter V. Principles and Criteria to Design Security of Supply Mechanisms

1. INTRODUCTION

We have seen in the previous chapters how the market is assumed to be unable to provide sufficient generation availability when needed without regulatory intervention (Chapter I and Chapter II), and also how the solution to the problem necessarily entails the development of additional signals to assure firmness and adequacy of supply (Chapter III). Indeed, we have seen that some regulatory mechanisms to secure long-term supply have been designed and implemented in a number of power markets around the world (Chapter IV).

We have also discussed in Chapter III the importance of properly defining the objectives so as to optimally design the incentives. Sometimes the objectives of maximizing the "net social benefit" at the power system level are nonetheless often modified by another series of aims, such as protection of an autochthonous natural resource, diversification into other sources of generation or the development of technologies with a lower environmental impact (what we have termed the fifth dimension in PART A). Theoretically, the desirability of such aims could be quantified in a function which, added to net social benefit, would yield a new target function to be optimized.

Once the regulator has decided to undertake the task of "helping" the market to reach what he/she considers to be the most efficient outcome, the next key question is how to introduce the necessary adjustments in the market designs in place so as to achieve the objective pursued in the long term. This is particularly complicated and controversial, because in the end, all medium to 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.

Here we aim to identify, under the light of the different analyses carried out up to this point, which are the steps and design elements a regulator should consider when designing a mechanism aimed at tackling inefficiencies at the firmness and adequacy dimensions.

2. DESIGN ELEMENTS

The first step when evaluating the necessity to introduce a mechanism is to make sure that the barriers interfering with the well functioning of the market have been removed as far as possible. If this is already the case and the market is not providing an efficient outcome, then the second necessary step is to analyze the causes behind this failure. Usually the most relevant cause, as seen in Chapter I and Chapter II, is the inefficient allocation of risk.

Generally speaking, all the mechanisms which aim at solving the security of supply problem involve the following four major decisions on the regulator's side:

- To determine the counterparties (buyers and sellers).
- The reliability product definition, that is, the product that the demand-side has to purchase from the generation-side.

- How to set the price of the product.
- Other details, as whether the contracts are signed bilaterally or within an auction, and whether the purchasing process is centralized or not.

We first analyze each of these four decisions to later enter into deeper study of the reliability product design details since, as pointed out in Chapter III, this reliability product plays a very important role in the whole mechanism performance.

2.1. The counterparties: buyers and sellers

The buyers

That is, the part of the demand on behalf of which the regulator makes decisions. The regulator has to decide whether to act on behalf of all the demand or just a proportion of it.

Care will need to be taken so as not to create free riding issues and cross subsidies. Sometimes, those being represented by the regulator in the mechanism are not always the only ones enjoying a higher level of security of supply.

- This may happen when the mechanism consists in acquiring a product presenting characteristics of "public good", in other words, a product that cannot be exclusively used by those who acquired it. For instance, on many occasions, particularly in Latin American power markets, large consumers are exempted from long-term contracting or defraying the capacity payment. The problem is that often, in a scarcity event, there is no technical way to discriminate between consumers and therefore these large customers receive supply of the same quality. This could be easily fixed by simply better defining the reliability product, for instance by charging large customers an extremely high cost for their consumption in this situation.

The sellers

The regulator also has to define who is entitled to act as a seller in the mechanism. In some cases all types of units are allowed, in some other just new investments or some particular technologies. Depending on the particular case, discriminating among different units may create a market segmentation with undesired long term effects.

2.2. The reliability product

It is important to properly define what product the generating units sell in return for the additional instrument providing the risk hedge and/or the additional source of income introduced by the security-of-supply mechanism. As introduced in PART A, and then analyzed in this PART B, this is known as the "reliability product".

We have seen in Chapter III that determining the product to be purchased from the generation is of the utmost importance and complexity. In practice, there are many different alternatives: fixed or flexible long-term energy contracts, certificates on installed capacity (or reservoir capacity), certificates on available capacity (or available energy), certificates on a certain technology installed capacity, long-term reserves requirements, physical units to be operated by the SO under certain conditions, energy financial contracts, etc.

Defining an adequate product can determine the success or failure of the whole mechanism. In Chapter III we pointed out that the introduction of a mechanism consisting in buying a reliability product clearly affects the generator decisions. This way, 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. For instance, if the regulator decides to buy installed capacity, it will probably get the capacity which presents the lowest investment costs, but maybe with low availability rates (see description of the Peruvian experience in Chapter IV, section 5.2.6). 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[69]. Sometimes the consequences of the product definition are not evaluated beforehand, and highly inefficient situations are the result.

There is a certain consensus around the idea that the reliability product should remunerate the capability of producing energy at "reasonable" prices when the system is suffering a scarcity (whatever "reasonable" might mean, usually below the NSE value). But at the same time, it is also far from being obvious how to define a scarcity. In this respect, the market price seems to be one of the most adequate and transparent indexes, but although inferior, other possibilities have also been implemented, such as defining certain periods ex-ante, using the reserve margin measure (whatever the methodology to calculate it used), etc.

Some examples of reliability product extracted from the international experience are analyzed later on section 3.

2.3. Price versus quantity

The regulator has to decide whether a price-based, quantity-based or price-quantity based curve is going to be offered on behalf of the demand.

Resorting to a fixed-price mechanism may result in a security of supply that is either too large or too small. Analogously, resorting to a fixed quantity may result in too high a price (particularly if the regulator does not expect much competitive pressure).

Elastic requirements better reflect the utility each security-of-supply level provides to the buyer (the demand, in this case represented by the regulator). Additionally, they help to

[69] This was the case in the Colombian market back in the late nineties. 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).

reduce market power and also provide more information about how far the system is from suffering a scarcity.

2.4. Other details

The regulator also has to determine some characteristics of the purchasing process. Two relevant decisions in this respect are:

- Whether the purchasing process is centralized or not, i.e. if the retailers (Distributors or Load Serving Entities when part of the demand is regulated) are themselves who manage to do it or if it is the regulator who takes charge of it.
- Whether the purchasing process is arranged through bilateral agreements or by means of public auctions.

The regulatory intervention

As mentioned previously, in a purely market-based context, the interaction of demand and generation, in which both express their preferences, should lead to an efficient outcome. If a regulatory mechanism is implemented, demand preferences are partially substituted by the decisions that have just been described, and thus, the role that the regulator plays in the long-term equilibrium becomes evident.

It does not matter what market-based procedure is followed when the reliability product is purchased; the parameters defining the product and the offer curve determine to a large extent not only the security of supply level but also the type of generators that will provide it.

3. THE RELIABILITY PRODUCT: SOME EXAMPLES AND DESIGN ISSUES

Regulators have used a large number of approaches worldwide in their attempt to secure long-term security of supply. The main objective pursued by them all has been to provide generators with the extra income or the hedging instruments they require to proceed with investments (adequacy and strategic expansion policy dimensions), and at the same time provide incentives to make generators' resources available when most needed (firmness dimension). Obviously, this commitment has to be associated to a certain product (the so-called reliability product) that the generators give in exchange.

Experience worldwide shows that this product definition is the cornerstone of the whole mechanism: if it is not carefully designed, there may be a risk of imposing additional charges on demand in exchange of nothing or even in exchange for larger costs and greater inefficiencies –as for instance has been the case in the Colombian or Argentinean former capacity mechanisms (see Chapter IV, section 4.3).

Additionally, as was previously discussed, an inadequate definition of the product can also introduce fee riding issues.

The reliability product can take many different forms, and its definition can even be in terms of the delivery geographical location (nodal or zonal), see (Hogan, 2009).

Some of these products include:

- Certificates of installed capacity (e.g. the former ICAP in the Northeast USA). Experience so far has shown that paying for installed capacity ensures neither an efficient investment nor efficient resource management.

  The contribution that a newly installed MW provides to the system in terms of security of supply depends on the particular technology being considered. Thus, an additional MW of a coal-fired thermal plant is not comparable to an additional MW of a hydro plant in terms of the security of supply they provide. An installed-capacity-based product does not allow giving incentives to those MWs that can better increase security of supply.

  Indeed, highly inefficient "junk" generation plants can take advantage of such a definition of the product (as it has been the case in the Peru or Guatemala). The reason is that plants with low investment costs, although presenting extremely high production costs and even high failure rates, may find it profitable to enter the system with the sole objective of receiving the payment provided by the long-term mechanism.

- Certificates of firm supply: firm capacity or firm energy. They were introduced in a subsequent redesign of the ICAP, and also in Italy, Peru, Panama, Chile, the former mechanism in Spain, etc. The "firm production" is a wide concept which makes reference to the plant's production capability. Previous forced outages or actual production values are usually taken into account when assessing this production capability.

  Depending on the details of their implementation, these firm capacity (or energy) certificates may represent an average annual production capability or the plant's production capability under stress conditions.

  The evaluation of the firm capacity (or energy) of a new investment is usually calculated based on a benchmark of a number of similar plants (also, if possible, operated under similar conditions). This assessment is particularly difficult when a new technology is being introduced in a system (e.g. wind generation).

- Long-term energy supply contracts (e.g. Peru, Panama).

  These energy contracts may take many different forms, among others a base-load energy production commitment; a specific energy associated to a certain profile; a real-time percentage of an overall demand full-requirement consumption, and so on.

- Options to buy energy at a certain price (e.g. Brazil, New England or Colombia).

- Short-term operating reserves.

  As it has been previously pointed out, the requirements of operation reserves play a key role in the determination of short term energy prices, since spot energy prices and operation reserves prices are clearly connected (a scarcity in any of those markets should affect at the same time both spot and reserve prices).

Based on this idea, in (Stoft, 2003) and then in (Hogan, 2005) it is pointed out that regardless whether a price cap is in place or not, the regulator can obtain any reserve margin by just fine-tuning the operation reserve requirements. The reason is that in doing so, the regulator can alter the length of the scarcity periods[70], and thus, the resulting scarcity rents. In this way, by increasing these rents a higher reserve margin would be achieved.

The problem with this approach is the fact that it increases the demand for a product which is intended for another purpose, that is, short term system security. This is commented on in (Stoft, 2003), where it is pointed out that it would be inefficient "to have a coal plant providing 30-min spinning reserves if that is not needed for short-run reliability".

To overcome this inefficiency, in (Stoft, 2003) it is recommended to introduce a 24-h non-spinning reserve requirement. This latter approach is very close to a short term capacity (based on availability) market, which, on the one hand, has demonstrated that it is capable of providing extra remuneration with respect to that which is provided by an energy-only market, but on the other hand, fails to provide a stable income for the generator.

- Reserves that can be called upon by the system operator when needed:
  - To ensure enough ancillary services (e.g. U.K. or France).
  - To cover peak demand (e.g. Finland, Norway or Sweden).
- Reserves to be bid in the wholesale market at a certain price (e.g. New Zealand).
  - This price is usually above any technology operating costs.

Product characteristics: design issues

There are certain design characteristics of the reliability product that, to a large extent, condition the results provided by the mechanism. Among others, we find:

- Contract duration: large investments usually require long-term contracts in order to obtain the project finance conditions that will allow the plants involved to be competitive.
- Lag period: as previously defined, this is the time period between the moment the commitment of deliverability is signed and the moment it has to be delivered. The projects which take longer than the lag period to be built will not be able to participate in the mechanism. An extreme situation arises when the lag period does not allow any new investment to participate (this was the case for instance in the former short-term ICAP markets which were in place a few years ago in the Northeast USA).

[70] Note that a scarcity in the operation reserve margin would increase prices in both markets (energy and reserves) without implying a situation of energy supply rationing.

- Penalties: the clauses establishing the corresponding penalties that should be applied in the event that the generator does not fulfill its commitment. An appropriate range of penalties is essential to provide generators with incentives. Sometimes these penalties (and incentives) are implicitly defined within the product considered (e.g. an energy option penalizes the generator whose plants are not available when prices are reflecting a scarcity).

- Force majeure clauses: they are clauses that exempt a party from liability in the event that some unforeseen event beyond the control of that party prevents it from performing its obligations under the contract.

  Some of these unforeseen events may be explicitly mentioned in the contract, and if these clauses are not carefully designed, they may prevent the reliability objectives from being achieved. For example, in the former Chilean market scheme, generators that did not have enough energy to supply their contracts had to pay compensation at much higher prices than the average production cost (the cost of non-served energy as defined by the regulator). Nevertheless, the regulation before the 98-99 supply crisis also defined exceptional conditions where that compensation would not be paid. A drought that was not in the 40-year statistical record used by the regulator to calculate regulated prices was defined as a "force majeure" condition (defined by the old regulation article "99 bis"). After the 98-99 electrical supply crisis, which produced severe supply disruptions with scheduled load curtailment, the regulator introduced a change in the controversial article 99 bis, eliminating (among others) drought from the "force majeure" condition (Huber et al., 2006).

- Guarantees: they are usually required by the regulator as a mean of hedging against companies' credit risks. However, they sometimes may introduce entry barriers.

  A guarantee that is usually required for long-term energy financial contracts is a physical generation back-up (i.e. the physical capability of self-producing the quantity committed). This physical back-up is usually determined by means of the firm capacity (or energy) concept. We have analyzed this issue in depth below (in section 3.1).

Measuring the unit's compliance

In any case, the regulator needs 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 (or penalizations).

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[71], 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).

Two extreme alternatives: setting the quantity versus setting the price

Once the regulator has established the reliability product, the next step involves defining the price-quantity demand curve.

For the sake of simplicity we will classify the different experiences into two extreme approaches: the first one consists in the regulator determining the desired quantity of the product and the second in it directly setting the price of the product itself.

We will consider as quantity mechanisms those in which the main objective has been to acquire a certain quantity of a certain product (although this quantity may be flexible) and as price mechanisms, those in which the objective has been to provide a complementary payment that for the most part has been calculated and determined by the regulator (although it may also be flexible).

When the regulator determines the required quantity of the product this can be traded either bilaterally or within public auctions. The international learning process has led to a certain consensus about the desirability of using auctions for different reasons, among others, to increase competition, to avoid vertical integrated companies taking advantage of obscure agreements, to benefit from economies of scale, etc. In this auction, the agents' bids, along with the defined quantity, determine the resulting price of the product. The specific design characteristics of the auction may have some impact on the final results. Although some specific comments will be provided in the context of certain international experiences, it is not the objective of this present work to deal with the problems associated with the auction design (open versus sealed bid, ascending versus descending clock, single price versus pay as bid, etc.).

Some examples of this quantity approach are the so-called capacity markets or long-term (operating or peak-load) reserves purchasing.

On the other hand, when the regulator administratively sets the price of the product, the amount of the product that will be brought to the system is usually left in the agents' hands. Capacity payments (in all of their multiple implementation variations) are clear examples of this methodology.

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

3.1. Guarantees: the regulatory estimation of the so-called firm supply

As previously mentioned, the guarantees usually play a key role in the security of supply oriented mechanisms. In this section we analyze both the reasons behind the need of asking for guarantees, and we also describe a particularly relevant type of guarantee: the physical back-up.

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 previously in this chapter, 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, the regulator also determines a minimum limit to be offered in the mechanism.

The calculation of this firm supply limit represents an important challenge and also a very controversial issue, for it puts boundaries to a significant source of income for generators. One of the key difficulties lies in the necessary assumptions about the behavior of the generators (medium-term resource management, maintenances, etc.). Furthermore, the 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 the 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 so relevant 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, 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.
  - 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: 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[72]. 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 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).

4. 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.

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

CREG, Gas and Energy Regulatory Commission (2006). "Cargo por Confiabilidad, una visión de largo plazo". (In Spanish).

[72] 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.

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.

Huber, E. R., V. R. Espinoza, R. Palma-Behnke (2006). "Hydrothermal Coordination and Capacity Payment Schemes in Chile: Current Discussion and Future Challenges". Working Paper. Mimeo.

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

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.

Stoft, S. (2003) "The demand for operating reserves: key to price spikes and investment", IEEE Transactions on Power Systems, Volume: 18, Issue: 2, pp. 470- 477, May 2003.

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

PART D

PART D. SIMULATION MODELS TO SUPPORT THE DESIGN AND IMPACT ASSESSMENT OF SECURITY OF SUPPLY MECHANISMS

Chapter VI

Chapter VI. PPC Models: Classic Tools to Estimate the Reliability of a Power Generation System

1. INTRODUCTION

The ability to analyze the market performance, in order to set proper objectives and to anticipate and analyze the market response to potential new mechanisms is essential for regulators. In this context, simulation models represent powerful tools to assist the regulator in his duties.

In Chapter III, it was highlighted that the first necessary step in the complex task of achieving an optimal level of security of supply is to detect whether or not (and assess to what extent) there is a problem. In order to carry out such an assessment it is necessary to define metrics to evaluate the system and market performance. As discussed, these metrics, when referred to the firmness and adequacy dimensions, have been usually expressed in terms of "pure" reliability (for instance, based on the LOLP, the LOLE or the ENSE values[73]), assigning a secondary role to the actual cost of enjoying the level required. It was also discussed that these reliability measures should be reconsidered and redefined in a context in which there is a non-negligible demand elasticity (see Chapter III, section 2).

One of the simulation tools which has been traditionally used to carry out electric power generation system reliability assessments is the so-called Probabilistic Production Cost (PPC) model.

In this chapter, our purpose is to contribute to the development of this approach:

- We first face one of the challenges in this context: we propose to use an algorithm that allows extending the classic PPC models design to explicitly represent demand elasticity. The methodology proposed is a simple, robust, closed and consistent solution which clears up the task without complicating the original approach.
- Then, this redesign will serve us to illustrate how, in the presence of demand elasticity, the aforementioned traditional reliability measures are no longer a consistent proxy to estimate the system performance, since as we discuss, they leave aside relevant information. Indeed, in a fully elastic demand context, the LOLP or the ENSE value as traditionally defined, would take a zero value, no matter which the generation availability would be.

