SECURITY OF SUPPLY: WHAT ROLE CAN CAPACITY MARKETS PLAY?[1]

Fernando Barrera
Senior Consultant
NERA Economic Consulting
Paseo de la Castellana 13
Madrid 28043, Spain

Joseph Crespo
Senior Consultant
NERA Economic Consulting
One, Main Street, 5th Floor
Cambridge, Mass. 02142, USA

Abstract

This paper analyses the reasons behind the introduction of capacity markets, their performance and their future role in deregulated electricity markets. It emphasises the importance of having capacity markets that reduce the threat of regulatory intervention as it jeopardises the recovery of capacity costs. It looks at the experience of capacity markets in the Northeast US and highlights their conceptual difficulties and the reasons why they have not fared well while discussing two possible approaches for longer term capacity markets. It discusses which aspects of capacity markets will have to be improved upon to provide appropriate long and short-term signals for capacity expansion.

1. INTRODUCTION

A well-known result of competitive electricity markets is that, under marginal system pricing, the optimum amount of capacity will recover total costs provided the price of energy is allowed to rise to the level of the Value of Lost Load in those hours of the year where the capacity constraint is binding. While this may be efficient and, in contrast to any intervention in the market, will provide the right level of capacity, there are various reasons why explicit capacity remuneration methods tend to be imposed as additional supply remuneration on liberalised electricity markets.

Usual reasons advocated have to do with public good arguments about supply adequacy, investment incentives, and price volatility. However, the main reasons seem to be the imposition, real or threatened, of price caps in energy markets and the political difficulty of having the boom and bust cycles that may be a feature of the competitive paradigm outlined above. Price caps do not have to be explicit to discourage investment in supply, as even in pure energy markets the commitment not to impose price caps is, in many instances, unlikely to be credible. Boom and bust cycles are also common in other capital-intensive industries but, given the nature of electricity, these cycles can have a high political cost so politicians would like to shift the blame elsewhere.[2]

---

[1] The opinions expressed in this article can only be attributed to them and I no way represent NERA's statements on the issues. We would like to thank colleagues for discussions but the usual disclaimer applies.

[2] By either imposing capacity obligations, being able to pin the blame on specific retailers, or by paying explicitly for capacity, shifting the blame on generating companies.

Capacity remuneration in practice has taken the form of price uplifts or quantity requirements. As such, the debate over their relative superiority can easily be framed in terms of the traditional prices versus quantities discussion and a simple results seems to show that quantity requirements seem to be superior. However, the challenge of designing efficient capacity markets is no easy matter.

Many challenges have been revealed in the process of capacity market design. The first is determining the appropriate economic value of capacity and its link with energy prices. An even greater challenge can be designing a mechanism that delivers this value to the owners or the possible suppliers of capacity and provides a linkage between short-term price signals and long-term incremental cost of capacity. In the United States, for example, one of the first capacity markets ever employed produced prices that converged to zero within just a few months. This resulted from a design flaw, which can have fatal effects in capacity markets – even more than in energy markets.

Notwithstanding such setbacks, more recent versions of capacity markets may provide a means to ensure security of supply. In contrast to simpler mechanisms, such as administratively determined capacity charges, capacity markets can minimize the cost of achieving security of supply by facilitating competition among security of supply alternatives (i.e., generation capacity, transmission capacity or demand response). In this paper, we discuss the basic issues for designing these markets, which are now being considered in the United States and in other jurisdictions. We explain why the key to designing capacity markets with appropriate and reliable price signals depends on having time-consistent price signals and on the coherence of capacity requirements and regulatory commitment.

