VII SEPOPE
21 a 26 de Maio de 2000
May - 21 st to 26th - 2000
CURITIBA (PR) - BRASIL

VII SIMPÓSIO DE ESPECIALISTAS EM PLANEJAMENTO DA OPERAÇÃO E EXPANSÃO ELÉTRICA

VII SYMPOSIUM OF SPECIALISTS IN ELECTRIC OPERATIONAL AND EXPANSION PLANNING

Capacity Payments and Supply Adequacy in Competitive Electricity Markets[1]

Shmuel S. Oren[2]
University of California at Berkeley
USA

Abstract: This paper discusses alternative approaches that have been adopted around the world for guaranteeing the appropriate level of investment in electric generation capacity. We argue that the use of "capacity payments" is the least desirable approach that undermines the long-term efficiency objectives of the electric industry restructuring. We explain how in an energy only market, long term supply contracts in the form of call options with premiums that depend on the contracts' strike prices can meet the need for ensuring supply adequacy and the financial health of the generation sector.

Keywords: Capacity payments, Generation planning, Reliability, Adequacy.

1. INTRODUCTION

The reliability of electricity supply has been one of the overriding concerns guiding the restructuring of the electric power industry. The slogan "keeping the lights on" has been the principal motivation for many technical and economic constraints imposed on market designs. The term supply reliability, encompasses, however, a mix of system attributes that have diverse economic and technical implications under alternative market structures. NERC (National Electric Reliability Council) defines reliability as: "the degree to which the performance of the elements of the technical system results in power being delivered to consumers within accepted standards and in the amount desired". Imbedded within this definition is the notion of the "obligation to serve" which is arguably out of step with the notion of a deregulated industry with competitive supply. In fact, the concept of reliability as defined by NERC encompasses two attributes of the electricity system: *Security*, which describes the ability of the system to withstand disturbances (contingencies) and *Adequacy*, which represents the ability of the system to meet the aggregate power and energy requirement of all consumers at all times.

The notion of system security identifies short term operational aspects of the system which are characterized through contingency analysis and dynamic stability assessments. Security is provided by means of protection devices and operation standards and procedures that include security constrained dispatch and the requirement for so called *ancillary services* such as: voltage support, regulation (AGC) capacity, spinning reserves, black start capability etc.. The notion of adequacy on the other hand represents the systems ability to meet demand on a longer time scale basis in view of the inherent fluctuation and uncertainty in demand and supply, the non-storability of power and the long lead time for capacity expansion. Generation adequacy has been traditionally measured in terms of the amounts of planning and operable reserves in the system and the corresponding loss of load probabilities (LOLP) that served as criteria for planning and investment decisions.

From a technical perspective security and adequacy are clearly closely related since a system with abundance of reserve capacity provides more flexibility in handling unforeseen disturbances. However, while a system with limited planning reserves may experience shortages it can still be operated in a secure manner while a system with ample reserve can be operated insecurely.

[1] This paper benefited from discussions with Dr. Ignacio Perez- Arriega of the Spanish Electricity Regulatory Commission who organized a workshop on the subject of capacity payment that motivated this work and from the presentations of Larry Ruff and Harry Singh at that workshop. The positions expressed in this paper,however, are the author's own views.

[2] Department of IEOR, University of California at Berkeley, Berkeley, CA 94720 USA Oren@IEOR.Berkeley.Edu

From an economic point of view security and adequacy are quite distinct in the sense that the former is a *public good* while the latter is a *private good.* Security is a systemwide phenomenon with inherent externality and free ridership problems. For instance, it is not possible to exclude customers who refuse to pay for spinning reserves from enjoying the benefits of a secure system. Hence, like in the case of other public goods such as fire protection or military defense, security must be centrally managed and funded through some mandatory charges or self-provision rules. The resources for such central provision, however, can be procured competitively through ancillary service markets, long term contracts or other procurement mechanisms. Adequacy provision on the other hand, as will be explained later, amounts to no more than insurance against shortages, which in a competitive environment with no barriers to entry translate into temporary price hikes. Such insurance is clearly a private good of which the adequate quantity to be provided can be decided through customer choice. In an environment in which "obligation to serve" is replaced by "obligation to serve at a price", the concept of loss of load probability is not well defined unless a distinction is made between probability of lost load due to system collapse vs. lost load due to inadequate supply. It is a prerogative of consumers and producers to decide what is the appropriate level of price insurance they wish to procure and how much they are willing to pay for it as long as they are able from a technical point of view to bear the consequences of their decisions without affecting others. In the remainder of this discussion we will only focus on adequacy provision.

