

Escuela Técnica Superior de Ingeniería (ICAI)

Policy and regulatory design issues on security of electricity generation supply in a market-oriented environment

Problem fundamentals and analysis of regulatory mechanisms

Carlos Batlle & Pablo Rodilla IIT Working Paper IIT-09-057A

Madrid, April 2010

First version: July 2009. This version 14.4 April 2010

[**iiT**]-

Santa Cruz de Marcenado 26, E-28015 Madrid - Tel: +34 91 542 2800 - Fax: +34 91 542 3176 - http://www.iit.upcomillas.es/

Contents

Executive summary 1	
1	The different dimensions of security of supply
2	Security of electricity supply at the generation level: problem analysis122.1 Is the market capable of ensuring a reliable supply?122.1 Ideally the market solves the problem132.2 Market imperfections and flawed regulatory rules142.2.1 The effect of the existence of the lumpiness problem in generation142.2.2 Agents risk aversion and the consequences of an inefficient allocation of risk162.2.3 The consequences of introducing regulatory flawed rules that distort market signals212.3 The need for regulatory intervention23
3	A critical assessment of the different approaches aimed to secure electricity generation supply
4	Principles and criteria to design security of supply mechanisms464.1 Design elements464.2 The reliability product: some examples and design issues48
5	Conclusions
6	References
Ar	A. i Theoretical results under ideal hypotheses

Title to the copyright shall remain with the members of the research team, who shall be cited in any use whatsoever of the results of their work. In accordance with the practice of the scientific community, the conclusions and points of view reflected in the reports and results belong solely to the authors of said reports and results and do not bind or commit in any way the Universidad Pontificia Comillas de Madrid or its Centers and Institutes or other professors and researchers.

The energy industry, and particularly the electricity sector, has been subject to major reforms over recent decades. Until these processes started, the activity that we now call "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 feasible level. Ideally competition is useful because it sends sound economic signals –market prices- both to electricity consumers and suppliers that are supposed to drive market agents' 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). Indeed, the term "restructuring" has sometimes been preferred to denote this process in some systems (particularly in the United States).

The discussion that we develop next, the need for the regulator's intervention to complement the electricity market in order to guarantee 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 (or produce) inefficient (or efficient) outcomes in different places and timescales which might (or might not) otherwise occur¹. 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.

Note from the authors. 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 et al.) 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 references are provided at the end).

¹ In the liberalized context, these regulatory objectives have to be carried out while at the same time trying to minimize as far as possible the intervention and distortion of the purely market based mechanisms. Determining the exact level of intervention is probably one of the most controversial challenges regarding market regulation.

Executive summary

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 a reliable supply represents a major concern for regulators.

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, 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, 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.

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-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 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.

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 adverse effect that risk aversion has for investment), but also alleviating the effect of some other problems and market imperfections. In principle, both generators and demand should have clear incentives in participating in these markets. 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.)

But 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 market².

The regulator alternatives

In this context the regulator has two alternatives to deal with long-term security of supply:

- 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. This is usually known as the "energy-only market" approach.
- Take an active role and try to represent the demand's best interests.

The core of the everlasting controversy

It is important to bear in mind that if demand revealed its utility function (i.e. its preferences) and would enter into long term contracts (in any time term) to hedge its own supply, it would be directly determining its security-of-supply preferences, and thus, not only the current and future generation mix (adequacy) but also the generating units planning (firmness).

Conversely, if it is the regulator the one that acts on behalf of the demand, it will be the regulator decisions those clearly and inevitably conditioning the market outcomes. Every regulator's decision, and consequently every level of intervention, is usually justified as having been carried out in the demand's best interest. This usually leads to another well-known argument, which is that a regulators' decision corresponds to what the demand would do if it understood the underlying problems. But here a contradiction arises, since this is indeed the perfect argument for a return to the traditional centralized scheme, which was considered to be highly inefficient.

Nowadays it seems clear that neither totally regulated nor purely market-based schemes are generally capable of providing optimal results. This has led to hybrid alternatives where the regulator takes an active role using several market-based tools. This is what we have analyzed in the present study.

The energy-only markets: is there any?

The energy-only approach 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. As we review in the present work, in practice it is very difficult to find electricity

² Several experiences, as the case of OMIP in the Iberian market serve to illustrate this reality.

markets in which the regulator does not resort to any explicit (or implicit) safety measure to ensure system reliability. Although it is true that there exist some cases that can be given the title of "greater market believers" (as for instance NEM, ERCOT, etc.), to some extent almost all systems depend on some kind of additional explicit or implicit mechanism that can "help" the market to enhance security of supply.

Designing regulatory mechanisms to ensure security of supply

If the regulator opts for taking action, in the light of the international experience, the major issues to be tackled are enumerated and briefly discussed next.

The counterparties: buyers and sellers

The regulator has to define the part of the demand on behalf of which he will make decisions. Thus, 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 regulator has also 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 case, discriminating among different units may create a market segmentation with undesired long term effects.