Next in the remainder of this introduction, we describe the basic PPC model methodology. Then in section 2 the algorithm to model demand elasticity is developed. Finally, in section 3, we discuss the reasons that lead us to state that reliability measures are not suitable metrics in a context in which demand elasticity is significant. This will lead to the discussion on the ways to better define a proper metric to estimate to what extent the market outcomes are adequate.

[73] Lost of Load Probability (LOLP), Lost of Load Expectancy (LOLE) and Expected Non-Served Energy (ENSE).

1.1. Probabilistic production cost models: a classic tool to measure the reliability of a power system

Probabilistic Production Cost models (PPC models) have been traditionally used as a support tool in the centralized long-term decision-making process in electric power systems. These models are characterized for centering all the computational efforts in representing the random nature of some of the most relevant variables involved in the long-term planning problem (typically the demand values and the forced outage rates of each generating unit). They allow for reliability assessments of real-size electric power systems with little computational effort. However, this is achieved at the cost of making strong simplifications regarding short- and medium-term operational and planning constraints of the generation plants.

This approach has attracted considerable efforts from academia since the late 60's. The basic model corresponds to its application to a non-constrained thermal system; see the pioneering works of (Baleriaux et al, 1967) and (Booth, 1972).

The major outputs that were first calculated using these models were:

- Reliability measures: the loss of load probability, the loss of load expectancy, the expected value of the non served energy, etc.
- Expected production schedules, that is, the expected energy generated by each generating unit.
- Expected production costs.

Although for the purposes of the present chapter, this basic approach will serve as the starting point to illustrate the effect of demand elasticity, is it important to note that there are dozens of papers where several developments have been introduced to the classic approach. For instance, there are remarkable works focused on introducing simplified alternatives to include hydraulic units (Billington & Harrington, 1978), (Finger, 1978), (Ramos et al., 1991) or (Malik, 2004), storage units (Conejo, 1987) or (Invernizzi et al., 1988), time dependent units (Conejo et al., 1985), etc.

One of the most popular extensions to the basic model is the so-called frequency and duration method. This extension introduces information on frequency and durations of the different states (demand interval, outage rates, etc.). This allows calculating additional information, as for instance the mean time existing between two consecutive events (typically scarcities). There are different approaches to introduce this frequency and duration data, see for instance (Ayoub & Patton, 1976) or (Finger, 1979).

There are also different approaches to compute some additional results, for instance in (Leite da Silva et al., 1988) or in (Lee et al., 1990) a means to calculate the underlying variance of the results is provided. A description on how to estimate derivatives (e.g. marginal values), can be found in (Ramos et al, 1994) and also in (Maceira & Pereira, 1996).

These PPC models have been applied to determine the marginal contribution of each generating unit to the regulator's reliability objectives. One of the first works in this respect is the one developed in (Garver, 1966). A more recent work trying to determine this contribution to reliability objectives (in this case, the contribution of wind energy) by means of a PPC model can be found in (Kahn, 2004). This sort of calculations have served for instance to set the remuneration for each generating unit in some real systems in which a capacity payment mechanism had been implemented (this was for example the case of the former Chilean mechanism or the Panamanian case[74]).

However, in the literature the way to address electricity demand elasticity in PPC models is still a pending issue, since no close solutions to deal with this issue can be found.

Malik (2001) aimed at representing the impact of demand-side programs, but the chief objective of these approaches has been to introduce and assess the effect of load shifting programs, which are demand side management programs that seek to move the load consumption from peak hours to off-peak hours. This is carried out by breaking down the load sifting operation into two operations that can be separately modeled within the PPC context by means of limited energy generators. This way the model separates peak clipping (reduction of the consumption on the peak) and valley filling (increasing it on off-peak hours), and models the first operation as an equivalent hydro unit and the second as the process of loading a pump storage unit (for more details, see any of the references provided before on how to perform such operations). However, in this model, the energy to be shifted from peak load to base load is introduced as an exogenous parameter, thus, in rigor no explicit response to prices is modeled.

As stated, our objective is not to model load shifting programs, but introducing explicitly demand response to prices. Here we propose a methodology that solves the problem in a simple and compact way. In order to ease and clarify the conclusions, we will present the proposed new formulation on the basis of PPC a model design (representing just non-energy limited thermal generating plants, modeled through their maximum output and their forced outage rate). As it is latter shown, thanks to the consistency and straightforwardness of the algorithm proposed, the methodology allows further complication of the system representation on exactly the same basis as the traditional approach itself.

1.2. Description of the Basic PPC Model

The basic PPC model is built upon the assumption that all generation plants can produce at full capacity at any time unless when they are out-of-order due to a forced outage. Hourly demand is considered to be inelastic and stochastic.

[74] The Centro Nacional de Despacho (CND) of the Empresa de Transmisión Eléctrica S.A. (ETESA) in Panamá uses the FLOP model (see www.iit.upcomillas.es/aramos/flop.htm), a PPC model to calculate the so-called Firm Capacity according to which generating plants are paid in the context of the capacity mechanism in force.

These models were conceived to check a basic reliability condition: whenever the system's (inelastic) demand exceeds the available generating capacity a loss of load takes place. The probability of such an event happening (the Loss of Load Probability or LOLP) and the corresponding expected non-served energy (ENSE), have been the main reliability results obtained from these PPC models.

In such a context, the loss of load probability distribution can be evaluated by means of the distribution of the difference between two random variables: the demand and the total generation available[75]. This difference is usually evaluated in a generic random hour. Longer term results (e.g. the ENSE in a whole year) are calculated by directly extending the results obtained when computing this generic hour.

If all variables (demand and failure rates in the most simple case) are statistically independent, then the computation of the former difference considerably simplifies, since the probabilistic distribution of the sum (or difference) of two independent random variables is equal to the convolution of their probability distribution functions.

We next present how the demand and the thermal generating units are modeled and the order followed in the convolution operation to simulate the generating units' scheduling. Then we explain how to interpret the results obtained when performing this operation.

1.2.1. Hourly load probability distribution

The probability curve for the electricity demand in a generic random hour should ideally be calculated by means of probabilistic forecasting techniques. However, it is commonly accepted as a well suited proxy to take a large set of historical data and then assign the same probability to each one of the historical realizations of the demand, that is, each realization is supposed to have a probability of $1/n$, being n the number of hourly data considered. This way, the percent of time that a given load level (or a greater than a given load level) occurs in the set of data considered will be interpreted as a probability. Thus, at any given time (hour) there will be a probability of 1 that the load will be higher than the minimum load being considered.

Under the latter assumption, and as illustrated in Figure 14, we can calculate the Load Complementary Distribution Function (LCDF) just by rotating the axes of the load duration curve corresponding to the historical horizon considered, and then normalizing the time period so that the vertical axis gives the percent of time (the probability) that a certain value of demand level is exceeded. This is the reason why the hourly Load Complementary Distribution Function (LCDF) is sometimes referred to as the Inverted Load Duration Curve (ILDC) or just Load Duration Curve (LDC).

[75] If only loss of load is being evaluated, just positive values of such difference would be of interest.

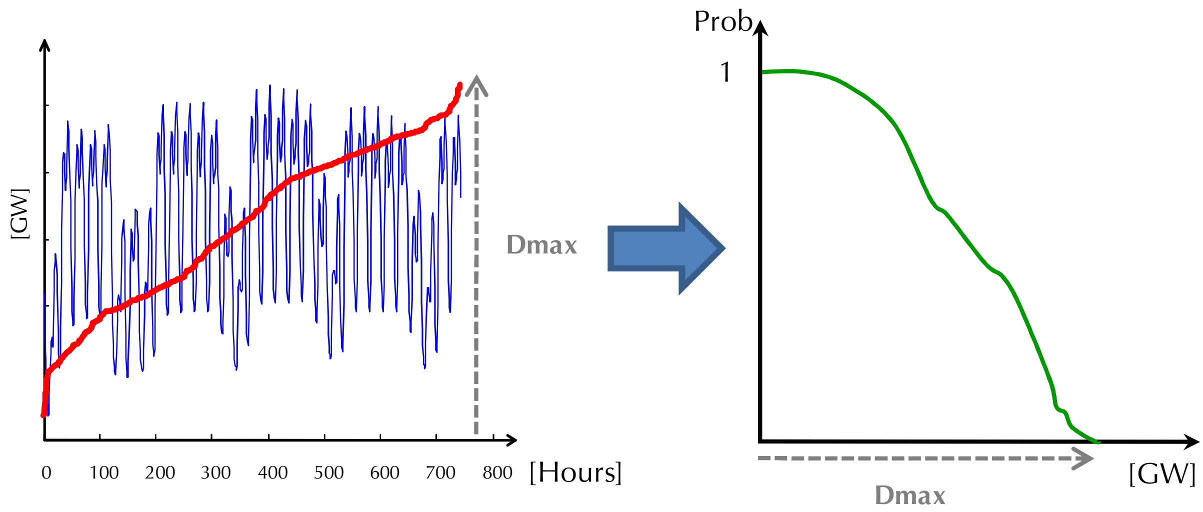


Figure 14. The chronologic demand (blue), the corresponding load duration curve (red) and the estimated load complementary distribution function (green).

However, it is important to bear in mind that the LDCF curve does not represent anymore a demand monotone, but a complementary distribution function of the demand in a generic hour.

1.2.2. Thermal plants modeling

Each generator's available capacity is modeled as an independent discrete random variable. The simplest representation would be the two-state model, where the plant either is able to produce at maximum capacity (probability p) or it cannot produce because of a forced outage (probability $q = 1 - p$). The complementary distribution function of the available capacity under this modeling assumption takes the form represented in Figure 15.

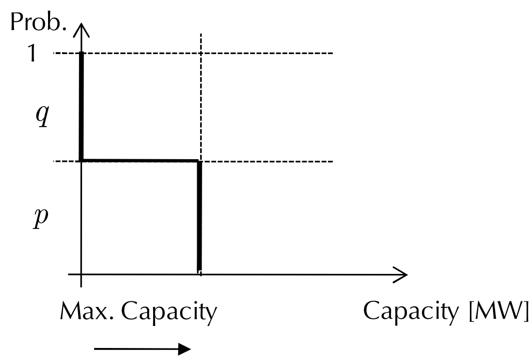


Figure 15. Complementary distribution function of the thermal unit available capacity

1.2.3. Dispatch criteria: the merit order

When operating constraints are not considered, the dispatch that results in the minimum operating cost is the one in which generators are dispatched in order of increasing marginal cost[76]. This ranking of the generators is usually known as the *merit order* or the *loading order*. This way, the convolution operation is performed following this merit order.

1.2.4. The equivalent load and the results provided by the basic model

As described in (Booth, 1972), let us introduce the concept of the "equivalent load after dispatching the first n units", denoted by EqL_n. This EqL_n represents the distribution function of the non-served load after having dispatched the first n generating groups in the merit order. This equivalent load can be computed by carrying out the convolution of the variables involved, as expressed next:

$$EqL_n = L - \sum_n C_n, \qquad n = 0,1,2,3,... \tag{16}$$

The first equivalent curve (n =0) represents the complementary distribution function of the load consumption (L) of the system, when no generator has been dispatched yet. The successive equivalent loads represent the load yet to be covered after dispatching each generator in the system. The last curve, EqL_N, represents the complementary distribution function of the demand left uncovered once all the system generators have been dispatched. Figure 16 illustrates this procedure, where the successive equivalent loads are calculated as a result of the successive probabilistic dispatch of the units.

[76] In practice, based on heuristic algorithms, this merit order can be modified in order to consider approximately some operating constraints (high start-up, on-line or shutdown costs, network transmission constraints, etc.).

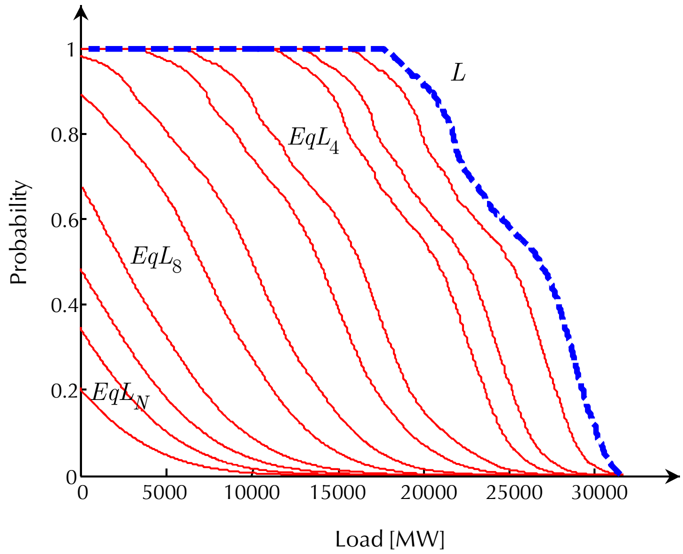


Figure 16. Equivalent load curve

The last equivalent load curve represents the distribution function of the non-served energy. From this last equivalent curve we can extract the two following valuable pieces of information (see Figure 16):

- The loss of load probability (LOLP) is the point where this last curve intercepts the y-axis (probability). It represents the probability that there will still be non-served demand left to be satisfied after all the generators have been dispatched.
- The non-served energy expectation (ENSE) is the area beneath this last curve. It represents the expected amount of energy in MWh that is left unsupplied (in the generic random hour being represented) in the system after all the generators have been dispatched.

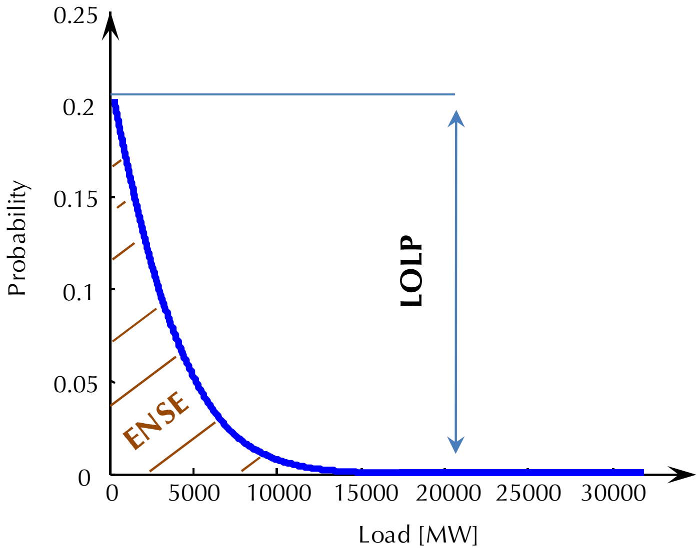


Figure 17. The last equivalent curve: the non-served energy distribution function

2. INTRODUCING DEMAND ELASTICITY IN PPC MODELS

To introduce explicitly demand response to prices in PPC models we opt for modeling elastic demand bids by means of a set of equivalent thermal units, which as we show next, allows us to solve one of the pending issues in the literature regarding PPC modeling, introducing demand price elasticity in an extremely simple and robust way[77].

2.1. Redefining the merit order: modeling demand offer bids as equivalent generators

For the sake of clarity we have opted for illustrating the basic idea of the algorithm proposed making use of both a deterministic demand and a deterministic set of the thermal units being available to produce. Once presented the main idea, it will be straightforward to introduce an analogous reasoning in the PPC methodology.

Let us consider that demand marginal utility (the demand offer curve) is given by the red step-wise curve presented in Figure 18. The available generators' capacities (MW) and their corresponding marginal costs (€/MWh) have been represented as an aggregated step-wise curve (dotted in blue) in the same chart.

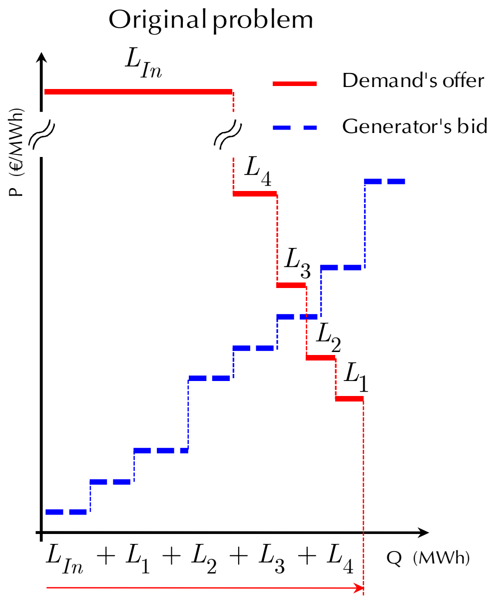


Figure 18. Demand and generation offer curves

Note that in the demand representation there are two differentiated parts:

[77] This solution has been previously used in other modelling approaches, as for instance in the context of a long-term deterministic simulation tool representing market agents' strategic behavior, see (Batlle & Barquín, 2005), as well as in former market equilibrium models designed by Prof. Barquín.

- The inelastic load consumption, denoted here by L_{In}. The associated offer price representing this portion of the consumption is assumed to be much higher than the variable cost of any of the generators. This price should ideally represent the value of loss of load (VOLL), also known as the Non-Served Energy Cost (NSEC).
- The offers corresponding to elastic demand consumption; each step of the elastic fragment has been denoted by L_i, where i represents the elastic offer index, offered at a price, p_i^L.

The algorithm proposed consists in solving an equivalent problem (see Figure 19), in which the elastic demand is substituted by:

- A new fictitious inelastic demand which is equal to the sum of quantities of all the consumption of the original demand curve (i.e. including both the inelastic and elastic consumption). This new fictitious inelastic demand, has been denoted in Figure 19 by L_{In}^F.
- A set of fictitious generators, each one representing an offer step from the original demand curve. This way, each L_i, becomes a fictitious generator G_{Li}^F The corresponding quantity (MWh) and price (€/MWh) are those defining the original demand offer.

