Working Paper IIT-15-088A. First version: March 2015. This version: June 2015

Submitted to Economics of Energy & Environmental Policy on April 17th, 2015

THE NEED FOR NON-PERFORMANCE PENALTIES IN CAPACITY MECHANISMS: CONCEPTUAL CONSIDERATIONS AND EMPIRICAL EVIDENCE

Paolo Mastropietro [a, *], Pablo Rodilla [a], Carlos Batlle [a, b]

[a] Institute for Research in Technology, Comillas Pontifical University, Sta. Cruz de Marcenado 26, Madrid, Spain.

[b] Also with MIT Energy Initiative, 77 Mass. Av., Cambridge, US and Florence School of Regulation, Florence, Italy

[*] Corresponding author. IIT, Comillas Pontifical University, Sta. Cruz de Marcenado 26, Madrid, Spain.
Tel.: +34 91 542 2800 ext. 2746. E-mail: <Paolo.Mastropietro@comillas.edu>

This article analyses international experience with performance incentives in capacity mechanisms. The existence of significant financial penalties for failure to comply fully with commitments to a capacity mechanism creates an incentive for resources to invest in measures to enhance their performance under scarcity conditions. Such investments, in turn, improve power system reliability. Performance incentives may be linked to constraints on tradable quantities or eligibility criteria, also studied in this paper, which limit the amount of reliability that a given resource may trade (and for which it may be penalised). The empirical evidence compiled in this article reveals current trends in international regulation and provides guidelines for implementing more effective capacity mechanisms.

Keywords

Capacity mechanisms; performance incentives; explicit penalty; reliability product; system adequacy.

1 INTRODUCTION

Capacity mechanisms are regulatory instruments designed to reinforce the economic signal provided by short-term electricity markets with additional remuneration to attract investment and ensure system adequacy in liberalised power sectors. These mechanisms are edging up on the political agenda in nearly all systems with some manner of electricity market. In North and South America (Batlle et al., 2015), where these incentives were included in the original market design or implemented soon after restructuring, capacity mechanisms are being intensely reworked to improve their outcomes in light of the new challenges facing these systems. The challenge in Latin America is the vulnerability to drought, given the large share of hydropower in the generation mix, and in the United States, certain new concerns around flexibility in the interaction between the power and gas industries. In Europe, where many regulators opted for the so-called energy-only market when liberalising the industry, capacity remuneration mechanisms are now being implemented or are under design in several countries (United Kingdom, France, Italy and Germany, among others) to respond to local or regional issues.

Capacity mechanisms are necessary to offset market failure, which detracts from the efficiency of the price signal (Neuhoff and De Vries, 2004; Joskow, 2008; Cramton et al., 2013). Market failure is generally identified with factors such as investors' risk aversion, lumpiness of investment in generation facilities and the inelasticity of electricity demand. In particular, demand inelasticity is believed to impede proper market clearing when the reserve margin is tight, and that, in conjunction with other more politically-oriented reasons, leads regulators to set price caps. The absence of the efficient signals that would be emitted by prices higher than such official thresholds also removes the natural incentive for generators to be available at times of system stress, when the reserve margin is narrow and the risk of demand curtailment consequently high. These "missing performance incentives" pose a threat to system reliability (ISO-NE, 2012).

In (imperfect) markets where price signals are distorted, the missing performance incentives problem can be minimised by precisely defining the capacity mechanism product to be purchased, especially as regards performance assessment. The importance of linking capacity mechanism remuneration to each resource's actual contribution to security of supply has been highlighted by authors working on capacity mechanism design (Vázquez et al., 2002; Bidwell, 2005; Finon and Pignon, 2008). A need has been identified for strong penalties for underperformance in scarcity conditions, firstly to discourage bids from non-firm generation, and secondly to enhance agents' incentive to manage and operate their resources in ways that raise their availability in such events.

These factors were given insufficient consideration in early capacity mechanism design. In Latin America, where initial capacity payments remunerated generation facility availability (with no further definition of what that meant), actual performance played almost no role in revenue flows. Long-term auctioning mechanisms introduced subsequently corrected some of the flaws in these early schemes (see Batlle et al., 2010, for details), but placed little emphasis on penalising underperformance. The pioneering reliability options mechanism implemented in Colombia, for instance, included no explicit penalty scheme, even though that element was the cornerstone of Vázquez et al.'s (2002) original proposal.

A slightly higher remuneration-performance correlation was gradually introduced in the capacity mechanisms implemented in the United States. ISO New England explicitly included a penalty for non-compliance in its capacity market. However, as explained in ISO-NE (2012), "at times of greatest need, many resources are delivering far below the performance ability represented in their supply offers" (quantified at up to 40% of the additional power required by the System Operator during contingencies). This situation resulted in a wave of reforms based on a "pay-for-performance" approach, which are being introduced at this writing[1].

A capacity mechanism that in addition to providing a long-term investment hedge for market agents also ensures security of supply calls firstly and foremostly for due definition of the

[1] The arguments put forward in these two paragraphs should be set against the backdrop of "real-life" regulation. The introduction of a capacity mechanism always entails negotiations among all the agents involved, whose conflicting interests must be integrated. In this context, the absence of a robust penalty scheme in a capacity mechanism is never the result of regulator "oversight", but rather the outcome of this "bargaining" between the CRM designer and the reliability providers, who normally oppose heavy penalties for non-performance.

reliability product to be acquired[2]. The design elements for this product have been identified by Batlle et al. (2015). This article focuses on the two major (non-conflicting) approaches to coping with scarcity conditions through capacity mechanisms, namely constraints on tradable quantities and eligibility criteria (analysed in section 2), and performance incentives (studied in section 3). The latter are analysed in some detail, identifying the major features of penalty schemes (critical period indicator, penalty rate, overperformance credits, exemptions, and penalty caps). The analysis covers the experience of several countries particularly prominent in this regard (Colombia, ISO New England, PJM, United Kingdom, and France) and includes very recent developments. Section 4 summarises the research findings.