The product

To properly define what generating units sell in return for the additional hedge instrument or source of income the security-of-supply mechanism aims to raise. This is known as the reliability product.

Determining the product to be bought from the generation is of the utmost importance and complexity. There are many different alternatives: fixed or flexible long-term energy contracts, certificates of installed capacity, certificates of available capacity (or available energy), certificates of a certain technology installed capacity, long-term reserves requirements, physical units to be operated by the System Operator under certain conditions, energy financial contracts, etc.

Defining an adequate product can determine the success or failure of the whole mechanism. In this sense, when defining the reliability product, the regulator has to be careful with the foreseeable response on the generators side, so as to analyze whether this response leads or not to an efficient result³.

There is a certain consensus around the idea that the reliability product should remunerate the capability of producing energy at "reasonable" prices (whatever "reasonable" might mean, usually below the NSE value) when the system is suffering a scarcity. 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 reasonable and transparent indexes, but although inferior, other possibilities have also

³ For instance, if the regulator decides to buy installed capacity by means of a public auction, it will probably get the capacity which presents the lowest investment costs, but maybe with low availability rates. If it decides to pay for the water reservoir level in the "dry season", it will fill reservoirs to their full capacity in that season. Sometimes the consequences of the product definition are not evaluated beforehand, and highly inefficient situations are the result.



been implemented, such as defining certain periods ex-ante, using the reserve margin measure (whatever the methodology used), etc.

Other relevant aspects of the contract associated to the reliability product include:

- The time parameters. There are two time parameters of the product that have to be carefully designed: the lag period and the duration of the commitment (contract).
 - The lag period is the existing time duration between the moment the commitment of deliverability is signed and the moment the electricity has to be delivered. The *lag period* indirectly determines whether (and which) new investments may participate in the mechanism. If the lag period is shorter that the time required for constructing and installing a new project, then this new project will not be able to participate. Thus the definition of this parameter has two major consequences:
 - · A proper definition gives generators the chance to bid their investment costs in the mechanism, and thus helps reducing generators' risks.
 - It also affects the competitive pressure in the process (the longer the lag period, the larger the technologies and plants that may compete in the mechanism).
 - The *contract duration* affects the generators' risk exposure. Those products which have a short contract duration do not help hedging generators' risks. They provide an additional source of income, which may help to solve some investment problems (such as the so-called missing money problem caused by price caps), but they do not help to reduce risk exposure, which is particularly relevant when large investments are evaluated.
 - The contract duration in price-based mechanisms (capacity payments) conditions the stability of the additional income provided. If the payment is recalculated every year (as is the case in Ireland, for instance), generators' risk exposure may be significant.

Both the lag period and the contract duration defined by the regulator condition the results of the mechanism (i. e. a seven-year lag period with long term contract "makes life easier" for large hydro plants, while a three-year lag period and short term contracts makes it close to impossible).

Capacity markets of the former PJM or NY-ISO systems are a good example of a mechanism with short lag periods and short contract duration (did not ensure enough competition, did not allow generators to bid investment costs and did not reduce the risk involved).

Other relevant characteristics to be taken into account are the penalties for non-compliance; force majeure clauses; whether the contract is indexed to reduce generators' risk exposure; or credit guarantees, which in this context are of the utmost importance.

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 which is either too large or too small. Analogously, resorting to a fixed quantity may result in too high a 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 to be more interesting.

Elastic requirements better reflect the utility each security-of-supply level provides to the buyer (the demand). Additionally, they help to reduce market power and also provide more information about how far the system is from suffering a scarcity.

Other details

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 regulation design problem, not the market problem

This report analyses the consequences of the so-called market failure in the context of the security of supply and discusses some possible alternative solutions. The market usually fails to provide adequate incentives for generators to properly manage, plan and operate their facilities and to make the investments necessary to achieve what the regulator deems to be the adequate level of security of supply

In the light of this evidence, 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. The main aim of our work has been to highlight the fact that 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 present⁵.

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, see (Batlle & Pérez-Arriaga, 2008B). While we have intensively reviewed almost all the cases in which a market-like approach has been implemented, providing a critical analysis of the failures, we have not looked at other electricity systems in which the reform has not been implemented since these have fallen outside the scope of this report. However, it should not be forgotten that these unreformed markets have not escaped similar or even worse problems. In this respect, the latest news from Venezuela or Mexico illustrates the fact that the formerly traditional centralized model also does not guarantee an "adequate" functioning of the electricity system.

⁵ It is difficult to better illustrate this statement than the way Alfred Kahn did in "The economics of regulation" (MIT Press, 1988): 'Continued deregulation is the proper way to go, to the extent feasible...The central institutional issue of public utility regulation remains finding the best possible mix of inevitably imperfect regulation and inevitably imperfect competition. All competition is imperfect; the preferred remedy is to try to diminish the imperfection. Even when highly imperfect, it can often be a valuable supplement to regulation. But to the extent that it is intolerably imperfect, the only acceptable alternative is regulation. And for the inescapable imperfections of regulation, the only available remedy is to try to make it work better'.