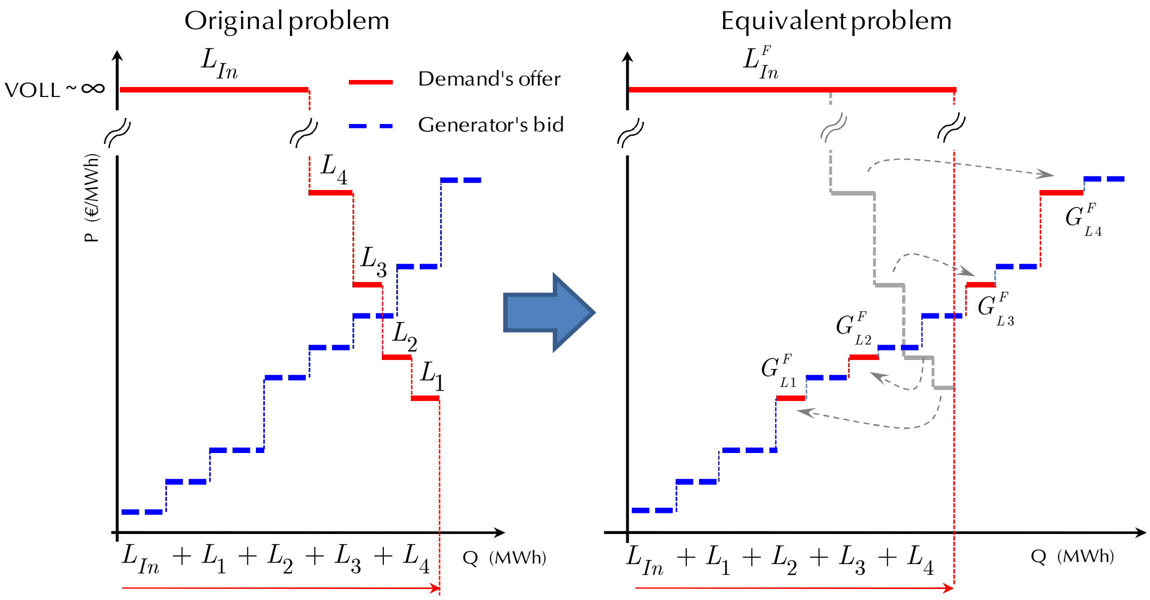


Figure 19. The equivalent model formulation

In the figure it can also be checked how the market outcome remains unchanged, since the resulting price, and the committed generators are exactly the same in both cases.

Note that in the equivalent problem, the production of the fictitious generators corresponds to those blocks of the demand which were not supplied, that is, it corresponds to the energy that was not purchased because the offer price was below the resulting market price[78].

2.2. The PPC model with an elastic demand

The aforementioned methodology allows including demand elasticity in the PPC framework by performing well-known operations (the dispatch of thermal units). As described, the algorithm consists in solving an equivalent problem, where the elastic demand is substituted by an inelastic demand and a set of fictitious thermal units.

In the following, for the sake of simplicity we consider a unitary value for all the availability rates of the fictitious thermal generators. However, note that by means of the availability failure rate of the fictitious thermal units we can represent stochastic demand elasticity (exactly in the same way that the thermal units are modeled).

As it has just been pointed out, the production of the fictitious generators corresponds to demand consumption not committed at the corresponding price. This production can be calculated by means of the resulting equivalent load before dispatching each one of the fictitious unit.

In Figure 20 it is shown how this production can be calculated using the equivalent load after dispatching the previous units in the merit order (i.e. EqL_{n-1}), and the capacity of the fictitious generator. Assuming that the failure rate is zero, the dotted line (in red) represents the complementary distribution function of the production of the fictitious unit. The dashed area (in each of the charts included in the figure), represents the expected production (i.e. the expected non-purchased energy at that price). In the figures it has also been represented the probability of the corresponding demand block not being committed, what will be termed here as non-purchased energy probability (NPEP)[79].

[78] In other words, this is a way to represent that when the market price rises up to the level of an elastic segment of demand, this energy is retired. This is modeled by artificially committing a virtual generator that covers with its fictitious production this "hole".

[79] Note that this non-purchased energy probability is nothing but what is usually known as the $LOLP_{n-1}$, that is, the resulting LOLP after having dispatched al the preceding units in the merit order.

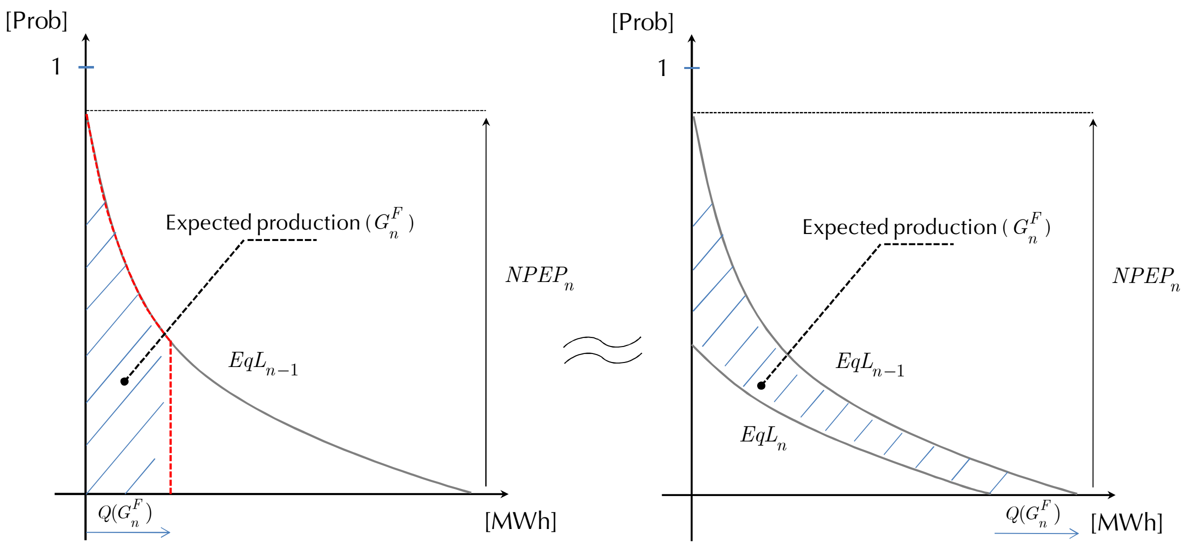
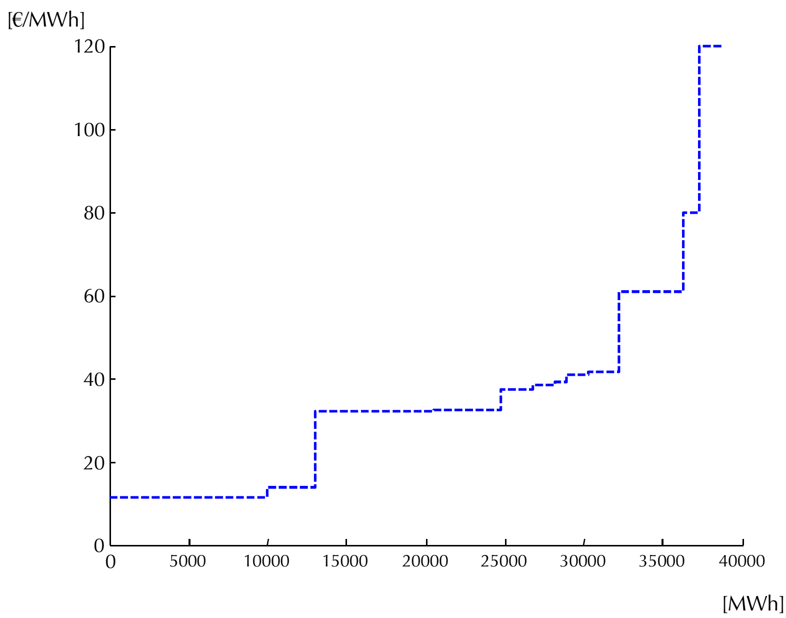


Figure 20. Expected production of the fictitious generator

2.3. Numerical case example

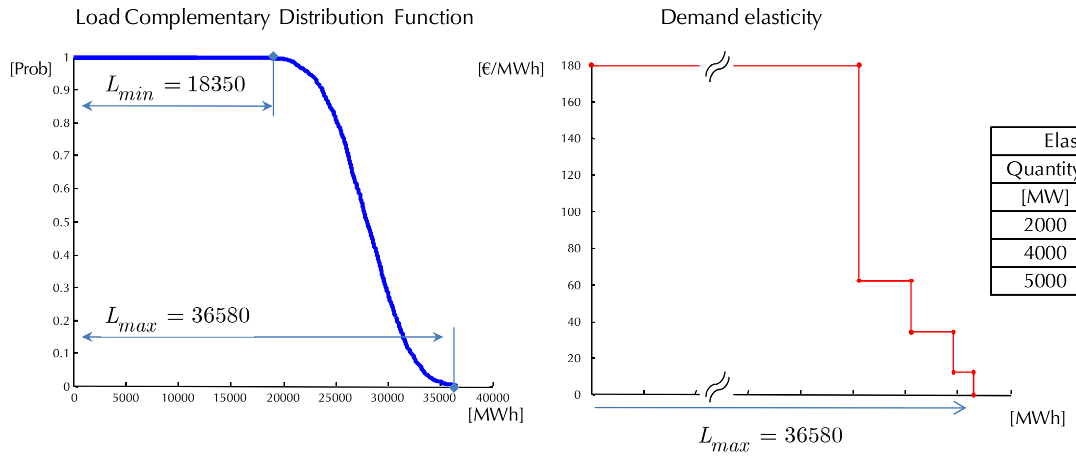
In order to better illustrate the procedure, we resort to a case example. In Figure 21 we have represented the thermal plants that are considered in the analysis as well as its characteristics (installed capacity, availability and variable costs).



Installed capacity	Availability	Variable costs
[MW]	[p.u.]	[€/MWh]
5000	0,95	11,50
5000	0,95	11,50
3000	0,93	14,00
2500	0,90	32,00
2500	0,90	32,00
2500	0,50	32,00
1500	0,90	32,29
1500	0,93	32,29
1250	0,93	32,29
1000	0,92	37,41
1000	0,92	37,41
700	0,92	38,45
700	0,92	38,45
700	0,92	39,05
700	0,92	41,00
700	0,92	41,00
2000	0,92	41,42
1000	0,78	60,85
1000	0,78	60,85
1000	0,78	61,00
1000	0,78	61,00
1000	0,78	80,00
1500	0,78	120,00

Figure 21. Thermal plants considered in the case study

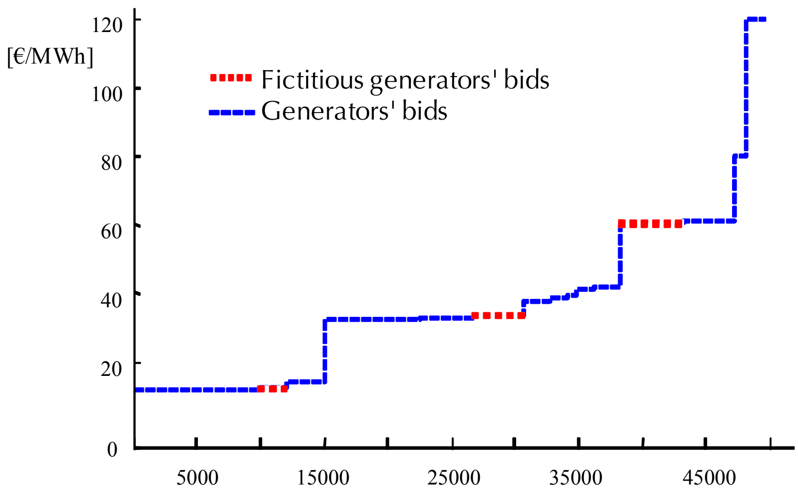
Next, in Figure 22 we present the Load Complementary Distribution Function as well as the demand elasticity[80]. The minimum demand value (including both the inelastic and elastic consumption) is equal to 18350 MWh, and the maximum 36580 MWh.



Elastic offers	
Quantity	Price
[MW]	[€/MWh]
2000	12
4000	31
5000	62

Figure 22. LCDF and demand elasticity function

We introduce the demand elastic offers in the model by means of the fictitious generators. The resulting generating mix considered, including these fictitious units (in red), have been represented in Figure 23.



Installed capacity	Availability	Variable costs
[MW]	p.u.	[€/MWh]
5000	0,95	11,50
5000	0,95	11,50
2000	**1,00**	**12,00**
3000	0,93	14,00
2500	0,90	32,00
2500	0,90	32,00
2500	0,50	32,00
1500	0,90	32,29
1500	0,93	32,29
1250	0,93	32,29
4000	**1,00**	**31,00**
1000	0,92	37,41
1000	0,92	37,41
700	0,92	38,45
700	0,92	38,45
700	0,92	39,05
700	0,92	41,00
700	0,92	41,00
2000	0,92	41,42
5000	**1,00**	**62,00**
1000	0,78	60,85
1000	0,78	60,85
1000	0,78	61,00
1000	0,78	61,00
1000	0,78	80,00
1500	0,78	120,00

Figure 23. The equivalent thermal unit problem

[80] Note that in order to simplify the algorithm understanding, implicitly we have assumed that demand elasticity remains unchanged during all the yearly blocks, i.e. roughly speaking it is the same in the off-peak hours than in the peak ones.

Results

In Figure 24 it has been represented the successive results obtained from dispatching probabilistically each one of generators considered in the equivalent problem. The areas dashed in red correspond to the dispatch of the fictitious generators.

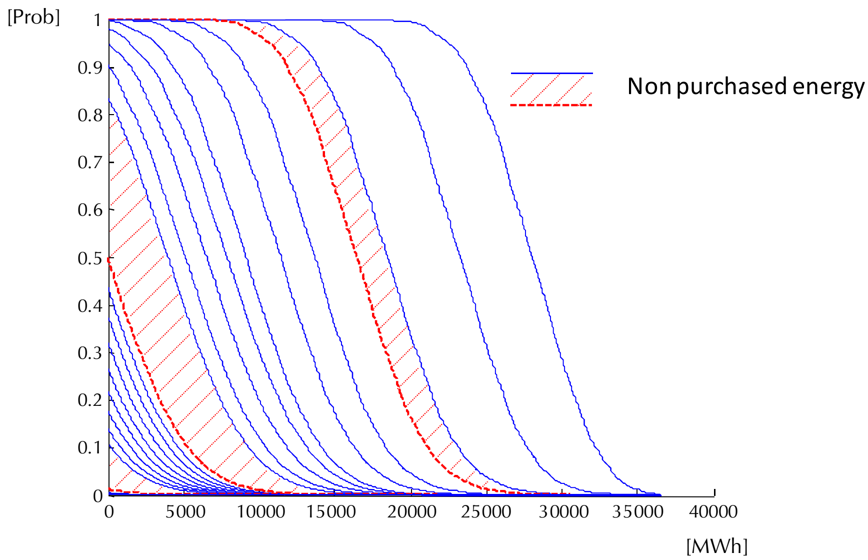


Figure 24. Case example results

Following the procedure described in Figure 20, we can calculate the probability that each one of the offers does not result accepted for falling below the market price. We have termed this probability associated to each offer as the Non-Purchased Energy Probability (NPEP). It is also possible to calculate the Expected Non-Purchased Energy (ENPE) corresponding to each one of the demand offers.

In the following table we have gathered these results.

Elastic offers		Results	
Quantity	Price	NPEP	ENPE
[MWh]	[€/MWh]	[p.u.]	[MWh]
2000	12	1.00	2000
4000	31	0.83	2718
5000	62	0.06	102

This way, the first block is never dispatched, since the NPEP value is 1, and the expected value of the non purchased energy (ENPE) is equal to the quantity being offered. On the other extreme, the block offered at a highest price, is not dispatched 6% of the time, and the expected energy not being committed is 102 MWh.

The classic reliability measures take the following values:

LOLP	1,8E-04
ENSE	0,2325

3. NON-SERVED ENERGY AND NON-PURCHASED ENERGY

In real markets, the regulator is particularly concerned about guaranteeing the electricity supply for the inelastic demand. This happens for several reasons, among which we can highlight the following:

- the inelastic consumption represents a large percentage of overall energy consumption,
- the underlying marginal demand utility (ideally, the demand's offer price) is considered to be much higher than the one corresponding to the elastic demand,

In Figure 25 we have a real example of both the system demand and supply curves in the Spanish spot market. It can be clearly observed the two types of consumption already mentioned (i.e. the inelastic L_{In} and the elastic consumption L_{El})[81].