2 CONSTRAINTS AND ELIGIBILITY CRITERIA: FIRM ENERGY AND FIRM CAPACITY

As Batlle et al. (2015) point out, market agents are theoretically better able than anyone else to estimate the expected contribution from their facilities in scarcity conditions. If a sound explicit penalty scheme is in place (including a demand for "collateral" to provide the system with a financial guarantee in case of curtailments), then, there should be no need for the regulator to impose limits on the amounts of reliability product that each resource can trade in a capacity mechanism or to establish criteria that must be met to earn capacity remuneration. Nevertheless, most capacity mechanisms implemented to date include such features, reflecting regulators' mistrust of market agents' estimates and fear of the concomitant power shortages.

[2] By reliability product is meant the product/service that agents are required to offer in the framework of the capacity mechanism and the concomitant obligations, regardless of the underlying security of supply dimension. The term capacity mechanism itself is becoming more and more misleading as the product traded under these mechanisms grows more complex and deviates from mere capacity.

Constraints on tradable quantities and additional eligibility criteria (requiring any manner of guarantees) consist in quantity caps assigned by the regulator to each agent willing to participate in the capacity mechanism. Such caps should reflect its firm energy or capacity, i.e., the energy and capacity that the resource can provide in scarcity conditions. Constraints, normally applied in a qualification or verification phase, are based on each resource's historical or statistical performance. Strategies differ depending on the technology and include, for example, de-rating of thermal plant capacity on the grounds of the equivalent forced outage rate, or using a stochastic optimisation medium-term planning modelling tool to calculate the maximum energy that a hydropower plant can produce in a dry year. An example of the eligibility requirements for participation mentioned earlier would be for gas-fired plants to be in possession of a long-term contract for the firm supply of natural gas.

If properly implemented, limiting the amount of reliability product that can be traded by each resource to the contribution expected by the regulator from that resource in scarcity conditions reduces the likelihood of hidden under-procurement and improves security of supply. Constraints on tradable quantities are particularly important where demand-response products are concerned, in which a baseline must be clearly defined to measure the actual contribution. Nonetheless, such ex-ante qualifications may also induce inefficiencies if the limits or warranties required by the regulator are overly demanding.

Another element to be borne in mind is that when constraints on tradable quantities reflect actual past performance, a resource subject to frequent outages could see its tradable volume diminished in subsequent capacity auctions. As this would lower the revenue obtainable by the resource from the capacity mechanism, it constitutes an "indirect" penalty for underperformance.

In Latin America, due to the high share of hydropower in the generation mix, the use of complex optimisation models to calculate the expected contributions is advisable. The region's systems have consequently always featured constraints on tradable quantities. A paradigmatic example of this approach can be found in Colombia, which is addressed in the subsection below. That analysis is followed by a discussion of experience in North America, specifically in ISO New England and PJM[3].

2.1 Constraints on tradable quantities and guarantees: Colombia

Colombia's hydro-dominated power system is particularly vulnerable in dry years such as induced in the region by *El Niño/La Niña* fluctuations, when low hydropower contributions must be offset by raising thermal plant output. Initially, this objective was pursued by introducing a capacity payment to remunerate generators based on their availability during the dry season. Batlle and Pérez-Arriaga (2008) showed that this design led to inefficient reservoir management by hydro-power resources. In an attempt to correct this and other flaws detected in the system, in 2006 the regulator introduced the reliability charge mechanism, in which market agents sell firm energy obligations (Spanish initials OEF, for *obligaciones de energía firme*)[4] on a centralised auction in exchange for a fixed annual payment.

---

[3] The conclusion that may be drawn from recent events in the US, where power systems suffered significant stress situations due to the lack of gas transmission capacity (see e.g., Babula and Petek, 2014), is that gas-fired plants may require similar treatment (i.e., a statistical analysis on the likelihood of fuel curtailment). Gas transmission constraints are not the only factor affecting generating unit performance, however. As illustrated by PJM (2014a), generator outages rise significantly as temperatures (measured as wind chill) decline. Support for that assertion came during the January 2014 polar vortex: PJM experienced tight operational conditions and a significantly higher number of forced generator outages than in a typical month of January.

[4] OEF are based on the reliability option contracts mechanism mentioned earlier, proposed by a team headed by Prof. Pérez-Arriaga (Vázquez et al., 2002) and ultimately commissioned from Cramton and Stoft (2007). The reliability option mechanism is based on financial call option contracts with a high strike price to be backed by

To be eligible to sell OEFs, resources participating on the Colombian market must be backed by ENFICC certificates (Spanish initials for *energía firme para el cargo por confiabilidad*, i.e., firm energy for the reliability charge), which establish constraints on tradable quantities. ENFICC certificates are assigned using methodologies that vary depending on the technology (see Mastropietro et al., 2014, for details). For hydropower plants, the expected contribution is calculated with an optimisation modelling tool that assumes inflows where the likelihood of being exceeded is high. For thermal plants, certificates are assigned on the grounds of installed capacity, track record of forced outages and fuel availability. As an additional guarantee, these plants are required to ensure the fuel supply (and, if necessary, the fuel transportation capacity) necessary to back the ENFICC assigned to them through advance contracts.