The traditional approach to ensuring generation adequacy in vertically integrated utilities was to build planning reserves based on load forecasts LOLP calculation and estimates of the value of lost load (VOLL) and assign the cost of the extra capacity as a rate uplift. More elaborate schemes, which will be discussed below, attempted to allocate the cost of capacity according to time of use so that peak consumption bears a larger portion of that cost. In an ideal competitive market were prices of energy vary continuously to reflect the equilibrium between supply and demand at each moment, payment to inframarginal generators (above marginal cost) should cover their capacity cost. Economic theory tells us that in a long-term equilibrium, the optimal capacity stock is such that scarcity payments to the marginal generators when demand exceed supply will exactly cover the capacity cost of these generators. Furthermore, the optimal generation mix (were generators are characterized by their fixed and variable cost) will be such that the operating profit of each generator type will exactly cover their capacity costs. This optimal equilibrium mix is achieved through exit of plants that do not cover their cost and entry of plants whose cost structure will yield them operating profits that exceed their capacity costs.

The critical role of electricity in the economy and the political ramifications of widespread electricity shortages have prompted many regulators around the world to take steps above and beyond reliance on market forces in order to ensure generation adequacy. While in theory, allowing the prices of energy to reflect short run supply and demand equilibrium will create market signals and provide adequate financing for proper capacity expansion, many regulators have been concerned that energy prices occurring in the various restructured systems are not sufficiently high to cover generators' capacity costs and to prompt adequate investment. The prevalence of regulatory intervention to suppress energy prices even when they reflect legitimate scarcity rents justifies the concern that indeed generators would not be able to cover their fixed costs through energy sales alone. Ruff [1] argues that "suppression of energy prices" is inherent in all current market designs that accept only hourly or half hourly energy prices that could not possibly reflect the second by second changes in the supply and demand balance. He claims that such mechanisms tend to suppress the price spikes that would arise in an idealized continuous double auction reflecting true spot prices. Furthermore, the absence of demand side bidding in most energy markets (that would determine the spot price in case of shortage) and the tendency to dispatch reserves to mitigate shortages obscure the scarcity rents. On the other hand one could argue to the contrary that in a continuous double auction with full demand side participation prices would be lower than when demand is treated as inelastic over hourly intervals regardless of changes in supply conditions. There is also the strong possibility that the measures taken to ensure generation adequacy have the effect of suppressing energy prices due to excess capacity or perverse incentives so that the necessity of such measures becomes self-perpetuating. This is clearly the case in Argentina, for instance, were a large capacity payment paid on the basis of generated energy induces generators to bid below marginal cost so as to increase production and capacity payment revenues.

2. Approaches to Ensuring Adequacy

There are currently three basic approaches to dealing with generation adequacy in restructured electricity markets.

i) Energy Only markets

This approach has been adopted in California, Nordpool and the Australian Victoria pool. Generators in such markets bid only energy prices and, in the absence of constraints, all bids below the market-clearing price in each hour get dispatched and paid the market-clearing price. The primary income sources for recovery of capacity cost is the difference between the market clearing price and the generators' marginal costs. When ancillary services are procured separately by the system operator, as in California, generators can earn additional revenue by selling ancillary services, such as regulation and spinning reserve capacity, through short term ancillary service markets or long term contracts.

ii) Capacity payments

This approach is used in the UK, Spain and several Latin American countries. Generators in such systems are given a per MW payment based on their availability (whether they get dispatched or not) or based on generated energy as an adder to the energy market clearing price. The capacity payments are collected from customers as a prorated uplift similarly to other uplift charges such as transmission charge. In some cases such as in Spain capacity payments are indistinguishable from stranded investment compensation, which are viewed as an additional source of revenue for the generators that is needed in addition to the competitive energy revenues in order to guarantee their profitability.

iii) Planning reserves requirement.