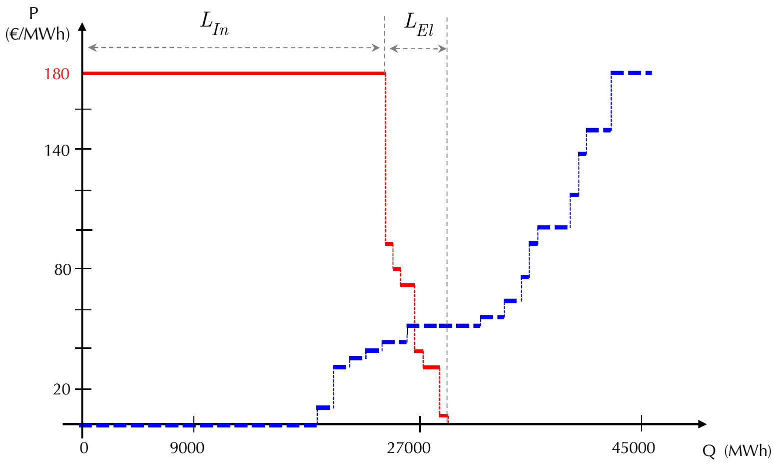


Figure 25. Demand and offer curves in the Spanish spot market

For the three reasons just mentioned, traditionally the regulator has defined its security of supply objectives in terms of the system capability to supply the inelastic consumption. This capability to provide this "critical" fraction of the demand is what is usually analyzed by the different so-called reliability measures. This way, the LOLP represents the probability of not being able to supply this inelastic consumption and the NSE the Non-Served Energy corresponding to this inelastic consumption.

[81] As pointed out in Chapter I, section 5.3, in many systems the regulator has imposed price limits avoiding the system to reflect the true marginal demand utility. This is the case in Spain, where a limit of 180€/MWh has been in force since the electricity market was introduced.

Therefore, the non-served consumption corresponding to the elastic demand blocks not being committed (which will be denoted in the following as non-purchased energy or NPE), does not intervene in the reliability measures.

To illustrate this, we have taken the previous example representing a tighter generation availability scenario (just three thermal groups are now available). In this new case we have represented the non-served energy (NSE) and the non-purchased energy (NPE) in both the real market situation and in the equivalent problem formulation.

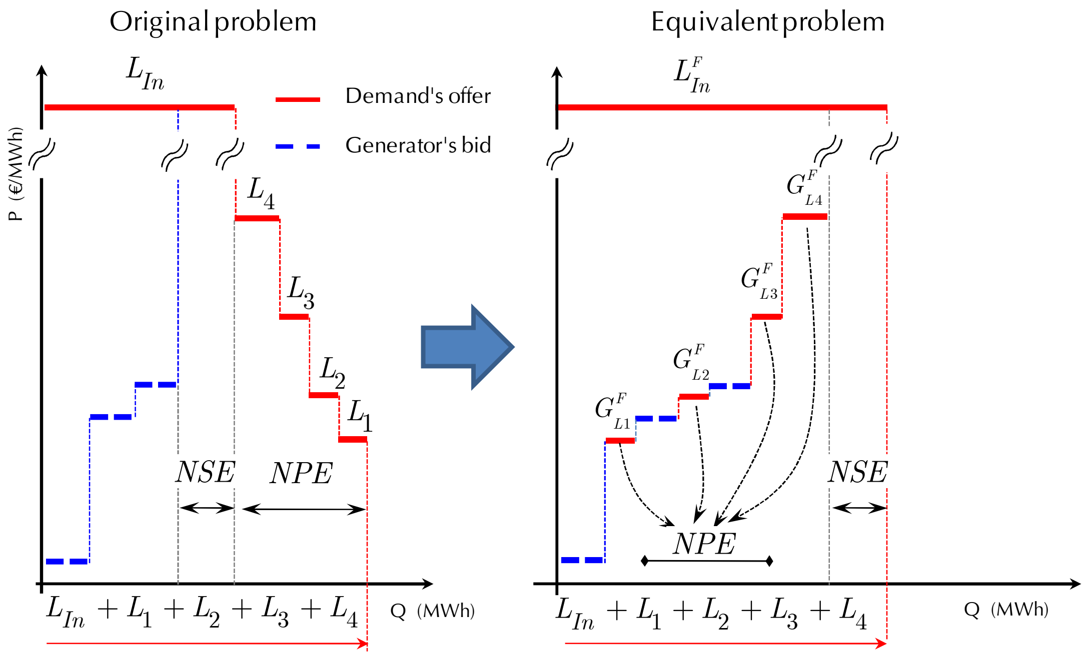


Figure 26. The Non Served Energy (NSE) and the Non Purchased Energy (NPE)

3.1.1. Reliability measures in the presence of demand elasticity

We next show the fact that, when assessing the system performance by means of the reliability measures, since we are just taking into account the inelastic consumption, we are leaving aside relevant information. This has been illustrated in the following figures, which serve us as the basis to give rise to the discussion about the weakness of these reliability measures in the presence of an elastic demand.

Three different scenarios of demand have been considered: a fully inelastic scenario, a partially elastic scenario and a fully elastic scenario. Each of these demand scenarios have been confronted with three deterministic scenarios of generation availability, one presenting a sufficient reserve margin, one in which several groups are unavailable and also one representing a severe scarcity (just a few groups are available). This is what we can observe:

- In the case of the inelastic demand, see Figure 27, there are two generation availability scenarios which lead to non-served energy.

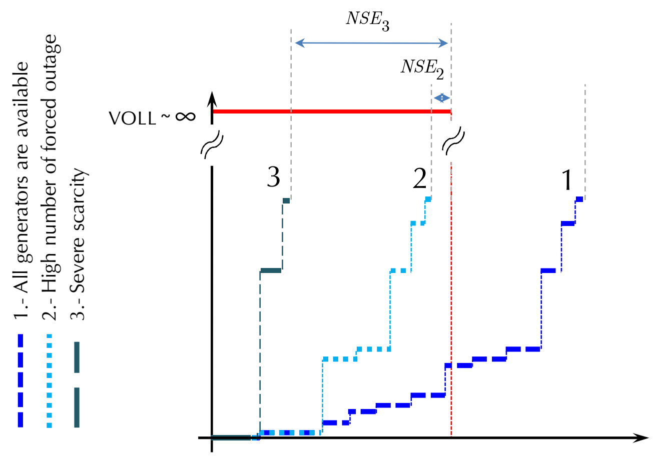


Figure 27. Fully inelastic demand scenario

- In the case of the partial elastic demand, see Figure 28, there is only non-served energy in the scenario presenting the tightest reserve margin.

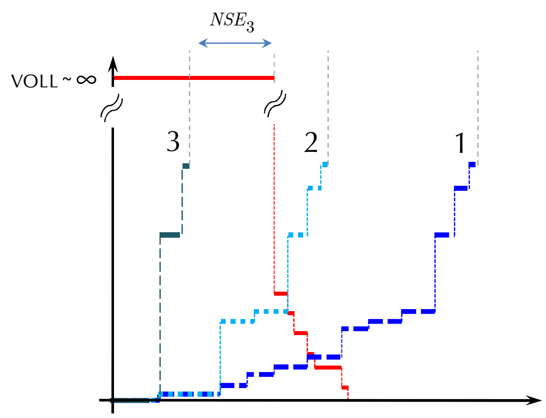


Figure 28. Partially elastic demand scenario

- In the case of the fully elastic demand scenario, see Figure 29, resorting to the classic definition of the aforementioned reliability measures, the non-served energy is equal to zero.

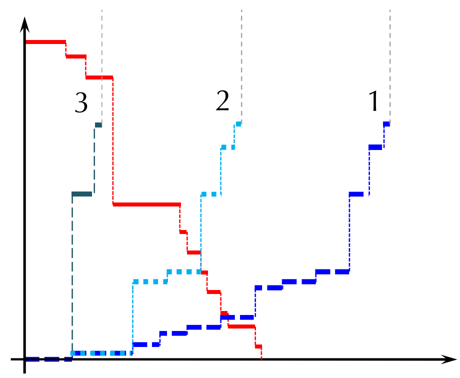


Figure 29. Fully elastic demand scenario

Does this mean that in a context where demand is fully elastic, reliability is not longer a problem? Resorting to the definition provided previously the answer is that effectively, reliability as classically defined is not longer a problem.

However, this does not mean that security of supply as defined in the present work is not longer a problem. Note that in this fully elastic demand case, a permanent scenario of severe scarcity will clearly be a problem from the regulator perspective. As previously discussed in Chapter III, the objective of the regulator is to maximize the net social benefit, and this implies evaluating how far the outcomes provided by the market are from an ideal benchmark (the most efficient investments, resource management, scheduling, etc.).

Indeed, comparing the market results with a benchmark is exactly what is done when the regulator seeks to achieve a reliability standard (like the one day in ten years). These standards are supposed to represent the optimal benchmark, and this benchmark is supposed to take into account investment costs, the value of loss of load, etc.

3.1.2. The metric to evaluate market performance

In order to take into account the elastic consumption in the performance measure, we propose to evaluate the distribution function of the total value of what we have termed the Non-Purchased Energy.

This Non-Purchased Energy Distribution Function would have to be compared with that of resulting from the benchmark. This comparison would provide a more precise idea of how well the market is performing its "job" than the mere reliability criteria. The comparison can be performed in terms of the expected value, the VaR, the CVaR, or in general, in terms of any measure of a distribution function. In Figure 30 in has been illustrated how this comparison could be performed in a generic case (the measure represented in the figure is the CVaR).

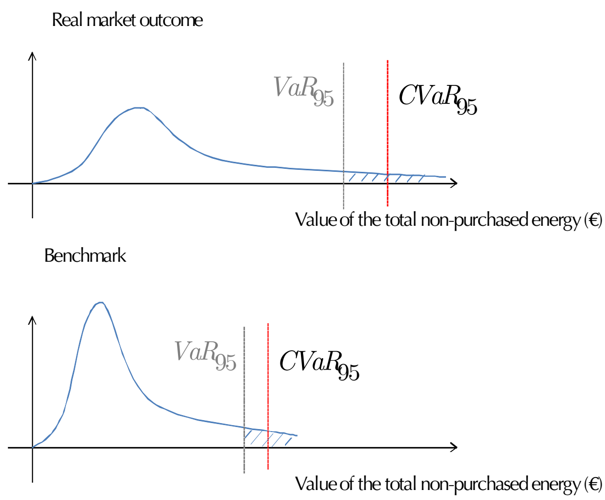


Figure 30. Measuring the market performance comparing with the benchmark solution

Note that the LOLP would no longer make sense in the fully elastic demand scenario. In this new context it can rather be calculated the probability of not supplying a certain demand block.

3.1.3. Numerical case example

The expected value[82] of the previous case example can be computed taking into account for each offer, the expected non-purchased energy and the corresponding price (we have considered a value of VOLL equal to 10000 €/MWh).

The expected value of the total non-purchased energy (ENPE):

$$EVNPE = ENSE{\cdot}VOLL + \sum_i ENPE_i{\cdot}p_i^L =$$
$$= 0,232 \cdot 10000 + 2000 \cdot 12 + 2718 \cdot 33 + 102 \cdot 60 = 122139\ €$$

This value should be compared with that of resulting from the benchmark scenario.

4. CONCLUSIONS

We have developed the Probabilistic Production Cost (PPC) model approach by extending the classic formulation of a novel algorithm that allows introducing demand elasticity in a simple and closed way.

Then, we have taken advantage of this model formulation to show how the traditional reliability measures, such as the LOLP or the ENSE, as a measure to assess the level of security of supply should be reconsidered in the presence of significant demand elasticity. This has led us to propose a better way to define a proper metric to estimate if the market results comply with the regulator's standards.

Acknowledgements

The author wants to thank Luiz Barroso and Andrés Ramos for their valuable comments and suggestions, which helped to improve the work developed in this chapter.

5. REFERENCES

Ayoub, A. K. & A. D. Patton (1976). "A Frequency and Duration Method for Generating System Reliability Evaluation". IEEE Transactions on Power Apparatus and Systems, vol. 95, iss. 6, part 1, pp. 1929-1933, 1976.

[82] In our case example, the distribution function of the non-purchased energy value could also be easily computed taking into account that if any energy offer is not committed, then any energy offer presenting a lower price will also be left uncommitted. This is true since we are considering that demand elastic offers have a failure rate equal to zero.

Baleriaux, H., E. Jamoulle and F. Linard de Guertechin (1967). "*Simulation de l'explotation d'un parc de machines thermiques de production d'electricité couplé à des stations de pompage*," Revue E, Vol. V, No. 7, pp. 225-245, 1967.

Batlle, C. & J. Barquín (2005). "A strategic production cost model for electricity market risk analysis", IEEE Transactions on Power Systems, vol. 20, iss. 1, pp. 67-74, Feb. 2005.

Bloom, J. A. (1984). "Generation Cost Curves Including Energy Storage". IEEE Transaction on Power Apparatus and Systems. Vol. PAS-103. No. 7, pp 1725-1731. July 1984.

Booth, R. R. (1962). "Power System Simulation Model Based on Probability Analysis," IEEE Trasaction on Power Appartus and Systems, PAS-91, January/February 1972, pp. 62-69.

Conejo, A. J. (1987) "Optimal Utilization of Electricity Storage Reservoirs: Efficient Algorithms Embedded in Probabilistic Production Costing Models". M. S. Thesis. Massachusetts Institute of Technology. August 1987.

Conejo, A. J., I. J. Pérez-Arriaga, A. Ramos and A. Santamaría (1985). "Evaluation of the Impact of Solar Thermal Generation on the Reliability and Economics of an Electrical Utility System", IEEE Mediterranean Electrotechnical Conference. MELECON '85 4: 167-173 A. Luque, A.R. Figueiras Vidal, D. Nobili (eds.) Elsevier Science Publishers B.V. Madrid, Spain October 1985

Finger, S. (1979) "Electric power system production costing and reliability analysis including hydro-electric, storage, and time dependent power plants". MIT Energy Laboratory Technical Report #MIT-EL-79-006, February 1979

Garver, L. L. (1966). "Effective Load Carrying Capability of Generating Units" IEEE Transactions on Power Apparatus and Systems, Vol. PAS-85, no. 8, August 1966, pp. 910-919.

Halperin, H. & H. A. Adler (1958). "Determination of reserve generating capability". Trans. AIEE, vol. 77, pt. III, pp. 530-544, August 1958.

Invernizzi, A., G. Manzoni and A. Rivoiro (1988) "Probabilistic simulation of generating system operation including seasonal hydro reservoirs and pumped-storage plants." Electrical Power and Energy Systems. Vol. 10, no. 1, pp. 25-35. 1988.

Kahn, E.P., (2004). "Effective Load Carrying Capability of Wind Generation: Initial Results with Public Data", Electricity Journal, December 2004.

Lee, F. N., M., Lin, and A. M., Breipohl (1990). "Evaluation of the variance of production cost using a stochastic outage capacity state model", IEEE Transactions on Power Systems, Vol. 5., No. 4, November 1990.

Leite da Silva, A.M., F. A. F. Pazo Blanco, and J. Coelho (1988). "Discrete Convolution in Generating Capacity Reliability Evaluation- LOLE Calculations and Uncertainty Aspects". Paper 88 WM 186-9. IEEE PES Winter Meeting New York, USA. January 1988.

Maceira, M. E. P & M-V. F. Pereira (1996). "Analytical modeling of chronological reservoir operation in probabilistic production costing [of hydrothermal power systems]", IEEE Transactions on Power System, Vol. 11, No. 1, February 1996

Malik, A. S. (2001). "Modelling and economic analysis of DSM programs in generation planning", International Journal of Electrical Power & Energy Systems, Volume 23, Issue 5, June 2001.

Malik, A. S. (2004). "Simulating limited energy units within the framework of ELDC and FD methods", International Journal of Electrical Power & Energy Systems, Volume 26, Issue 8, October 2004.

Ramos, A. & J. Arrojo (1991). Storage Plants Energy Optimization in Probabilistic Production Cost Models. IIT Working Paper IIT-91-008. January 1991.

Ramos, A., L. Muñoz-Moro & I. J. Pérez-Arriaga (1994). "Direct computation of derivatives in production cost models based on probabilistic simulation". Conference on Probabilistic Methods Applications in Power Systems, PMAPS-94. pp. 355-360. Rio de Janeiro, Brazil. September 1994.

Ringlee, R. J. & A. J. Wood (1969). "Frequency and Duration Methods for Power System Reliability Calculations. Part II. Demand Model and Capacity Reserve Model". IEEE Transactions, PAS-38, No. 4, April 1969.

Zahavi, J. (1985). "Probabilistic Simulation Incorporating Single and Multiple Hydroelectric Units with Stochastic Energy Availabilities". Electrical Power and Energy Systems. Vo. 7. No 4. pp 229-232. October 1985.

Chapter VII

Chapter VII. A System Dynamics Model to Analyze Generation Expansion in the Context of a SoS Mechanism Based on Long-Term Auctions

1. INTRODUCTION

Since the outset of electricity markets in the nineteen eighties, one of the key questions posed has been whether the market can ensure satisfactory security of supply from the standpoint of power generation.

When an additional regulatory mechanism is found to be necessary, the question posed is how to most suitably tackle the problem. No international consensus has been reached in this regard, but in almost every electricity market today (see Chapter IV), the regulator has designed some manner of rule to drive or delimit natural market developments in an attempt to guarantee short-, medium- and long-term supply.

Although designing a security-of-supply-oriented mechanism is no easy task, we have seen how lessons learned from international experience have helped to narrow the alternatives a regulator should consider. Broadly speaking, these mechanisms should provide the incentives the market is not providing to ensure security of supply. This translates into an extra income and/or hedge instruments for generators in exchange for an asset/instrument/product aimed to enhance security of supply. We have also seen that some mechanisms have implemented solutions to provide generators with a (not always appropriate) extra income (e.g. the former short-term capacity markets in the northeastern United States or the capacity payments in many Latin American electricity markets), but for a number of reasons[83] they have failed to provide the hedge instruments required by generators to efficiently guarantee security of supply.

Generally speaking, security of supply mechanisms have yielded suitable results when they have been adequately designed. Some of the more successful mechanisms in this respect are those based on long-term auctions.