The first testing ground for the reliability charge mechanism was the dry year that affected the country in 2009/2010.

The regulator deemed that hydro generators' approach to resource management (generating at the beginning of the dry year to honour their bilateral contracts, as reported in CSMEM, 2011) would have resulted in very low reservoir levels at the beginning of the actual dry season. As a result, the regulator decided to intervene, incentivising thermal generation with a change in dispatching rules and consequently constraining hydropower plants. At the same time, exports of both electricity and natural gas were blocked to guarantee the use of national resources to supply domestic demand.

---

physical resources. When the spot market price exceeds the strike price, reliability providers are required to deliver the committed contribution and to return any positive difference between the spot price and the strike price, receiving the option premium in exchange. As noted, the original proposal also included an additional non-performance penalty, not included in the design ultimately implemented.

While these measures ensured water storage in the reservoirs, they also revealed the limitations of the natural gas transportation network. Despite the firm contracts with gas suppliers (to meet the aforementioned OEF eligibility requirements), only thermal plants on the coast received the natural gas committed to under contract. All other units received a lower supply, due primarily to structural insufficiencies in some pipelines (these units consequently contended that their lack of availability and the high market prices were not attributable to them). Since many inland facilities were able to operate as dual-fuel plants, liquid fuels were used to meet obligations. The infrastructure for transporting liquid fuels had not been tested either, however, and severe problems were detected in some cases. Finally, of the 93 GWh per day of firm energy obligations (backed by ENFICC certificates) contracted with thermal plants, only 80 GWh per day were actually delivered (CSMEM, 2011), due mostly to constraints in the supply chain of natural gas and liquid fuels. Despite these setbacks, the Colombian electricity system managed to operate with no demand curtailment in the dry year.

OEF mechanism performance during the 2009/2010 *el Niño* prompted criticism in several respects. The lower contribution from natural gas- and liquid fuel-fired plants revealed flaws in the ENFICC calculation methodology. The Colombian Market Monitoring Committee (CSMEM, 2010) contended that the firm energy expected from thermal plants proved to be overestimated, with fuel supply the factor primarily conditioning their actual contribution. ENFICC calculation methodology proved to be unable to reflect the shortcomings in the gas transportation network (and the associated risk) when assigning firm energy certificates to thermal plants. Hydro generators, in turn, argued that the regulator's intervention prevented them from demonstrating that they would have been able to perform better than the ENFICC limit they had been assigned. This raises the question of whether regulators have the fullest information and the necessary expertise to calculate each resource's expected contribution in scarcity conditions or whether this task should be left to market agents, contingent upon the existence of properly designed penalty schemes[5].

The Colombian experience, which provides for no explicit penalties for underperformance, affords the opportunity to reflect on performance incentives. As noted, contrary to the regulator's expectations, during the 2009/2010 *el Niño*, hydropower plants continued to operate early in the dry year to meet their bilateral energy commitments, progressively emptying reservoirs. The Colombian Market Monitoring Committee (CSMEM, 2010) concluded that this was an indication that hydropower companies preferred the risk of future non-performance of their firm energy obligations over the immediate economic loss they would have incurred had they interrupted production and purchased power on the market to honour their bilateral contracts.

A strong explicit penalty might have harshened the consequences of future non-compliance with OEF contracts, prompting hydropower companies to store water to avoid potential charges[6]. Such an incentive might have also mitigated the fuel shortage-induced underperformance of thermal plants. The risk of a very high penalty might have encouraged thermal generators to sign "fully" firm (primary or alternative) fuel supply contracts, making

[5] As a clarification, in Colombia market agents calculate their own ENFICC, but through a model or a methodology established by the regulator, who is also in charge of verifying and approving the results of the calculation.

[6] The regulator does not appear to be inclined to implement such measures, inasmuch as the CREG (2014) "institutionalised" the interventionist approach applied in 2010: Resolution no. 26 introduces indicators to identify periods at risk of shortage, during which additional dispatching rules are applied, allegedly to guarantee the reliability of supply.

provision for the potential charge in the agreement, thereby motivating gas suppliers to reinforce the transportation grid.

The country underwent another dry year in 2014. Although no regulator intervention was required on this occasion, the behaviours reported resembled the 2009/2010 reactions (CSMEM, 2014b). Nonetheless, hydropower plants were observed to increase their bids prior to the dry season, paving the way for earlier thermal plant dispatching (CSMEM, 2014a).

2.2 ISO New England and PJM and eligibility criteria

Thermal generation prevails in the ISO New England and PJM power systems. In ISO New England, 44% of electric power consumption is presently met by natural-gas-driven plants (ISO-NE, 2015). With the rising availability of low-cost shale gas, the role of this fuel is expected to grow in the near future. Generation project proposals in the pipeline consist almost exclusively of natural gas thermal plants (57%) and wind turbines (42%). Peak demand in PJM is nearly six times larger than in ISO New England. According to Monitoring Analytics (2014), at 40.5% of installed capacity, coal-fired power plants continue to be predominant in the PJM generation mix, followed by natural gas with 30.1% and nuclear power with 17.8%. That same report notes that from 2011 to 2019, nearly 20.5 GW of coal-fired capacity will be decommissioned and replaced with around 40 GW of gas-fired plants that will come on stream in the interim. This pattern mirrors the shale gas revolution and is expected to significantly increase PJM's dependence on natural gas, which will eclipse coal as the prevalent fuel by 2015/2016 (PJM, 2014c).