The eastern pools in the US including PJM NYPP and New England have adopted this approach. In such systems load serving entities are required to have or contract with generators for a prescribed level of reserve capacity above their peak load within a certain time frame. The specific form of the reserve requirement and the time frame over which such obligation are determined varies among systems. New England for instance has separate requirements for installed capacity specified with respect to the annual peak and separate requirements for operable capacity specified relative to the monthly peak. Formal or informal capacity markets that allow trading of capacity obligations among the load serving entities have accompanied reserve capacity obligations. The reserve requirements and the capacity markets provide generators with the opportunity to collect extra revenue for their unutilized reserve generation capacity and provide incentives for the building of reserves beyond the reserves that meet the short term needs for ancillary services.

The calculation of planning reserve requirements and capacity payments in the second and third approaches listed above are typically based on engineering models of "loss of load probability" (LOLP) and on estimates of the "value of lost (unserved) load" (VOLL). The LOLP calculations take into consideration the quantity and mix of the available capacity in relation to the forecasted load and the probabilities of forced outages. In the UK design capacity payments are directly computed as the product of LOLP x (VOLL-SMP[3]) and vary each half hour. In systems with mandated planning reserves, the prescribed reserves requirement are based on a threshold criterion on the expected cost of lost load given by the product of LOLP and VOLL net of energy cost.

[3] System marginal energy price

The fundamental relationship between capacity and energy prices in a long run equilibrium is such that the expected social cost of unserved energy as reflected by the energy-only market prices should equal the marginal cost of incremental capacity. However, the separate capacity markets created for trading reserve capacity requirement set through engineering based methods may produce prices that are in disequilibria with the energy market prices. For instance, overestimating the expected cost of lost load would create artificially inflated demand for capacity and result in high capacity prices which in turn will lead to overcapacity that results in suppressed energy prices and socially inefficient production and consumption. Similarly, capacity payments based on such calculations would tend to suppress energy prices to or below marginal cost resulting in excess consumption and excess generation capacity.

The reliance of capacity payments and capacity requirement on engineering based calculation has been criticized repeatedly on the grounds that the VOLL used in these calculations is administratively set and has no market base. The usual remedy proposed for instance in [2] is to employ VOLL figures based on demand side bidding. Further criticism [3], [4] points to the fact that the LOLP calculations often employ simplistic models of probabilistic failure (e.g. Poison arrivals) and do not account for more complex phenomena such as the incentives of operators to keep plants running during peak price periods. Both the arbitrariness in the VOLL and the approximate nature of the LOLP calculation are likely to result in a mismatch between energy market prices and capacity values set directly or via a capacity market induced by capacity obligations. Furthermore, as the UK experience taught us, the predictability of calculated capacity payments can lead to gaming and manipulation of the payments.

An overriding question that must also be addressed in discussing LOLP calculation is: what is the meaning of lost load in a competitive market with no obligation to serve? A more appropriate statistic in such an environment that reflects scarcity would be some distributional information on the explicit or implicit spot prices that would clear the market. Such information could be interpreted as a market based statistics concerning the distribution of the quantity LLOP x VOLL.

In order to understand the meaning of such statistics and how it could be used to create market instruments that will facilitate generation adequacy in a competitive environment, we will first examine the origins of capacity payment in the traditional regulated electricity industry.

3. THE ORIGINS OF CAPACITY PAYMENTS.

The concept of capacity payment is rooted in the theory of peak load pricing whose application in the context of electric power was pioneered by Boiteux. According to this theory generation of electricity requires two factors of production, capacity and energy where the amount of energy that can be produced in any given time period is constrained by the available capacity. Consider a simple case of two consumption periods: peak and offpeak with two respective deterministic demand function and assume that the same fixed capacity is available in both periods. According to the basic theory, energy is priced at marginal cost in both periods and a capacity payment that would recover the fixed capacity cost is imposed on the peak energy users. The optimal capacity will be such that the incremental cost of a capacity unit equals the shadow price on the capacity constraint that is active during the peak. That shadow price reflects the incremental value of unserved load as measured by willingness to pay net of marginal energy cost. It is important to realize that the above approach to pricing has evolved in the context of a regulated monopoly whose primary objectives have been to recover cost and encourage consumption.