Long-term auctions as mechanisms to ensure security of supply

In the auction-based approach, specifically, the market authority calls a public auction to buy (or orders the Load Serving Entities to buy), on behalf of demand all (or at least acquire a partial hedge over) the expected consumption for a number of years. This (reliability) product may adopt any of several forms, ranging from a fixed payment in exchange for guaranteed production in times of scarcity, e.g. the so-called reliability option, see (Vázquez, 2002), to a more "plain vanilla" future power supply contract (i.e., a sort-of contract-for-differences).

[83] Due to regulatory uncertainty, for instance, most capacity payments are not viewed as a long-term hedge.

Main elements of auctioned reliability product

As we discussed in PART C, the design the reliability product critically conditions the final outcomes of the mechanism. Two key elements can be highlighted in this regard: time terms and the measure underlying asset (e.g. energy, capacity, etc.).

- First, since the objective is to provide investors with a sufficient hedge to implement their projects, a suitable definition of the time terms is of vital importance:
  - "lag period", i.e. the time lapsing from the date the deliverability commitment is signed to the actual delivery date. Allowing the investors sufficient time to build the generating unit after winning the auction (e.g., four years) is a key element.
  - contract duration, i.e. the duration in years of the commitment stemming from the auction (e.g. ten years).
- Regulators must also define a product based on an underlying asset that represents a measure of the actual contribution to the security of supply (see Chapter III). Two very simple (and simplistic) criteria would be "installed capacity" [MW] or monthly energy production [MWh/month]. This definition, however, is usually based on a measure of the availability of units during critical periods when the likelihood of scarcity is highest. Hence the appearance of the concept "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).

These long-term auction-based mechanisms are intended to largely guarantee the investments required to maintain an adequate and stable generation reserve margin. Agents' bids, along with the quantity of "firm supply" defined by the regulator to be necessary to ensure future supply, determine the price of the reliability contract, i.e. the hedge that allows investors to make project financing feasible.

As noted above, mechanisms of this nature are already in place in a significant number of electricity systems, namely in Brazil, Colombia, Chile and Peru in Latin America and New England and PJM in the US (see Chapter IV for further details)[84].

1.1. The need for models to analyze long-term security of supply auctions

In this context, the ability to analyze long-term investment dynamics is a key issue not only for market agents, but also for regulators. For the latter, the ability to suitably evaluate the dynamics driving system evolution may facilitate the mechanism design process. Simulation tools able to jointly address financial criteria and economic dispatching would provide support for such analysis.

---

[84] No European market has yet implemented a solution of this nature, but Directive 2005/89/EC explicitly envisages resort to them by regulators if needed ('the possibility of imposing public service obligations on electricity undertakings, *inter alia*, in relation to security of supply').

From the market agents' perspective, the need for tools to support decision-making in connection with these auction mechanisms is evident. The same simulation models can also be used in this case to analyze the suitability of investment in a new generating facility, as well as to help define the most appropriate auction strategies.

In this chapter we described a model developed to serve this purpose. General methodology has been designed to analyze these long-term auction mechanisms in the formats presently in place. This model has been designed not as a forecasting tool but instead as a tool aimed to gain insight into the possible market outcomes in a given context. The Colombian system is used as a full-scale example to illustrate model performance

The model draws from system dynamics methodology (Sterman, 2000). System dynamics is one of the techniques that has been shown to be capable of modeling the long-term dynamics of liberalized systems, particularly in terms of how new generation capacity enters the market. A number of approaches have been proposed for accommodating this theory to the long-term analysis of electricity systems, see for instance (Ford, 1997), (Ford, 2002) or (Bunn and Larsen, 1997). In particular, system dynamics-based models have proven to be useful in identifying and characterizing the so-called boom-and-bust investment cycles that tend to characterize generating investment decisions in the absence of any additional adequacy regulatory mechanism. They have also been used to analyze the effect of regulatory long-term security of supply mechanisms (Bunn and Oliveira, 2001) and (Park *et al.*, 2007). A detailed description of the state of the art of these models can be found in (Sánchez, 2009).

This chapter has been structured as follows:

Section 2 contains a description of the main features of the model, which is inspired by the guidelines and basic structure of the approach set out by (Sánchez, 2009). Certain additional refinements of key importance have been developed in order to better deal with some of the complexities that exist in real markets. In this regard, for instance, a multi-scenario analysis tool and an approach to assess new hydro plants investments have been introduced in the standard model. Another relevant refinement has been to introduce how each new investment modifies the financial structure of the company. This way, within the same auction it is considered the increasing costs of issuing new debt.

In section 3, we illustrate the potential of the methodology with a real-size case study. This exercise analyzes the long-term results obtained when the model is used to simulate long-term generation expansion in the Colombian system, in the presence of Reliability Charge Auctions. The reason for choosing this particular case is that it is the widest ranging example available, for it combines a market structure in which several technologies compete to enter the system (hydro, coal, fuel or wind) with the most sophisticated security of supply mechanism design (aimed to deal with such technological diversity).

2. METHODOLOGY

2.1. Model overview

As previously mentioned, the aim of long-term auction-based mechanisms is to maintain investment at the level required to ensure an adequate and stable generation reserve margin.

The *boom-and-bust* cycles that characterize generating investment decisions in the absence of any additional adequacy regulatory mechanism are therefore expected to disappear when such mechanisms are in place. By reducing the frequency of scarcity episodes, they also purpose to narrow price volatility. As we show this does not eliminate uncertainty altogether, by any means.

In this type of regulatory mechanisms market agents must still indispensably be able to estimate future market conditions in the presence of an auction-based security of supply mechanism. The ultimate objective for the regulator is to be able to anticipate the future structure of the expansion portfolio and predict system evolution. For market agents, in turn, the aim is to assess the profitability of their potential investments. It is in response to this need that forecasting tools can play an important role.

In light of the foregoing, a methodology has been developed in this study. Inspired by system dynamics, it also draws from a short-term game-theory model. By simulating electricity market agents' interaction in the spot market, this latter model reflects their strategies as well as supply function building and market clearing and predicts future market conditions (prices and output).

Figure 31 shows the conceptual overview of the model. It consists of three distinct but inter-connected modules.

- The first, which concerns regulator decision-making, serves to assess the need to call an auction for new reliability contracts with generators to hedge future demand requirements. This module also determines the demand curve, i.e. the price-quantity curve calculated by the regulator on behalf of demand to reflect the demand utility function (the maximum value/price of the contracts to be signed).

  The regulator may limit the quantity of the reliability product that each unit is entitled to commit in the auction (see the discussion on the firm supply in Chapter V).

- The second module calculates the generating companies (or GenCos) bid curves to be presented in the long-term auction. This second module constitutes the core of the model and, as shown in the figure, embeds the game-theory spot market model used to simulate the short-term market (i.e. the resulting spot prices and each unit's future production).

- Finally, the third module calculates long-term auction matching and updates all the relevant variables (new generating units entering the system as a result of the auction, new debt issued by GenCos, etc.).

These modules are run sequentially on a year-by-year basis. Consequently, the cyclical process simulated includes determining demand requirements, assessing potential new investments, calculating the respective bids in the auction, and performing the clearing operation.

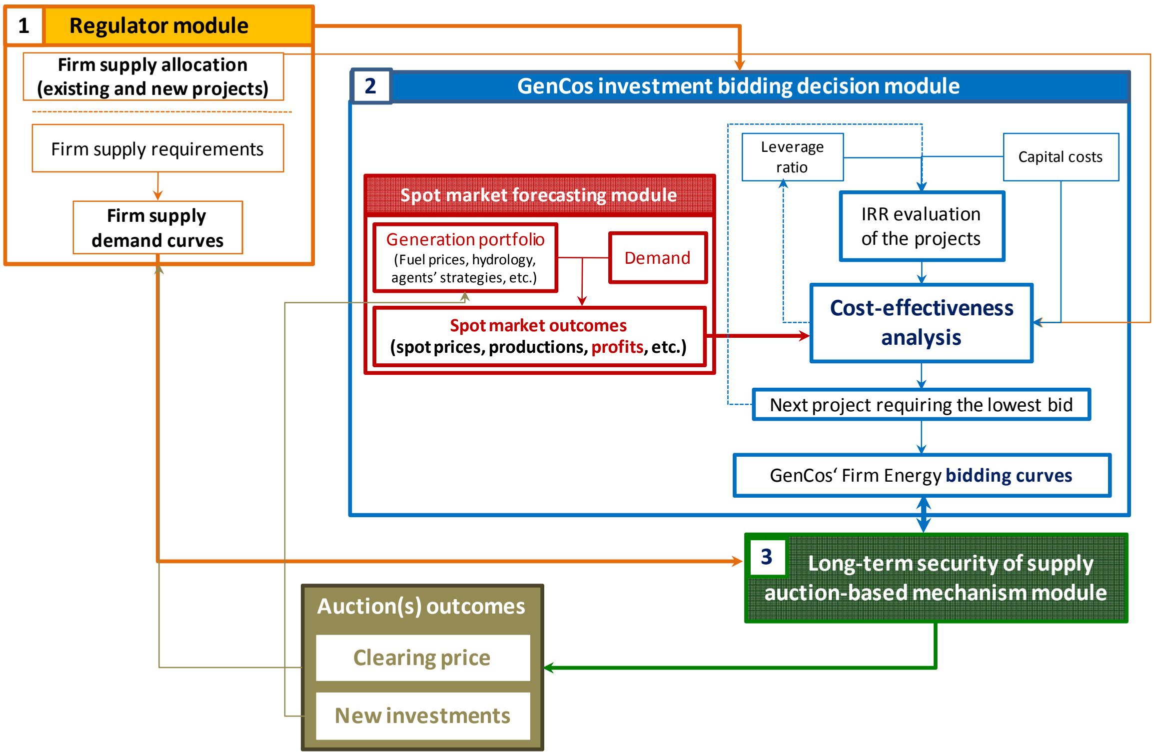


Figure 31. Model diagram

These three components are discussed in greater detail below.

2.2. The regulator module

2.2.1. The decision to call an auction

The model performs this assessment annually (as most regulators do in real life) bearing in mind the lag period defined in the contracts that are being offered to generators in the mechanism. The regulator calls a long-term auction to acquire the reliability product contracts needed (for instance: energy swaps, financial options of energy delivery or any other long-term instrument) to hedge the future supply of electricity. These auctions only make sense in the event of a negative imbalance between the quantity covered under reliability contracts in effect (from former auctions) and expected future requirements. In other words, if at any given time the regulator deems that the market, left to its own devices, is able to provide sufficient generation availability when needed, no auction is required.

2.2.2. Firm supply allocation criteria

The firm supply assigned to each plant is the amount of the reliability contract[85] that the regulator regards each unit to be capable of providing. This entails evaluating how the unit may behave under scarcity conditions, this can be carried out based on historical records or estimates of future values.

In the Peruvian system, for instance, the quantity the regulator allows hydroelectric units to commit corresponds to the 95 percentile of total unit output during critical periods (assuming that the unit operates at full capacity seven hours a day every day).

2.2.3. Reliability product demand curve

After estimating future needs, i.e. the quantity to be purchased in the next auction, the regulator proceeds to calculate the demand curve.

In some systems (Colombia, NE, PJM, etc.) this curve is expressed as a piecewise linear function. The simplest demand curve consists in merely determining a fixed quantity requirements (usually with the so-called reserve price, such as in the long-term auctions held in the Peruvian system).

2.3. Investment bidding decision module

The five main assumptions used to simulate the generator bid-building process are:

- First, from the generators' perspective, only two sources of income are considered: the spot market and the long-term security of supply auction-based mechanism. Other possible sources of income such as markets for ancillary services are excluded.
- Second, only two sources of direct costs are taken in to account[86]: variable and investment costs. In other words, this disregards many other financial planning costs such as tax exposure and other cash flow drivers.
- Third, in the simulation new generating plants are implicitly regarded to be able to enter the system only after winning a long-term auction. Although under certain conditions a plant may recover both its operating and investment costs with spot market revenues only, it seems implausible that investors would waive the stable additional payment entailed in the security of supply mechanism. Assuming that the possibility of calling these auctions is considered annually, this is not regarded to be an over-simplification, inasmuch as

[85] For instance, energy produced in an hour in which the spot price exceeds a certain threshold, as in NE-ISO, or in a full day in the Colombian system.

[86] The penalization envisaged for non-compliance (e.g., failure to produce when so provided in the reliability contract) is an indirect cost that has been taken into account when calculating generators' bids in the auction

generators would be unlikely to enter the market with no intention to collect the reliability charge.

- Fourth, generators' bids are calculated to obtain the so-called break-even point. In other words, bids are computed to fully recover the cost of capital plus an adequate rate of return. Therefore, strategic bidding (such as withdrawing potential investments in new capacity) is regarded to be non-existent in long-term auctions.
- Fifth, the rate of return required to new a project depends on the leverage ratio of the company carrying out the investment.

When modeling long-term dynamics under an energy-only market scheme, generation investments have usually been assumed to be driven primarily by the profitability expected from each project, typically measured in terms of the Internal Rate of Return (hereafter IRR). That approach obviously calls for an analysis of the expected cash flows throughout a project's life span. Under the long-term auction scheme, such a perspective does not apply because cash flows cannot be determined until the auction has been cleared. Here the traditional methodology is inverted: the IRR required by a new project is determined first, to subsequently back-calculate the bid that would ensure such profitability.

2.3.1. Internal rate of return (IRR) required of a new project

The profitability that each agent requires of a potential new project is calculated with an empirical function that relates the IRR required to the company's leverage ratio (hereafter, the IRR function), i.e. the result of combining the expected returns both on equity provided by the generator and debt invested by the banks involved in the project. All the investment costs associated with a new project are assumed to be financed partially by issuing new debt, i.e. the model is aimed at representing the common situation in the Latin American region, and more in particular in the Colombian one, investments are debt financed as special projects, but in which generators (often incumbents) assume a significant portion of the investment by injecting equity. Therefore, as we later describe, although in most cases we deal with off-balance sheet projects, we consider that to some extent, entering into an additional investment project affects the generators leverage, in proportion to the equity destined to it[87].

The characteristics of the IRR functions used in the model are shown in Figure 32, where the required rate of return grows with the company's leverage ratio. Note that it has been defined a critical leverage ratio that cannot be exceeded[88]. The function has to be adjusted for each agent separately to accommodate the differences in their respective investment potential (in view of their size and financial structure). The advantage of using empirical

[87] Note that the target IRR is assumed to be independent of the technology under evaluation, which is a simplification, since in practice the risk involved in the technology also affects the IRR required.

[88] The strong theoretical support for such dependence can be explained by credit risk theory, see Sánchez (2009).

functions lies in their ability to represent most of the variables influencing companies' decisions simply and in a compact way.

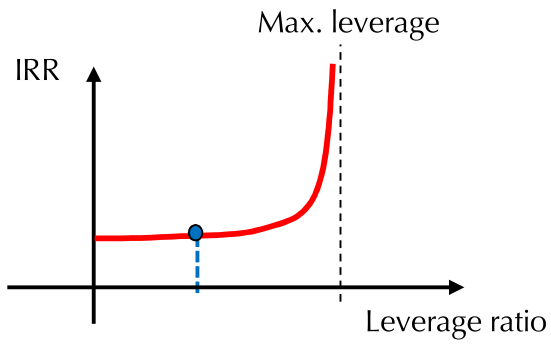


Figure 32. IRR function for a new project

In the simulation model this function is continuously updated to enable each company to assign a different IRR to each project submitted in any given auction, since undertaking a new project alters a company's financial structure.

Representation of new projects

Each potential new project has been characterized by a number of parameters, including: maximum capacity (MW), variable cost ($/MWh) throughout the simulation period, the firm supply value assigned by the regulator (MWh or MW, depending on the case), capital costs, construction period and life span and its unforced failure rate.

In the specific case of new hydroelectric projects, the expected monthly energy output (MWh) must also be entered in the model. Output is expressed as a linear function of the energy produced by hydro plants presently operational. The reason for adopting this approach is that data on the hydrological conditions historically prevailing at existing sites are readily available and are essential to computing future scenarios. In the model, several energy availability scenarios can be simulated in the same run.

2.3.2. Bidding curve algorithm

Generators' bidding curves were built using the iterative procedure shown in Figure 33.

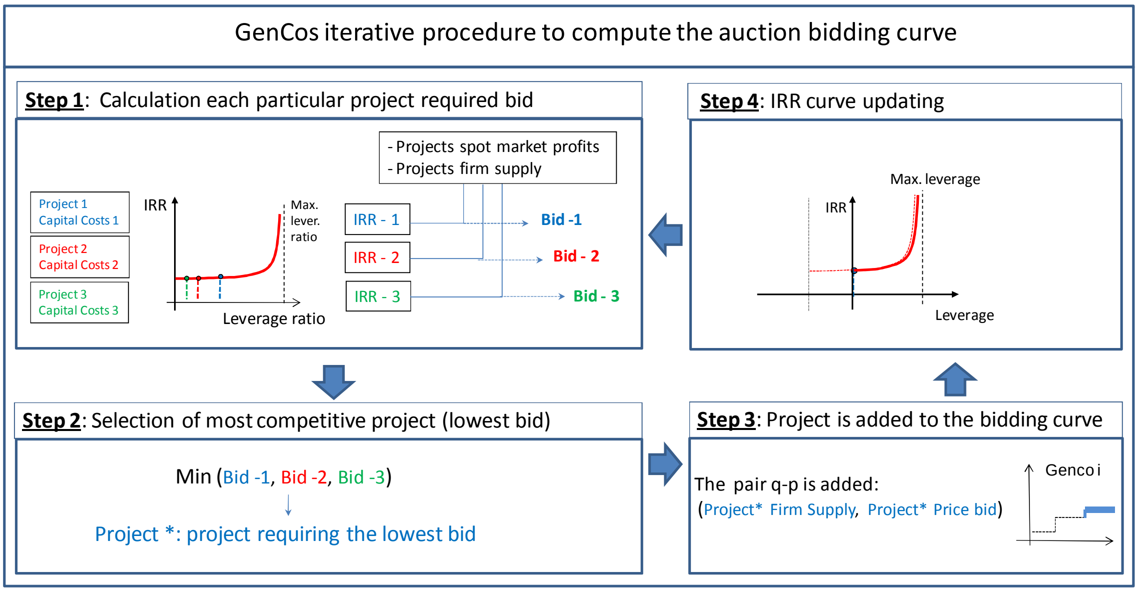


Figure 33. Iterative procedure for computing auction bidding curves

- Step 1: calculate for each potential new project the individual bid that would ensure the break-even position.