These data imply a challenging transition for these systems, partially explaining the reforms presently being introduced in their capacity mechanisms. The growing share of intermittent renewable energy will need to be backed by a thermal fleet that is also changing dramatically, with old (nuclear and coal) plants being replaced by new natural gas units. The major threat for ISO New England and PJM, as well as for many other power systems in the United States, is probably related to the actual reliability of the supply of this natural gas. Shale gas is revolutionising the country's energy consumption, with a significant proportion of demand shifting to natural gas to benefit from low prices. This shift renders the current infrastructure inadequate and calls for fast-paced reinforcement of the gas transportation network.

Moreover, this growing demand also intensifies competition for the supply of natural gas, particularly as regards the use of networks. MITEI (2013) notes that prior to the shale gas revolution, power generators did not secure firm gas supplies because interruptible contracts were less expensive and supply was almost never suspended. According to the same report, at this time "power plants in many parts of the country are unable to recover the higher cost of long-term firm contracts through their electricity market bids, and therefore rely on interruptible contracts for gas supply". This lack of firmness obviously constitutes a threat to the reliability of power sectors so heavily reliant on natural gas. The fact that plants burning natural gas purchase fuel on a day-to-day basis from LDCs (local distribution companies) increases the risk that generators may not be able to obtain the gas supply needed to run their plants when most needed. That problem is intensified by the current regulation on gas pipeline network expansion. As explained by MITEI (2013), "gas pipelines (are) built if and only if LDCs signed sufficient contracts for firm gas supplies".

Therefore, even if the right incentives were in place, a generating unit could find it impossible to conclude a firm gas supply contract unless a coalition of plants requested the installation of a new pipeline (and even then, guaranteed firm supply could be delayed).

Such concerns obviously affected deliberations on ISO New England's and PJM's capacity markets (the forward capacity market or FCM and the reliability pricing model or RPM, respectively). Both systems feature constraints on tradable quantities assigned during a qualification phase. Proposals were put forward to supplement these constraints with an eligibility criterion, which would allow gas-fired units to sell their power on the capacity market only where backed by a firm contract for natural gas supply. At this writing, however, the focus is apparently shifting from eligibility criteria to performance incentives (analysed in section 3). This choice is openly mentioned in PJM (2014b), which states that "based on stakeholder input, PJM also has eliminated many of the stated eligibility requirements and instead depends on stringent performance standards to allow resources to manage their own participation approaches". Although the interaction between the gas and power industries transcends this discussion, be it said that a number of amendments to the regulatory scheme will be necessary to ensure optimal exploitation of their synergies.

3 PERFORMANCE INCENTIVES

Performance incentives aim to ensure that agents assuming capacity mechanism commitments manage their resources in ways enabling them to meet their obligations when the system is tight or pay the consequences of underperformance. A critical period indicator must first be selected to identify scarcity conditions, during which the contribution committed to must be delivered and each resource's performance is assessed. Penalties or incentives are applied based on recorded performance during critical periods. Performance incentives and constraints on tradable quantities are not incompatible and are usually applied in parallel.

The discussion of the importance of performance incentives and of the difference between implicit and explicit penalty schemes in the subsections below is followed by a classification of performance incentives by their design elements based on recent experience on both sides of the Atlantic (ISO New England, PJM, United Kingdom, and France).

3.1 The importance of performance incentives

As noted in the introduction, the central role of performance incentives in the design of effective capacity mechanisms was identified, among others, by Vázquez et al. (2002), who contended that the implementation of explicit penalties is the only way to guarantee the installation of a sufficient level of firm capacity and its delivery in scarcity conditions. The consequences of the lack of an explicit penalty scheme or of the existence of a scheme with a limited impact on resources were felt by some of the systems that reformed their capacity mechanisms in the first decade of the century (Colombia, ISO New England, and PJM). The inefficiencies resulting from the lack of an explicit penalty for underperformance in the Colombian mechanism were described in the preceding section.

Contrary to that experience, the design of the more recent capacity mechanisms implemented in the ISO New England and PJM systems did include explicit performance incentives. ISO New England's FCM established so-called availability penalties, based on the calculation of shortage event availability scores. PJM's RPM, in turn, applies a peak-hour period availability (PHPA) charge based on actual obligor performance during 500 peak hours. Nevertheless, these performance mechanisms, which provided for many exemptions from penalties, did not yield the expected results and their limitations have become increasingly visible in recent years.

During recent scarcity events, many resources in the ISO New England system failed to deliver the full capacity specified in their forward capacity market supply offer, with average underperformance quantified as 40% of the additional power required by the System Operator during contingencies (ISO-NE, 2012). That obviously placed system reliability at risk. The System Operator attributed such significant underperformance to the fact that "capacity resources rarely face financial consequences for failing to perform, and therefore have little incentive to make investments to ensure that they can reliably provide what the region needs: energy and reserves when supply is scarce" (FERC, 2014b).

The road test for PJM's reliability pricing model came with the "polar vortex" event in January 2014. Extremely low temperatures not only prompted high electricity demand, but affected the generation capability of some plants rather directly. During the record-breaking winter peak

(141 846 MW), PJM experienced an equivalent forced outage rate (EFOR) of 22%, far in excess of the 7% historical average. The capacity shortfall relative to obligations (PJM, 2014c) amounted to 40 200 MW, 47% of which was accounted for by gas and 34% by coal-fired plants. In much the same vein as ISO New England, PJM concluded that "a capacity resource committed in RPM currently faces only limited and attenuated adverse consequences for failing to provide energy and reserves when needed" (FERC, 2014a).