Subsequent developments of peakload pricing theory focused on two important aspects of electricity supply: uncertainty and technology mix (see [5] for the most general treatment of these two aspects.) The affect of uncertainty leads to redefining the basic ingredient of electricity service as energy and reliability where reliability Is manifested by LOLP calculation as a function of available capacity relative to load. The distinction between peak and offpeak than becomes a matter of degree. This perspective rationalizes levying a time varying capacity charge on all consumption and the payment to generation capacity that is not utilized for production of energy on the ground that such capacity provides added reliability. The capacity adders employed in the UK system to augment energy prices and compensate available nondispatched capacity are based on the above perspective.

Another perspective motivating capacity payments focuses on cost recovery in a system with optimal technology mix serving a load profile characterized by a load duration curve. In the following we adopt a deterministic interpretation of the load duration curve. However, a similar argument can be developed by interpreting the load duration curve as a cumulative probability distribution on load level and using average availability in determining the technology mix.

Consider a set of generation technologies characterized by a fixed and variable cost per capacity increment (the variable cost defined with respect to load factor). The lower envelope of the different cost functions creates a nonlinear technology mix cost curve per capacity unit as function of operating duration. That curve can be interpreted as the system's cost of serving any horizontal load slice under the load duration curve, as illustrated on the left part of Figure 1.

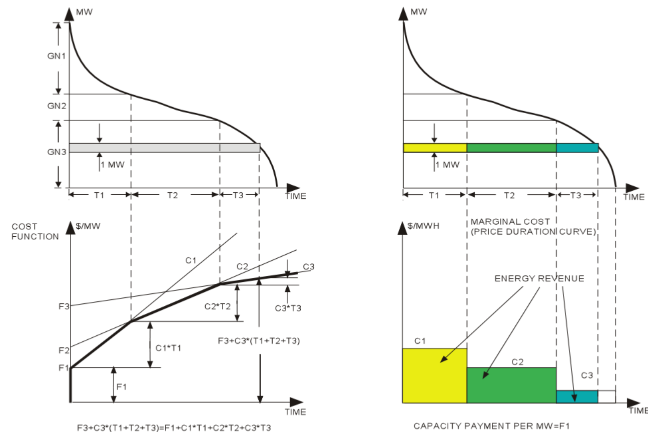

Figure 1 - Load Slice vs. Marginal Cost Pricing

This interpretation is the basis for Wright tariffs that price load slices nonlinearly based on load factor. In a system with coincident peaks, pricing each load slice according to the load slice nonlinear cost curve will exactly recover the total cost of generation. Furthermore, that nonlinear function coincides with the technology specific cost function in the relevant duration interval. Hence, compensating generators based on the load slice nonlinear cost curve is equivalent to paying generators their technology specific capacity and energy costs.

An alternative approach illustrated on the right hand side of Figure 1 is to price consumption and compensate generation of energy at each point in time at the corresponding marginal energy cost, that is the variable cost of the most expansive energy dispatched at that time. As we can see from Figure 1. the sum $\sum_{i=1}^{3} C_i \cdot T_i$ of marginal costs times the duration during which they are applied produces the same payment as the variable portion of the nonlinear duration-based cost function. Thus if each generator is paid the uniform system marginal cost for their energy at each point in time they end up with a shortfall in the amount of F_1, the fixed cost of the peaking technology, per each unit of capacity.

This argument rationalizes awarding generators a uniform capacity payment based on the fixed cost of the peaking technology (typically CT's) to supplement energy revenues based on marginal cost. Under optimal capacity planning the marginal cost of incremental capacity equals the marginal cost of unserved load which can be approximated by the marginal value of unserved load (VOLL) times the probability or fraction of time that load must be curtailed due to insufficient capacity. Hence, two alternative methods for capacity payment calculation (which are, in theory, equivalent under optimal capacity configuration) are to base the payment on the cost of peaking technology (e.g. CT) construction or to use the expected value of unserved load estimated by VOLL x LOLP.