1 The different dimensions of security of supply⁶

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 a reliable supply 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, electricity networks and 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, and metering, control and system protections functioned correctly.

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 –generators of a diversity of technologies, transmission and distribution networks or the means of provision of primary fuels-.

Previously, we stated that regulation has to cover different time scales. Indeed, as with any other problem, decoupling the security of supply problem into its major components facilitates its understanding and the design of proper technical procedures and regulatory measures. 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 are 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 reliability of electricity supply

Thus, from the "time" perspective one can distinguish four dimensions of reliability of electricity supply:

- Security, a short-term issue, 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.

⁶ A previous version of the first two sections of this report is included in the first two chapters of the Report for the Florence School of Regulation "Policy and regulatory issues concerning security of electricity and gas supply", which was written for the Training course for senior regulatory staff from MEDREG MPC countries, September 2009.

- 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 term⁷.
- 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.





The main objective of the reform of the electricity system consisted 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 power system benefit" can be decomposed.

Following the reform in the new environment, 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 midterm, 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.

⁷ The NERC defines adequacy as the "ability of the system, at any time, to supply totally the demand of the electrical system and the requirements of energy of the consumers, considering programmed outages of components and non-programmed outages, but reasonably expected" (NERC, 1997). The definition provided in this document also includes the issue of taking steps to guarantee sufficient future capability of providing security services, since they are essential if demand is to be met efficiently. In the same way that short-term energy prices may not be enough to ensure new investments to cover demand requirements, short term ancillary markets may not provide enough incentives to ensure security requirements. When this situation arises, the regulator should consider introducing security criteria in the long term security of supply mechanism.



The energy sector plays a key role in the overall development of a country/state, so more often than not, the previously enumerated dimensions are affected by higher level criteria. Thus, there is an additional dimension that conditions the previous ones and that falls outside of the scope of the present work:

• 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, the European Community 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.

Here, we will focus on the analysis of the different dimensions exclusively from the point of view of the sub-problem, that is, the maximization of the net benefit of the electricity system. Next, we introduce a more detailed description of each one of these security dimensions.

Reliability of electricity supply

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 stability. 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- are a good hybrid (market & regulation) alternative for ensuring security⁸. But although these market

⁸ 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

mechanisms are functioning adequately, the intervention of the regulator at this level is also present.

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 engineers as a percentage of the consumer demand, indirectly interacts with the long-term investment incentives. The importance of this effect has been analyzed in (Stoft, 2003) and (Hogan, 2005) and will be later discussed in section 0.

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 market⁹ 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.

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

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.

Buying reserves with of 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.

⁹ When this dual imbalance pricing is coupled the fact that imbalances are evaluated on a portfolio basis, a clear entry barrier for new entrants is introuduced.



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.

Therefore, 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 so as to reduce the expectation of undelivered energy. This would involve 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 order to perform such an assessment, the regulator has to define a methodology to evaluate the production capability of the different generating plants, taking into account current market incentives. In such a definition, the regulator introduces (its own) risk aversion criteria. This production capability is often estimated by means of historical data regarding availability, of historical production data, of historical availability or of production under scarcity conditions. In other cases, it is provided by a mathematical model (which has to be fed again with historical data), etc.

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.

Connecting with the next dimension, adequacy, if the regulator considers that there is a market failure in generation investment and that somehow a certain adequacy mechanism has to be introduced, then a measure of the production capability of the different plants (both existing and possible new investment) should be used in order to ensure an effective margin of production capability or simply to establish certain physical guarantees for the participants. In this sense, the production capability of the existing plants is usually considered to be a reasonable benchmark when trying to predict the future performance of potential new investments.

Adequacy

Again, the regulator's objective in terms of adequacy is to guarantee "adequate" incentives to attract new entrants (i.e. to attract new efficient generating units).

The regulatory instrument basically consists in assuring new entrants an extra source of income and/or the hedging instruments they require to proceed with efficient investments. As we will see later, the product that is received form generators in exchange may present many different forms. The definition of the time terms (lag period¹⁰, the duration of the incentive, etc.) and the volatility associated with this additional income 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.

¹⁰ 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.

There has been much discussion about the convenience of introducing regulatory measures to enable electricity markets to provide a reliability level with which the regulator feels comfortable.

In the next section, we discuss in detail how market signals are not enough and for several different reasons, but before delving in the existence of this market failure, we briefly comment on the last dimension, strategic expansion policy, which is very close related to the one that has just been presented.

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 (fuel price spikes, fuel exhaustion, etc.). Nevertheless, it may be appropriate to consider the introduction of some very long-term criteria in the adequacy mechanisms; criteria which will reflect the long-term view of the regulator.

The introduction of such criteria is usually supported 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.

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 sustainable¹¹.

Document structure

After this brief introduction and discussion of key terminology, the remainder of the document has been organized as follows: in section 2 we discuss the market failure on the security of electricity supply issue at the firmness and (mainly) adequacy level. Then in section 3, based on the international experiences so far, we analyze the different mechanisms aimed to deal with this problem. Later, in section 4 we highlight the different elements to be taken into account in the design of a long term security of supply mechanism. Finally, in section 5 the major conclusions of the present work are summarized. Some references have been provided at the end.