With a view to correcting such flaws, both ISO New England and PJM are in the process of introducing new performance incentives, which tie each resource's capacity mechanism remuneration more tightly to its performance in scarcity conditions. In May 2014, the regulator partially approved ISO New England's pay-for-performance mechanism (FERC, 2014b). In December 2014, PJM presented its proposal for a capacity performance product, under study at this writing[7]. These schemes are analysed in greater detail in the breakdown of performance incentive design elements.

The importance of explicit penalties and performance incentives in the framework of capacity mechanisms is apparently being acknowledged in Europe (EEAG, 2015), where regulators are introducing a new wave of market-based capacity mechanisms. Performance schemes proposed by or implemented in the United Kingdom (DECC, 2014) and France (RTE, 2014) are also addressed in the design element analysis below.

[7] At this writing, it is difficult to say what the new performance mechanism will look like. PJM's proposal as reflected in the FERC filing has changed substantially relative to the drafts posted by PJM. The amendments required by the regulator are also difficult to predict. The design discussed in this article is based on FERC (2014a).

Before proceeding with the analysis, however, it should be noted that the design of capacity mechanisms, including any penalty schemes, is the outcome of negotiations between the regulator (and/or the team of experts selected by the regulator) and power industry agents. Some of the latter object to the introduction of heavy penalties, which could mean significant financial loss for underperformance. The implementation of a "soft" penalisation scheme or none whatsoever in CRM design is, then, never the result of regulator "oversight", but rather the outcome of this "bargaining" between opposing interests.

3.2 Implicit and explicit penalties

In the context of capacity mechanisms, care must be taken to clearly distinguish between implicit and explicit penalties. As explained by Batlle et al. (2015), an implicit penalty requires an underperforming agent to resort to the electricity market to acquire the reliability product it is unable to provide with its own means to honour its entire expected contribution. Explicit penalties, which adopt the form of an extra charge for non-performance, are applied in addition to implicit penalties. Here, underperforming resources are explicitly "fined" for failing to comply. Some documents written on this topic do not properly recognise the difference between implicit and explicit penalty (e.g., EEAG, 2015), but this is of paramount importance for the effectiveness of the capacity mechanism. Implicit penalties, in any event, are not performance incentives, for agents assuming an obligation subject to an implicit penalty have exactly the same incentive to produce as agents with no such obligation.

The following example may make this clearer. Imagine a system with reliability option contracts in place (see footnote 4) in which the strike price for reliability contracts is 100 $/MWh and the current spot price is 500 $/MWh. An agent with a reliability option and delivering its contribution obtains 500 $/MWh from the market but has to return 500 – 100 = 400 $/MWh to the system operator, for a net profit of 100 $/MWh (excluding variable costs). If the same agent is unable to meet its commitment, it will be required to pay the implicit penalty of 400 $/MWh, for a net loss of 400 $/MWh. The difference between delivering and not delivering is a net loss of 500 $/MWh. Agents with no reliability option contract, in turn, earn 500 $/MWh if they produce and 0 $/MWh if they do not. The difference between delivering and not delivering is also a net loss of 500 $/MWh. Clearly, the implicit penalty affords no incentive. This example shows how only a properly defined explicit penalty scheme can encourage agents to honour their obligations and ensure system reliability. Each design element of such explicit penalties are analysed in the subsections that follow.

3.3 Design elements of performance incentives

The remainder of this article focuses on the design elements for performance incentive schemes in capacity mechanisms. A thorough review of the past and present experience identified the relevant design elements to be the critical period indicator (to identify scarcity conditions during which the capacity obligation must be met and resource performance is assessed), the penalty applied for underperformance, overperformance payments[8], exemptions and penalty caps.

3.3.1 Critical period indicator

Batlle et al. (2015) discussed the advantages of using the market price as the critical period indicator. Of the systems analysed here, however, scarcity conditions are defined by the short-term market price under the Colombian firm energy obligations arrangement only[9]. In

---

[8] By way of clarification, here and throughout this article, "overperformance" should be construed positively, i.e., to mean an extra contribution to power system resources that improves reliability.

[9] The contracts signed in ISO New England's FCM are also financial options (when the real-time locational marginal price exceeds a certain strike price, obligation holders must return the difference between the former and the latter), although the physical obligation is linked to shortage events identified by the System Operator.

contrast, in many capacity mechanisms scarcity conditions are identified via technical grid parameters or emergency actions taken by system operators. Under such an approach, establishing a criterion that denotes the existence of actual scarcity conditions is of paramount importance, for otherwise committed resources would never, or very rarely, be required to perform their contract obligations.

In ISO New England, for instance, shortage events were initially defined as periods when the system exhibited shortfalls in 10-minute reserves for over 30 minutes. According to the FERC (2013), such a strict criterion had not induced a single shortage event at any time since the advent of the capacity market. This flaw was corrected by amending the definition of "capacity deficiency", which is now deemed to exist where 30-minute, i.e., operating, reserves are short.

In the new design for the PJM's reliability pricing model, outstanding approval at this writing, resources assuming capacity performance obligations commit to being available to supply energy and reserves whenever PJM determines that an emergency condition exists. Such conditions are "locational or system-wide capacity shortages that cause pre-emergency mandatory load management reductions or a more severe action" (FERC, 2014a).

A resource committed in the UK capacity market, in turn, is required to deliver during stress events, defined as "any settlement periods in which either voltage control or controlled load shedding are experienced at any point on the system for 15 minutes or longer" (DECC, 2014). Stress events are announced 4 hours in advance via capacity market warnings.