  - For each firm's potential new project it is calculated the required IRR on the grounds of its financial structure and the involved investment costs. This is done with the IRR function introduced above.

  - The IRR value is used together with the cash flow expected from the spot market (computed using the game-theory simulation model described below) to determine the bids that would ensure the break even position.

- Step 2: identify the project with the lowest bid.

  The most competitive project, the one with the lowest bid, is the next project included in the bidding curve.

- Step 3: add the aforementioned project to the bidding curve.

  The firm supply for the most competitive project and its respective bidding price represent the next quantity-price pair on the generator's bidding curve.

- Step 4: update the IRR function and continue determining the bidding curve.

  The IRR function is updated to take into consideration the rise in IRR attendant upon company investment in another project

The entire process is then repeated from the first step on for the project with the next lowest bid, and so on until all potential new projects[89] have been entered in the model or the critical leverage ratio (the maximum leverage ratio allowed for each company in the simulation) is reached.

2.3.3. Short-term market income forecasting: spot market price module

The spot market price is modeled using the SPCM (Strategic Production Costing Model) described in (Batlle and Barquín, 2005). This model further develops the classic Production Costing Model approach to more accurately simulate agents' strategies in a liberalized context by considering the slope of the residual demand function in bid building.

Compared to other oligopolistic models, the main advantage of the SPCM is its computational speed. This is a key feature in long-term modeling, where simulations are very time-consuming when several scenarios are computed. The three main characteristics of the SPCM model are:

- The expected demand in each period (e.g. week or month) is represented as a load duration curve.
- The energy to be dispatched by the hydro units in each period is determined by an exogenous model that provides hydro output scenarios. Plants' production profiles in each period are calculated with an algorithm that peak-shaves the demand monotone[90].

  Although such hydroelectric dispatching constitutes perfect competitive behavior, hydro output is subsequently regarded (on a company-by-company basis) to be inframarginal production when calculating the strategic (oligopolistic) bids submitted by thermal plants.

  Each hydro plant's maximum (and also minimum) output is calculated as a function of the weekly power produced. Historical data are required to calibrate such functions.
- The thermal units are ranked along the monotone by order of merit on the basis of their bid prices. These bids are calculated by internalizing the slope of the residual demand to be faced by each agent, considered as an exogenous variable in the SPCM model.

The market model is run for different values of two uncertain variables that play a key role in price formation in hydro-thermal system markets: the hydrological scenario and thermal plant availability (using Monte Carlo).

---

[89] The number of projects in which a firm can invest per year can be limited. Such limits can be imposed on a technology-by-technology basis.

[90] The peak-shave algorithm is applied to the equivalent demand monotone. This equivalent demand is calculated by summing the power demand and the energy corresponding to non-available units submitting bid prices below the resulting system's marginal price. This serves to show that identical load levels may lead to different marginal prices due to thermal plant failures. Unit availability is computed on a time block-by-time block basis (where each time block consists of 10 hours of similar demand over the monotone).

2.4. Auction module

Once both the demand and the bid curves for the generating companies are plotted, the auction is cleared. The information delivered by this module is, on one hand, the price to be paid (the auction marginal price) and on the other, the awardees, i.e. the new investments entering the system.

After simulating the auction, each GenCo leverage ratio must be updated. The IRR curve is thereby modified to provide for future assessments. The new plants must be entered as part of the generation mix in the system in accordance with the lag period established in the regulations.

3. LONG-TERM SIMULATION OF THE COLOMBIAN SECURITY OF SUPPLY MECHANISM

This section describes a full-scale simulation designed to study the expected growth trend in Colombian generation capacity in the next few decades in the presence of an auction-based security of supply mechanism. The discussion first addresses the main elements of the regulatory design, then the input data defining the scenario considered and finally the results and main conclusions.

3.1. Ensuring long-term resource adequacy in the Colombian system: a brief description of Reliability Charge Auctions

So-called Reliability Charge Auctions were introduced in Colombia in 2006 (the first auction was called in May 2008). The mechanism ensures generators a fixed payment in exchange for producing in critical periods (also known as scarcity periods) at a pre-established or strike price. Since a period is defined to be critical whenever the spot price exceeds the strike price, the reliability contract (called the Firm Energy Obligation) implies a commitment that adopts the form of a financial call option.

If the generator fails to comply with the production committed to in such periods, it must purchase energy on the spot market (to make up for the shortage created by its failure to comply with the commitment acquired). Thus, the incentive to be available during tight conditions is high, for the mechanism institutes an implicit penalization for non-compliance.

Both the generators that receive such payments and the payments themselves are determined by market forces in a public auction. The quantity (the total amount of Firm Energy Obligations) to be purchased in the auction, i.e., the amount of power that would have to be produced whenever a critical period is declared, is determined by the regulator (acting on behalf of regulated demand).

Prior to the auction, the regulator certifies the maximum quantity that each unit is entitled to offer and hence to commit in the auction. This ceiling, previously referred to as firm supply and here as Firm Energy for the Reliability Charge[91], is calculated on the basis of preset rules and algorithms that are laid down by the regulator.

The auction is held four years (the lag period) before the power commitment may be enforced by the regulator. The Energy and Gas Regulatory Commission (CREG) must determine the existence or otherwise of an imbalance between the Firm Energy already contracted (in former auctions) and the expected demand for the year evaluated. Where this imbalance is negative, the Regulatory Commission announces its decision to call an auction.

The conditions applicable to each unit depend on its characteristics:

- Existing plants: plants already in operation at the time of the auction. The term of the contract in this case is just one year, and these plants may neither set the price nor be withdrawn during the auction (unless the auction price sinks to below a given threshold).
- New plants: plants whose construction has not begun at the time of the auction. The term of the contract may be up to 20 years. These units submit quantity-price bids and consequently determine the marginal clearing price stemming from the auction.

In Colombia, most new large-scale hydro projects need more than four years to become fully operational. To enable such projects (known as GPPS projects) to participate in the mechanism as new entrants, special conditions are applied. After a standard auction is held, a second auction is called for these GPPS projects, which are eligible for lag periods of up to eight years. This auction is subject to a reserve price (a price cap), which is the clearing price established in the standard auction[92].

A comprehensive description of the rules governing the Reliability Charge mechanism can be found in CREG (2006), XM (2008) and XM and BBVA (2007).

3.2. Definition of scenarios

The simulation studied spot market and Reliability Charge Auction trends across a 25-year horizon, starting in 2010. Auction periodicity was assumed to be fixed, with all standard auctions being called annually.

The Firm Energy demand function (see Figure 34) was based on the definition of two parameters, M1 and M2, whose definition is considered to remain constant throughout the simulation period. The CONE (cost of the new entrant) was updated after each auction was cleared in accordance with the formula provided in the regulation, i.e., as the weighted value of the last CONE considered and the auction clearing price.

---

[91] In Spanish "Energía Firme para el Cargo por Confiabilidad" (ENFICC)

[92] The sequential nature of these two auctions was taken into account in the simulation to enter all the information deriving from the standard auction (new investments and resulting price) in the GPPS auction. Certain simplifications were introduced with respect to these auctions: all the potential GPPS entrants were assumed to be granted the same lag period (8 years) and the same contract term (20 years).

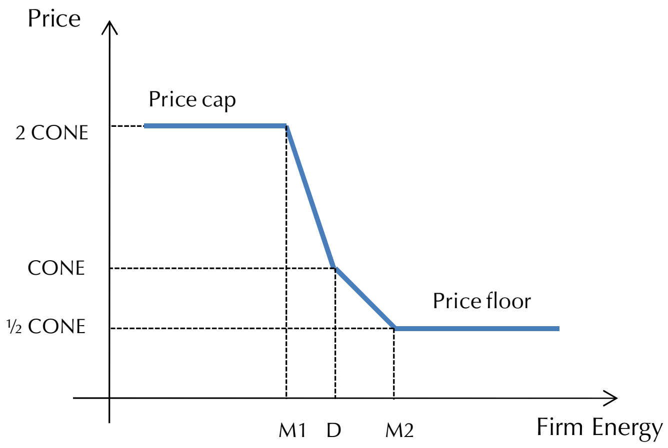


Figure 34. Firm Energy demand

Demand growth was assumed to be as forecast by the Mining and Energy Planning Unit (UPME, 2008). As shown in Figure 35, Firm Energy Obligation demand targets are assumed to follow a similar pattern.

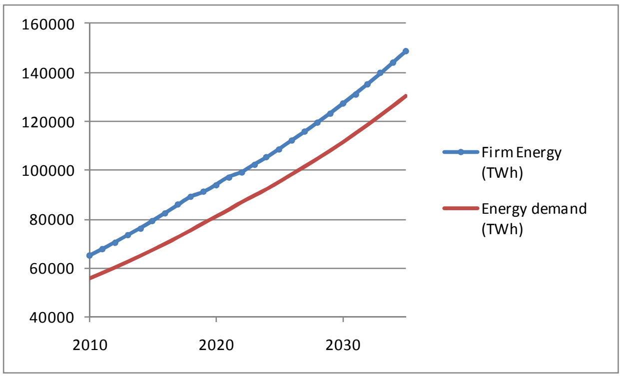


Figure 35. Expected energy demand and reference values for the Firm Energy Obligations to be purchased by the regulator

The ratio between maximum and average demand was regarded to rise at a yearly rate of 3%.

The fuel price series followed the baseline scenario described by the Energy Information Administration (EIA, 2008). Fuel price trends are shown in Figure 36. No additional environment-related costs were entered at any time in the simulation period.

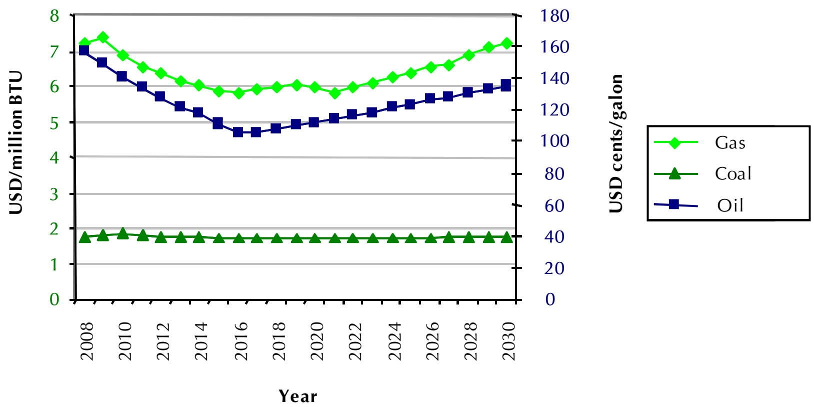


Figure 36. Fuel price forecasting scenario (source: EIA baseline scenario)

When modeling the Colombian electricity market, it is of utmost importance to reflect the system's heavy dependence on the availability of water resources. Moreover, the presence of severe (and difficult to predict) droughts during phenomena such as the El Niño-Southern Oscillation, plays a relevant role in investment decision-making. Three hydro resource scenarios were considered to take account of such hydraulic variability.

Expected yearly production for the hydro plants already installed in 2010 in the scenarios studied is shown in Figure 37.

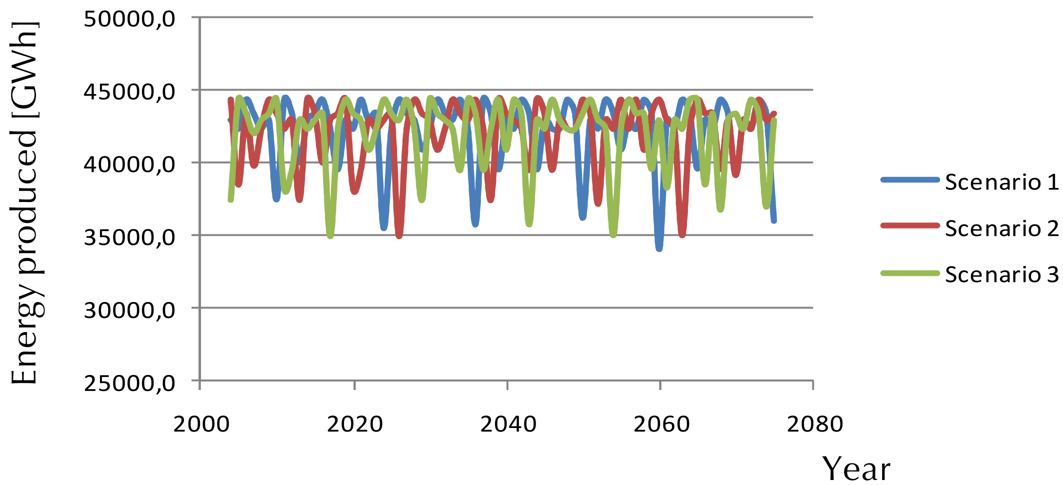


Figure 37. Hydrology data used in the simulation

The annual production of hydro power plants to be installed after 2010 was expressed as a linear function of the output of the plants already operating in the first year of the simulation (see Table i).

Table i. Coefficients for the linear function determining new generation investments

	ALBAN	BETANIA	CHIVOR	CALIMA	GUATAPE	GUATRON	GUAVIO	MIEL	JAGUAS	TASAJERA	PAGUA	PLAYAS	PORCE II	PRADO	SALVAJINA	SAN CARLOS	URRA	Constant
Medium size project	0	0	0	0	0,030	0,030	0	0	0,030	0,030	0	0,030	0,030	0	0	0,030	0	0
Amoya	0,051	0	0,019	-0,047	-0,020	0	0	0	0	0	0,126	0	0,014	0	0	0	0	-15,860
Sogamoso	1,386	0	0,192	0	-0,951	0	0	0	-2,164	0	0	0	0	-3,021	0	0,858	0	47,120
Pescadero	3,078	0	0,398	0	0	0	0	0	0	0	0	0	0	0	3,020	0	0	269,800
Quimbo	-0,191	0,507	0,024	0	0	0,106	0	0	0	0	0	0	0	0	0	0	0,161	-18,816
Cucuana	0,124	0	0,021	-0,060	-0,022	0	0,008	0	0	0	0	0	0	-0,152	0	0	0	-0,508
Porce4	0	0	0	0	0,097	0	0	0	0	-0,043	0,308	0	0,722	1,466	0	-0,102	0	-27,115
Miel2	0,346	0	0	0	-0,054	-0,093	0,019	0	0	0	0	0,206	0	0	0	0	-0,131	28,193
Andaquí	0	0	0,159	0	-0,236	0,860	0,099	0	0	0	0	0	0	0	0	0,144	0	-22,720
El Neme	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fonce	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cabrera	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Porce3	0	0	-0,036	0	0,2019	0,718	0	0	0	0,319	0	-1,052	1,155	0	0	0	0	3,010

The model calculated short-term spot market prices for a 40-year horizon. The first 10 years were computed directly with the spot market model, while all subsequent years' prices were estimated with an interpolation function in which prices were made to converge on the very long-term system prices. These very long-term prices were also calculated with the short-term model, assuming a perfect competitive environment and the generating mix shown in Figure 38.

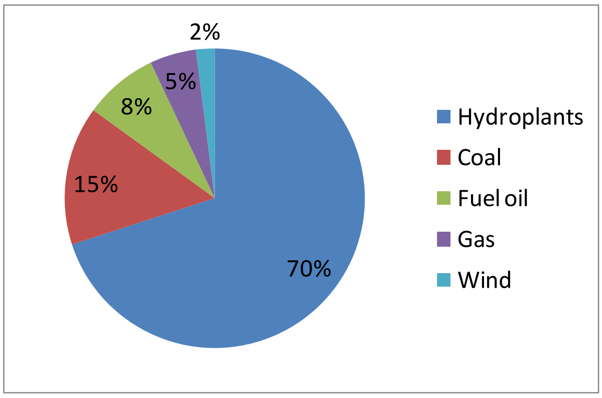


Figure 38 Very long-term generating mix

The expected evolution of the generating mix must also be roughly estimated for the spot market simulation. This roughly estimated growth was assumed to be the same for all market participants. The expected expansion hypothesis was updated after each auction to include the information on the latest investments.

The new projects evaluated included coal, fuel-oil, wind and hydro plants (of different sizes and in different locations[93]). The information on plausible future investments was taken from UPME (2008) as well as from a number of firms' expansion plans. The data on small-scale projects are given in Table ii.

[93] The location affects hydro inflows.

Table ii. Small-scale project characteristics

Project	Construction period	Life Span	Capacity (MW)	Investment cost (M€/MW)	Variable cost (€/MWh)	Average availability factor	ENFICC (GWh/Year)
Fuel-oil	3	20	200	140	59	0,85	1577
Coal	3	20	325	400	22	0,95	2800
Medium size hydro plant	3	40	120	122	0	0,55	250
Wind	3	30	150	290	0	0,4	262

The possible entry of new generating companies into the business was not contemplated because several small companies with growth potential and characteristics similar to the features typical of a new entrant are in fact presently operating. Each firm was modeled in accordance with the function that relates the IRR required of a new investment to its financial structure (debt-to-equity ratio).