2. RECOVERING THE COST OF CAPACITY: FOUR ALTERNATIVES

In theory there are four alternative ways to remunerate capacity. The first one is the competitive paradigm in which during the hours of unserved energy the price can rise to the Value of Lost Load (VOLL) and the revenues collected by generators in those hours should be equal to the exact capacity revenue. The second way consists of awarding generators a uniform capacity payment based on the fixed cost of the peaking technology. Thirdly, in equilibrium the marginal cost of incremental capacity is equivalent to the marginal cost of unserved energy (which can be proxyed by the VOLL times the loss of load probability, LOLP). In practice, the three price-based methods have been complemented with the setting of capacity obligations or a quantity requirement:

The Pure-energy Approach: The first method, the pure energy method, is used in Australia, in England in the new NETA framework, in the Nordpool and in the, now defunct, Californian market. This method tests the commitment of the authorities and in the only case where operating margins have been tight, price caps were duly imposed as market power situation complicated the operation of the market. The method is also very sensitive to the calculation of the VOLL.[3]

The Cost-based Approach: Under this approach, generators are paid a monthly or annual capacity payment that is based on the cost of the peaking technology. It has been used in South America (Argentina, Chile, and Colombia), and in Spain. There are some differences in the way it is applied but its main drawback is that it lacks the benefit of a true market price and has many opportunities to be adjusted arbitrarily. The regulator has to decide the cost of the peaking plant, the total revenue requirement, and the units that qualify for remuneration, but above all the problem is that remuneration depends on expected contribution to supply adequacy so the

[3] In Australia the price has risen to VOLL in many occasions.

incentives of generators are not clear. Notwithstanding these difficulties, a cost-based approach is often preferable to a poorly designed capacity market, of which there have been a few.

The (VOLL×LOLP) approach: This was used in the old electricity pool in England and Wales. Its premise was to base capacity payments on the "value of lost load" or the maximum willingness that customers were willing to pay to keep the lights on.[4] Using this basic parameter and a scalar that varied with the probability of lost load, a capacity charge was added to the energy price charged to retailers and paid to generators. Determining the probability of lost load was therefore very important to determining the appropriate charge. In practice, the method proved to be very conditioned by short-term fluctuations and by manipulations of the availability of generators that could have a large impact on the LOLP.

The Capacity Market Approach: This approach has been popular in the US but has lead to varied results. The basic premise is that generators should compete to supply capacity that is required to meet reserve margins. However, since these markets were first utilized ISOs continue to have difficulty with designing and implementing markets for capacity. Most likely, this difficulty has resulted mainly from a misunderstanding of capacity's unique characteristics, relative to other energy products. For example, once a plant is built, the owners of installed capacity cannot withhold their installed capacity from the market when the capacity price does not meet their costs. This is analogous to forcing a generator to run regardless of the energy price, whenever that generator available.

It has been shown that the debate about capacity uplifts and capacity requirements is just another example of the prices versus quantities debate. Given that the supply for capacity tends to be more elastic than the demand for capacity, mistakes in the setting of the capacity requirements will have a lower efficiency loss than mistakes on the setting of prices.[5]

In practice, capacity markets have been created to decentralise capacity planning and have taken the form of additional remuneration rather than an explicit capacity remuneration. Capacity markets are also capped by means of an explicit penalty as short-term capacity shortages are quite common and unlikely to be resolved quickly given the long lead times in generation investment.

3. THE CHALLENGES OF DESIGNING CAPACITY MARKETS

The design of capacity markets has to be sustainable politically and economically sound. To be politically sustainable, energy market prices should not rise to extremely high levels and, to be economically sound, this restriction should have as a counterpart an adequate capacity remuneration for the optimal level of capacity. The recent examples of capacity markets in the US have not managed to achieve this outcome given their extremely short-run emphasis and . Thus, in practice, capacity remuneration has been synonymous with operating reserve remuneration (i.e. short run marginal cost of expansion). This is not the same as long run incremental cost of capacity and with the use of price caps in capacity markets these two costs are unlikely to converge. The design challenge of capacity markets, apart from the many implementation problems that arise, is to devise a long-run signal that is traded in long-term markets that can give policymakers a capped energy price that is not out of line with capacity remuneration.

---

[4] Ironically, although there have been several studies that attempt to value lost load, this parameter is sometimes set arbitrarily. This undermines the economic principles that motivated the approach.

[5] See Oren (2000) for a description of an argument made by Larry Ruff.