¹¹ 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.



2 Security of electricity supply at the generation level: problem analysis

2.1 Is the market capable of ensuring a reliable supply?

Though motivations may differ in each specific case, the universal *leit motiv* in electric power industry reform is the need to seek new vehicles, new regulatory models, to channel the necessary strategic expansion of electric infrastructure in general and generation facilities in particular.

The changes in the regulation of the electric power industry worldwide have modified the traditional reliability issues and approaches drastically. In the vertically integrated utility, under cost-of-service regulation, reliability 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 document 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 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 needed¹².

Next we discuss this issue, demonstrating that the answer to the question is that actual markets need some regulatory intervention, reviewing first the major results stemming from the marginal theory applied to electricity markets, and showing how short-term prices, under ideal hypothesis, are supposed to drive and efficient operation, planning and investments.

¹² Indeed, as shown in (Batlle & Rodilla, 2010), 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.

2.1 Ideally the market solves the problem

Under a market-based scheme, driven by demand and supply laws, an equilibrium price¹³ and an equilibrium quantity are determined as the result of generators and demand interaction in the market¹⁴.

In perfect competitive short-term markets all plants' supply bids should reflect their actual production costs, and consequently, the equilibrium price (also known as the system's *marginal price*) should ideally represent the demand's marginal utility, which except in the case of an scarcity in the generation resources, 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 ideal hypotheses are:

- Generators' costs functions are convex¹⁵.
- Risk is allocated efficiently. That is, there is a well-functioning long-term market. Indeed, most of the theoretical analyses are based on the risk neutrality assumption.
- 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.
- There is a perfect competitive market with perfect information¹⁶.

It can be demonstrated that under 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). Thus, if less generation than the optimal and efficient amount is installed, then the market provides higher profits for

¹³ 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 lower prices, 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).

¹⁴ In simple auctions the market equilibrium is directly determined by the intersection of the supply's bid curve and the demand's offer curve.

¹⁵ 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).

¹⁶ 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.



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 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.).

2.2 Market imperfections and flawed regulatory rules

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

• The first imperfection is the non-existence of the ideal conditions to introduce the "reform" at the generation level. We will 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 (no economies of scale, vertical unbundling, an adequate horizontal structure, etc.).

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 imperfection is the lack of short-term demand elasticity. Although ideally, the most efficient result would be achieved if prices were perceived by the demand-side in real time, 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 a scarcity, the price could be determined by the demand offer curve. Otherwise markets will not provide effective incentives for efficient resource management, either to new generators to come or to generators that are already present in the market.

Unless indicated the opposite, it will be assumed that there is not demand elasticity in the short term.

Apart from these two previous imperfections, we next analyze the impact of the nonaccomplishment 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.

2.2.1 The effect of the existence of the lumpiness problem in generation

Investments in generation are lumpy, that means 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 described in section 2.1 and 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 section 3), leading to a dash for extremely expensive 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 an (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

2.2.2 Agents risk aversion and the consequences of an inefficient allocation of risk

Power generation investment decision-making risk is high and failures²⁰ 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 centralised co-ordinator 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.

In this context, and from the point of view of the required generation investments, incentives to make decisions efficiently are weaker, and errors in planning are paid for by (not always all) customers via tariffs²¹ 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 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 for itself and profitability is in principle not guaranteed ex-ante. Analogously, each agent has to decide the medium term

²⁰ 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.

²¹ The nuclear moratorium in Spain is a good example.

resource management of its generating plants, and again the profitability of its decisions is not guaranteed.

From 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 management²² (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 requirements²³, 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 will affect the availability of electricity in the future, and thus, the system's reliability. 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 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 presented²⁴. 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%, for 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

²² The importance and the effect of generators' risk aversion depend on the particular structure and characteristics of the system.

²³ Some contracts may imply rigid constraints as is the case with the "take or pay" or "use it or lose it" modalities.

²⁴ We take Brazil as an example, but we could have also resorted to the Colombian or New Zealand cases.



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).

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 markets²⁶, 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.

²⁶ Indeed this risk should lead to the development of well-functioning, efficient and liquid long-term markets which make possible the determining of 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.

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.
 - 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 long-term contracts, it reduces generators' risk exposure, and consequently their required expected rate of return.
 - By entering into long-term contracts a more efficient medium-term resource management can also be achieved, see (Rodilla et al., 2010).
 - Hence, even if the demand was 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) of 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 maximisation 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 forces²⁷.

²⁷ 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) a long-term 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



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"²⁸.

The need of learning processes of immature electricity markets

Real 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. Most consumers are not mature enough to realize the risks involved in electricity markets and in these cases they tend to make their decisions using only very short-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 nothing³¹. 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.

transparent framework for trade, but if the market structure conditions are not adequate, artificially implementing a trading floor makes no difference.

²⁸ 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).

³¹ Although some educational programs which provide information about the potential consequences of not contracting may help to reduce the potential impact.