The need for a capacity payment to make up for generation cost recovery shortfall can be eliminated by introducing into the technology stack demand curtailment as an equivalent supply technology with zero fixed cost and marginal cost equal to VOLL. The supply curve describing cost per capacity unit as function of operating duration for the augmented technology stack starts continuously through the origin with a slope of VOLL. Hence, if we set a spot price to marginal cost in each duration interval, the spot price during the period where demand is curtailed should be set to VOLL as illustrated in Figure 2. Paying generators that spot price during supply scarcity periods will provide them with the same income as capacity payment. There is, however, an important difference between the two alternative forms of compensation. Capacity payments set to the value of peaking technology capacity cost fully compensates such technology even if it is idle and consequently may induce excess capacity. On the other hand paying the VOLL for energy produced during scarcity periods only compensates generators that can sell their power at that price and will hence avoid the incentive for over investment. Furthermore, capacity payments are usually paid to generators whereas curtailed load can only avoid the peak technology marginal cost of energy. On the other hand, when the capacity cost is collected by generators in the form of a scarcity spot price, the curtailed load avoids the full VOLL payment and hence such an approach incents demand side participation in shortage mitigation.

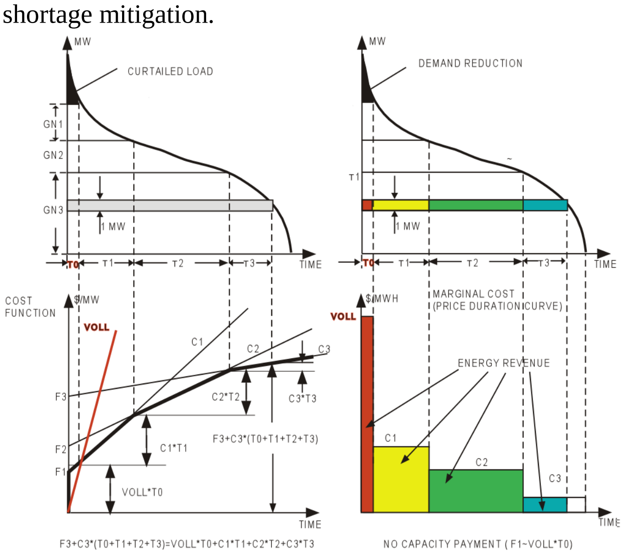

Figure 2 - Adding Demand Curtailment to the Technology Stack

Setting the spot price at VOLL during curtailment period is a proxy to demand side bidding where true values of lost load would be manifested. Thus VOLL attempts to represent an average of the value of lost load distribution. With demand side bidding the full distribution (rather than a uniform approximation set to VOLL) is included in the supply stack. This could be depicted by replacing the straight line representing curtailment on the bottom left of Figure 2 with a concave curve whose average slope equals VOLL. The resulting spot prices during curtailment periods will at time go below the VOLL level and consequently more demand side displacement of peak generation capacity will occur.

In the absence of demand side bidding it is often the case that involuntary curtailments are averted by the system operator through dispatch of reserves whose energy is priced based on their marginal operating cost. Such practices give rise to an important practical question: what should be considered as a curtailment period during which the price is raised to VOLL. If the right amount of reserves were procured such deployment of reserves impacts security and that impact should be reflected in the price of energy. The use of reserves to mitigate energy shortages at prices reflecting the incremental energy costs of the reserves amounts to a subsidy between security and adequacy. This is analogous to using the army to mitigate labor shortages and charging the employers variable hourly incremental cost for the soldiers' time. A pricing scheme that would reflect the scarcity that led to deployment of reserves should augment the energy price during such periods with some prorated portion of the reserve capacity payments (of the ancillary service market) that would have otherwise been levied on all customers as an uplift. Intuitively that adder should increase gradually as more reserves procured for security purposes are being deployed to meet energy shortages and price of energy plus the adder should approach the VOLL when involuntary load curtailments are invoked. A rigorous determination of how to set the real time spot price when reserves are being deployed would require a model that asses the effect of such deployment on system security[4].