3.1 Capacity Markets First Generation Designs

In the Northeastern US several capacity markets have been utilized to compensate generators for the fixed cost of their capacity. The characteristics of the initial capacity markets are described in Table 1 below.[6]

Table 1 *The Features of Northeastern Capacity Markets*

	PJM	NEPOOL	NY
Size of Obligation	In November, the SO forecasts peak demand in the following summer (+20% margin)	Defined in real time. Actual peak demand plus an operating reserve margin	For two periods May-Oct; Nov-Apr, forecasted peak demand + 18%.
LSE requirements	Show every day sufficient capacity to meet peak demand plus margin	Monthly comparison of actual peak demand with LSE's capacity rights	Twice a year LSEs have to provide proof of adequate coverage.
Time frame	6-7 months	One month after	6 months
Penalties	Daily charge US$176/MW and double that when the market as a whole is deficient	No penalties	Incentive to build peaking generator
Capacity Credit Market	2 markets: daily and monthly. Supply and demand side bidding. Single marginal price.	ICAP: monthly trading of the difference between peak demand and obligation.	Six-month auctions

Source: NERA (2002).

The mechanism is quite simple, (i) a reserve requirement is established, (ii) that requirement is decomposed into an obligation for each Load Serving Entity LSE, (iii) the owners of generation unit are assigned a specific amount of "capacity credits" based on a generation unit's capacity capabilities. Finally, LSEs are required to obtain entitlements to enough "capacity credits" to meet their capacity requirement by either purchasing these credits in advance of the monthly market, or in the case of ex-post markets, they could wait until the end of the month when their actual capacity requirement would be determined.[7] However, if they were found to be deficient in "capacity credits" they were forced to buy the amount of their deficiency in an ex-post market (or pay a penalty charge).

These credits could be traded in daily markets (PJM) or in monthly markets (PJM, and the new NY monthly market). In the ex-post auction, the owners of capacity credits could offer the credits at stated prices to the auctioneer and the amount of credit deficiencies would set the clearing price. This uniform price is what all LSEs would pay and all generators would receive.

3.2 Basic Design Flaws and Challenges in Capacity Markets

So far very few restructured electricity markets have operated in the US. The only experiences are those of California and the Northeast (PJM, NY an New England) markets that, by contrast, have operated quite well. The latter markets have attracted investment and have avoided the price spikes and shortages that were a feature of the Californian fiasco. However, the designs of the capacity markets in the Northeast may not provide the right signals for two basic reasons: (i) the short-run nature of the capacity market, (ii) the adoption of price caps in the capacity markets.

3.2.1 Short-run price of capacity

To understand the basic design flaws in the early US capacity markets it is helpful to compare the bidding incentives that the same plant manager might have for bidding the two different

[6] NEPOOL operates two capacity markets the Installed Capacity Market ICAP and the daily Operable Capability market OPCAP.

[7] In the even that they had extra "capacity credits" they could sell those into the ex-post market.

products provided by a generation facility – energy and capacity. For example, each day a plant manager must face the decision of how to bid in the day-ahead energy market. The manager knows that its marginal cost per MWh is $40. Thus, if the price is below $40 the manager knows that it will not recover its fuel costs and the manager will not run the plant. If the price is above $40, then the manager can at least cover its fuel costs and anything that it collects above fuel costs can be used to pay for fixed costs including the cost of capital. Thus, the plant is usually run at any price above $40.

The same manager must decide how to price or whether to supply its "installed capacity". This is a far more complicated decision. "Installed capacity" refers to the plant's generation capability. Winning bidders are not required to provide operating or spinning reserves so the additional cost of supplying installed capacity can be negligible depending on a plant's planned maintenance cycle. That is, once capacity is installed the cost of supplying installed capacity is zero.

In most situations, a manager would be willing to accept anything for its installed capacity. Moreover, when there is more capacity than is needed to meet minimum reserve requirements, there is a risk of bidding too high and getting nothing. This would be the worst result for the owner of capacity because capacity credits for September are worthless in October.

In New England, generators eventually understood this risk and bid zero in order to assure being selected in the market. When too many generators did this, the price converged to zero. That is, until there was less generation than was needed to meet the minimum reserve requirements. At this point, generators knew that they could bid any price and be selected. For these reasons, prices in capacity markets, often oscillate between the minimum and maximum price.