In France, a decentralised capacity market based on bilateral transactions is being implemented. The new mechanism seeks primarily to ensure that growing peak winter loads resulting from indoor heating systems can be met. According to the most recent detailed proposal (RTE, 2014), the days constituting peak periods are announced by the System Operator 1 day in advance. Electric power suppliers' actual demand and capacity providers' real performance are assessed on such days.

Lastly, depending on the critical period indicator, obligations may be fixed or load-following. When fixed obligations enter into effect, resources must deliver the entire contribution committed. Under load-following arrangements, the enforceable contribution is calculated by applying to the total commitment the ratio of the demand at issue to the peak demand covered by the mechanism. In other words, if a shortage event occurs when the demand is 60% of peak demand, the units committed must deliver at least 60% of their capacity obligation or pay the penalty. Of the mechanisms analysed in this article, Colombia and France have fixed obligations, whereas ISO New England, PJM, and the United Kingdom have opted for a load-following design. The potential for overperformance is higher when load-following arrangements are in place. In off-peak shortage events in which obligors are required to deliver 60% (for instance) of their committed capacity, they may try to overperform (supplying up to 100% of their capacity obligation) to earn extra revenue, if provided for in the mechanism. Conversely, load-following arrangements may hamper demand response, for such resources may find it difficult to contribute in off-peak times.

3.3.2 Penalty rate

The penalty rate established for non-compliance must be high enough to encourage investment in performance improvements, the ultimate aim of capacity mechanisms. Such improvements may consist in shorter start-up times, continuous staffing, more reliable fuel supply arrangements or dual-fuel capability. The lower the firmness of a resource, the greater the investment necessary to avoid very high penalties. The rate applied constitutes a signal for attracting firm capacity and impacts the merit order of the capacity auction heavily. It can be calculated either on the grounds of the capacity price cleared in the auction (which would link it to capacity remuneration) or of the net cost of new entry (net CONE)[10]. Another data item often used in this calculation is the expected number of scarcity hours during the year at issue.

In ISO New England, the System Operator established a high performance payment rate (based on the net cost of new entry) subject to a phase-in period. The rate will be increased from $2 000/MWh (2018/2021) to $3 500/MWh (2021/2024) and subsequently to $5 455/MWh (from 2024 onwards), as reported in FERC (2014b).

The penalty rate in PJM's proposed capacity performance product is also based on the net CONE. A comment to the Independent Market Monitor's proposal (FERC, 2015) estimated the non-performance charge rate to amount to $3 625/MWh.

In France, negative imbalances are settled at a so-called unit price (representing the penalty rate), which assumes different values depending on the security-of-supply status of the system. When security of supply is at risk, the unit price rises to an administratively-set price based on the net cost of new entry (RTE, 2014), for which no estimate has yet been provided.

In the United Kingdom's capacity market, in contrast, penalties are linked to the capacity remuneration. The penalty rate is 1/24 of the respective auction clearing price, adjusted for inflation (DECC, 2014).

[10] Methodologies to calculate the penalty rate based on net cost of new entry may vary. For instance, both FERC (2014c) and ISO-NE (2013) propose calculating the "performance payment rate" as the yearly net CONE (based on CCGT technology) divided by the expected number of hours with scarcity conditions and the expected performance during scarcity conditions: e.g., $106 394/MW-year / (21.2 hours/year x 0.92) = $5 455/MWh.

3.3.3 Overperformance payments

Overperformance payments may be envisaged to supplement underperformance charges. Depending on the design, performance payments may constitute significant revenues over and above the base capacity auction payment. This clearly strengthens the performance signal. However, as mentioned above, this approach is more effective in capacity mechanisms based on load-following obligations, in which the potential for overperforming is higher. If in an off-peak shortage event a resource is required to provide 60% of its committed capacity, for example, and delivers 100% of its obligation, its overperformance payment would amount to 40% of the total contribution. Further potential for overperforming lies in constraints on tradable quantities. If a resource were allowed to trade only 80% of its installed capacity on the capacity market due to de-rating, it could generate electricity at full load and receive overperformance payments for 20% of its installed capacity.

The payment rate can be set to equal the penalty rate to establish completely symmetrical performance incentives. It might otherwise be calculated for each separate shortage event by providing that charges collected from underperforming resources are to be distributed among overperforming resources. The second approach is preferred by some regulators (such as PJM and the United Kingdom) because it guarantees that payments never exceed charges (self-balancing mechanism). However, as shown in ISO-NE (2012), designs based on symmetrical charge and payment rates can achieve the same result if implemented properly. Examples of this latter approach are ISO New England's pay-for-performance scheme and France's capacity imbalance settlements.

The matter of who is entitled to overperformance payments is another factor that must be determined. In ISO New England, United Kingdom, and France, only resources participating in the capacity mechanism are eligible for performance credits. In PJM, in contrast, both RPM resources that exceed their load-following obligations and resources that did not sign on to the capacity market qualify for such payments. In the latter case, the entire contribution is regarded as extra performance.