4. Revisiting the Role of Capacity Payment and Generation Adequacy

Theoretical rationale and practical experience suggest that energy-only markets with spot prices that are allowed to reflect scarcity rents will generate sufficient income to allow capacity cost recovery by generators. The massive influx of new planned generation capacity in California indicates that investors do believe that they will be able to recover their investment and make a profit. Hence from a supply

[4] Ruff [1] argues that the capacity adder in the UK system was designed to accomplish this objective. In the UK there is no separate procurement of reserves by the system operator but rather all the capacity that is bid and not dispatched is regarded as reserves. The extent to which the load use of reserves impacts security is reflected by the LLOP calculation, which determines the capacity adder to the spot price.

adequacy point of view a well functioning energy-only market can provide the correct incentives for generation adequacy. Yet there may be good reasons for some form of capacity payment and even for regulatory intervention to ensure generation adequacy. Legitimate concerns for failure of the energy markets to reflect scarcity rents or failure of the capital market to produce proper levels of investment in response to such rents may justify some intervention. In some cases regulatory intervention in adequacy assurance is needed to compensate for regulatory interference in the energy market. The supply resource stack of electricity generation in systems with significant amounts of thermal generation exhibits an inherently steep rise in cost around the capacity limit. This phenomenon combined with the typically low short-term elasticity of electricity demand tends to produce high price volatility in fully competitive energy spot markets. Spot markets that clear on an hourly or half hourly basis tend to average out some of the volatility but even in such markets it may be politically infeasible to allow the energy spot prices to fully reflect scarcity rents. Consequently, energy prices are often suppressed through regulatory intervention (price caps) and by the market design, which in turn creates revenue deficiency for the generator that may cause insufficient investment in generation capacity. Often the threat of regulatory interference to curb scarcity rents is sufficient to inhibit capital formation and raise the capital cost for investment in generation capacity. Such interference is due to misperceptions and difficulties in distinguishing between market power abuse and legitimate scarcity rents. Thus, capacity payments or capacity obligations that stimulate capacity markets are largely viewed as remedial measures needed to offset suppression of energy prices and to ensure generation adequacy.

A useful perspective in addressing the generation adequacy problem is to view the regulatory intervention as a form of insurance against price volatility. Rather than considering the intervention as a reaction to the failure of the energy spot prices to properly reflect scarcity rents, one may regard the regulatory intervention as a proactive measure in the form of a mandatory hedge or insurance that will assure that prices stay within a socially acceptable range. Such an insurance-based view recognizes the private good nature of generation adequacy. It lays the foundation for introducing customer choice in selecting the appropriate level of price protection and for establishing a relation between the capacity payment awarded to a generator and the responsibility that such payment entails. For instance rather than setting a uniform capacity obligation or payment whose cost is evenly distributed among consumers, load serving entities, direct access customers and generators may be able to select their desired level of exposure to price risk and pay or receive an appropriate premium. Thus, generators receiving a capacity payment will guarantee the availability of their capacity to produce energy at a prespecified strike price so the capacity payment is interpreted as premium for a call option on that capacity. The higher the payment the lower the strike price and vice versa.

5. MARKET BASED PROVISION OF SUPPLY ADEQUACY

The underlying principles for a market based provision of generation adequacy are:

- "Obligation to serve" is replaced by "obligation to serve at a price"
- Energy prices are determined by supply and demand and consist of production cost + scarcity rent
- Consumers (or their load serving entities) are free to choose level of eposure to price risk through risk management and contractual agreements
- Reserve generation capacity beyond security needs is just a hedge against high prices.
- Forward markets and hedging instruments provide competitive market alternatives to capacity payments or mandatory planning reserve requirements.

The following features would characterize an idealized market based provision of adequacy that is based on the above principles:

- Customers decide how much they want to pay for capacity according to the price risk they are willing and able to assume.
- Generators can diversify their investment risk through physical forward contracts or hedge their risk through financial instruments.
- Generation gets built if and only if market value of capacity (as reflected by the financial markets) exceeds the cost of new generation.
- Equal opportunity for demand side participation in mitigating price risk.
- Administratively set VOLL are replaced by demand side response to price signals
- Theoretical probabilistic models for calculating LLOP are replaced by empirically calibrated stochastic price models underlying the pricing of physical generation capacity and of hedging instruments.