3.3. Calibration of the spot market model

The parameters used in the spot market model were adjusted using historical market data. The agents' strategic behavior was calibrated to match market data for 2007. The average monthly prices obtained with the calibrated model and actual market prices are plotted in Figure 39.

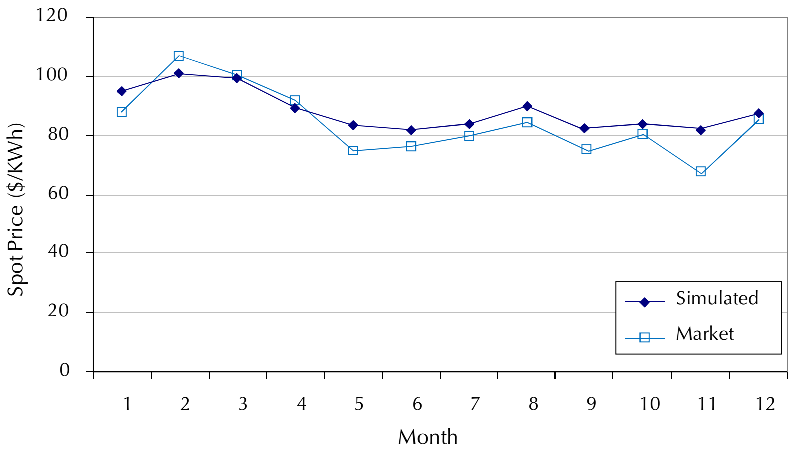


Figure 39. Mean simulated and market prices for 2007

The accuracy of hydro plant dispatch modeling (i.e. the peak-shaving algorithm) was also verified. Figure 40 shows simulated and actual dispatching for two months. While the accuracy of the approach may differ from one period to another, the trend was always fairly well reflected.

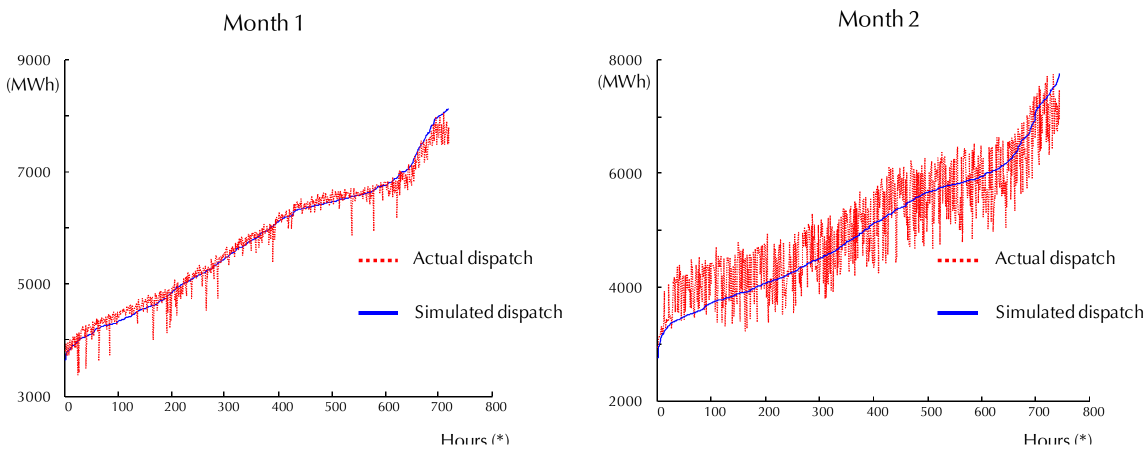


Figure 40. Verification of the suitability of the hydro plant dispatching hypothesis

4. SIMULATION RESULTS

The result of agents' investment decisions is plotted in Figure 41. Since the negative deviations from the regulator's Firm Energy targets were relatively small (see the graph on the right), the mechanism clearly fulfilled its main objective, i.e. to ensure sufficient long-term generating resources.

The mechanism in fact provided for a more stable investment scenario, in which boom-and-bust cycles were almost non-existent. Nevertheless, an analysis of the technologies entering the system in each year (graph on the left) revealed that each technology was subject to an investment cycle. In other words, the most profitable technology varied across the period simulated.

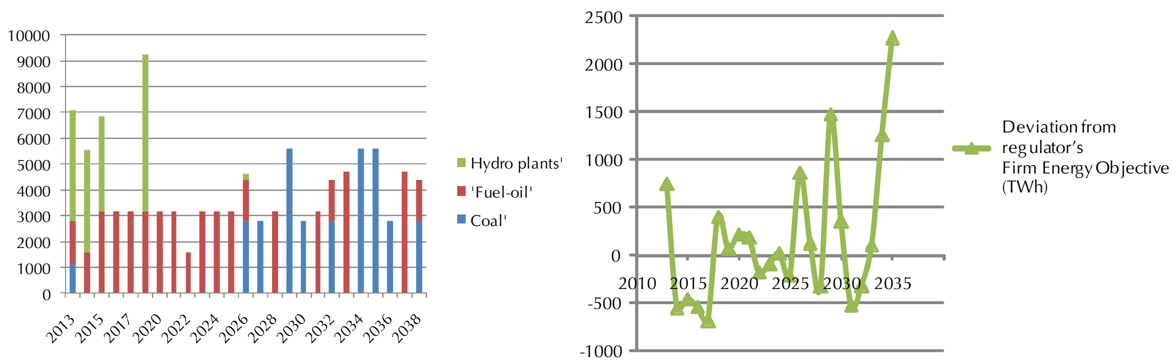


Figure 41. New investment trends

These cyclical investment trends can be explained by effect of the entry of new plants on both market prices and companies' financial structure.

4.1.1. Investment cycles

Model computations for some of the simulated years are discussed below to illustrate the long-term dynamics simulated by the investment decision module. The objective is to show how expected market income and the required IRR (which depends on each company's financial structure) determine the optimal investment.

The expected market cash flow for a given plant can be used to calculate its annualized payment (the premium resulting from the bidding price) required to fully recover its investment costs. This premium (expressed in $ per MWh of Firm Energy) was plotted versus the required IRR for some of the simulated years and for all the technologies studied (see Figure 42).

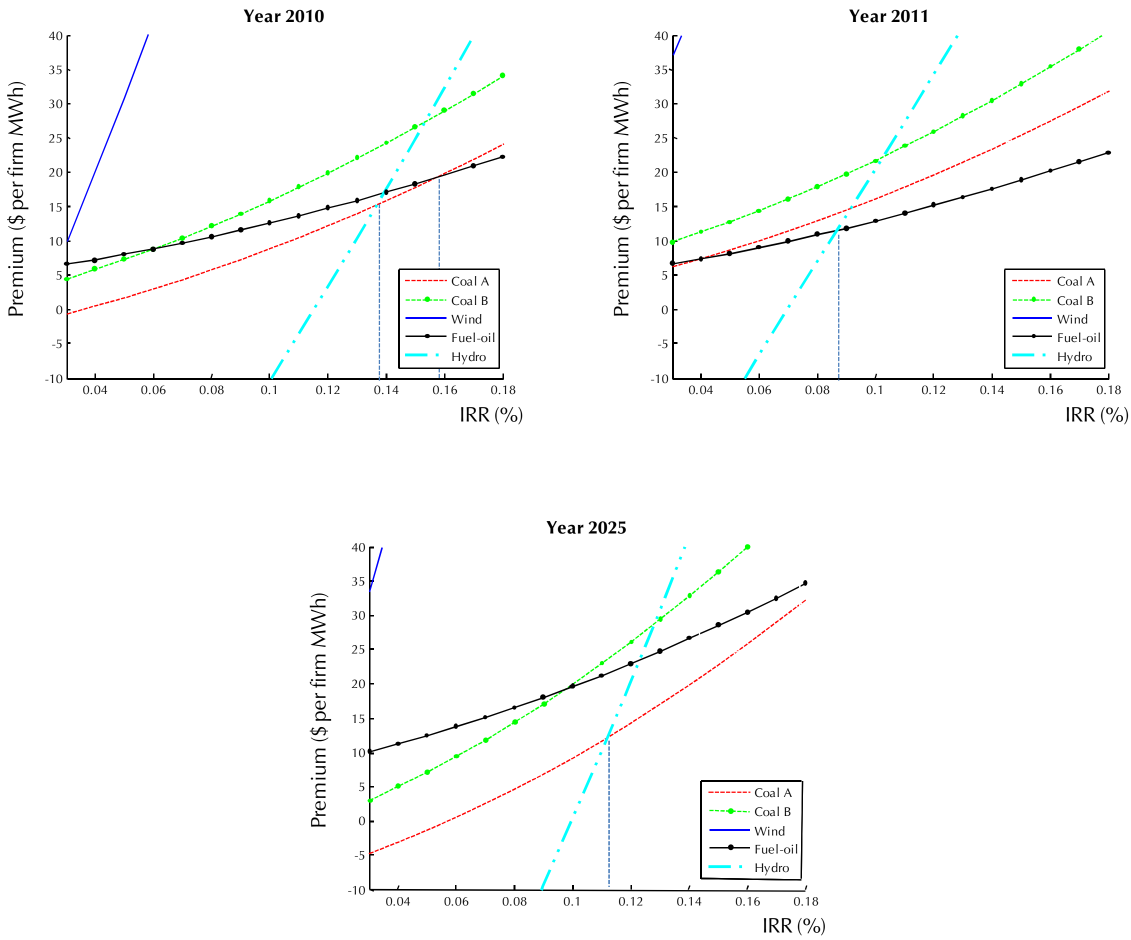


Figure 42. Premium by technology vs IRR

Although the required IRR typically ranged from 10% to 14,5%, depending on the company and its leverage ratio, in the first year of the simulation (upper-left graph in Figure 42) hydroelectric plants constituted the most profitable investment for most of the firms studied (followed first by coal-plants and then fuel-oil-fired plants). Remarkably, for certain values of the IRR the spot market prices ensured full recovery of investment costs.

Since a considerable amount of new hydro energy was committed in the first auction, spot prices could be expected to decline in subsequent periods. This led to an entirely different situation the following year (see the upper-right graph in the same figure). Note that coal-fired plants were never the most profitable technology regardless of the internal rate of return required of a new project, while fuel-oil was the optimal technology for almost the full range of IRR values.

The entry of fuel-oil plants in subsequent years caused a rise in spot prices, and as a result the reversion of the investment curves to the initial situation (see the lower graph in the figure). Under these circumstances, the expected variation in fuel prices, coupled with the expected spot market prices, made coal-fired plants the most competitive technology within the range of IRR values in question.

4.1.2. Regulatory intervention

Although this market-based mechanism has been claimed to leave the determination of the most profitable technology entirely in market agents' hands, there can be no question that certain regulatory decisions largely condition the final results (we already analyze this effect back in Chapter III).

In the present case study, the aforementioned results would be significantly modified if the criteria for allocating the ENFICC were slightly different for hydroelectric plants. By way of illustration, the ENFICC allocated to new medium-size hydro plant projects was raised from 250 GWh to 350 GWh, a value that is still well below the average expected plant output. The entry of new Firm Energy (GWh) under this hypothesis is graphed in Figure 43, which shows the increase in the hydro power plant share compared to the baseline scenario presented above.

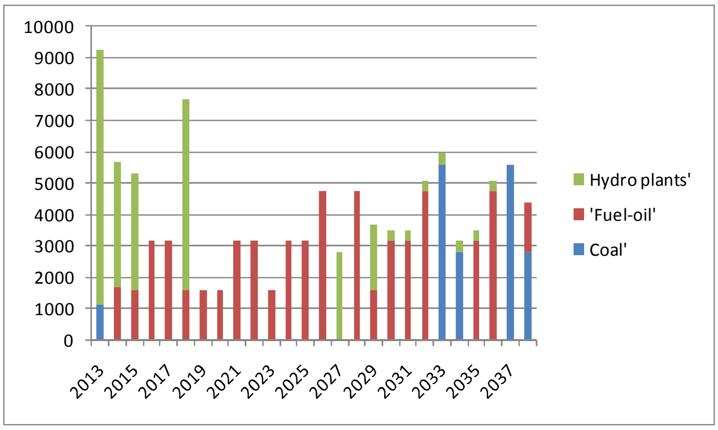


Figure 43. Entry of new investments in terms of Firm Energy

5. CONCLUSIONS

A brief description of the long-term auction-based security of supply mechanism was followed by the introduction of a methodology to simulate generation growth dynamics in such a context.

The simulation tool developed can be regarded to be a system dynamics-inspired model, that draws from a strategic production cost model (Batlle and Barquín, 2005) to better represent agents' interaction on the short-term spot market.

A full-scale simulation based on the Colombian system was developed to illustrate the capabilities of the model. The hydro-dominated Colombian electricity system is subject to variable hydraulic resources (due to the presence of difficult-to-forecast cyclical climate phenomena such as El Niño).

The findings led to a number of interesting conclusions:

- The present mechanism, the so-called Reliability Charge, clearly fulfils its objective of capturing new investment resources for generation. Indeed, the Firm Energy installed during the period studied tracks the regulator's objectives very closely. The price cap and the definition of the demand curve entail no entry barrier for the investments needed to meet Firm Energy requirements.
- The mechanism ensures a more stable investment environment. Nonetheless, investment cycles for each technology can be clearly identified. These cycles are occasioned by the delayed response that usually characterizes long-term investment dynamics. For example, high spot prices are a magnet for hydro plants, leading to a spot market decrease in subsequent years and changing the long-term signal conveyed by the mechanism.
- Although this market-based mechanism is purported to leave the decision on the most profitable technology entirely in market agents' hands, certain regulatory decisions are bound to condition the final results. The example given here shows that the results may vary significantly if the criteria for allocating the ENFICC were less strict for hydro plants.

Electric power system regulation is currently pursuing the optimal balance between free market initiative and centralized (sometimes called indicative) planning; see (Pérez-Arriaga and Linares, 2008). In this context, the scope for security of supply mechanisms is broadening for they should enable regulators to recover some control over the future development of their electricity systems. As a result, simulation tools such as the one discussed in this chapter, able to jointly address financial criteria, economic dispatching and regulatory design for full-scale case studies, will acquire major significance in the near future.

Acknowledgements

The author wants to thank José Salazar and Juan José Sánchez for their on-line support throughout the development of this chapter, as well as to Julián Barquín for his valuable comments and suggestions, which helped to improve the work developed in this chapter.

6. REFERENCES

Batlle, C. and J. Barquín (2005). "A strategic production costing model for electricity market price analysis". IEEE Transactions on Power Systems. 20 (1), 67-74.

Batlle, C. and 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, vol. 16, iss. 3, pp. 184-193, September 2008. doi:10.1016/j.jup.2007.10.004.

Batlle, C. and 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, July 2009. Available at: http://www.iit.upcomillas.es/batlle/Publications.html

Bunn, D. W. and E. R. Larsen (1997). "Systems Modelling for Energy Policy". Wiley: Chichester, UK.

Bunn, D. W. and F. S. Oliveira (2001). “Agent-Based Simulation- An Application to the New Electricity Trading Arrangements of England and Wales”, IEEE Transactions on Evolutionary Computation, Vol. 5, No. 5, October 2001.

Cramton, P. and S. Stoft (2006), “Colombian Firm Energy Market”. Document prepared for the CREG (Gas and Energy Regulatory Commission).

CREG, Gas and Energy Regulatory Commission (2006). “Cargo por Confiabilidad, una visión de largo plazo”. (In Spanish)

Ford, A. (1997). "System Dynamics and the Electric Power Industry". System Dynamics Review. 13,53-86.

Ford, A. (2002). "Boom and Bust in Power Plant Construction: Lessons from the California Electricity Crisis". Journal of Industry, Competition and Trade. 2 (1-2),59-74.

Park, J., N. Ahn, Y. Yoon, K. Koh, D. Bunn (2007). “Investment incentives in the Korean electricity market”, Energy Policy 35 5819–5828, June 2007.

Pérez-Arriaga, I. J. and P. Linares (2008). "Markets vs. Regulation: A role for indicative energy planning," Energy Journal. vol. 29, no. Special #2, pp. 149-164, Octubre 2008.

Salazar, J. E- (2008). “Simulación del Cargo por Confiabilidad y de la expansión del mercado de generación en Colombia”, Master Thesis, Universidad Pontificia Comillas, September 2008.

Sánchez, J. J. (2009). “Strategic Analysis of the Long-Term Planning of Electric Generation Capacity in Liberalised Electricity Markets”. Ph. D. Thesis, Instituto de Investigación Tecnológica, Escuela Técnica Superior de Ingeniería (ICAI), Universidad Pontificia Comillas, Madrid, Spain.

Sánchez, J. J., D. W. Bunn, E. Centeno, J. Barquín (2009). "Dynamics in forward and spot electricity markets", IEEE Transactions on Power Systems. vol. 24, no. 2, pp. 582-591, May 2009.

Sterman, J. (2000). Business Dynamics: System Thinking and Modeling for a Complex World. McGraw-Hill/Irwin, Boston.

UPME, Unidad de Planeación Minero Energética (2008). “Proyección de demanda de energía eléctrica y potencia, revisión julio 2008” Documento UPME.

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.

XM (2008). "*Subasta para la asignación de OEF*" (In Spanish).

XM and BBVA (2007). "*Infomemo, promoción de la primera subasta de energía firm*e" (In Spanish).

PART E

PART E. CONCLUSIONS AND FUTURE RESEARCH

1. CONCLUSIONS

The analyses developed in the thesis have been gathered around three differentiated but interrelated areas of the security of supply regulatory problem: first, the diagnostic of the problem (PART B), second, the analysis and design of mechanisms (rules and incentives) to tackle the problem (PART C), and third, the development of simulation modeling tools to analyze both the performance of the system and also the impact derived from the implementation of these regulatory mechanisms (PART D).