3.3.4 Exemptions and penalty caps

Incentive arrangements may also envisage exemptions from performance and hence from penalties. Specific pre-established circumstances may partially or entirely exempt some resources from their obligations. Exemptions should be minimised, however, so as not to significantly attenuate the performance incentive signal. Some authors have called for the application of a strict no-exemption policy (Peter Cramton in FERC, 2014c). The principle that informs this approach is that performance incentives in a CRM should mimic the incentives in an ideal energy-only market. If a generator fails to deliver electricity to the grid during a price spike occurring in a perfect energy-only market, it forfeits that revenue regardless of the reason for its unavailability and of where the responsibility for such unavailability lies. This raises the question of which risks should be assumed by the resources taking part in the capacity market. Holding agents financially responsible for risks far beyond their control could be extremely counterproductive. In the opinion of the authors of this article, penalising resources is questionable when their non-performance is due to a transmission facility outage that disconnects the unit; or, where new plants are concerned, when the commitment cannot be fulfilled because of connection delays attributable to bureaucratic issues or to a delay in the construction of a planned transmission line. Nonetheless, these exemptions should be valid for the explicit penalty, while the implicit penalty, which is more related to the market cash flow, can consider fewer or no concessions.

Penalty caps may be applied to reduce the risk exposure of resources committed to a capacity mechanism scheme (particularly when scant or, better still, no exemptions are considered). Here also, exposure to the risk of large penalties is what prompts resources to make the investments necessary to improve their performance. Consequently, penalty caps must be designed to have a minimum impact on this efficient signal. Some schemes limit the duration of punishable underperformance events, beyond which no charge is applied. More frequently, designs envisage monthly or yearly caps. Such ceilings may be linked to capacity remuneration (so as to avoid penalties exceeding the initial revenue from the mechanism) or other parameters. Defining the penalty cap in terms of time (monthly, seasonal, yearly) has been shown to have a heavy impact on the strength of the efficient signal. Inasmuch as scarcity conditions are usually concentrated in a very few months, setting a monthly cap and setting it too low (e.g., one-twelfth of the yearly cap) all but eliminates the performance incentive. These parameters usually lie at the core of the aforementioned negotiations between regulator and power sector agents.

ISO New England's pay-for-performance scheme provides for barely any exemptions. Whether a resource underperforms due to an outage or scheduled maintenance, it is not eligible for an exemption in the calculation of its performance score. If the agent believes that this creates significant financial risk, it should include the risk in its FCM bid. According to Peter Cramton (FERC, 2014c) "a policy of no exemptions creates a level playing field. Responsibilities are clear and settlement is straightforward. Suppliers do bear greater performance risk, but it is precisely this risk that motivates performance-improving investments".

The pay-for-performance approach does, however, provide for a stop-loss mechanism that caps underperformance charges. According to FERC (2014b), the monthly stop-loss limit is the resource capacity obligation times the starting price of the capacity auction (which could be considerably higher than the clearing price), while the annual stop-loss limit is three times the resource's maximum potential net loss per month.

For participants in the PJM capacity market, the only exceptions to the delivery obligation (and the consequent penalty) are planned and approved maintenance outages or System Operator decisions that prevent commitment of the resource. As in ISO New England, PJM has a stop-loss provision that sets monthly (0.5 times the yearly net CONE) and yearly (1.5 times the yearly net CONE) caps on non-performance penalties (FERC, 2014a).

In the United Kingdom, delivery exceptions are limited to *force-majeure* situations. According to DECC (2014), resource delivery obligations are suspended and no penalties are applied only in the event of National Grid-imposed transmission constraints, suspension of the electricity market or similarly severe or emergency situations. Contrary to the approaches adopted by the US systems, however, in the United Kingdom, penalty caps are linked to capacity market revenues (through the capacity auction clearing price). Penalty caps apply on a monthly (200% of the monthly capacity market revenue) and yearly (100% of the yearly capacity market revenue) basis.

Lastly, a word is in order about the interaction between overperformance payments and penalty caps. The primary drawback to penalty caps is that, once the ceiling is reached, the resource committed lacks any incentive to continue to meet its obligations. The incentive is restored, however, if provision is made for overperformance payments.

4 CONCLUSIONS AND POLICY IMPLICATIONS

The developments described in this article highlight the central role played by performance incentives in enhancing the effectiveness of capacity mechanisms and ensuring power system reliability. Without properly designed penalty/credit schemes, resources committed in a capacity mechanism are not subject to any significant financial consequences for underperformance. They consequently have insufficient incentives to invest in measures able to improve their performance during scarcity conditions, when their contribution is much needed. In this context, explicit penalties and overperformance payments are of paramount importance for solving what has been defined in this article as the “missing performance incentives” problem.

The experience of electricity markets in the United States, where severe underperformance was observed despite the existence of explicit penalty schemes built into capacity mechanisms from the outset, stands as proof of the utmost importance of performance incentive details and their implementation. Only careful refinement of the design elements discussed in this article can guarantee the expected performance of resources participating in a capacity mechanism. This constitutes a lesson worth learning for European regulators, especially in the present context of widespread institution of capacity mechanisms.

5 ACKNOWLEDGMENTS

The authors would like to thank José Salazar from Empresas Públicas de Medellín (Colombia) and Luiz A. Barroso from PSR (Brazil) for their support and insights, as well as two anonymous reviewers for their fruitful comments that enhanced this article.

6 REFERENCES

Babula, M., Petak, K. (2014). "The Cold Truth: Managing Gas-Electric Integration: The ISO New England Experience". *IEEE Power and Energy Magazine*, vol. 12, iss. 6, pp. 20-28.

Batlle, C., Mastropietro, P., Rodilla, P., Pérez-Arriaga, I. J. (2015). "The System Adequacy Problem: Lessons Learned from the American Continent". Chapter of the book *Capacity Mechanisms in the EU Energy Markets: Law, Policy, and Economics*, 2015, Oxford University Press.

Batlle, C., Barroso, L. A., Pérez-Arriaga, I. J. (2010). "The Changing Role of the State in the Expansion of Electricity Supply in Latin America". *Energy Policy*, vol. 38, iss. 11, pp. 7152-7160.