6. SOME CAVEATS AND IMPEDIMENTS TO MARKET BASED PROVISION OF GENERATION ADEQUACY

An important concern that is often voiced in countries where there is no well developed institutional infrastructure that can enforce financial liability of corporation is that load serving entities or generators may assume more risk than they could handle reliably. So for instance, hydro generators may oversell their water in the present market and not be able to meet their generation adequacy obligations for which they collected capacity payments through premiums on private contracts. Likewise, load-serving entities left to their own devices may not hedge their supply sufficiently in order to reduce their capacity payments and may go out of business or default on their obligation to their customers if the spot prices for electricity skyrocket due to supply shortages. Such problems, however, face any commercial entity that is involved in underwriting risk. This is true for banks, savings and loans and insurance companies that require some form of regulation which will protect the customers from default. In the case of electricity it may be necessary to set some minimum contracting or hedging level on load serving entities. The premium payment for meeting such requirements through contracting with generators will produce the capacity payments that generators need to insure the stable income stream for financing adequate generation investment. In exchange for a stable source of income the generators will forgo some of the opportunity to collect high scarcity rents. However, there is no need for a "one size fits all" approach that awards a uniform capacity payment to all generators and imposes a uniform capacity charge on all the loads. A market based approach, which allows parties to trade energy price risk, and investment risk through different contractual arrangements can achieve better efficiency in risk sharing and investment. Regulatory intervention can then be limited to enforcement of minimal hedging requirement and oversight of commercial liability standards and adherence to contractual arrangements.

A system of capacity payments that is linked to assumption of energy price risk can also address the problem of over or under compensation of generators based on simulated market conditions. In Colombia for instance capacity payments to generators are based on simulation results of hydro scarcity and forecasted need for dispatch of thermal plants under such scarcity conditions. Generators that are not "dispatched" by the simulation are not entitled to capacity payments although they may still be dispatched in reality whereas a generator that received the capacity payment may be unavailable. A system were the capacity payments represent a call option would require generators that receive capacity payments to be available to produce energy at the strike price, or purchase it and provide it at that price. On the other hand, generators that did not receive capacity payments should be allowed to collect up to the VOLL for their power. The short term inelasticity of demand and steep supply curve may necessitate the setting of a price cap at an administratively chosen VOLL. That cap value will then serve as both, a penalty for unmet availability obligation and as a cap on the scarcity rents collected by generators who did not receive capacity payments. Further extension of this approach would allow generators to select among different levels of capacity payment in exchange for being available to provide energy at corresponding strike price levels, or buyout of their obligation at VOLL.

Another problem that may arise in a market based capacity payment system concerns possible failure of the capital market to provide long term financing for generation investments at rates that commensurate with the associated risk. Such market failure may arise since supply contracts that will provide the equivalent capacity payments as option premiums are typically of short duration (no longer than five years) whereas generation investment requires fifteen to thirty years of financing. The practice of securitizing long term investment by rolling over short term contracts is prevalent in many industries (e.g. using short term savings to finance thirty year mortgages). However, lack of experience with commodity trading in the electricity industry and the perceived regulatory intervention risk (especially in developing countries) may raise the cost of capital to levels that will reduce investment below the efficient adequacy level. Capacity payments are often viewed as a means of income stabilization that would enable generators to obtain financing for adequate investment level. If this indeed were the concern that capacity payments address a more appropriate mechanism would be some form of loan guarantees by the regulator. Since regulatory intervention is one of the important risks factors concerning investors in this business such loan guarantees may inspire confidence in the regulators commitment to uphold free market principles.

7. CONTRACT DURATION.

Viewing capacity payments as premium for call options at corresponding strike prices requires the specification of contract duration. Locking in the capacity payment for a longer duration has the effect of averaging out price volatility thus, providing the security of a stable income stream for the generator and stable energy prices for the consumers. However, the argument for diversity of choices in strike prices also applies to diversity of choice in contract terms. As contracts get shorter the corresponding option premium constituting the capacity payment becomes more volatile and starts to behave as a spot market for capacity. At the limit the capacity payment becomes an energy adder, which is, indistinguishable from energy payments for dispatched generators or from ancillary services payments to generators providing spinning reserves. Ideally the capacity adder should be rolled into the energy bids and reflected in the hourly or half-hourly energy market clearing prices. When a subsequent ancillary service market exist as in California, equilibrium between the energy and ancillary service market dictates that energy bids are raised by the opportunity cost of selling capacity in the ancillary service market. Hence, the market-clearing price for reserves is a good estimate of the capacity component contained in the market clearing prices for energy. In the old UK system that equilibrium condition is enforced administratively by calculating a capacity adder based on LOLP x (VOLL-SMP) which is paid to dispatched generators on the top of the system marginal energy price (SMP) and to non dispatched generators that declare availability. Excess availability will depress the capacity adder but all the available capacity receives that payment regardless of the price that they bid for energy. An option premium based calculation of the capacity adder would adjust the capacity adder according to the energy price bid by the generator. Thus dispatched generators would receive an option premium based on the hourly SMP serving as strike price while generators whose bids exceeded the SMP should be paid a call option premium according to their energy bid serving as strike price.