The basic problem with these markets is that, in the short run, the commitment to supply is made before the market provides any price signal. In fact, in ex-post markets bids are not submitted until *after* the capacity is supplied. For practical purposes, ex-ante monthly markets are not much better. That is because once a plant is installed, the cost of maintaining its generation capability is small relative to the initial investment and decisions regarding this are likely to be made on an annual basis. Moreover, installed capacity is immobile. Thus, once a plant has been built most of the cost of supplying capacity has already been incurred and opportunity cost of supplying installed capacity to the market is zero. As a result, capacity prices in these short-term markets do not signal the costs of building new capacity. Short-run marginal cost and long-run incremental cost are unlikely to be in line.

3.2.2 The level of penalties

Given that prices in short-run capacity markets tend to be equal to zero or equal to the penalty, the choice of penalty ends up determining whether to build additional capacity or not. In practice, the choice of penalty resembles a capacity price like those based on the cost of a peaking plant and so far it seems the choice of penalty may not give sharp-enough incentives to build capacity.[8] Low levels of the penalty act just like a low energy cap and may reduce the incentives to build capacity.

Capacity tends to have a very low or very high scarcity price. Given the high volatility they result in price spikes that are uncoupled from the price of energy. The reason is simple; the capacity obligation is only the commitment to offer the plant for short-term balancing, not to offer the plant for a given price. Thus, they do not reduce overall (energy + capacity) volatility, only insofar as the payment makes shortages less frequent, and the price of capacity can be high or low in the presence of different energy prices.

In sum, capacity markets, as they have so far operated, have been developed to decentralise capacity decisions and to signal that energy price caps may not be necessary. In reality, they have not reduced price volatility nor can they do without some form of price caps; energy price caps

[8] NERA (2002) has estimated that, in PJM, as long as the penalty is incurred for less then 170 days a year there is no incentive to build a peaking plant.

are traded for capacity price caps generating a regulatory intervention that may hinder the prospects of recovering capacity costs and discouraging entry.

4. TWO NEW APPROACHES TO CAPACITY MARKETS

The volatilities in capacity markets and the concern that they may not be providing a good signal for capacity expansion have meant that their use has been called into question. We have found two strands of market-based capacity payments, the current study of resource adequacy markets in the Northeast US and the various approaches to link energy and capacity prices by means of a financial option.

4.1 The Northeast US: The Resource Adequacy Market

A new market is currently under design in the Northeast US. The intention of this market is to improve price signals by giving greater consideration to the timing of investment decisions and synchronizing these with market decisions. The decision to build a new plant or retrofit an old plant to increase generation capacity takes place years in advance of when the capacity is available to the market. It is important that credible price signals and/or contracts for capacity are available at this point in time. This is what the market requires to finance generation projects.

The fact that current price signals in many capacity markets are volatile and often meaningless from an economic perspective is one reason why investment has slowed in the Northeastern US.

In the proposal, the ISO forecasts the load and establishes an unforced capacity obligation for the future-operating year (defined as a June-May period) of a region/locality. The commitment of resources to meet the unforced capacity obligation is up to several years in advance of the operating year (say, for example, 3 years in advance of the operating year) through a centralized auction. Each ISO conducts its own separate centralized auction; however, the timing of the auctions will be coordinated.[9] Other features include:

- Products that resource providers can offer into the centralized auction include: existing generation; planned generation; bilateral contracts for capacity resources; load management products, and, if applicable, transmission upgrades.
- Bilateral transactions are intended to be an integral part of the market to permit LSEs to self-supply their own generation or as a way for an LSE to hedge against a potentially higher resource price in the centralized auction.
- The centralized auction-clearing price would be the price charged to all LSEs serving load during the operating year.
- The resource provider that offered and cleared MWs through the centralized auction would receive the centralized auction clearing price for their cleared MWs during the operating year.
- Deficiency charges will be assessed during the operating year to those resource providers that fail to actually provide the capacity commitments made through the centralized auction.