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 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.

2.2.3 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,
- 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 in the determining of marginal market pricing formation have taken many different forms:

Price or offer 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", i. e. codes defining constraints 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 generating units 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.
- Operating reserves contracting. Often the System Operator (SO), following the instructions of the regulator, purchases a certain number, or all, of the peaking units in the system or guarantees them an additional out-of-the-market payment. Examples of this approach are the Swedish case, in which the SO directly bought peak generators, or New Zealand's Dry Year Reserves mechanism (MED-NZ, 2003), in which the SO contracts strategic hydro capacity reserves to be dispatched at its own discretion, particularly when the SO considers the reserve margin is tight. Such market segmentation ensures peaking units' remuneration, but it distorts the marginal signal that the other generating units require to recover their investment. In the long term, this policy will impact on investment in base-load plants, and the regulator, through the System Operator, will have to steadily increase its intervention. Later (in section 5) a more detailed description of these mechanisms is provided.
- "Operating reserve shortage" actions. In other 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). These types of measures complicate the price formation process in conditions of scarcity, and again affect the proper and expected recovery of generation investments.

All these 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. 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. With respect to the latter, 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.

2.3 The need for regulatory intervention

Up to this point, we have demonstrated 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"³², 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.

³² 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).



3 A critical assessment of the different approaches aimed to secure electricity generation supply

In section 2 we have seen how, for many different reasons, the market if left to its own devices does not seems to be capable of ensuring an efficient supply in the long term. The objective of this section is to provide a categorization and a critical review of the different approaches that have been used in the different contexts to deal with this problem. This analysis will help to set the conceptual background that is later used in section 4, where the different design features a regulator should consider when introducing a long-term security-of-supply-oriented mechanism are presented.

When the so-called "market reform" started, the expectation was that little by little market agents, especially demand, would be able to learn the market game and therefore no additional mechanism would be needed to guarantee an adequate security of supply level. But, the reality nowadays is that security of electricity supply is more and more turning into a priority in the agendas of electricity regulators.

Indeed, (Ofgem, 2010) and (CEER, 2009) are two of the consultation processes opened at the time of writing this work that represent two good illustrative examples of the importance of this concern at the present time. Both initiatives are aimed to receive feedback from the different stakeholders on how to ensure security of supply at all levels³³. Another clear example is the number of systems worldwide (as it is for instance the case in Ireland, Panama, Peru, etc.) that have been recently or are currently in the process of revisiting their long-term mechanisms design.

As we introduced in section 1, 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 energy policy (a very long-term issue). As previously pointed out, there is a certain consensus around the idea that, on one extreme, the security dimension can be tackled by means of operation reserves markets³⁴, 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-market" mechanisms (e.g. feed-in tariffs or cap and trade mechanisms). But in between these two dimensions, the debate on firmness and adequacy (particularly this latter) has always been, and still is, quite intense.

3.1.1.a Security of supply mechanisms in deregulated electricity markets

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:

³³ 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).

³⁴ Nevertheless, some attention should be devoted to the interaction between reserves requirement and long term signals, see (Stoft, 2002) and (Hogan, 2005).

- Do nothing; in the belief that the market will provide the efficient 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", as we will call it hereafter) aimed to ensure system security of supply (i.e. 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:

- 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.

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 itself³⁵.

- 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"³⁶.
- 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 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.

³⁵ 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).

³⁶ For a detailed analysis on this reliability product, see (Rodilla et al., 2010).



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.

We first 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 quantitybased 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.

3.2 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 presented in section 2.1 and also in the Annex, 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). In this way, inframarginal energy revenues (the so-called scarcity rents being particularly important³⁷) 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

³⁷ 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.

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 the 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 scarce situations and described later in deeper 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 take 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.
- 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 longterm 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 opt 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 the incumbent.

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 (France³⁹, Italy or Portugal are some examples)⁴⁰.

Indeed, in the European case there are "latent" security of supply mechanisms thanks to Directive 2005/89/EC⁴¹, 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.

³⁹ 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'.

⁴⁰ 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.

⁴¹ 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.

- 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 these requirements, above the actual needs to face exclusively very short-term security issues⁴² 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).

3.3 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 strategy energy 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). See (Rodilla et al., 2010) for a detailed discussion on this issue.

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

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.

⁴² Note that a scarcity in generation supply is not a very short-term issue.



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 provide 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).

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)⁴³ 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).

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 PPAD⁴⁴) 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 lay in the fact that it was only provided to those generators producing in each hourly interval. The problem that rapidly

⁴³The 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.

⁴⁴ PPAD (*Potencia Puesta A Disposición* in Spanish) makes reference to the capacity being committed in the market by the generation unit.
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.

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 requirement⁴⁵) 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 of 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 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 2006⁴⁶.

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 operation⁴⁷ an annually capacity payment (expressed in euros) per installed megawatt.

This investment incentive depends on the value of a so-called "reserve margin index" ("*indice de cobertura*" in Spanish, or IC) that has to be calculated by the system operator. While the value of

⁴⁵ 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 force being committed to receive the payment.