8. PRICES VERSUS QUANTITIES.

In the above discussion it was argued that in a market based approach generators get compensated for the cost of capacity either through scarcity rents that would arise in the energy market or through option premiums for availability at agreed strike prices. Yet many regulators around the world feel that energy and contract markets are not mature enough and cannot be trusted to produce the desired adequacy outcome so that some supporting mechanism is need, at least in the near future, in order to ensure generation adequacy. Given this attitude there is still a choice to be made between capacity payments versus enforcement of reserves obligations. Ruff [1] makes the case that among these two options the later is preferred on the ground of the classic prices vs. quantities argument. According to that argument, the supply function for capacity is relatively flat while the demand function is steep as illustrated in Figure 3. Hence a small error in the set capacity price may result in a large error in level of investment whereas controlling the quantity directly will produce a fairly accurate price. Thus, imposing planning reserves requirements that can than be traded in a capacity market will ensure that the desired adequacy is achieved and the capacity markets will produce the correct capacity price signal.

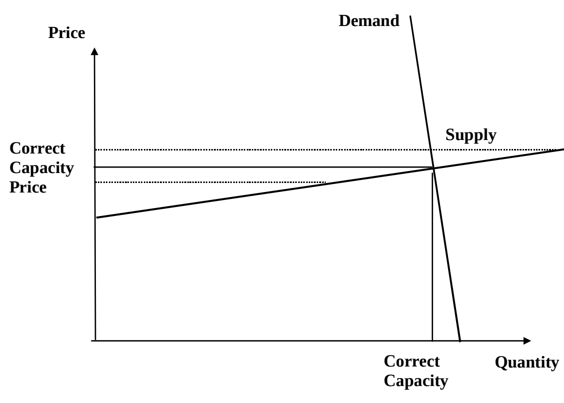

Figure 3 - Characteristic Shape of Supply and Demand Functions for Capacity

9. SUMMARY

The role of capacity payments in ensuring adequacy of supply can be fulfilled by risk management approaches and hedging instruments that permit diverse choices and promote demand side participation. The market should determine the value of capacity as a hedge for price risk If capacity payment are intended to correct failures of capital markets then regulatory intervention should address directly the availability and cost of long-term financing for capacity expansion secured by short-term contracts (e.g., through loan guarantees) and focus on promoting market confidence and rules that facilitate liquid markets for energy futures and other risk management instruments

When energy markets are not sufficiently developed to provide correct market signals for generation investment, setting capacity requirements with secondary markets that enable trading of capacity reserves is the preferred approach. It is more likely to produce correct market signals for investment than administratively set capacity payments which are likely to distort energy prices and result in over investment.

REFERENCES

[1] Ruff Larry, "Capacity payment, security of supply and stranded costs"' presentation to the National Spanish Electricity System Commission, Madrid, (June 7, 1999).

[2] Chuang Angela and Felix Wu, "Capacity payments and pricing of reliability in competitive generation markets", Proceedings of the 33rd Hawaii International System Science Conference" (January 2000)

[3] Graves Frank and Philip Hanser and Robert Earl, "The death and resurrection of electric capacity pricing", Working Paper, The Brattle Group, (October 1998)

[4] Hirst Eric, Brendan Kirby and Stan Hadly, "Generation and transmission adequacy in a restructured US electricity industry", Report to the EEI (March 1999)

[5] Chao Hung Po, "Peak load pricing and capacity planning with demand and supply uncedrtainty", Bell Journal of Economics Vol 14, (1983) pp. 179-190