We next briefly summarize the major conclusions derived from this work following this same division:

PART B: SECURITY OF ELECTRICITY GENERATION SUPPLY: PROBLEM DIAGNOSIS

We have seen how decoupling the security of supply concept into its major components facilitates both the problem understanding and the proper design of additional technical procedures and regulatory measures (if considered necessary). These components, from the time dimension perspective, are:

- **Security**, a very short-term issue (close to real time), defined by the NERC (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 or suddenly disconnection"(NERC 1997).
- **Firmness**, a short to mid-term issue, which can be defined as the ability of the already installed facilities to provide generating resources efficiently (especially when most needed). This dimension is linked to both the generating units' technical characteristics and also their medium-term resource management decisions (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 demand efficiently 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.

There is a certain consensus around the idea that the security dimension can be tackled by means of operation reserves markets, where the reserves requirements are prescribed by the System Operator. However, we have seen that there are still some open and important discussions as whether to use a single or a dual pricing mechanism to clear balancing markets or how to optimally determine the operation reserves requirements (introducing not only reliability but also cost efficiency criteria). Another problem detected at this dimension, is that sometimes the operational reserves are used with other purposes different from guaranteeing short term security (for instance, to cover peak load when generation is scarce).

On the other hand, it is also commonly agreed that the strategic expansion policy has to be solved through the implementation of additional "out-of-the-electricity-market" mechanisms (e.g. feed-in tariffs or cap and trade mechanisms).

This way, regulatory intervention is accepted to be necessary at the security and strategic expansion policy dimensions. However, in between these two dimensions the debate on the necessity of intervening to ensure security of supply has always been, and still is, open and quite intense. For many different reasons, ensuring a secure generation supply is still far from being an evident matter as it seems that a system driven by purely market-based mechanisms ensures neither an efficient resource management nor the required long to very long-term expansion.

We have shown how the inefficient allocation of risk plays a key role, but it is not the only issue hampering long-term security of supply in electricity markets. Some flawed regulatory rules which cap short-term signals, coupled with the lumpiness investment problem, economies of scale or the lack of short-term demand elasticity, result in a supply that is far from being perfectly optimal (i. e. efficient from the net social benefit point of view).

It seems that a well-functioning long-term market would provide the most suitable solution for overcoming most of the hurdles, not only helping to allocate risk efficiently (and thus removing the severe adverse effect that risk aversion has for investment), but also alleviating the effect of some other problems and market imperfections -providing a safety valve, as pointed out by Joskow (2007)-. In principle, both generators and demand should have clear incentives in participating in these markets. Indeed, on the demand side, we have analyzed how there are two different incentives:

- To hedge against its own risk.
- To help the generators to hedge their risk, since by doing so, its own welfare increases.

However, the demand is not actually taking part for many different reasons (regulatory tariff protection, the belief that "somebody will ensure reliability in the future", lack of rationality, etc.).

Well-functioning long-term markets cannot be seen as an option which is within the control of the regulator. Efficient long-term markets arise because of the willingness of market participants. Regulators can just help by creating a suitable transparent framework for trade, but implementing a trading floor does not ensure a well-functioning long-term market.

After the general analysis we have focused on the firmness dimensions, analyzing by means of a mathematical model the potential inefficiencies caused by the market failure in the medium term resource management. We have illustrated how when generation side is risk averse and there is no forward market, the market equilibrium can deviate significantly from what we have defined as the efficient benchmark solution (the solution a central planner with perfect information would devise). The reason is that generating companies use their limited production resources to hedge against low-profit scenarios.

We show how, under a number of hypothesis, a well functioning long term market (modeled by means of a forward market for the sake of simplicity) would replicate the efficient benchmark solution. The problem is that up to now, due to a number of different possible reasons demand does not yet actively participate in long term markets. This lack of participation supports the regulators intervention to avoid inefficient market outcomes.

PART C: SECURITY OF SUPPLY MECHANISMS DESIGN: OBJECTIVES, ELEMENTS, INCENTIVES AND OUTCOMES

When the market is assumed to be unable to provide an adequate security of supply level, the solution to the problem necessarily entails the development of additional mechanisms.

Once the regulator has decided to undertake the task of "helping" the market to reach what he/she 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 short to long-term scheduling and 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.

In this respect, PART C of the thesis has been aimed at contributing at the discussion on the mechanisms design. Three steps have been followed in this analysis

- We have first reviewed the fundamental criteria that support the analysis on how the corresponding price incentives should be designed, calculated and managed in order to provide optimal signals. We have highlighted the importance of properly defining a metric to evaluate the market performance. This metric will serve to set the objectives pursued by the mechanism (if required), thus this represents a fundamental first step towards the correct design of the whole mechanism. A conceptual simplified mathematical model of the electricity market has been used to illustrate the optimal pricing criteria behind a correct design of these security-of-supply mechanisms. We have shown 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 have introduced some more detailed in the formulation to deeply analyze the particular case of the mechanisms aimed at the firmness and adequacy dimensions. We have illustrated how when implementing this kind of mechanisms, the regulator conditions the market functioning thus, partially recovering its former central planner role.
- Second, we have analyzed the different mechanisms that have been implemented worldwide to tackle the security of supply problem at the firmness and adequacy

dimensions. In this context the regulator has two alternatives to deal with long-term security of supply: to do nothing (in the belief that the market will provide an efficient result, hopefully sooner rather than later, given the possibility of periods of scarcity in the meantime) or to take an active role trying to represent its own view about demand's best interests by introducing a long-term mechanism. 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.

- Third, based on the two previous chapters, we have detected and discussed several key design elements, namely:
  - The determination of the counterparties, i.e. which generating units are allowed to sell and how much, and also the part of the demand on behalf of which the regulator makes decisions. The regulator has to decide whether to act on behalf of all the demand or just a proportion of it. Care will need to be taken so as not to create free riding issues.
  - The way to set the reliability product price, i.e. whether the regulator administratively sets the price or just the quantity and allows for a market-based mechanism to reveal the price. This decision depends mainly on the existing market structure and the expected level of competition and absence of entry barriers. If these conditions are adequate, a market-based solution appears as more interesting. In this case, the regulator has also to decide whether the product is bought in an auction or bilaterally and finally if the purchasing process is centralized or left to the retailers' initiative. The international learning process has led to the conclusion that it is desirable to use centralized auctions for different reasons, among others, to benefit from economies of scale increasing competition, to avoid vertical integrated companies taking advantage of obscure agreements, etc.
  - The reliability product characteristics. Often the consequences of the product definition are not sufficiently evaluated beforehand, and as we have thoroughly discussed, highly inefficient situations are the result. Therefore, the regulator has to determine the time terms (lag period and contract duration), which as we have discussed determine critically the type of generators that will enter the system, as well as a non-arbitrary rule to identify near-rationing conditions so as to assess each unit's contribution to system reliability. The spot market price is the best indicator of the existence of critical situations. There are many other product design issues that deserve themselves lengthy discussion that exceed the scope of work, such as product optionalities (forward or option contracts), penalties, financial guarantees, force majeure clauses, etc.

Be it said in summary, that although it can be observed a certain convergence in long-term security of supply mechanisms design criteria worldwide, we are still far from obtaining a definite consensus on the subject. The reason lies on the fact that each market's particularities make it very difficult, for what could have been considered as a successful mechanism in one system, to be directly exported with guarantees to another.

The regulation design problem, not the market problem

In the light of the evidence discussed throughout PART B and PART C, one might conclude (as is often the case) that the market resulting from the reform of the electricity carried out over recent decades is not the right alternative. However, we consider that the final problem is not the market approach itself, but the lack of adequate regulatory mechanisms to deal with the complications that real life markets may present. These regulatory flaws have resurrected and encouraged numerous lines of argument in favor of a step back towards the traditional centralized (even nationalized) model; for instance, in the case of Ecuador.

But just a quick look at some electricity systems in which the market reform has not been implemented shows that they have not escaped similar or even worse problems. In this respect, the latest news at the time of this writing from Venezuela or Mexico illustrates the fact that the formerly traditional centralized model also does not guarantee an "adequate and sufficient" functioning of the electricity system.

PART D: SIMULATION MODELS TO SUPPORT THE DESIGN AND IMPACT ASSESSMENT OF SECURITY OF SUPPLY MECHANISMS

The first simulation tool, a Probabilistic Production Cost (PPC) model, has been developed with the objective to illustrate and deal with the proper definition of metrics to evaluate market performance. We have first extended the formulation of the classic PPC model by including a novel algorithm that allows to introduce demand elasticity and then on this basis we have illustrate how traditional reliability measures are not suitable metrics to be used when a non-negligible part of the demand is elastic.

The second simulation tool developed can be regarded to be a system dynamics-inspired model, that draws from a strategic production cost model (Batlle and Barquín, 2005) to better represent agents' interaction on the short-term spot market.

The model is inspired by the guidelines and basic structure of the approach set out by (Sánchez, 2009). Certain additional refinements have been developed in order to better deal with some of the complexities that exist in some particular markets. In this regard, for instance a multi-scenario analysis tool and an approach to assess new hydro plants investments have been introduced in the standard model. Another relevant refinement has been to include how each new investment modifies the financial structure of the company. This way, within the same auction it is considered the increasing costs of issuing new debt.

A full-scale simulation based on the Colombian system was developed to illustrate the capabilities of the model. The hydro-dominated Colombian electricity system is subject to variable hydraulic resources (due to the presence of difficult-to-forecast cyclical climate phenomena such as El Niño).

The findings led to a number of interesting conclusions:

- The present mechanism, the so-called Reliability Charge, clearly fulfils its objective of capturing new investment resources for generation. Indeed, the Firm Energy installed during the period studied tracks the regulator's objectives very closely. The price cap and the definition of the demand curve entail no entry barrier for the investments needed to meet Firm Energy requirements.
- The mechanism ensures a more stable investment environment. Nonetheless, investment cycles for each technology can be clearly identified. These cycles are occasioned by the delayed response that usually characterizes long-term investment dynamics. For example, high spot prices are a magnet for hydro plants, leading to a spot market decrease in subsequent years and changing the long-term signal conveyed by the mechanism.
- Although this market-based mechanism is purported to leave the decision on the most profitable technology entirely in market agents' hands, certain regulatory decisions are bound to condition the final results. The example given here shows that the results may vary significantly if the criteria for allocating the ENFICC were less strict for hydro plants.

Electric power system regulation is currently pursuing the optimal balance between free market initiative and centralized (sometimes called indicative) planning, see (Pérez-Arriaga and Linares, 2008). In this context, the scope for security of supply mechanisms is broadening for they should enable regulators to recover some control over the future development of their electricity systems. As a result, simulation tools such as the one discussed in Chapter VII, able to jointly address financial criteria, economic dispatching and regulatory design for full-scale case studies, will acquire major significance in the near future.

2. FUTURE RESEARCH

This work has focused on contributing to the security of generation supply problem from a good number of perspectives, through the microeconomic analysis of the problem, the assessment of the complexity of the actual regulatory design of the proper mechanisms and the development of two different sorts of simulation models to support the impact assessment of these designs. This varied scope of approaches of the different analyses, developments and contributions carried out in the context of this thesis lead to a significant number of future lines of research. This section tries to briefly outline them and to introduce and discuss the relevance of continuing those lines in the future.

These lines of proposals for further research have been organized again around the three core parts in which this thesis has been classified: the problem diagnosis, the design of regulatory mechanisms and the simulation models as support tools for the assessment of the different alternatives.

PART B: SECURITY OF ELECTRICITY GENERATION SUPPLY: PROBLEM DIAGNOSIS

Although a number of mathematical models can be found in the literature to analyze the adequacy problem from the regulator's point of view, a relevant effect is usually forgotten: how risk alters the conditions of the project finance (mainly the internal rate of return asked to generators). This relevant effect is in many cases the main driver behind the introduction of additional mechanisms based on long-term hedging instruments, but little has been analyzed from a theoretical point of view in this respect. It would be of great interest to model such dependence to analyze the expected result that was already pointed out in Chapter I from a qualitative point of view: the demand has two clear incentives to take an active role in long-term markets, one direct (to hedge against the price volatility) and one indirect (to benefit from the more efficient situation that results when the demand helps the generators to hedge their risk).

In Chapter II, we have shown how the medium-term resource management depends both on the generators' risk aversion and on the existence of a well-functioning long-term market. Although perfect competitive market conditions have been assumed along the different analyses carried out, it would be interesting to analyze how market power would modify the medium-term resource management in the different cases (with and without long-term markets). The effect of market power on the medium-term resource management has been previously studied from the regulatory perspective, see for instance (Bushnell, 1998), but as far as we know, there are not studies including both market power and risk aversion to analyze how in such a context the resulting outcome could deviate from what we have defined as the benchmark problem (the central planner problem with perfect information).

PART C: SECURITY OF SUPPLY MECHANISMS DESIGN: OBJECTIVES, ELEMENTS, INCENTIVES AND OUTCOMES

In Chapter III, we have presented a stylized model to analyze the optimal incentives of security-of-supply-oriented mechanisms. Then we have focused on those mechanisms aimed at ensuring firmness and adequacy by purchasing a single product (known as the reliability product in this context). The next step, not considered in this thesis, would be to introduce in this model the following developments:

- Generators' risk aversion.
- The definition of different objectives on the regulator's side, and also the definition of different products aimed at fulfilling these objectives.

In Chapter V, we have pointed out the necessity of asking for guarantees in the auction to generators. Sometimes these guarantees translate into requiring some sort of physical back-up (the firm supply). A deeper analysis on the optimal design of these guarantees is a hot topic in the current regulatory design of long-term mechanisms, particularly since once compromised the future expansion of the system relying on new investments to be entering the system in a number of years, which leaves little ability to react if anything fails. Such an analysis should take into account not only the default risk perceived from the regulator's point of view, but also the additional costs that these guarantees may imply in the end and the entry barriers that the guarantee may introduce.

Finally, we saw that the security of supply oriented mechanisms should ideally remunerate each generating unit based on its contribution to the regulator's objectives. In this context it would be relevant to ponder how the product should take into consideration the effect of the network, a consideration that was considered to fall outside of the scope of the present work, but whose influence may be relevant.

PART D: SIMULATION MODELS TO SUPPORT THE DESIGN AND IMPACT ASSESSMENT OF SECURITY OF SUPPLY MECHANISMS

The analysis of long-term generation expansion is an issue that has received much attention during the last decades. As we have described in this work, the capability to anticipate the market evolution in the long run is essential both from the point of view of generators and from the point of view of the regulator.

In the particular approach adopted in the expansion simulation model developed in this thesis (the system-dynamics-inspired methodology), there are some developments which would constitute interesting modeling advances. We briefly describe them next:

- The income an investor may face when dealing with the problem of deciding new investments is subject to certain risk. We have faced the random nature of two of the most relevant variables involved in the spot market price calculation (hydro production and the units' availability) through a multi-scenario analysis. The problem is that just a few scenarios can be computed in reasonable times. In order to better represent the spot market prices, a Probabilistic Production Cost (PPC) model similar to those presented in Chapter VI could be introduced in the spot market module. As discussed in Chapter VII, when analyzing long-term expansion dynamics based on the selected approach, the computational efficiency of the spot market module is an essential feature. This PPC models would represent a suitable alternative to introduce some more detail in the price determination, since it allows for representing all possible states of availability of the system in a compact and computationally efficient manner. This way, a good estimation of the prices probability distribution function could be achieved. This could help to more precisely characterize the so-called scarcity periods, which play a key role from the regulator's and agent's perspective. In the generators point of view these periods are relevant because they not only imply a much higher potential remuneration in the spot market, but also because in the context of long-term regulatory mechanisms, in one way or another, the units that do not produce in those periods are penalized[94]. From the regulator's point of view, it is clear that the chief objective is to reduce the frequency of those periods (taking into account economical criteria).

- In the model we have introduced risk aversion criteria in the generator's decision making process, since they choose investments based on their expected rate of return and the variance of the results. But the decisions are taken on a unit by unit basis, without taking into account the mean-variance income of the whole portfolio of the company. This way, the model does not take into account the advantages stemming from diversifying among different technologies, since it represents a means to hedge against risk. To make decisions in order to optimize the mean-variance of the whole portfolio results would represent a quite interesting development in this regard.
- In the proposed model, the very long-term prices are calculated assuming a fixed generating mix, which is introduced as an input data. A future development that could help to refine the results would be to update this very long-term generating mix hypothesis based on the results of the auctions that the model simulates.
- Finally, the model design and the discussion built around it has been primarily focused on assessing the potential outcomes and the future market agent's dynamics that could result from the regulatory scheme implemented. The model allows, and it would be of great interest, to analyze the sensibility of the results with respect to the mechanism design elements, for instance, the time terms involved in the contracts, the lag period or the scarcity price level (in the context of the reliability option scheme).

3. References

Batlle, C. and J. Barquín (2005). "A strategic production costing model for electricity market price analysis". IEEE Transactions on Power Systems. 20 (1), 67-74.

Bushnell, J. (1998), "Water and Power: Hydroelectric Resources in the Era of Competition in the Western US". PowerWorking Paper PWP-056r

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.

North American Electric Reliability Council (1997) "NERC Planning Standards", September 1997.

[94] This expected future penalization a unit may suffer has to be taken into account when calculating the generators' bids to be submitted in the auction process, thus an accurate prediction of the probability of the scarcity events would be of interest.

Pérez-Arriaga, I. J. and P. Linares (2008). "Markets vs. Regulation: A role for indicative energy planning," Energy Journal. vol. 29, no. Special #2, pp. 149-164, Octubre 2008.

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.