⁴⁶ 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.

⁴⁷ 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 mainly nuclear and hydro plants (installed under the former regulated context) are receiving in the new market scheme.



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.

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.

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.

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 lost 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 units⁴⁸ 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).

⁴⁸It 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.

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.

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 (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/kWh⁴⁹) 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.

3.3.1.a 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).

3.4 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

 $^{^{49}}$ As for year 2006, 1 euro \sim 1200 won.



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

We next review 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.

3.4.1 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).

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)⁵¹.

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.



Figure 5 Average capacity prices in PJM - Source: PJM

⁵¹ 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.



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 periods⁵².

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

⁵² 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.

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.

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" consists 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.

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.

3.4.1.a 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 of 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.

3.4.2 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.

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 review 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 periods⁵³ but also did not constitute a stable and trustworthy long-term signal to potential investors⁵⁴. 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.

The reliability option⁵⁵

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 situations⁵⁶. 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.

⁵³ 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.

⁵⁴ A capacity payment is nothing but a regulatory compromise. In a context like the Latin American, in which regulatory risk perception for investors is large, 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.

⁵⁵ In the finally implemented mechanism, commissioned to Cramton & Stoft (2007), the reliability product was named Firm Energy Obligation.

⁵⁶ In practice, this results in a strike price that is set at a level slightly higher than the most expensive unit's marginal costs.



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).

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 riding⁵⁷.

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).

3.4.2.a 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

⁵⁷ 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.

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 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 & Pérez-Arriaga, 2008).

ISO New England

In ISO-NE, the so-called Forward Capacity Market (FCM) replaced the previous ICAP mechanism, see (Cramton & Stoft, 2005), (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.

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 6, 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 6 Timing of the RPM Auctions. Source: (PJM, 2007)

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.

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.

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.

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).

3.5 Strategic reserves as the reliability product

This last category has been left to include 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 SO considers, according to more or less objective criteria, that it is necessary.

3.5.1.a 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 accepted⁵⁸ 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 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.

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 years⁵⁹. 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

⁵⁸ 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).

⁵⁹ 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

scarcity situations⁶⁰, 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 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)

⁶⁰ For instance, the Whirinaki reserve energy trigger price was set at 38.7c/kWh (\$387/MWh) in its December 2008 update.



4 Principles and criteria to design security of supply mechanisms

We have seen in the previous sections how the market, left to its own devices, is assumed to be unable to provide sufficient generation availability when needed without regulatory intervention, and also how the solution to the problem necessarily entails the development of additional mechanisms to assure firmness and adequacy of supply. 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.

In all cases, regulation aims to achieve prices that lead to socially "optimal" system performance. These prices are the "signals" used by generators as the basis for operating their plants and planning their investments. Some type of regulatory objective (what is to be optimized?) is, of course, necessary to achieve this aim. The target or objective is usually assumed to be the so-called "net social benefit": the sum of generation and demand surplus.

This main objective of maximizing the "net social benefit" at the power system level is 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. 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 it considers to be an 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 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.

4.1 Design elements

Generally speaking, all the mechanisms which aim to solve 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 to 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 this reliability product plays a very important role in the whole mechanism performance.

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; that is, those being represented by the regulator in the mechanism are not always the only ones enjoying a higher level of reliability. 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 an equal supply. 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 has also 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.

The reliability product

It is important to properly define what 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. This is known as the "reliability product".

Determining the product to be bought from the generation is of the utmost importance and complexity. 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. 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 section 2.2.1). 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 season⁶². 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

⁶² 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).



(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 below.

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 which 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 these case represented by the regulator). Additionally, they help to reduce market power and also provide more information about how far the system is from suffering a scarcity.

Other details

The regulator also has to determine how the product will be bought, that is, 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.
- arranged through bilateral agreements or 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 three 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 that a 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 level of reliability but also the type of generators that will provide it.

For example, if the regulator opts for auctioning the long-term premium as, for example, in the case of the New England Forward Capacity Market (or the Colombian or Brazilian cases; indeed, generally speaking, all the mechanisms discussed next in section 5), it is implicitly and significantly conditioning the future generation technology mix, see discussion in next session.

4.2 The reliability product: some examples and design issues

Regulators have used a large number of approaches worldwide in their attempt to secure longterm 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 payment 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 section 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 reliability that a newly installed MW provides to the system 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 the giving of incentives to those MWs that can further increase reliability.

Indeed, highly inefficient "junk" generation plants can take advantage of such a definition of the product (as 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 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 no matter if a price cap is in place, 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 periods⁶³, 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 stability. 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 nonspinning 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.

⁶³ 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.

- 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).
- 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).
- 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, which has just been described.

• 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 (Ernesto et al., 2006).

Two extreme alternatives: setting the quantity versus setting the price

Once the regulator has established the reliability product, the next step involves defining the pricequantity 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.

5 Conclusions

Under the market-oriented paradigm, the new regulation must make sure that the appropriate incentives exist so as to ensure an efficient long-term security of supply level.