4.2 Capacity remuneration as a call option

Several authors have recently proposed another market instrument for the remuneration of capacity.[10] The idea is very simple and follows simple options theory. As the delivery of electricity requires two complementary inputs (capacity and energy), one with a fixed component and

9 ISOs with locational requirements may conduct separate auctions for each location

10 See for example Oren (2000) and the proposals made to the Colombian energy regulator by private consultants TERA and by the Colombian association of Electricity generators, ACOLGEN, by Rivier and Pérez-Arriaga (2000).

the other with a variable component, capacity can be remunerated by means of an options premium and energy by means of a strike price.

This allows the capping of energy prices without failing to provide the right incentives for capacity expansion. As in all capacity markets, the reserve margin is determined by the authorities and a capacity obligation is imposed on the retailers. Through an organised auction both current generators and possible entrants can compete on the premium and the strike price.[11] Some propose (TERA) that options contracts can be traded in secondary market while others propose that contracts are not traded as it could reduce the commitment of generators (see Rivier and Pérez-Arriaga, 2000). Generators would have to produce energy at the strike price, or purcahes it and make it available at that price.

The concept comes from insurance markets; the regulator sets capacity obligations as insurance against future price spikes and replaces the obligation to supply energy with an obligation to supply energy at a price (the strike). The contract duration in this framework is very important, locking in the capacity payment for a long time in advance reduces the risk of generators, but as contracts get shorter they start to behave like reserve remuneration.

5. IS THERE A FUTURE ROLE FOR CAPACITY MARKETS IN PROVIDING SECURITY OF SUPPLY?

Capacity markets are a clever way of dealing with the threat of regulatory intervention in energy markets. It is more acceptable to have markets decide how to trade price and investment risk through different contractual agreements than imposing prices and limiting the ability of market players to come up with solutions. Thus, this should be a promising area for solutions to the problem of encouraging capacity expansion but, in practice, the challenge has proved to be tough.

First, to be politically feasible the scheme has to provide a way of linking energy and capacity markets. This should avoid the problems of volatility and price spikes that are the source of the threat of regulatory intervention. Second, they should provide long-term signals that are set concordant with the lead times of adding generation capacity. This is necessary so that entrants would decide to enter into capacity obligations when the remuneration of capacity is above the cost of entry. Third, the long-term price of capacity should converge in the short run to the price of operating reserves. This would require some short-term trading of the capacity obligations. Fourth, there should be some form of local market price signal so that transmission constraints do not distort the price signal.

This is no easy task, it could get bogged down in all the practical details that are so common in designing electricity markets. Things like retail competition, the setting of penalties for non-compliance, deciding the firmness of generating plants, transmission constraints, demand forecasts, default obligations, etc, could prove to be complex. .

Added to this implementation issues, there are other more structural problems. Things like failures of capital markets, the underwriting of commercial risk in new markets, the rolling of short duration capacity obligations compared to the long-term investment in generation, the little experience with long-term electricity trading, the few experienced agents in long-term trading.

However, even if the challenges are great, the benefits from pursuing this option are likely to be large. This form of insurance would reduce the possibility of regulatory intervention, would give signals to expansion which are, so far, subject to many arbitrary decisions, and could provide the market expansion that has not yet been forthcoming in many countries.[12]

---

[11] Some propose that the regulator should set the strike price of the option (see Rivier and Pérez) and others propose that the auction should be based on both strike prices and premia (see TERA).

[12] Most liberalised electricity markets have been introduced in the midst of large reserve margins and when the margin narrows down market expansion has not been forthcoming. In many cases governments take the initiative by giving long-term guarantees and in others the lights have gone out.

REFERENCES

NERA (2002) “Electricity Markets and Capacity Obligations” A NERA Report for the UK Department of Trade and Industry.

Oren, S. (2000) “Capacity Payments and Supply Adequacy in Competitive Electricity Markets” Proceedings from the Symposium of Specialists in Electric Operational and Expansion Planning Curitiba, Brazil.

Rivier, Michel and I. Pérez-Arriaga (2000) “Estudio cargo por capacidad en Colombia” available at www.acolgen.com.co.