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

Bidwell, M. (2005). "Reliability Options: A Market-Oriented Approach to Long-Term Adequacy". *The Electricity Journal*, vol. 18, iss. 5, pp. 11-25.

Cramton, P., Ockenfels, A., Stoft, S. (2013). "Capacity Market Fundamentals". *Economics of Energy & Environmental Policy*, vol. 2, iss. 2, pp. 27-46.

Cramton, P., Stoft, S. (2007). *Colombia Firm Energy Market.* Proceedings of the Hawaii International Conference on System Sciences, 2007.

CREG, Comisión de Regulación de Energía y Gas (2014). *Resolución 26 de 2014, Por la cual se establece el Estatuto para Situaciones de Riesgo de Desabastecimiento en el Mercado Mayorista de Energía como parte del Reglamento de Operación.* Resolution from the Regulator, issued on 7 March 2014.

CSMEM, Comité de Seguimiento del Mercado Mayorista de Energía Eléctrica (2014a). *Mercado de energía mayorista: aspectos relevantes del Niño 2014-2015.* Report no. 93/2014, released on 21 August 2014.

CSMEM, Comité de Seguimiento del Mercado Mayorista de Energía Eléctrica (2014b). *Cargo por confiabilidad, precio de escasez y comportamiento de las ofertas de precio.* Report no. 92/2014, released on 24 July 2014.

CSMEM, Comité de Seguimiento del Mercado Mayorista de Energía Eléctrica (2011). *Abastecimiento adecuado de gas natural: un tema sin resolver.* Report no. 60/2011, released on 14 July 2011.

CSMEM, Comité de Seguimiento del Mercado Mayorista de Energía Eléctrica (2010). *Experiencias de la intervención del MEM bajo efecto del Niño 2009-10.* Report no. 53/2010, released on 13 October 2010.

DECC, Department of Energy & Climate Change (2014). *Implementing Electricity Market Reform (ERM) – Finalised policy positions for implementation of EMR*. Policy document, released in June 2014.

EEAG, Environmental and Energy State Aid Guidelines of the European Commission (2015). *Designing appropriate obligations and penalties.* Capacity Mechanisms Working Group. Working document released in April 2015.

FERC, Federal Energy Regulatory Commission (2015). *Comments of the Independent Market Monitor for PJM.* Docket No. ER15-623-000-001, issued January 2015.

FERC, Federal Energy Regulatory Commission (2014a). *Reforms to the Reliability Pricing Market ("RPM") and Related Rules in the PJM Open Access Transmission Tariff ("Tariff") and Reliability Assurance Agreement Among Load Serving Entities ("RAA").* Docket No. ER15-623-000, issued on 12 December 2014.

FERC, Federal Energy Regulatory Commission (2014b). *Order on Tariff Filing and Instituting Section 206 Proceeding*. Docket No. ER14-1050-000, issued on 30 May 2014.

FERC, Federal Energy Regulatory Commission (2014c). *Testimony of Peter Cramton on Behalf of ISO New England Inc.* Docket No. ER14-1050-000-001, issued in January 2014.

FERC, Federal Energy Regulatory Commission (2013). *Order on Proposed Tariff Revisions.* Docket No. ER13-2313-000, issued on November 1, 2013.

Finon, D., Pignon, V. (2008). "Electricity and long-term capacity adequacy: The quest for regulatory mechanism compatible with electricity market". *Utilities Policy*, vol. 16, iss. 3, pp. 143-158.

ISO-NE, Independent System Operator of New England (2015). *2015 Regional Energy Outlook.* Annual report released in January 2015.

ISO-NE, Independent System Operator of New England (2013). *FCM Performance Incentives - Performance Payment Rate*. Memorandum released on 4 September 2014.

ISO-NE, Independent System Operator of New England (2012). *FCM (Forward Capacity Market) Performance Incentives*. Working document released in October 2012.

Joskow, P. L. (2008). "Capacity payments in imperfect electricity markets: Need and design". *Utilities Policy*, vol. 16, iss. 3, pp. 159-170.

Mastropietro, P., Batlle, C., Barroso, L.A., Rodilla, P. (2014). "Electricity Auctions in South America: Towards Convergence of System Adequacy and RES-E Support". *Renewable and Sustainable Energy Reviews*, vol. 40, pp. 375-385.

MITEI, MIT Energy Initiative (2013). *Growing Concerns, Possible Solutions: The Interdependency of Natural Gas and Electricity Systems*. Report released in April 2013.

Monitoring Analytics, LLC (2014). *2014 Quarterly State of the Market Report for PJM: January through September*. Monitoring report released on 13 November 2014.

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

PJM, Pennsylvania New Jersey Maryland Interconnection (2014a). *Wind Chill versus Forced Outages. Capacity Performance*. Working document released on 11 September 2014.

PJM, Pennsylvania New Jersey Maryland Interconnection (2014b). *PJM Capacity Performance Updated Proposal*. Working document released on 7 October 2014.

PJM, Pennsylvania New Jersey Maryland Interconnection (2014c). *Problem Statement on PJM Capacity Performance Definition*. Working document released on 1 August 2014.

RTE, Réseau de Transport d'Électricité (2014). *French capacity market. Report accompanying the draft rules.* Document released in April 2014.

Vázquez, C., Rivier, M., Perez-Arriaga, I.J. (2002). "A Market Approach to Long-Term Security of Supply". *IEEE Transactions on Power Systems*, vol. 17, iss. 2, pp. 349-357.