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.

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

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. Throughout the paper, 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 this paper, 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.

5.1.1.a The regulation design problem, not the market problem

In the light of the evidence discussed throughout this paper, 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. The main aim of our work has been to highlight the fact that 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 present⁶⁴.

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, see (Batlle & Pérez-Arriaga, 2008). While we have intensively reviewed almost all the cases in which a market-like approach has been implemented, providing a critical analysis of the failures, we have not looked at other electricity systems in which the reform has not been implemented since these have fallen outside the scope of this report. However, it should not be forgotten that these unreformed markets have not escaped similar or even worse problems. In this respect, the latest news 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.

⁶⁴ It is difficult to better illustrate this statement than the way Alfred Kahn did in "The economics of regulation" (MIT Press, 1988): 'Continued deregulation is the proper way to go, to the extent feasible...The central institutional issue of public utility regulation remains finding the best possible mix of inevitably imperfect regulation and inevitably imperfect competition. All competition is imperfect; the preferred remedy is to try to diminish the imperfection. Even when highly imperfect, it can often be a valuable supplement to regulation. But to the extent that it is intolerably imperfect, the only acceptable alternative is regulation. And for the inescapable imperfections of regulation, the only available remedy is to try to make it work better'.

6 References

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

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, 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., C. Vázquez, J. Barquín (2007B). "A critical analysis of the current balancing mechanism in France". IIT Working Paper, 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., I. J. Pérez-Arriaga (2008B). "The changing role of the State in the expansion of electricity supply in Latin America". IIT Working Paper IIT-08-061A, 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.

Barroso, L. A. (2009). "Power sector deregulation in Brazil. Ensuring supply adequacy through energy auctions". Master on the Electric Power Industry. Universidad Pontificia Comillas, Madrid, 2009.

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

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

Bidwell, Miles. (2005). "Reliability Options". Electricity Journal, Vol. 18,-5, pp. 11-25.

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

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

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

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.

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

Chao, H. (2008), "Overview of the Capacity Markets in the United States", APEX Conference, Sydney, Australia, October 2008

CND (2008), "Reglas Comerciales" (In Spanish), Centro Nacional de Despacho (CND), 2008

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

Concept Consulting Group (2004). "Developing Emergency Security of Supply Provision", A discussion paper prepared for the Energy Commission (New Zealand), November 2004.

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

Cramton, P. & S. Stoft (2006). "The convergence of market designs for adequate generating capacity". University of California Energy Institute Working Paper, 25 April.

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.

Hobbs, B. F. (2005). "Affidavit of Benjamin F. Hobbs on Behalf of PJM Interconnection. L.L.C." FERC Docket No. ER05-1410-000, 2005.

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", October 2009.

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.

IMO (2009). "Electricity industry (wholesale electricity market) regulations 2004. Wholesale electricity market amending rules (1 February 2009)", Electricity Industry Act, 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.

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

MED, Ministry of Economic Development (2003). Electricity Supply Security: Questions and Answers. 20 May 2003. Available at <u>www.med.govt.nz/templates/Page 17418.aspx</u>.

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 <u>www.boe.es/boe/dias/</u> 2007/09/29/pdfs/A39690-39698.pdf (in Spanish).

Ministerio de Economía, Fomento y Reconstrucción (2008). Nº 39.048, Diario Oficial de la República de Chile. Abril 2008.

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

NGET, National Grid Electricity Transmission (2010). The Grid Code. Issue 4 Revision 2. 22nd March 2010. Available at <u>www.nationalgrid.com</u>

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), "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. (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, 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.

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.

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

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

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

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

Shanker, R. J. (2003). "Comments on Standard Market Design: Resource Adequacy Requirement". Federal Energy Regulatory Commission, Docket RM01-12-000, January 10, 2003.

Sioshansi, F. P. (2008) "Competitive Electricity Markets: Design, Implementation, Performance", Elsevier, ISBN: 978-0-08-047172-3, 2008.

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", available at <u>http://www.terna.it/LinkClick.aspx?fileticket=Qmyh8X33TCQ%3d&tabid=2578</u> (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., C. Batlle, M. Rivier and I. J. Pérez-Arriaga (2003). Security of supply in the Dutch electricity market: the role of reliability options, for The Office for Energy Regulation (DTe) of The Netherlands. Presented at the workshop CEPR Competition & Coordination in the Electricity Industry, Toulouse, January 2004.

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" (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.

Wolak, F. (2004). "What's Wrong with Capacity Markets," Mimeo, Stanford University.



Annex A Optimal short-term prices under ideal hypotheses

A.i Theoretical results under ideal hypotheses

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

The application of the 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 that of 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 an efficient operation, planning and investments.

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

- 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.
- A perfect competitive market has been assumed.
 - In a perfectly competitive market, short-term generation bids should equal marginal production costs⁶⁵. This is actually a theoretical concept, which never materializes completely in the real world. 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.

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:

⁶⁵ This is, however, subject to certain conditions. Firstly, the various agents must be able to communicate readily (plant owners must be able to convey to possible buyers their willingness to produce at a price lower than the going market price), must not be able to affect the market price, and so on.

Policy and regulatory design on security of electricity generation supply in a market-oriented environment Problem fundamentals and analysis of mechanisms

$$\begin{array}{l} The \ central \ planner's \ problem \\ Max \ \sum\limits_{\substack{q_{ih} \\ q_{ih}}} \sum\limits_{\substack{h}} [U_{dh}(\sum\limits_{i} q_{ih}) - \sum\limits_{i} C_{i}(q_{ih})] \\ s.t. \\ q_{ih} \leq \overline{q}_{ih} \ \perp \psi_{ih} \\ R(q_{ih}) = 0 \ \perp \zeta_{ih} \end{array}$$
(Eq.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 $Q = \sum_i q_{ih}$.

 \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:

$$\begin{array}{l} The \ central \ planner's \ problem \ optimality \ conditions \\ \\ \frac{dU_{dh}(\sum\limits_{i}q_{ih})}{dq_{ih}} - \frac{dC_{i}(q_{ih})}{dq_{ih}} + \psi_{ih} + \frac{dR(q_{ih})}{dq_{ih}}\zeta_{ih} = 0 \Rightarrow \\ \\ \Rightarrow \frac{dU_{dh}(Q_{h})}{dQ_{h}} = \frac{dC_{i}(q_{ih})}{dq_{ih}} - \psi_{ih} - \frac{dR(q_{ih})}{dq_{ih}}\zeta_{ih}, \forall i,h \end{array}$$

$$(Eq.2)$$

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:

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:

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

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.

Note that short-term prices should always be determined by the marginal demand utility. These short-term prices are also equal 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 result that will be used in the following: 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:

$$\begin{split} & \underset{\overline{q}_{ih}}{Max} \quad NSB(q_{ih},\overline{q}_{ih}) - \sum_{i} IC_{i}(\overline{q}_{ih}) \\ & NSB = \begin{cases} & \underset{q_{ih}}{Max} \sum_{h} [U_{dh}(\sum_{i} q_{ih}) - \sum_{i} C_{i}(q_{ih})] \\ & s.t. \\ & q_{ih} \leq \overline{q}_{ih} \quad \perp \psi_{ih}, \; \forall i,h \\ & R(q_{ih}) = 0 \quad \perp \zeta_{ih}, \; \forall i,h \end{cases} \end{split}$$

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:

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

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:

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

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

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

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

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:

Demand's optimality conditions Generators' optimality conditions

$$\begin{split} \frac{dU_{dh}(Q_h)}{dQ_h} &= \pi_h, \forall h \\ \pi_h &= \frac{dIC_i(\overline{q}_{ih})}{d\overline{q}_{ih}} \Rightarrow \psi_{ih} = \frac{dIC_i(\overline{q}_{ih})}{d\overline{q}_{ih}}, \forall i, h \\ \pi_h &= \frac{dC_i(q_{ih})}{dq_{ih}} - \psi_{ih} - \frac{dR(q_{ih})}{dq_{ih}} \zeta_{ih}, \forall i, h \end{split}$$

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.

The upper part of Figure 7 (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 are increase rapidly with the hours of use. As indicated above, 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.

Policy and regulatory design on security of electricity generation supply in a market-oriented environment Problem fundamentals and analysis of mechanisms



Figure 7. "Perfectly adapted" generation system

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)⁶⁶, 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 7 shows the most efficient alternative for a determined mode of operation (hours of use) of a generating unit. 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

⁶⁶ As is discussed in below, there are many markets in which generators are only allowed to bid their plain variable costs, which leads to lack of income for these peaking units. This feature leads to the need to implement an additional mechanism which is aims to guarantee the recovery of the investment costs of these plants.



 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. not to do it.

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 demand monotone 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 7. These represent 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, the technology 3 sets the system'S marginal price, which equals its variable cost C_3^V (see the lower graph in Figure 7). 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 and 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 existing difference between technology 2's variable cost and its own variable cost. Graphically, this is equal to the difference between the slope of the overall costs curve, in other words, the tangent of the angle. Thus, group 3 obtains a profit equal to the price spread for the





Figure 8. Detail of Figure 7: 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 $tg\beta \cdot (T_1 - T_0)$, which is the segment b, and analogously the segment c, for the interval from ⁶⁶
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 to aid the understanding, 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 costs⁶⁷ 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 9, 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.

⁶⁷ Including depreciation and a rate of return on debt and equity capital



Figure 9. Short-term impact on investment cost recovery of regulatory interventions.

But the impact of these kinds 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 9 as a reference, Figure 10 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.

Policy and regulatory design on security of electricity generation supply in a market-oriented environment Problem fundamentals and analysis of mechanisms



Figure 10. Long-term impact of regulatory interventions on generation structure, reliability and prices



Santa Cruz de Marcenado 26, E-28015 Madrid - Tel: +34 91 542 2800 - Fax: +34 91 542 3176 - http://www.iit.upcomillas.es/

